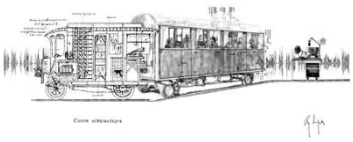
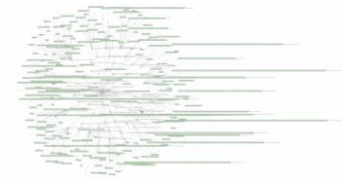
université
Paris Ouest
Nanterre La Défense

Observatologie :

Vers une science de l'adéquation observationnelle en linguistique





Olivier Baude

Habilitation à diriger des recherches

Volume des travaux et publications

Observatologie :

Vers une science de l'adéquation observationnelle en linguistique

Olivier Baude

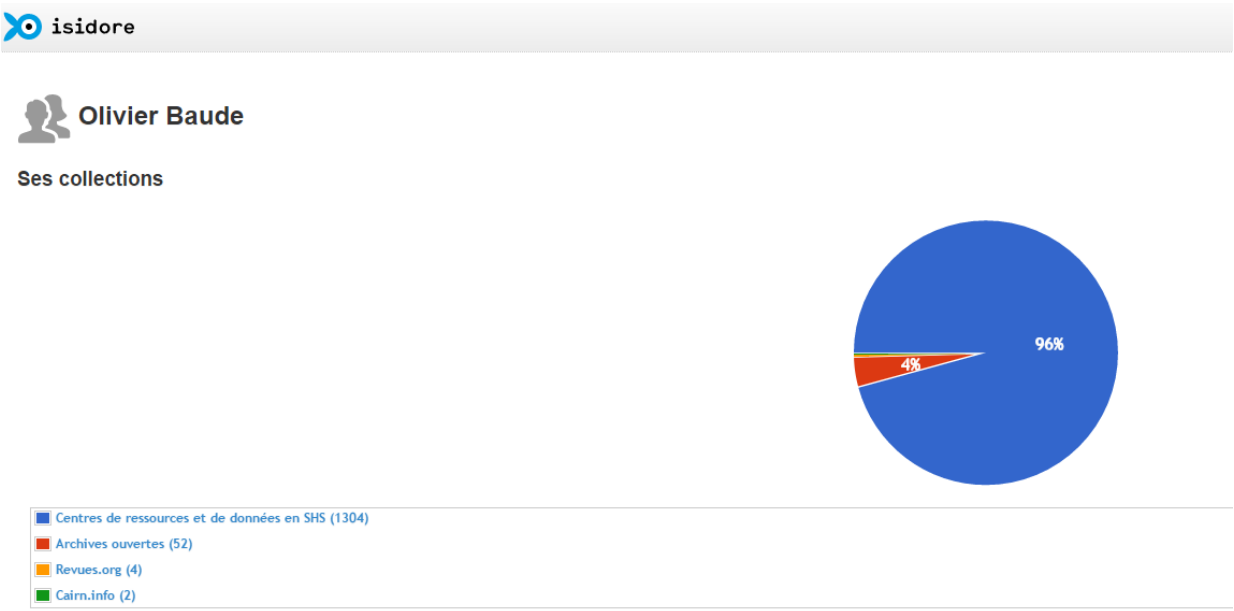
Habilitation à diriger des recherches

Volume des travaux et publications

Travaux signalés dans ISIDORE

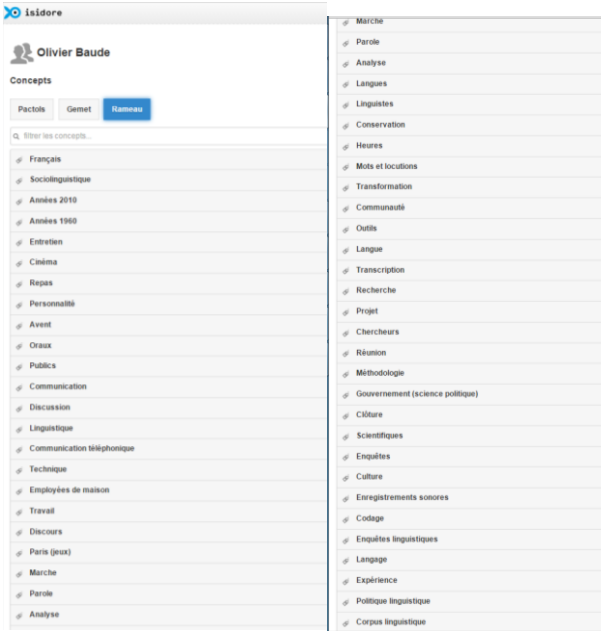
http://m.rechercheisidore.fr/auteur.php?familyName=Baude&givenName=Olivier

Note : la baisse brutale entre 2014 et 2015 correspond simplement au temps nécessaire pour décrire les productions. Il convient donc de ne pas tenir compte de la courbe sur la dernière année.



http://m.rechercheisidore.fr/auteur.php?familyName=Baude&givenName=Olivier

Référentiel Rameau :



Bibliographie Olivier Baude

DO

Baude, O., coord. (2012) *Corpus oraux, guide des bonnes pratiques*, Korea Pagijong Press. Version coréenne du guide Baude, O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

Le livre a également donné lieu à des versions électroniques : anglaise (2009) et allemande (2010).

Baude, O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

ACL/ACLN

Baude, O., Dugua, C., (2015 à paraitre, accepté pour publication) « Les ESLO, du portrait sonore au paysage digital », *Corpus*.

Baude, O., Bergouniuox, G., (à paraitre), chapitre « L'ESLO : une enquête en son temps » in *Linguistique de corpus : une étude de cas La recette de l'omelette*, Champion

A. Lacheret, P. Pietrandrea, O. Baude, A. C. Simon (accepté pour publication) "The collection of data for the Rhapsodie Treebank: typological criteria and ethical issues" in Lacheret A., Kahane S., Pietrandrea P., *Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French*, col Studies in Corpus Linguistics, Amsterdam, Benjamins.

Eshkol-Taravella I., **Baude O**., Maurel D., Hriba L., Dugua C., Tellier I., (2012) Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012. in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 17-46.

Jacobson M, **Baude O.**, (2011) « Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France », in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 47-69.

Baude O., Duga C. (2011) « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? » *Corpus 10*, *Varia*, 99-118.

Baude, O., (2007) « Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux », in *Corpus : état des lieux et perspectives.* RFLA XII-1, éditions De Werelt, Amsterdam, 71-84.

(act)

Baude, O. Dugua, C., (accepté pour publication), « Jean Zay et la mémoire orale : du politique au scientifique, de l'État à la ville, d'hier à aujourd'hui », Actes du colloque *Jean Zay : la culture et les langues Invention / Reconnaissance / Postérité*.

Baude, O. Dugua,C. (2015) « Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ? » in Dostie, G., Hadermann, P ., *La dia-variation en français actuel*, collection "Sciences pour la communication", Peter Lang ed, Berlin, pp 349-372.

Abouda, L., **Baude, O.** (2009) «Du Français Fondamental aux ESLO», in Bruxelles, Mondada, Simon, Traverso «Grand corpus de français parlé, Bilan historique et perspectives de recherche, *Cahiers de Linguistique Revue de sociolinguistique et de sociologie de la langue française 33/2*, EME, Louvain, 131-146.

Baude, O. (2008) **«**Le droit de la parole», in M. Bilger (ed), *Données orales, les enjeux de la transcription*, Presses universitaires de Perpignan, p 23-34.

Abouda, L., **Baude, O.** (2007) « Constituer et exploiter un grand corpus oral, choix et enjeux théoriques le cas des Eslos »**,** in actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification juin 2006, Presses universitaires de Toulouse: 161-168.

Baude, O. (2007) « Corpus oraux les bonnes pratiques d'une communauté scientifique »**,** in actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification, juin 2006, Presses universitaires de Toulouse, 61-66.

Baude, O. (2004) « Les corpus oraux entre science et patrimoine », l'expérience de l'Observatoire des pratiques linguistiques », Colloque du GEREC, PUG ed, Grenoble.

INV

Baude, O., (2015), « Mutualiser et diffuser des corpus : vraiment ? Pourquoi ? Comment ? », Colloque ICODOC *Corpus complexes et enjeux méthodologiques : de la collecte de données à leur analyse*, 18 et 19 mai, Lyon.

Baude, O., (2015) « Les langues de France et le programme corpus de la parole », Colloque *Technologies pour les Langues Régionales de France,*19-20 févr. 2015 Meudon (France)

Baude, O. Dugua, C., (2014), « Jean Zay et la mémoire orale : du politique au scientifique, de l' État à la ville, d'hier à aujourd'hui », Colloque *Jean Zay : la culture et les langues Invention / Reconnaissance / Postérité*, Université d'Orléans – BnF 25-26 novembre.

Baude, O., (2014) « Les ESLOs, du portrait sonore au paysage digital », Colloque international, *Corpus de français parlés et français des corpus* 08-09 mai 2014**,** Université de Neuchâtel, ILCF.

Baude, O., (2014) « Archivage de corpus oraux : Etat des lieux à partir de l'exemple du corpus ESLO », Conseil scientifique IRCOM, Paris, 22 septembre.

Baude,O. Dugua, C., Kanaan-Caillol, L., (2014), « Le corpus des ESLO à l'ère des Digital Humanities » *Journées FLORAL*, Paris, 5-8 décembre.

Baude, O. Guerin, E. (2014), « *ESLO : du portrait sonore à la "ville"*» Journées FLORAL, Paris, 5-8 décembre.

Baude O., Dugua, C., (2013) « Regards croisés sur la liaison dans le corpus ESLO », Journées PFC, Paris, 5-7 décembre.

Baude, O. (2013), « Actualité des Archives de la Parole : les corpus de la parole aujourd'hui », colloque *1913 – 2013, Centenaire des enquêtes de Ferdinand Brunot en Berry et en Limousin*, Bibliothèque nationale de France- Site François Mitterrand Tolbiac, Vendredi 8 novembre

Baude O, (2012) « Paroles élues, la liaison dans les discours des hommes politiques ». Colloque international *Paroles d'en-haut*, 6-7 décembre, Université d'Orléans-Sénat.

Baude O., (2012) « Corpus de référence : homogénéité, hétérogénéité et représentativité » Initiative Corpus français de référence. ILF, Paris, 14-15 juin 2012.

Baude, O. (2007) « Contributions des corpus oraux à la linguistique de corpus : *une démarche réflexive intégrée* » 5èmes Journées de la Linguistique de Corpus, Lorient, 13 – 15 septembre.

Baude O., (2006) **«** Pierre Encrevé et la réforme de l'orthographe : «le champ du linguiste », colloque international *Faire signe*, 16-18 octobre, Paris.

Baude, O., (2005), « Transcrire, les bonnes pratiques des linguistes », Journées d'étude de l'Association des Amis de Jacques Lacan, L'interprétation dans la cure et dans la transcription des textes, 11, 12 juin.

COM

Baude, O. Guerin, E. (2014) « Pourquoi et comment dresser le portrait sonore d'une "grande ville"? L'exemple d'ESLO2 », colloque international *Les Métropoles Francophones en Temps de Globalisation*, Nanterre, 5-7 juin.

Baude, O. Dugua,C. (2013) « Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ? », colloque international DIA du français actuel, *La dia-variation en français actuel des corpus aux ouvrages de référence (dictionnaire, grammaire),* 29 au 31 mai 2013 Université de Sherbrooke, Québec, Canada.

Baude, O. Dugua, C. (2011), « La variation en réserve », Colloque AFLS *Regards nouveaux sur les liens entre théories, méthodes et données en linguistique française,* 8-10 septembre, Nancy.

Baude, O. (2011), « Le programme Corpus de la parole », Colloque AFLS *Regards nouveaux sur les liens entre théories, méthodes et données en linguistique française*, 8-10 septembre, Nancy.

Baude, O. Dugua, C., Hriba L., (2011), « Transcrire : la norme, la variation, le linguiste », Colloque CERLICO *Transcrire, écrire, formaliser 2,* 27,28 mai, Orléans.

Baude, O. (2011) *Mai 68 dans le corpus ESLO,* Colloque Langage, Discours, Evénements, 31 mars -2 avril 2011, Florence.

Baude, O. (2009) « Les Eslos, un corpus variationniste représentatif d'une *communauté d'auditeurs* ? » , Colloque international du CERLICO, L'exemple et le corpus, quel statut ? 5 & 6 juin 2009, Poitiers.

Baude, O., Perrot, M-E., (2009) « Les Enquêtes Sociolinguistiques à Orléans (1970-2009) : l'entretien en questions », colloque international *Pour une épistémologie de la sociolinguistique,* Montpellier, 10, 11 et 12 décembre 2009, Université Paul-Valéry.

Baude, O., Hriba, L., (2009) « Sociolinguistique et transcription», colloque international *Pour une épistémologie de la sociolinguistique,* Montpellier, 10, 11 et 12 décembre 2009, Université Paul-Valéry.

Baude, O. (2008), «Un grand corpus de référence du français parlé : état des lieux et perspectives», Colloque international AFLS, *Les voix du français*, Oxford, 3-5 septembre.

Baude, O., Hriba, L. (2008), « De la variation à la norme, effets de codage dans les ESLOs », colloque international CATCOD 2008, Université d'Orléans.

Baude, O., Eshkol, I (2007), «Entrer dans l'anonymat, Etude des «entités dénommantes» dans un corpus oral. Colloque international *Proper Names in Spoken Language*, 22 et 23 novembre, Université de Bâle (Suisse).

Baude, O., Jacobson, M., Tchobanov, A., Walter, R. (2006), « interoperability of audio corpora : the case of the french corpora, LREC 2006, Genova, Italy.

Baude, O., Eshkol, I (2006,) « Constitution et exploitation d'un grand corpus de «données situées» *Problèmes et solutions pour les Enquêtes Socio-Linguistiques à Orléans (1968-2008) »*, Colloque international 3ème rencontre Fribourgeoise de la linguistique sur corpus appliquée aux langues romanes. Fribourg Allemagne.

Baude, O. (2005) «"Exploiter" l'oral : aspects juridiques et éthiques », Colloque *Transcription de la langue parlée. Aspects théoriques, pratiques et technologiques*, 27-30 juin, Perpignan.

Baude, O. (2002) « Les discussions autour de la réforme de l'orthographe de 1990 en France : une approche sociopragmatique cognitive », Colloque international Variation, catégorisation et pratiques discursives 12, 13, 14 septembre, Paris.

JOURNEES d'ETUDES, ECOLES THEMATIQUES, ATELIERS

Eshkol-Taravella I., Kanaan-Caillol L., Baude O., Dugua C., Maurel D., (2014) « Procédure d'anonymisation et traitement automatique : l'expérience d'ESLO » *Journée ATALA Ethique et TAL*, 22 novembre.

Baude, O. (2014), « Une très grande infrastructure pour les humanités numériques », Journée d'études Cascimodot, Tours, 19 juin.

Baude, O. (2014) Séminaire MSH Val de Loire, *Pratiques numériques en SHS*, « (Bonnes) pratiques du document sonore : l'exemple du programme *Corpus de la parole »*, 13 février 2014.

Baude, O. (2014) Séminaire doctoral ICAR-2 *Pratiques de linguiste : du juridique à l'éthique, de la collecte à la diffusion*, 17 février

Baude, O. (2013), « La TGIR Huma-Num », Symposium franco-japonais Analyse des données sonores et corpus de référence– Expériences croisées –, LLL-BnF18 octobre

Baude, O. (2013), « Les linguistes de l'oral et les données numériques » Séminaire Centre d'Alembert 2012/2013 *: Sensibles, empêchées, inaccessibles, ouvertes, archivées, revisitées… Les données de la recherche en questions*, 20 mars

Baude, O. (2013), « Quelques remarques sur la "publicisation" d'un corpus de référence du français », Journées Initiative corpus de référence, ILF, Paris, 29 mars

Baude, O., Jacobson M., (2013), « Conserver et exploiter les corpus de parole, Atelier général du labex Passés dans le Présent, Les archives de l'ethnomusicologie : mettre en commun, mettre à disposition », Paris-Nanterre, 27 Mars

Baude, O. (2012) Séminaire TGE ADONIS "Le document sonore" – « Eslo, instrument de comparaison pour le recueil de données linguistiques », 9 février.

Baude, O. (2008) «Les enquêtes sociolinguistiques à Orléans, Base et corpus». Ecole thématique CNRS I_DOCORA Interaction : DOnnées, CORpus, Analyse 23 au 27 juin 2008, Fourvière – Lyon.

Baude, O. (2008) «Outiller la sociolinguistique : une démarche réflexive autour du corpus des ESLOs (1968-2008)», Ecole thématique du CNRS / tge ADONIS *Préservation et diffusion numériques des sources de la recherche en sciences humaines et sociales* 19-24 octobre 2008, Fréjus (Var)

Baude, O. (2008) «Valorisation de corpus oraux : du terrain au portail», Ecole thématique du CNRS / tge ADONIS *Préservation et diffusion numériques des sources de la recherche en sciences humaines et sociales* 19-24 octobre 2008, Fréjus (Var).

Baude, O. (2008) Hriba, L. 2008, «*Les Enquêtes SocioLinguistiques à Orléans (1968-2008)* : choix méthodologiques pour un corpus prototypique», Journée d'étude *Parole*, FORELL, Poitiers.

Baude, O. (2007) « Mutualiser des corpus oraux, aspects juridiques et déontologiques» Journées d'étude CORPAFROAS, Corpus Oral en langues Afroasiatiques : Analyse Prosodique et Morphosyntaxique, 15 février, Paris.

Baude, O. (2007) «Constituer et exploiter un corpus d'interactions- aspects juridiques et éthiques **-** «, Ecole thématique du CNRS CONTACI, Lyon.

Baude, O. (2007) «Le corpus d'Orléans», Journée d'études de la BnF, *Autour du Français Parlé : de Brunot à nos jours. De l'archivage à l'exploitation*.

Baude, O. (2006) «Diffusion des corpus oraux, problèmes juridiques et déontologiques »**,** Ecole thématique du CNRS ELCO, Nantes.

Baude, O. (2003) «Réflexions empiriques pour une sociopragmatique cognitive », CURAPP

Séminaire de l'école doctorale, « L'analyse du discours médiatique », Usages sociolinguistique et sociologique des données journalistiques.

AP

Baude, O., Alessio, M., (2010) «Diversité des langues et plurilinguisme», in *Culture & Recherche*, n° 124, Ministère de la Culture et de la communication, Paris. Pp 4-5.

Baude, O., M., Sibille, J. (2010) «L'observatoire des pratiques linguistiques» et « entretien avec pierre Encrevé, in *Culture & Recherche*, n° 122 et 123, Ministère de la Culture et de la communication, Paris. Pp 82-83.

Baude, O. (2009) « découvrir les langues de France : le site corpus de la parole » Colloque international du CERLICO, *L'exemple et le corpus, quel statut ?* 5 & 6 juin 2009, Poitiers.

Baude, O., Alessio, M., (2008) «Les corpus de la parole, patrimoine immatériel et langues de France», in *Culture & Recherche*, n° 116 et 117, Ministère de la Culture et de la communication, Paris. Pp 42-43.

Baude, O. (2008) «Les «bonnes pratiques» de constitution et d'exploitation de corpus oraux, un exemple d'initiative fédérative pour une communauté spécifique», Atelier ANTHROPONET *: champ documentaire et champ scientifique* : *Quelles pratiques et quels standards dans l'indexation de corpus scientifiques multimédia ?,* Orléans, 26-27 juin 2008.

Baude, O. (2008) « l'oral un domaine à exploiter » conférence au salon Expolangues, 7 février, Paris.

Baude, O., Sibille, J. (2003) «L'observatoire des pratiques linguistiques», in *Culture & Recherche*, n° 96 Ministère de la Culture et de la communication, Paris. Pp 7-8.

Baude, O., (1999) «L'observatoire des pratiques linguistiques», in *Culture & Recherche*, n° 75 Ministère de la Culture et de la communication, Paris. Pp 6-8.

Baude, O. (2004-2012) Rédacteur en chef de *Langues & cité*, bulletin de la Délégation générale à la langue française et aux langues de France, Ministère de la Culture.

Bibliographie dans HAL avec liens

Voir la notice d'autorité sur Virtual International Authority File (VIAF): http://viaf.org/viaf/39685504

Références dans idref

- LE SENS SOUS PRESSE. UNE APPROCHE COGNITIVE ET SOCIOLOGIQUE DE LA CONSTRUCTION DU SENS D'UN TERME LEXICAL AU COEUR D'UN EVENMENT MEDIATIQUE ; UN EXEMPLE : "LA REFORME DE L'ORTHOGRAPHE DE 1990" / OLIVIER BAUDE ; SOUS LA DIR. DE PIERRE ENCREVE / [S.l.] : [s.n.] , 1998
- Le sens sous presse : une approche cognitive et sociologique de la construction du sens d'un terme lexical au coeur d'un évènement médiatique : un exemple, la réforme de l'orthographe de 1990 / Olivier Baude / Paris : [s.n.] , 1998
- La transmission intra-familiale de l'arabe marocain en France : Etude comparative des pratiques linguistiques déclarées et effectives de deux familles / Maha Abourahim ; sous la direction du professeur Dominique Caubet / [S.l.] : [s.n.] , 2011
- Le coaching strategico-linguistique : vers une science du changement ? / Maxence Lureau ; sous la direction de Bernard Laks et de Isabella Pezzini / [S.l.] : [s.n.] , 2014
- Corpus oraux [Texte imprimé] : guide des bonnes pratiques 2006 / coordonné par Olivier Baude / Paris : CNRS éd. , [2006]

Références dans HAL

- Iris Eshkol-Taravella, Olivier Baude, Denis Maurel, Layal Kanaan-Caillol. Recherche des indices permettant une identification: l'anonymisation des transcriptions du corpus ESLO. TALN2015, Jun 2015, Caen, France. 2015, Actes de la 1e Ethique et TRaitemeNt Automatique des Langues (ETeRNAL'2015), Caen (France). <https://taln2015.greyc.fr/articlesenlignetaln/>. https://hal.archives-ouvertes.fr/hal-01174647
- Olivier Baude. CORPUS ORAUX : LES BONNES PRATIQUES D'UNE COMMUNAUTE SCIENTIFIQUE. Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation, 2006, Albi, France. pp.61-66, 2007, Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation. https://halshs.archives-ouvertes.fr/halshs-01162487

- Lotfi Abouda, Olivier Baude. CONSTITUER ET EXPLOITER UN GRAND CORPUS ORAL : CHOIX ET ENJEUX THEORIQUES. LE CAS DES ESLO. Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation, 2006, Albi, France. 2007, Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation. https://halshs.archives-ouvertes.fr/halshs-01162506
- Olivier Baude. Les corpus oraux entre science et patrimoine. L'expérience de l'Observatoire des pratiques linguistiques. Publicisation de la science, 2004, Grenoble, France. https://halshs.archives-ouvertes.fr/halshs-01162520
- Lotfi Abouda, Olivier Baude. Du Français Fondamental aux ESLO. Grand corpus de français parlé, Bilan historique et perspectives de recherche, 2005, Lyon, France. 33 (2), pp.131-146, 2005, Cahiers de linguistique. https://halshs.archives-ouvertes.fr/halshs-01162533
- Olivier Baude. Le droit de la parole. Données orales : les enjeux de latranscription, 2005, Perpignan, France. Presses universitaires de Perpignan, 2008, Données orales : les enjeux de la transcription. https://halshs.archives-ouvertes.fr/halshs-01162543
- Olivier Baude. Transcrire : les bonnes pratiques des linguistes. Journées d'étude de l'Association des Amis de Jacques Lacan L'interprétation dans la cure et dans la transcription des textes, 2005, Paris, France. 2005. https://halshs.archives-ouvertes.fr/halshs-01162548
- Olivier Baude, Jean Sibille. L'observatoire des pratiques linguistiques. Culture et recherche, Paris : Ministère de la Culture et de la Communication, 2003, pp.7-9. https://halshs.archives-ouvertes.fr/halshs-01184590
- Olivier Baude, Michel Alessio. Les corpus de la parole : patrimoine immatériel et langues de France. Culture et recherche, Ministère de la Culture, 2008, pp.42-43. https://halshs.archives-ouvertes.fr/halshs-01184592
- Olivier Baude. Corpus oraux Un guide des bonnes pratiques. Culture et recherche, Ministère de la Culture, 2006, pp.2. https://halshs.archives-ouvertes.fr/halshs-01184593
- Olivier Baude, Jean Sibille. L'observatoire des pratiques linguistiques, suivi d'un entretien avec Pierre Ecrevé. Culture et Recherches, 2010, pp.82-83. https://halshs.archives-ouvertes.fr/halshs-01184595
- Michel Alessio, Olivier Baude. La diversité des langues. Culture et Recherches, 2010, pp.4-5. https://halshs.archives-ouvertes.fr/halshs-01184597
- Olivier Baude. L'observation des pratiques linguistiques en France. Culture et recherche, Ministère de la Culture, 1999, pp.6-8. https://halshs.archives-ouvertes.fr/halshs-01184281
- Olivier Baude, Céline Dugua. (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ?. Corpus, 2011, Varia, 10, pp.99-118. <http://corpus.revues.org/2036>. https://hal.archives-ouvertes.fr/hal-01162479

- Olivier Baude. Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux. Revue française de linguistique appliquée, De Werelt, 2007, Corpus état des lieux et perspectives, XII (1), pp.85-98. https://halshs.archives-ouvertes.fr/halshs-01163043
- Olivier Baude, Céline Dugua. Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ?. Gaétane Dostie et Pascale Hadermann. La dia-variation en français actuel. Etudes sur corpus, approches croisées et ouvrage de référence, 2013, Sherbrooke, Canada. Peter Lang, 2015, La dia-variation en français actuel. Etudes sur corpus, approches croisées et ouvrage de référence. https://halshs.archives-ouvertes.fr/halshs-01163047
- Iris Eshkol-Taravella, Olivier Baude, Denis Maurel, Linda Hriba, Céline Dugua, et al.. Un grand corpus oral « disponible » : le corpus d'Orléans 1 1968-2012. Traitement Automatique des Langues, ATALA, 2011, Ressources Linguistiques Libres, 53 (2), pp.17-46. https://halshs.archives-ouvertes.fr/halshs-01163053
- Olivier Baude. Les langues de France dans le programme Corpus de la parole. Technologies pour les Langues Régionales de France, Feb 2015, Meudon, France. 2015. https://halshs.archives-ouvertes.fr/halshs-01165904
- Olivier Baude. Mutualiser et diffuser des corpus : vraiment ? Pourquoi ? Comment ? . « Corpus complexes et enjeux méthodologiques : de la collecte de données à leur analyse », May 2015, Lyon, France. 2015. https://halshs.archives-ouvertes.fr/halshs-01165906
- Olivier Baude. Archivage de corpus oraux : Etat des lieux à partir de l'exemple du corpus ESLO. Journées scientifiques du consortium IRCOM-TGIR-Huma-nun. 2014. https://halshs.archives-ouvertes.fr/halshs-01165908
- Olivier Baude, Layal Kanaan. Le corpus des ESLO à l'ère des Digital Humanities. Premières Rencontres FLORAL, Dec 2014, Paris, France. 2014. https://halshs.archives-ouvertes.fr/halshs-01165909
- Olivier Baude, Guerin Emmanuelle. « ESLO : du portrait sonore à la "ville". Premières Rencontres FLORAL, Dec 2014, Paris, France. 2015. https://halshs.archives-ouvertes.fr/halshs-01165911
- Gabriel Bergounioux, Olivier Baude. ESLO, UNE ENQUÊTE EN SON TEMPS : ENJEUX, MÉTHODES ET RÉSULTATS. Gabriel Bergounioux. LINGUISTIQUE DE CORPUS UNE ÉTUDE DE CAS, LA RECETTE DE L'OMELETTE DANS L'ENQUÊTE SOCIO-LINGUISTIQUE À ORLÉANS (ESLO) , Champion, pp.7-13, 2015. https://halshs.archives-ouvertes.fr/halshs-01165934
- Céline Dugua, Olivier Baude. Regards croisés sur la liaison dans le corpus ESLO. Journées PFC 2013, Dec 2013, Paris, France. 2013. https://halshs.archives-ouvertes.fr/halshs-01165936
- Olivier Baude. Actualité des Archives de la Parole : les corpus de la parole aujourd'hui. 1913 – 2013 Centenaire des enquêtes de Ferdinand Brunot en Berry et en Limousin, Nov 2013, Paris, France. 2013. https://halshs.archives-ouvertes.fr/halshs-01165937

- Olivier Baude. Paroles élues, la liaison dans les discours des hommes politiques . Paroles d'en-haut, Dec 2012, Orléans / Paris, France. 2012. https://halshs.archives-ouvertes.fr/halshs-01165938
- Olivier Baude. Corpus de référence : homogénéité, hétérogénéité et représentativité . Journées ILF : Initiative Corpus de référence. 2012. https://halshs.archives-ouvertes.fr/halshs-01165939
- Olivier Baude. Contributions des corpus oraux à la linguistique de corpus : une démarche réflexive intégrée. 5èmes Journées de linguistique de corpus, Sep 2007, Lorient, France. 2007. https://halshs.archives-ouvertes.fr/halshs-01165940
- Olivier Baude. « Pierre Encrevé et la réforme de l'orthographe : "le champ du linguiste ». Faire signe, pour Pierre Encrevé, Oct 2006, Paris, France. 2006. https://halshs.archives-ouvertes.fr/halshs-01165941
- Olivier Baude. Transcrire, les bonnes pratiques des linguistes . L'interprétation dans la cure et dans la transcription des textes, Jun 2005, Orléans / Paris, France. https://halshs.archives-ouvertes.fr/halshs-01165942
- Olivier Baude, Céline Dugua. Jean Zay et la mémoire orale : du politique au scientifique, de l' État à la ville, d'hier à aujourd'hui. Jean Zay : la culture et les langues Invention/Reconnaissance/Postérité, Nov 2014, Orléans / Paris, France. 2014.https://halshs.archives-ouvertes.fr/halshs-01165944
- Olivier Baude, Emmanuelle Guerin. Pourquoi et comment dresser le portrait sonore d'une "grande ville"? L'exemple d'ESLO2. Les métropoles francophones en temps de globalisation, Jun 2014, Nanterre, France. 2014. https://halshs.archives-ouvertes.fr/halshs-01165945
- Olivier Baude, Céline Dugua. La variation en réserve . Colloque AFLS, Regards nouveaux sur les liens entre théories, méthodes et données en linguistique française, Sep 2011, Nancy, France. 2011. https://halshs.archives-ouvertes.fr/halshs-01165946
- Olivier Baude. Le programme Corpus de la parole. Colloque AFLS, Regards nouveaux sur les liens entre théories, méthodes et données en linguistique française, Sep 2011, Nancy, France. 2011. https://halshs.archives-ouvertes.fr/halshs-01165947
- Olivier Baude, Céline Dugua, Linda Hriba. Transcrire : la norme, la variation, le linguiste . 5 e Colloque international du CerLiCO "Transcrire, Ecrire, Formaliser 2", May 2011, orléans, France. 2011. https://halshs.archives-ouvertes.fr/halshs-01165948
- Olivier Baude. Les Eslos, un corpus variationniste représentatif d'une communauté d'auditeurs ? . L'exemple et le corpus, quels statuts ?, Jun 2009, Poitiers, France. 2009. https://halshs.archives-ouvertes.fr/halshs-01165949

- Olivier Baude, Marie-Eve Perrot. Les Enquêtes Sociolinguistiques à Orléans (1970-2009) : l'entretien en questions . Pour une épistémologie de la sociolinguistique, Dec 2009, Montpellier, France. 2009. https://halshs.archives-ouvertes.fr/halshs-01165950
- Olivier Baude, Linda Hriba. « Sociolinguistique et transcription. Pour une épistémologie de la sociolinguistique, Dec 2012, Montpellier, France. 2009. https://halshs.archives-ouvertes.fr/halshs-01165951
- Olivier Baude. Un grand corpus de référence du français parlé : état des lieux et perspectives. Colloque international AFLS, Les voix du français, Sep 2008, Oxford, United Kingdom. 2008. https://halshs.archives-ouvertes.fr/halshs-01165952
- Olivier Baude, Linda Hriba. De la variation à la norme, effets de codage dans les ESLOs . CATCOD, 2008, Orléans, France. 2008. https://halshs.archives-ouvertes.fr/halshs-01165953
- Olivier Baude, Iris Eshkol. Constitution et exploitation d'un grand corpus de "données situées"Problèmes et solutions pour les Enquêtes Socio-Linguistiques à Orléans (1968-2008). Corpus et pragmatique L'interaction verbale dans son contexte situationnel à la lumière des corpus et des bases de données, Sep 2006, Fribourg, Germany. 2006. https://halshs.archives-ouvertes.fr/halshs-01165954
- Olivier Baude. Les discussions autour de la réforme de l'orthographe de 1990 en France : une approche sociopragmatique cognitive . Colloque international Variation, catégorisation et pratiques discursives, Sep 2002, Paris, France. https://halshs.archives-ouvertes.fr/halshs-01165955
- Olivier Baude. Pratiques de linguiste : du juridique à l'éthique, de la collecte à la diffusion . Communication séminaire doctorants ICAR. 2014. https://halshs.archives-ouvertes.fr/halshs-01165956
- Iris Eshkol, Olivier Baude, Layal Kanaan, Denis Maurel, Céline Dugua. « Procédure d'anonymisation et traitement automatique : l'expérience d'ESLO ». Journée d'études ATALA, Ethique et TAL, 2014, Paris, France. 2014. https://halshs.archives-ouvertes.fr/halshs-01165957
- Olivier Baude, Michel Alessio. Les corpus de la parole : patrimoine immatériel et langues de France. Culture et recherche, Ministère de la Culture, 2008, Le patrimoine culturel immatériel, pp.42-43. https://halshs.archives-ouvertes.fr/halshs-01165963
- Olivier Baude, Michel Alessio. Diversité des langues et plurilinguisme, introduction. Culture et recherche, Ministère de la Culture, 2011, Diversité linguistique et plurilinguisme, 124, pp.4-5. https://halshs.archives-ouvertes.fr/halshs-01165964
- Olivier Baude. Une très grande infrastructure pour les humanités numériques . Journée d'études Cascimodot. 2014. https://halshs.archives-ouvertes.fr/halshs-01165990

- Olivier Baude. Pratiques numériques en SHS, « (Bonnes) pratiques du document sonore : l'exemple du programme Corpus de la parole . Séminaire MSH Val de Loire. 2014. https://halshs.archives-ouvertes.fr/halshs-01165991
- Olivier Baude. La TGIR Huma-Num . Symposium franco-japonais Analyse des données sonores et corpus de référence– Expériences croisée.. 2013. https://halshs.archives-ouvertes.fr/halshs-01165992
- Olivier Baude. Les linguistes de l'oral et les données numériques . Séminaire Centre d'Alembert 2012/2013 : Sensibles, empêchées, inaccessibles, ouvertes, archivées,.. 2013. https://halshs.archives-ouvertes.fr/halshs-01165993
- Olivier Baude. Quelques remarques sur la "publicisation" d'un corpus de référence du français . Journées Initiative corpus de référence, ILF, Paris. 2013. https://halshs.archives-ouvertes.fr/halshs-01165994
- Olivier Baude. Les enquêtes sociolinguistiques à Orléans, Base et corpus. Ecole thématique CNRS I_DOCORA Interaction : DOnnées, CORpus, Analyse, Lyon. 2008. https://halshs.archives-ouvertes.fr/halshs-01165996
- Olivier Baude. Outiller la sociolinguistique : une démarche réflexive autour du corpus des ESLOs (1968-2008). Ecole thématique du CNRS / tge ADONIS Préservation et diffusion numériques des sources de la rech.. 2008. https://halshs.archives-ouvertes.fr/halshs-01165998
- Olivier Baude. Valorisation de corpus oraux : du terrain au portail. Ecole thématique du CNRS / tge ADONIS Préservation et diffusion numériques des sources de la rech.. 2008. https://halshs.archives-ouvertes.fr/halshs-01165999
- Olivier Baude, Linda Hriba. Les Enquêtes SocioLinguistiques à Orléans (1968-2008) : choix méthodologiques pour un corpus prototypique. Journée d'étude Parole, FORELL, Poitiers. 2008. https://halshs.archives-ouvertes.fr/halshs-01166000
- Olivier Baude. Mutualiser des corpus oraux, aspects juridiques et déontologiques. Journée d'étude CORPAFROAS, Corpus Oral en langues Afroasiatiques : Analyse Prosodique et Morphos.. 2007. https://halshs.archives-ouvertes.fr/halshs-01166001
- Olivier Baude. Constituer et exploiter un corpus d'interactions- aspects juridiques et éthiques. Ecole thématique du CNRS CONTACI, Lyon. 2007. https://halshs.archives-ouvertes.fr/halshs-01166002
- Olivier Baude. Le corpus d'Orléans. Autour du Français Parlé : de Brunot à nos jours. De l'archivage à l'exploitation, 2007, Paris, France. https://halshs.archives-ouvertes.fr/halshs-01166003

- Olivier Baude. Diffusion des corpus oraux, problèmes juridiques et déontologiques . Ecole thématique du CNRS ELCO, Nantes. 2006. https://halshs.archives-ouvertes.fr/halshs-01166004
- Olivier Baude. Découvrir les langues de France : le site corpus de la parole. Colloque international du CERLICO, L'exemple et le corpus, quel statut ?, Jun 2009, Poitiers, France. https://halshs.archives-ouvertes.fr/halshs-01166005
- Olivier Baude. L'oral un domaine à exploiter . Conférence au salon Expolangues. 2008. https://halshs.archives-ouvertes.fr/halshs-01166006
- Olivier Baude. Les «bonnes pratiques» de constitution et d'exploitation de corpus oraux, un exemple d'initiative fédérative pour une communauté spécifique. Atelier ANTHROPONET : champ documentaire et champ scientifique : Quelles pratiques et quels stand.. 2008. https://halshs.archives-ouvertes.fr/halshs-01166007
- Isabelle De Lamberterie, Olivier Baude, Claire Blanche-Benveniste, Marie-France Calas, Paul Cappeau, et al.. Corpus oraux. Guide des bonnes pratiques 2006. Paris, France. CNRS Editions, 2006. https://halshs.archives-ouvertes.fr/halshs-00078730
- Olivier Baude, Claire Blanche-Benveniste, Marie-France Calas, Paul Cappeau, Pascal Cordereix, et al.. Corpus oraux, guide des bonnes pratiques 2006. CNRS Editions, Presses Universitaires Orléans, pp.203, 2006. https://hal.archives-ouvertes.fr/hal-00357706
- Olivier Baude, Christiane Marchello-Nizia, Lorenza Mondada, Claire Blanche-Benveniste, Marie-France Calas, et al.. Corpus oraux : guide des bonnes pratiques 2006. Baude, Olivier;. CNRS Éditions, pp.203, 2006. https://halshs.archives-ouvertes.fr/halshs-00355472
- Olivier Baude, Michel Jacobson, Atanas Tchobanov, Richard Walter. INTEROPERABILITY OF AUDIO CORPORA : THE CASE OF THE FRENCH CORPORA. 5 th International Conference on Language Ressources and Evaluation, May 2006, Genoa, Italy. 2006.https://halshs.archives-ouvertes.fr/halshs-01162927
- Michel Jacobson, Oliver Baude. Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France. Traitement Automatique des Langues, Lavoisier (Hermes Science Publications), 2011, 52 (3), pp.47-69. <http://www.atala.org>. https://halshs.archives-ouvertes.fr/halshs-01163037
- Michel Jacobson, Oliver Baude. Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France. Traitement Automatique des Langues, ATALA, 2011, Ressources libres, 52, pp.47-69.https://halshs.archives-ouvertes.fr/halshs-01165884
- Olivier Baude, Michel Jacobson. Conserver et exploiter les corpus de parole . Atelier général du Labex Passés dans le Présent, Les archives de l'ethnomusicologie : mettre en c.. 2013. https://halshs.archives-ouvertes.fr/halshs-01165995

- Olivier Baude, Claire Blanche-Benveniste, Marie-France Calas, Paul Cappeau, Pascal Cordereix, et al.. Corpus oraux, Guide des bonnes pratiques 2006. Pagijong Press, 2012, 978-89-6292-311-7. https://halshs.archives-ouvertes.fr/halshs-01165889
- Olivier Baude, Claire Blanche-Benveniste, Marie-France Calas, Paul Cappeau, Pascal Cordereix, et al.. Corpus oraux, Guide des bonnes pratiques 2006. Version allemande. 2010. https://halshs.archives-ouvertes.fr/halshs-01165896
- Olivier Baude, Claire Blanche-Benveniste, Marie-France Calas, Paul Cappeau, Pascal Cordereix, et al.. Spoken Corpora Good Practice Guide 2006. 2010. https://halshs.archives-ouvertes.fr/halshs-01165893

TRAVAUX

Liste des publications

Liste des publications

1. Article vulgarisation :
   Baude, O., (1999) «L'observatoire des pratiques linguistiques», in *Culture & Recherche*, n° 75 Ministère de la Culture et de la communication, Paris. Pp 6-8.

2. Article de vulgarisation :
   Baude, O., Sibille, J. (2003) «L'observatoire des pratiques linguistiques», in *Culture & Recherche*, n° 96 Ministère de la Culture et de la communication, Paris. Pp 7-8.

3. Actes
   Baude, O. (2004) « Les corpus oraux entre science et patrimoine », l'expérience de l'Observatoire des pratiques linguistiques », Colloque du GEREC, PUG ed, Grenoble.

4. Livre
   Baude, O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

5. Baude, O., Jacobson, M., Tchobanov, A., Walter, R. (2006), « interoperability of audio corpora : the case of the french corpora, LREC 2006, Genova, Italy.

6. Actes
   Baude, O. (2007) « Corpus oraux les bonnes pratiques d'une communauté scientifique »**,** in actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification, juin 2006, Presses universitaires de Toulouse, 61-66.

7. Actes
   Abouda, L., **Baude, O.** (2007) « Constituer et exploiter un grand corpus oral, choix et enjeux théoriques le cas des Eslos »**,** in actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification juin 2006, Presses universitaires de Toulouse: 161-168.

8. Article
   Baude, O., (2007) « Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux », in *Corpus : état des lieux et perspectives.*RFLA XII-1, éditions De Werelt, Amsterdam, 71-84.

9. Article de vulgarisation

Baude, O., Alessio, M., (2008) «Les corpus de la parole, patrimoine immatériel et langues de France», in *Culture & Recherche*, n° 116 et 117, Ministère de la Culture et de la communication, Paris. Pp 42-43.

10. Chapitre

    Baude, O. (2008) **«**Le droit de la parole», in M. Bilger (ed), *Données orales, les enjeux de la transcription*, Presses universitaires de Perpignan, p 23-34.

11. Actes

    Abouda, L., **Baude, O.** (2009) «Du Français Fondamental aux ESLO», in Bruxelles, Mondada, Simon, Traverso «Grand corpus de français parlé, Bilan historique et perspectives de recherche, *Cahiers de Linguistique Revue de sociolinguistique et de sociologie de la langue française 33/2*, EME, Louvain, 131-146.

12. Livre (traduction)

    Baude, O., coord. (2009) *Corpus oraux, guide des bonnes pratiques*, Traduction anglaise.

13. Livre (traduction)

    Baude, O., coord. (2010) *Corpus oraux, guide des bonnes pratiques*, Traduction allemande.

14. Article de vulgarisation

    Baude, O., M., Sibille, J. (2010) «L'observatoire des pratiques linguistiques» et « entretien avec Pierre Encrevé, in *Culture & Recherche*, n° 122 et 123, Ministère de la Culture et de la communication, Paris. Pp 82-83.

15. Article de vulgarisation

    Baude, O., Alessio, M., (2010) «Diversité des langues et plurilinguisme», in *Culture & Recherche*, n° 124, Ministère de la Culture et de la communication, Paris. Pp 4-5.

16. Article

    Baude O., Duga C. (2011) « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? » *Corpus 10*, *Varia*, 99-118.

17. Article

    Eshkol-Taravella I., **Baude O**., Maurel D., Hriba L., Dugua C., Tellier I., (2011) Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012. in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 17-46.

18. Article

    Jacobson M, **Baude O.**, (2011) « Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France », in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 47-69.

19. Livre

    Baude, O., coord. (2012) *Corpus oraux, guide des bonnes pratiques*, Korea Pagijong Press. Version coréenne du guide Baude, O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

20. Actes

    Baude, O. Dugua,C. (2015) « Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ? » in Dostie, G., Hadermann, P ., *La dia-variation en français actuel*, collection "Sciences pour la communication", Peter Lang ed, Berlin, pp 349-372.

21. Article

    Baude, O., Dugua, C., (2015 à paraitre, accepté pour publication) « Les ESLO, du portrait sonore au paysage digital », *Corpus*.

22. Chapitre

    Baude, O., Bergouniuox, G., (à paraitre), chapitre « L'ESLO : une enquête en son temps » in *Linguistique de corpus : une étude de cas La recette de l'omelette*, Champion

23. Actes

    Baude, O. Dugua, C., (accepté pour publication), « Jean Zay et la mémoire orale : du politique au scientifique, de l'État à la ville, d'hier à aujourd'hui », Actes du colloque *Jean Zay : la culture et les langues Invention / Reconnaissance / Postérité*.

24. Chapitre

    A. Lacheret, P. Pietrandrea, O. Baude, A. C. Simon (accepté pour publication) "The collection of data for the Rhapsodie Treebank: typological criteria and ethical issues" in Lacheret A., Kahane S., Pietrandrea P., *Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French*, col Studies in Corpus Linguistics, Amsterdam, Benjamins.

Titre	***L'observatoire des pratiques linguistiques***
Type	Article de vulgarisation
Editeur	MCC
Année	1999
Référence	Baude, O., (1999) «L'observatoire des pratiques linguistiques», in *Culture & Recherche*, n° 75 Ministère de la Culture et de la communication, Paris. Pp 6-8

culture & recherche

Ministère de la culture et de la communication
Mission de la recherche et de la technologie - 3, rue de Valois 75042 Paris cedex 01 - Tél. : 01 40 15 80 45

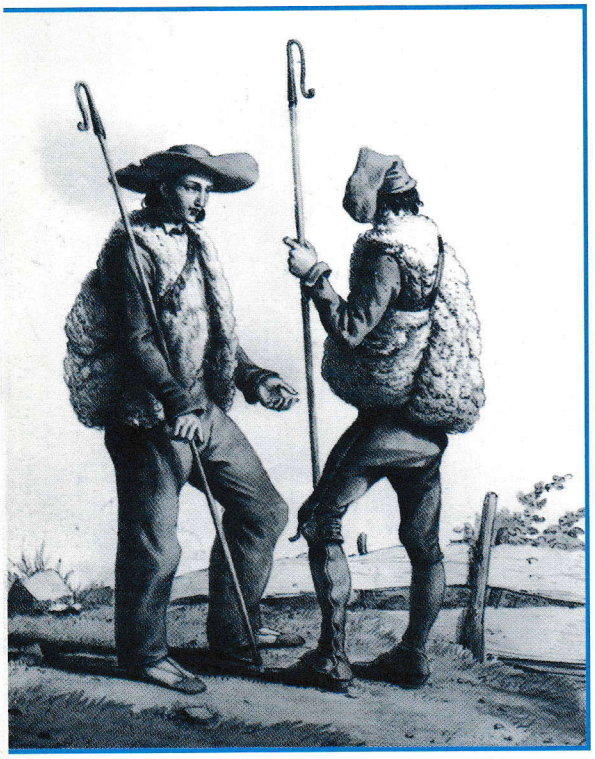
novembre
décembre
1999
N°75

sommaire

Ministère Culture Communication
Direction de l'administration générale

peut être amené à penser que l'arabe dialectal parlé en France a pour correspondant écrit l'arabe commun (celui de la presse, de la radio et de la télévision), qui n'est la langue maternelle de personne. Enfin, le nombre de langues "sans territoire", pour reprendre le terme assez discutable (le territoire d'une langue est le cerveau de ceux qui la parlent) de la Charte est frappant : le berbère et l'arabe dialectal, le yiddish, le romani chib et l'arménien occidental sont communément parlés en France (ou pratiqués, si l'on ajoute la langue des signes). Ce qui invite à prendre quelque distance avec la territorialisation des langues que la Charte privilégie.
Une telle inscription géographique, qui semble évidente pour certains idiomes (basque, alsacien, corse, etc.), n'en reste pas moins discutable. Elle est mise à mal par la réalité sociolinguistique, qui rappelle que la mobilité sociale contemporaine est telle que l'on parle les différentes langues "régionales" un peu partout. Le créole est une réalité linguistique bien vivante de la région parisienne. Elle s'oppose en outre aux principes républicains français, qui tiennent que la langue, élément culturel, appartient au patrimoine national ; le corse n'est pas propriété de la région de Corse, mais de la Nation. Le rayonnement de la langue française, langue de la citoyenneté, parlée aujourd'hui par tous et partout, est tel que l'on proposera la seule dichotomie suivante : le français d'une part, langue de la République, les langues minoritaires, de statut divers.

Cet ensemble constitue le patrimoine linguistique de la France. Il convient de le protéger, comme tout autre patrimoine, en conservant sa pratique, en favorisant son illustration. Mais il importe également de le décrire. Le retard en ce domaine est notable, si l'on pense aux autres domaines patrimoniaux. La connaissance de nombreuses langues que parlent des citoyens français est parfois très faible. Suggérons que la France se donne l'intention et les moyens d'une description scientifique de ses langues, aboutissant à une publication de synthèse. La dernière grande enquête sur le patrimoine linguistique de la République, menée il est vrai dans un esprit assez différent, fut celle de l'abbé Grégoire (1790-1792).



Pâtres de la Vallée d'Aran.
Exposition "Bergers de France". Cliché : Service historique. Musée national des arts et traditions populaires.

Bernard Cerquiglini
Directeur de l'Institut national de la langue française (C.N.R.S.)
Le rapport de Bernard Cerquiglini est disponible sur le site Internet de la Délégation générale à la langue française : http://dglf.culture.fr

NOTES

1- Bernard Cerquiglini, Les langues de la France. *Rapport au Ministre de l'Éducation Nationale, de la Recherche et de la Technologie, et à la Ministre de la culture et de la communication. Paris : Délégation générale à la langue française, 1999.*

2- Les pays européens qui ont retenu le plus grand nombre de langues régionales ou minoritaires sont l'Allemagne (sept : danois, haut sorabe, bas sorabe, frison septentrional, frison satera, bas allemand, rom) et la Croatie (sept : italien, serbe, hongrois, tchèque, slovaque, slovène, ukrainien).

L'observation des pratiques linguistiques en France

Dans le courant de l'année 1998, Mme Catherine Trautmann, ministre de la culture et de la communication, a demandé à la Délégation générale à la langue française (DGLF) de réfléchir à la définition et aux conditions de création d'un Observatoire des pratiques linguistiques. Afin de répondre à cette nouvelle mission, la DGLF a procédé, au cours de l'automne 1998, à de nombreuses consultations, notamment des milieux universitaires et de la recherche spécialisés en linguistique et sociolinguistique.

L'observatoire, cellule de la DGLF a pour mission d'étudier les pratiques linguistiques en France ainsi que les modalités et les effets du contact entre les langues, afin d'apporter des informations utiles pour l'élaboration des politiques sociales, éducatives et culturelles en permettant de prendre en compte les expériences linguistiques des individus et des groupes. Il s'agit de travaux sociolinguistiques sur l'usage actuel du français et des langues utilisées en France.
Après une période de préfiguration, l'Observatoire des pratiques linguistiques a lancé ses premiers travaux au printemps 1999 avec le soutien de la Mission de la recherche et de la technologie.
Le premier acte de l'Observatoire a été d'entreprendre simultanément un inventaire des études et des travaux de recherche dans le domaine des pratiques actuelles des locuteurs en France et la réalisation d'une base de données permettant de gérer l'ensemble de ces informations. Cette base de données comprendra la liste des centres de recherche et des autres organismes impliqués dans l'observation des pratiques linguistiques, des informations sur les travaux réalisés et conduits par ces équipes, et la liste des centres de documentation possédant des ressources en ce domaine.

Après consultation des spécialistes du domaine, la Délégation générale à la langue française a lancé par ailleurs, en avril 1999, un appel à propositions sur la description et l'analyse de l'hétérogénéité des pratiques linguistiques sur l'ensemble du territoire national : analyse des variations et description des variétés du français, analyse des variations et description des variétés des autres langues utilisées en métropole et dans les DOM-TOM, situations, nettement circonscrites à un lieu donné, de contacts entre ces langues. Ce premier appel à propositions a permis de soutenir des études sur des thèmes aussi divers que le palikur (langue de Guyane), les pratiques linguistiques des Grenoblois, les pratiques d'adolescents parlant le turc et le français, le répertoire verbal des enfants en situation de jeux, le français de Marseille, etc.

Observer les pratiques linguistiques: enquêtes et corpus.

L'intérêt d'une approche sociolinguistique de la langue réside dans la description et donc la prise en compte non pas d'une langue figée dans une norme idéale et standardisée, mais d'une langue

attestée, d'une langue telle qu'elle est utilisée par ceux qui la parlent, d'une langue "pratiquée". Il est alors possible de mettre en évidence les différentes

La chanson de Maître Amboise in *Mireille, Burnand, 1884. Exposition MNATP/BNF: Mireille. Le chef d'œuvre de Mistral dans l'histoire littéraire et dans son cadre provençal. Cliché : Service historique. Musée national des arts et traditions populaires*

variétés et les différentes variations de la langue. Même circonscrites au territoire français ces variations sont importantes. Le lexique, la prononciation et la syntaxe sont, par exemple, soumis à des variations régionales, sociales et culturelles. L'ensemble de ces variations n'est pas un phénomène marginal, bien au contraire ce sont les usages et les pratiques qui constituent la langue. Pour les sociolinguistes, la linguistique doit prendre en compte l'hétérogénéité de la langue et doit donc décrire toutes les formes de variations qui ne sont pas d'ordre strictement individuel. A la suite des travaux de William Labov, il a été démontré qu'il existe une variation sociale et une variation stylistique mais également une variation inhérente chez un même locuteur. L'analyse de ces variations est intéressante pour la contribution qu'elle peut apporter à l'étude des structures de la langue et du changement linguistique. Ainsi, grâce à l'observation directe, les signes du changement linguistique ont pu être repérés avant même qu'ils n'apparaissent à la conscience des locuteurs (c'est le cas par exemple de la liaison sans enchaînement décrite par Pierre Encrevé il y déjà une vingtaine d'années, alors que ce phénomène linguistique échappait à tous, linguistes compris). De plus l'analyse sociolinguistique permet d'étudier la structuration sociale des variations et donne donc des informations sur les groupes sociaux, acteurs de la diffusion de l'innovation linguistique et indicateurs de la direction du changement linguistique.
L'analyse et la description des pratiques linguistiques imposent de s'appuyer sur des données attestées recueillies de façons systématiques avec une méthodologie d'enquête sociologiquement contrôlée depuis le choix du terrain, la construction de l'échantillon jusqu'à l'étude qualitative et quantitative des données. L'observation et l'analyse ne sont donc possibles qu'avec la réalisation d'enquêtes complétées par un travail sociologique sur la situation d'enquête.

La volonté d'observer et de décrire les pratiques linguistiques existe depuis longtemps en France (même si les linguistes n'ont pas toujours prêté une oreille attentive et scientifique à celle-ci). Il faut reconnaître que la tâche est plutôt complexe et demande d'avoir recours à une méthodologie très contraignante pour qui veut décrire scientifiquement les pratiques linguistiques : l'enquête de terrain. Les enquêtes déclaratives apportent des informations utiles, notamment sur les représentations psychologiques liées à la langue, mais ne permettent pas une réelle description des pratiques et des expériences des individus. Seules les enquêtes et les enregistrements en situation procurent des corpus apportant des données attestées et incontestables sur des phénomènes linguistiques qui le plus souvent échappent à la conscience des locuteurs eux-mêmes.

Les premières enquêtes linguistiques importantes en France ont été réalisées par correspondance dans les années qui suivirent la Révolution. Tout au long du dix-neuvième siècle des séries d'investigations locales ont eu lieu dans le but de recueillir les patois avant que l'unification linguistique ne les fasse disparaître. A la fin du dix-neuvième siècle, Gaston Paris plaida pour une enquête qui concernerait toutes les communes du territoire français afin de relever les variations géographiques, supposées fort nombreuses, mais surtout dans le but de recueillir des informations sur le contact des patois avec le français et d'accéder ainsi à une meilleur connaissance de celui-ci. Au même moment l'avancée tech-

La cueillette in *Mireille, Burnand, 1884. Exposition MNATP/BNF: Mireille. Le chef d'œuvre de Mistral dans l'histoire littéraire et dans son cadre provençal. Cliché : Service historique. Musée national des arts et traditions populaires*

nologique permit de graver sur rouleau le premier échantillon du français commun. La possibilité fort enthousiasmante de conserver le son ouvrit alors de nouvelles perspectives. Les enquêtes qui suivirent eurent pour but principal de dresser un atlas linguistique de la France en prenant en compte les variétés régionales le plus souvent au détriment des variations sociales et culturelles de la langue. Depuis 1945, les enquêtes linguistiques correspondent à trois motivations le plus souvent distinctes: les atlas linguistiques, les méthodes pédagogiques d'enseignement du français standard et la sociolinguistique.

Dans le cadre de l'Observatoire des pratiques linguistiques récemment créé, la mise en réseau des équipes de recherche

Titre	***L'observatoire des pratiques linguistiques***
Type	Article de vulgarisation
Editeur	MCC
Année	2003
Référence	Baude, O., Sibille, J. (2003) «L'observatoire des pratiques linguistiques», in *Culture & Recherche*, n° 96 Ministère de la Culture et de la communication, Paris. Pp 7-8.

culture & recherche

n°96
mai-juin 2003

Ministère de la culture et de la communication - Mission de la recherche et de la technologie - 3, rue de Valois 75033 Paris cedex 01 - Tél. : 01 40 15 80 45

SOMMAIRE

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Ministère Culture Communication

La langue française et les langues de France

Les enjeux de la recherche en sociolinguistique

La plupart des analyses linguistiques du français se basent encore sur un ensemble de données assez hétéroclites qui en obère fortement la pertinence. Faute d'enquêtes de grande envergure et d'observations précises des usages linguistiques de locuteurs concrets saisis en situations réelles d'interlocution, on se fonde en général sur un ensemble de faits établis par la tradition grammaticale, tenus pour avérés et intangibles. Ces faits, jamais réellement attestés, renvoient à un imaginaire de la langue qui n'est autre que la norme et la tradition académique telle qu'elle s'est progressivement dégagée et construite au fil des siècles. Ce français de référence, pour utiliser l'appellation neutre forgée par Yves-Charles Morin, n'a que peu de rapports avec les usages réels attestés par les locuteurs francophones. Il est marqué par un conservatisme natif encore renforcé par la prégnance de la norme orthographique.

L'analyse des pratiques linguistiques suppose le recours systématique à l'enquête et à l'observation in situ, telle est donc la thèse centrale défendue par la sociolinguistique. Or, précisément parce qu'elle est toujours prise dans des interactions sociales spécifiques impossibles à suspendre, la langue ne se livre pas immédiatement à l'observateur et un ensemble de techniques sociolinguistiques particulières doivent être mises en œuvre pour l'atteindre et en documenter la phénoménologie.

La deuxième dimension centrale de la recherche en sociolinguistique, singulièrement de la linguistique variationniste, concerne l'attention portée aux variations de tous types. Dès que l'on observe la langue dans son contexte social et écologique, elle apparaît en effet comme un phénomène social profondément hétérogène, instable et variable, variable dans l'espace géographique comme dans l'espace social, dans l'espace historique comme dans l'espace stylistique. Comme l'a souvent défendu William Labov, pour une langue lien social de sociétés humaines toujours extrêmement structurées, divisées, hiérarchisées et en constante évolution, ce serait la stabilité, l'homogénéité et la constance dans le temps qui seraient surprenantes et en définitive contre-productives.
Ces variations d'usages mettent en doute l'existence d'une langue qui serait le français. Les deux forces antagoniques à l'œuvre dans l'espace géographique, baptisées par Ferdinand de Saussure « esprit de clocher » et « force d'intercourse », morcellent la langue livrant un camaïeu dialectal sans délimitations internes précises. Dans l'espace historique, le changement, à courte et longue échelle, générationnel et trans-générationnel travaille continûment la langue que viennent encore perturber les contacts, emprunts et échanges interculturels. Dans l'espace social, la stratification est aussi linguistique et les effets de champ, de marché et de distribution inégale des différentes espèces de capitaux qui sont au cœur de la sociologie des biens symboliques de Pierre Bourdieu induisent une hétérogénéité interne aux systèmes linguistiques eux-mêmes. Dans l'espace stylistique, la variété des situations écologiques, la diversité des relations sociales d'interlocution, encore redoublée par celle des contenus informationnels et des situations pragmatiques induisent de nouvelles stratifications et fragmentations linguistiques. Enfin, la grammaire intériorisée sous forme de compétence sociale et linguistique, bien loin d'apparaître stable et homogène, est aujourd'hui unanimement reconnue comme labile, instable, intrinsèquement variable et hétérogène parce que largement sous spécifiée et sous déterminée, relativement floue dans ses inscriptions cognitives, et donc profondément plastique, déformable, adaptable et évolutive.
Pour autant, la langue n'est pas une masse informe, livrée au chaos et à l'absence de régulations tant internes qu'externes. La variation et l'hétérogénéité sociolinguistiques sont limitées et contraintes par les nécessités de l'intercompréhension et par les dynamiques sociales qui poussent à des homogénéisations au moins partielles. Aux forces centrifuges qui émiettent les communautés linguistiques s'opposent ainsi d'autres forces sociales et linguistiques, centripètes, qui favorisent les stabilisations et les standardisations, encore renforcées par un apprentissage scolaire de longue durée de la norme académique et de l'arbitraire figé de la codification orthographique.

Sans qu'il soit nécessaire de la souligner plus avant, on mesure toute la distance qui sépare cette mise en perspective sociolinguistique et variationniste du tableau compassé que livre la tradition grammaticale. Mais au-delà du ressourcement empirique et phénoménal, au-delà des conséquences sur la théorie des grammaires et sur leurs modes d'implémentation mentale, au-delà même de la définition d'une perspective socio-cognitive nouvelle, cette analyse de la langue in situ et cette observation directe des pratiques ont nécessairement des conséquences au plan des politiques linguistiques et culturelles. Conséquences patrimoniales pour ce qui concerne la description et la conservation des variantes locales, culturelles pour ce qui concerne l'appréciation de la diversification des pratiques, de leur transmission dans le cadre familial et extra familial, culturelles également pour ce qui concerne les dynamiques identitaires, communautaires et leur diffusion, conséquences de politique linguistique enfin pour ce qui concerne la mise à niveau des standards officiels, la sensibilité aux variantes socio-dialectales, la planification linguistique et la définition précise de politiques de remédiation sociale et culturelle. Tels sont les enjeux multiples de la recherche en sociolinguistique, telles sont les préoccupations de l'Observatoire des pratiques linguistiques de la Délégation générale à la langue française et aux langues de France qui les soutient.

Bernard Laks
Université de Paris X,
directeur du laboratoire Modèles dynamiques corpus (CNRS)
membre du conseil scientifique de l'Observatoire des pratiques linguistiques

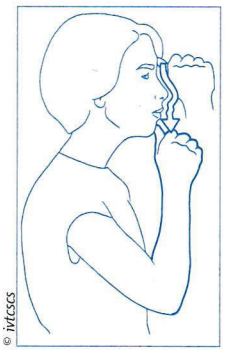
> Il y en a un une fois, je ne sais plus à quel propos, au cours d'une bouffe quelconque, il a dit en parlant de mon mari, mais tout fort, il a dit, « mais il n'a pas inventé le fil à couper le beurre » et je pense que ça, ça lui a fait beaucoup de mal.
>
> (P. Bourdieu, *La misère du monde*, 1993 [1991], p. 669)
> (Français familier *in* : http://www.inalf.fr/richlex/Richlex.htm)

L'Observatoire des pratiques linguistiques

Usage du français, langues de France, « langue des banlieues », enseignement des langues, plurilinguisme... : les questions relatives aux pratiques linguistiques réelles sont nombreuses et soulèvent des débats qui agitent l'actualité. Pour pouvoir répondre à ces questions, il est nécessaire de connaître la situation de ces pratiques dans leur ensemble et de s'appuyer sur un savoir scientifique.

L'Observatoire des pratiques linguistiques a été créé en 1999 auprès de la Délégation générale à la langue française et aux langues de France (DGLFLF) avec pour objectifs de recenser, de développer et de rendre disponibles les savoirs relatifs à la situation linguistique en France ; ceci afin que soit mieux connu un patrimoine linguistique commun, constitué de l'ensemble des langues et des variétés de langues parlées en France, qui participent de la diversité culturelle nationale, et afin également d'apporter des informations utiles pour l'élaboration des politiques culturelles, éducatives, sociales... L'Observatoire n'effectue pas directement des recherches mais œuvre en impulsant, en soutenant et en coordonnant des programmes de recherche sur des sujets qui intéressent non seulement le ministère de la Culture et de la Communication, mais aussi, plus largement, les pouvoirs publics, les élus, les décideurs, les acteurs culturels...

Le champ de l'observation est celui des pratiques linguistiques actuelles sur le territoire français. Sont donc concernés aussi bien le français et ses variétés que l'ensemble des langues utilisées en France, que ce soit les langues régionales de France métropolitaine et d'outre-mer, ou les langues issues de courants migratoires récents. Les données rassemblées proviennent d'enquêtes de terrain, et rendent compte des expériences langagières réelles des individus et des groupes. Elles portent aussi bien sur l'hétérogénéité des usages (variations géographiques ou sociales), que sur les questions de contact de langues, de transmission ou d'acquisition ; sur l'évolution des usages réels et de la norme, sur les modalités du plurilinguisme comme sur les évolutions en cours (féminisation, déplacement des normes, effets des supports de l'écrit sur la langue…).



« un musée » *(langue des signes française).* In : La langue des signes, *t2.* Dictionnaire élémentaire bilingue.

Gallo. Bretagne. http://www.ac-rennes.fr
gnolle (liseron), seü (sureau), castilles (groseilles), brou (lierre), caeüdes (noisetiers)

Le rôle de l'observatoire est aussi de favoriser la collaboration et l'organisation en réseau des équipes et centres de recherche qui travaillent sur les pratiques linguistiques sur l'ensemble du territoire et dans les pays francophones.

L'Observatoire s'est adjoint un comité scientifique qui détermine, avec la DGLFLF, les orientations et les axes de travail. La première tâche a été d'entreprendre un inventaire des centres de recherche qui travaillent sur les pratiques linguistiques actuelles ainsi que des recherches et des travaux en cours dans ce domaine.

Dorne un dischel stäche sehr, falschi zùnge noch viel meh
Epines et chardons piquent fort, mais mauvaises langues bien plus encore. *Alsacien*

Pour ce qui est des recherches proprement dites, l'Observatoire procède soit en lançant des appels à propositions sur des thèmes préalablement définis, soit en sélectionnant des projets correspondant aux orientations déterminées par le comité scientifique, soit, le cas échéant, en retenant des projets qui lui sont soumis spontanément. À l'heure actuelle, trois appels à propositions ont été lancés.

Des appels à propositions

Hétérogénéité des pratiques linguistiques

Ce premier appel à propositions a porté sur la description et l'analyse de l'hétérogénéité des pratiques linguistiques de l'ensemble du territoire national, les travaux devant concerner l'analyse des variations et la description des variétés du français, l'analyse des variations et la description des variétés des autres langues utilisées en métropole et dans les DOM-TOM, les situations, nettement circonscrites à un lieu donné, de contacts entre ces langues. Parmi les 32 projets présentés, 20 ont été retenus qui ont fait l'objet de subventions, 16 portant sur les variétés et variations du français, 1 sur les langues régionales (langues amérindiennes de Guyane), 3 sur les contacts entre les langues (corse, turc, contact de langues en région parisienne).

Corse. http://www.ac-corse
parolla, parulla, parola, paroda (parole), bremu, corda di i nassi, capuzzu, chjapuzzu, brimu, calomu, calamentu, calomu (le cordage qui relie les nass

Ces premières actions ont suscité un vif intérêt dans les milieux universitaires et les administrations qui témoigne du bien-fondé et de l'utilité de ces entreprises même si l'utilisation des résultats des études pointues ne va pas sans poser quelques difficultés. De septembre à décembre 1999, la DGLFLF a en effet entrepris une consultation des départements ministériels chargés de la recherche, de l'emploi, de l'action sociale et de la ville afin de leur présenter l'Observatoire des pratiques linguistiques et d'examiner avec eux la question de l'utilisation des observations de la recherche dans l'élaboration de politiques publiques, culturelles, sociales ou éducatives. Ces consultations ont fait ressortir la nécessité de créer un lien spécifique entre la recherche et l'administration pour tirer tout le bénéfice escompté des études entreprises. Par ailleurs, sur le plan de la thématique, le centre d'intérêt le plus fréquemment cité a été la problématique qui sous-tend ce qu'on appelle désormais la langue de jeunes.

Observation du contact linguistique

Le deuxième appel à propositions a porté sur l'observation du contact linguistique dans une situation géographique et sociale précise, le contact pouvant être aussi bien celui du français et d'une autre langue que celui de variétés ou variantes du français ou encore celui de l'écrit ou/et de l'oral, les situations de contacts étudiées pouvant concerner notamment des groupes de locuteurs d'âge scolaire particulièrement en milieu urbain.

À la suite de cet appel à propositions, 30 projets ont été présentés par 27 centres de recherche. Sur proposition de la commission scientifique réunie en mai 2000, 16 projets ont été retenus. Une très grande majorité de projets concerne des jeunes locuteurs scolarisés. 9 des 14 projets retenus relèvent de la problématique des langues de France : *Pratiques linguistiques d'élèves de zones suburbaines en Bretagne Gallo ; Contacts entre gascon, aragonais, français et castillan ; Vécu et représentations linguistiques après une scolarisation en « calandreta » (école occitane d'immersion) ; Pratiques langagières dans les familles issues de l'immigration ; Parlers jeunes à la Réunion ; Pratiques linguistiques et représentations en Alsace ; Picard, français, immigration ; Langues de Guyane française.*

Si ni artrouve pi, na écrire
Si on ne se revoit pas d'ici-là, on s'écrira. Créole réunionnais (http://www.ac-reunion.

Transmission familiale et acquisition non didactique des langues

Le troisième appel à propositions a porté sur la transmission familiale et l'acquisition non didactique des langues. 6 projets sur les 11 proposés ont été retenus :

Transmission de langues entre pairs dans les cours d'école ; Analyse des productions linguistiques en créole des enfants de grande section de maternelle ; Transmission du créole mères/enfants et acquisition non didactique dans les établissements scolaires de la Réunion ; Étude comparative de la transmission familiale et de l'acquisition non didactique du vietnamien dans les communautés niçoise et lyonnaise ; Transmission familiale et acquisition non didactique des langues, arabe maghrébin : dynamisation de la transmission familiale par la visibilité dans le domaine public (musique, comédie, reconnaissance dans le système scolaire) ; Transmission des langues : pratiques linguistiques dans les familles bilingues d'origine étrangère.

Titre	***« Les corpus oraux entre science et patrimoine », l'expérience de l'Observatoire des pratiques linguistiques***
Type	Actes, colloques du GEREC 2004
Editeur	PUG
Année	2004
Référence	Baude, O. (2004) « Les corpus oraux entre science et patrimoine », l'expérience de l'Observatoire des pratiques linguistiques », Colloque du GEREC, PUG ed, Grenoble

Les corpus oraux entre science et patrimoine

L'expérience de *l'observatoire des pratiques linguistiques*

Les premières initiatives de constitution de "corpus oraux" (recueils ordonnés d'enregistrements à des fins scientifiques) datent à peine d'un siècle et les nouvelles technologies permettant le traitement informatique de données sonores sont à l'aube de leur développement. Il serait cependant erroné de réduire le peu de reconnaissance que l'on concède aux corpus oraux à un problème exclusivement technique (de diffusion). En effet, c'est bien plus du statut de la voix et de la langue orale[1] dans le monde social dont il s'agit.

Depuis quelques mois, le conseil scientifique de l'Observatoire des pratiques linguistiques de la Délégation Générale à la Langue Française et aux Langues de France[2] développe une initiative en faveur *de la constitution, l'exploitation et la diffusion des corpus oraux en France*. Cette initiative en est à ses prémices, il ne s'agit donc pas de proposer un bilan, mais d'apporter simplement quelques réflexions sur une action qui mêle la *diffusion* de recherches et *la mise en public* d'objets scientifiques.

1 *L'oral*, du champ scientifique à l'espace social: la disparition du *locuteur*

L'intérêt porté à "ce que parlent" les Français est lié historiquement aux marchés linguistiques dès la première enquête de l'Abbé Grégoire et son *Rapport sur la nécessité et les moyens d'anéantir les patois et d'universaliser l'usage de la langue française* (1794). Toutefois, la constitution du champ scientifique va complexifier cette approche. Ainsi, la dialectologie se restreindra à un objectif de description des patois et des dialectes avant que l'unification linguistique ne les fasse disparaître.

Par la suite, la méthodologie des Atlas linguistiques qui consiste à reproduire sur fonds de carte des variétés régionales en disparition (par simples isoglosses sans prendre en compte la notion de système linguistique seul accès à la langue et donc à l'identité du témoin)[3] exclura le locuteur et son espace social au profit d'une représentation géographique. Pourtant dès 1911, Ferdinand Brunot avait créé à la Sorbonne les Archives de la Parole, première initiative de publicisation d'enregistrements de variétés régionales et de *français parlé*, modifiant à la fois le champ scientifique et l'espace publique soumis à la norme unificatrice. Cette ambition ne survécut pas à la guerre, et l'on abandonna l'idée "*d'étude, d'archivage et d'analyse de parlers d'hommes et de femmes parlant comme à l'auberge ou à la fontaine*"[4] pour se consacrer principalement à la conservation d'une culture folklorique. En 1938 la création de la Phonothèque nationale permettra à Roger Dèvigne de développer les "croisières folkloriques" pour constituer une *Encyclopédie nationale des parlers, patois et vieux chants de France*. La phonothèque intégrée par décret en 1977 au sein d'un département spécialisé de la Bibliothèque nationale s'orientera vers une activité de recueil[5], et sera en 1995, totalement incorporée au département de l'audiovisuel de la BNF.

La linguistique de l'oral ne saurait se réduire à la dialectologie. Si cette dernière s'est construite en réaction à une norme unificatrice (le français), la linguistique de l'oral s'opposera à l'académisme normatif prépondérant à la linguistique (de l'écrit).

Cette contrainte normative trouve un renfort historique dans la réception des concepts fondamentaux de la linguistique moderne élaborés par Ferdinand de Saussure. La lecture des dichotomies fondatrices synchronie/diachronie, signifiant/signifié et langue/parole par les linguistes provoqueront le rejet de la description des variations en générale et des productions orales en particulier par une science vouée à la recherche d'invariants structurés en système. En France, seule la linguistique variationniste essaiera de refuser cette séparation entre invariants et variations, en proposant une méthodologie de l'enquête et du recueil de données *attestées* et *situées*. Cette approche, la plus rigoureuse et novatrice d'une sociolinguistique qui souhaite reconstruire la

[1] Plus exactement des *manifestations orales de la langue*.
[2] DGLFLF, direction du ministère de la Culture et de la Communication
[3] Laks 2003
[4] Callas 1992, Vecken 1984
[5] convention avec le CNRS en 1979 pour l'archivage des données des Atlas

linguistique moderne sera théorisée principalement par Encrevé[6], restera fortement dominée dans le champ, tout comme les initiatives françaises de publicisation de grands corpus de référence (le français fondamental, le corpus du Groupe Aixois de Recherche en Syntaxe et les corpus d'applications).
Ainsi, le français et surtout sa forme normative (l'écrit littéraire) apparaît comme le seul objet scientifique légitime. Cela sera aussi le seul objet légitime de l'espace social, reconnu comme capital culturel par excellence.

2 Frémissements et bouleversements du champ

Depuis quelques années la situation se modifie tant dans le champ scientifique que dans l'espace social.

2.1 Une politique linguistique fondée sur la diffusion scientifique

Le premier frémissement provient de l'élaboration de nouvelles politiques linguistiques. La DGLF dont *le rôle est de contribuer à la mise en œuvre de la politique linguistique* est transformée, en 2001, en DGLF***LF*** (*et aux Langues de France*), avec une volonté de développer deux axes. L'un est traditionnel (promotion du français comme langue et norme de l'état et de la république, loi Toubon de 1994, commission de terminologie), l'autre est novateur (prise en compte des variétés et des variations linguistiques dans leurs usages sociaux, réforme de l'orthographe de 1990, féminisation, charte européenne de protection des langues régionales et minoritaires, assises des langues de France,…). Autre indice, l'installation en 1999 de l'Observatoire des pratiques linguistiques dont la mission est *de recenser et de rendre disponible les savoirs issus de la recherche scientifique relatifs à la situation linguistique en France*.
L'Observatoire, rattaché à la Mission de langues de France est doté d'un conseil scientifique qui propose des actions de diffusion d'informations scientifiques auprès de deux cercles distincts : les acteurs des politiques linguistiques (direction des ministères de la culture, de l'éducation, les collectivités territoriales,…) et un public plus large (acteurs scientifiques, éducatifs, sociaux, culturels,…). L'observatoire valorise la recherche fondée sur les données attestées et situées par la diffusion de "synthèses vulgarisées" (en ligne et par un bulletin) et par l'aide financière à certains projets de recherche (appel d'offre).
Depuis peu, le ministère de la Culture connaît un autre bouleversement : la volonté de rendre accessible des masses de données reconnues comme patrimoine. Ainsi, pour l'oral, l'INA a commencé la numérisation de 60 ans de radios (et 50 ans de télé) soit plus de 700 000 heures d'archives sonores et a annoncé, en février 2004 le lancement, de *la première banque mondiale d'archives numérisées*.

2.2 L'ère de la mise en public des masses de données

La notion du traitement informatique de masse de données est également au cœur du champ scientifique avec l'émergence du Traitement Automatique du Langage et *des linguistiques de corpus*[7].
Les requêtes complexes sur un corpus très vaste (plusieurs centaines de millions de mots) apportent une méthodologie très puissante mobilisée par les linguistiques de l'écrit. Plus récemment, la numérisation de données sonores permet d'envisager un engouement similaire. Il ne s'agit donc plus d'enregistrements isolés de témoignages sonores mais de la constitution de bases de données de grande ampleur. Les outils se développent, qu'il s'agisse de numérisation, de transcription, d'annotation et de balisage, mais aussi de conservation et de diffusion, ce qui modifie la notion d'accès aux données. En effet, une tendance très forte se dessine autour de *l'interopérabilité*. La normalisation permet de concevoir les corpus comme des objets scientifiques

[6] Encrevé 1976, 1983, 1992.
[7] Habert Les linguistiques de corpus, 1998.

ré-exploitables (le fait le plus marquant est la création du métalangage XML[8] qui facilite l'échange des outils et des documents, et les initiatives internationales de normalisation comme la TEI[9]).
Cette interopérabilité est aussi à la pointe des réflexions sur la conservation et la mise à disposition des politiques culturelles et patrimoniales. La Commission Européenne a financé le projet Minerva (Réseau ministériel pour la valorisation des activités de numérisation) et le Groupe des représentants nationaux (GRN) des Etats membres de l'UE sur la numérisation du patrimoine culturel a adopté en 2001 les principes de "Lund" pour faciliter l'accès aux ressources numérisées. Cette même année la France créait le *comité de concertation pour les données en sciences humaines et sociales* auprès des ministres chargés de l'économie, de l'emploi, de l'éducation nationale et de la recherche, dont le rôle est de promouvoir la diffusion des données ayant un intérêt scientifique.

3 L'initiative de l'Observatoire

Le traitement automatique des corpus oraux apparaît alors comme l'occasion d'apporter la reconnaissance souhaitée à l'hétérogénéité des pratiques linguistiques et à la légitimité des corpus oraux comme objet scientifique et patrimonial. Le conseil scientifique de l'Observatoire ne pouvait être insensible à cet enjeu[10] et fut donc à l'initiative de la création d'un comité de pilotage *pour la constitution et l'exploitation des corpus oraux*. Outre les membres du conseil scientifique, celui-ci est composé de *conservateurs* dépendant du ministère de la culture (INA, Archives, BNF) de représentants des institutions scientifiques (CNRS), de linguistes avec le statut d'experts et de juristes spécialistes de la propriété intellectuelle et du droit de la science.
Le travail de réflexion de ce comité est en cours et je ne peux donc que mentionner l'ébauche de celle-ci qui, en tout état de cause, porte sur la valorisation des corpus oraux et la mise en public de cette légitimité scientifique et patrimoniale. Je me bornerai à présenter trois aspects de cette réflexion: les aspects juridiques, la normalisation et les outils techniques.

3.1 Les aspects juridiques.

Le comité de pilotage a créé un premier groupe de travail composé de linguistes, de conservateurs et de juristes, sur *les aspects juridiques*. Si ceux-ci forment le premier verrou évoqué par les chercheurs et les conservateurs c'est justement parce que les nouveaux outils de traitement et de diffusion de ces données modifient considérablement la donne. Ainsi la diffusion des enregistrements implique une procédure d'anonymisation, qui ne peut se réduire, à l'ère du croisement des bases de données, à une opération simpliste comme le bippage des noms propres. De même l'interopérabilité rend imprévisible *les finalités* et complique donc le recueil de consentement du témoin. Enfin la diffusion et l'utilisation industrielle posent la question de la propriété de la voix. Il faut donc reconnaître un *auteur*, un *propriétaire*, un *responsable*, aux données orales. Le comité de pilotage a ainsi décidé de rédiger un *guide des bonnes pratiques* dont l'objectif n'est pas de fournir des solutions clés en mains, mais de construire au sein de la communauté des chercheurs une éthique qui oblige à repenser les liens entre *données* et *donneur* et à reconnaître au locuteur un *corps socialement situé.*

[8] eXtensible Markup Language www.w3c.org
[9] Text Encoding Initiative
[10] Le président du conseil scientifique, P. Encrevé, à l'origine de la linguistique variationniste est le principal acteur des politiques linguistiques novatrices en France (membre du cabinet du 1er Ministre M. Rocard en charge de la langue, puis avec les mêmes fonctions au cabinet de C. Trautman Ministre de la culture) Le second membre B. Laks, issu lui aussi du champ de la linguistique variationniste et de la phonologie cognitive est directeur d'un grand laboratoire de recherche universitaire et coresponsable du projet actuel sur la phonologie du français contemporain (grand corpus oral).

3.2 Normalisation et métadonnées : pour des données situées

Cette reconnaissance passe aussi par les enjeux d'une *normalisation* dans un but d'interopérabilité. Les normes internationales de stockage et d'échange de données, en cours de constitution, définissent la structure même des objets scientifiques. Techniquement les standards actuels séparent la représentation physique et logique des documents (les données et les métadonnées). Tout document XML comporte l'identification des éléments possibles et leurs relations possibles (Définition de Type de Document) *et* les données identifiées selon cette DTD. C'est alors la notion même de données brutes qui est redéfinie. Ainsi la TEI rend obligatoire la constitution d'un header (en-tête) en début de corpus qui recense les informations sur le contexte de production des données. Un certain nombre d'informations sont donc fournies sur le locuteur avec la possibilité de reconnaître celui-ci comme être *socialement situé*. Tout l'enjeu réside ici : quels vont être les choix de standardisation? Et qui participent à ces choix? Actuellement ces normes sont soumises à deux forces, celle des utilisateurs (aucune recommandation ne peut se pérenniser sans l'assentiment des utilisateurs, c'est encore vrai dans le champ scientifique) et celles des producteurs (W3C Word Wide Web consortium, ISO–TC37 pour les ressources linguistiques,…).

3.3 Diffusion de l'oral et transcription

L'exemple de la transcription des enregistrements permet également d'expliciter les enjeux de l'interopérabilité. Le travail scientifique sur la langue parlée passe toujours par la *transcription* de la parole. Ce n'est pas le moindre des paradoxes pour des linguistes qui souhaitent travailler sur l'oral que d'être confronté au plus fort des effets de normalisation de l'écrit : l'orthographe. Pendant longtemps les transcripteurs devaient choisir entre l'orthographe normée, un alphabet phonétique, des conventions pour les formes orales et/ou l'ajout d'informations supplémentaires (prosodie,…). Ces choix fondamentalement liés aux enjeux théoriques de la recherche ont eu pour conséquence de rendre confidentiels les corpus transcris avec le plus de finesse (quand les marques du locuteur situé étaient présentes) au profit de corpus transformés pour les besoins des outils de traitement automatique de l'écrit et de la norme sociale.
Les logiciels[11] développés pour l'aide à la transcription sont fondés sur l'interopérabilité et ils risquent bien de modifier le champ scientifique et par delà même l'espace social.
Premièrement ceux-ci reposent sur l'alignement du signal reliant définitivement la *transcription* à la *voix* en un unique objet. Deuxièmement, la majorité de ces logiciels est prévue pour la multitranscription et proposent donc une granularité modulable et une représentation hétérogène (une première ligne de transcription peut contenir une transcription orthographique, une seconde phonétique, une troisième être aménagée pour une recherche spécifique, etc.).
C'est l'impact même de la représentation graphique et de sa norme qui est estompé et il devient bien plus difficile d'oublier que derrière les mots il y a une *voix* et un *locuteur*.

Ces trois aspects de la publicisation des corpus oraux offrent l'opportunité de replacer la *nature sociale* de la langue au centre de la linguistique et de rappeler par la même à la communauté scientifique comme à l'espace social qu'il n'y a pas de langue sans locuteur. Considérer l'hétérogénéité des usages de l'oral comme inhérent à la constitution de l'objet d'étude comme du patrimoine doit permettre de respecter les deux principes qui définissent, selon Labov, la *dette du chercheur* : la responsabilité sociale envers le terrain et le rendu scientifique à la communauté observée.

[11] Praat, Tasx, Anvil, etc.

Olivier Baude
MCF Université d'Orléans, **C**entre **O**rléanais de **R**echerche en **A**nthropologie et en **L**inguistique
Celith, EHESS, Paris.
DGLFLF

Les corpus oraux entre science et patrimoine.
L'expérience de *l'observatoire des pratiques linguistiques*

mots-clés: corpus oraux, interopérabilité, politique linguistique, publicisation, normalisation.

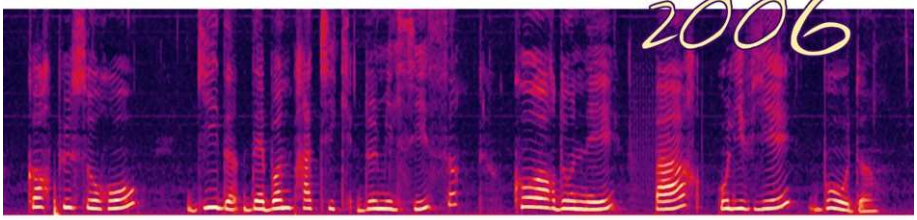
Bibliographie:

Baude O. 1999, *L'observation des pratiques linguistiques en France*, Culture et Recherche n°75, Ministère de la Culture et de la Communication.
Biber D. 1988, *Variation across speech and writing*, Cambridge University Press.
Bergounioux G. (dir) 1992, *Enquêtes, corpus et témoins*, Langue Française n°93, Larousse.
Blanche-Benveniste C. 2000, *Corpus de français parlé*, in Corpus méthodologie et applications linguistiques, édité par M. Bilger, Champion.
Calas M-F. 2002, *Le statut documentaire de la source orale*, in de la source à l'archive, actes des journées d'études, AFAS.
Encrevé P. 1976, *Présentation*, in Sociolinguistique, W. Labov, Minuit
Encrevé P. 2001, La langue de la République, in La République, Pouvoirs n°100, Seuil, Paris
Gibbon D, Moore R, Winski R., 1997*, Handbook of standards and ressources for spoken language ressources*, Mouton de Gruyter. New-York.
Habert B. 2000, *Des corpus représentatifs: de quoi, pourquoi, comment?* In Cahiers de l'Université de Perpignan, N° 31, Presses universitaires de Perpignan.
Laks B. 2003, *Les grandes enquêtes phonologiques en France*, in la prononciation du français dans sa variation, La tribune internationale des langues vivantes n°33.

Titre	***Corpus oraux, Guide des bonnes pratiques 2006***
Type	Livre
Editeur	CNRS éditions et PUO
Année	2006
Référence	Baude, O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

CORPUS ORAUX

Guide des bonnes pratiques 2006



coordonné par **Olivier BAUDE**

PRESSES UNIVERSITAIRES ORLÉANS

CNRS EDITIONS

CORPUS ORAUX

Guide des bonnes pratiques 2006

CORPUS ORAUX

Guide des bonnes pratiques 2006

coordonné par **Olivier BAUDE**

PRESSES UNIVERSITAIRES ORLÉANS

CNRS EDITIONS

Délégation générale à la langue française et aux langues de France
6, rue des Pyramides75001 PARIS
http://www.dglflf.culture.gouv.fr

ISBN 2-271-06425-2 (CNRS ÉDITIONS)
ISBN 2-913454-30-5 (PUO)
EAN 9782271064 257 (CNRS ÉDITIONS)
EAN 9782913454 309 (PUO)

Cet ouvrage est le résultat des travaux d'un groupe de réflexion
réuni autour d'Isabelle **de LAMBERTERIE**.
Il a été coordonné par Olivier **BAUDE**.

Olivier **BAUDE** (*DGLFLF et CORAL – Université d'Orléans*)
Claire **BLANCHE-BENVENISTE** (*EPHE et Université de Provence*)
Marie-France **CALAS** (*DMF*)
Paul **CAPPEAU** (*Université de Poitiers*)
Pascal **CORDEREIX** (*BnF)*
Laurence **GOURY** (*CNRS – CELIA*)
Michel **JACOBSON** (*CNRS – LACITO*)
Isabelle **de LAMBERTERIE** (*CNRS-CECOJI*)
Christiane **MARCHELLO-NIZIA** (*CNRS-ILF et ENS-LSH-Lyon*)
Lorenza **MONDADA** (*ICAR, CNRS, Université Lyon2*)

Avec la collaboration de :

Gilles **ADDA** (*pour le collectif COPTE LIMSI-CNRS*), Michel **ALESSIO** (*DGLFLF*), Alain **CAROU** (*BnF*), Ibrahim **COULIBALY** (*CDF – Université de Grenoble*), Valérie **GAME** (*BnF*), Fabrice **MOLLO** (*CNRS-CECOJI*), Michel **RAYNAL** (*INA*), Jean **SIBILLE** (*DGLFLF*), Dominique **THERON** (*BnF*), Luc **VERRIER** (*BnF*).

PRESENTATION DES AUTEURS

OLIVIER BAUDE
Maitre de conférences en sciences du langage à l'Université d'Orléans, membre du Centre Orléanais de Recherche en Anthropologie et Linguistique (EA-3850). Secrétaire du conseil scientifique de l'Observatoire des pratiques linguistiques, Délégation générale à la langue française et aux langues de France.

CLAIRE BLANCHE-BENVENISTE
Professeur émérite, École Pratique des Hautes Études à Paris et à l'Université de Provence. Recherche dans le domaine de la linguistique française : langue écrite et langue parlée, syntaxe, morphologie, constitution de corpus de langue parlée.

MARIE-FRANCE CALAS
Conservateur général du Patrimoine. Inspecteur général des musées, Direction des Musées de France. Spécialiste du domaine sonore, compris comme un vaste domaine pluridisciplinaire incluant l'histoire, la gestion, la conservation et la valorisation des enregistrements parlés, musicaux, des sons de l'environnement, aujourd'hui partie intégrante du patrimoine immatériel.

PASCAL CORDEREIX
Conservateur en chef des bibliothèques. Chef du service des documents sonores au département de l'Audiovisuel de la Bibliothèque nationale de France ; par ailleurs vice-président de l'Association française des détenteurs de documents audiovisuels et sonores (AFAS). L'essentiel de son activité est orienté vers les questions d'archivistique du son.

LAURENCE GOURY
Chargée de Recherche à l'IRD (Institut de Recherche pour le Développement), membre du CELIA (Centre d'Etude des Langues Indigènes d'Amérique), linguistique de terrain et typologie (en particulier langues créoles).

MICHEL JACOBSON
Ingénieur informaticien au laboratoire des « Langues et Civilisations à Tradition orale » du Centre National de la Recherche Scientifique. Co-responsable du programme « Archivage ». Spécialiste de la gestion de corpus oraux.

ISABELLE DE LAMBERTERIE
Directrice de recherche au CNRS, responsable de l'équipe « Normativité et société de l'information » du Centre d'études sur la coopération juridique internationale (CECOJI – UMR 6224), membre du Comité d'éthique du CNRS.

CHRISTIANE MARCHELLO-NIZIA
Professeur en Sciences du langage à l'ENS-LSH (Lyon), Directrice de l'Institut de Linguistique Française (CNRS) : Linguistique historique, histoire du français, théories de l'évolution des langues.

LORENZA MONDADA
Professeur en Sciences du Langage à l'Université Lyon 2 et membre du Laboratoire ICAR (UMR CNRS 5191). Travaille en linguistique interactionnelle sur les corpus de langue parlée en interaction ainsi que sur l'analyse multimodale de corpus vidéo.

PREFACE DE XAVIER NORTH,
DELEGUE GENERAL A LA LANGUE FRANÇAISE ET AUX LANGUES DE FRANCE

Rares sont les moments, dans l'histoire des sciences ou des politiques culturelles, où un ensemble de données brutes et de matériaux incertains se convertit en objet de savoir. La publication de ce guide est de ceux-là, puisqu'il offre à tout chercheur les outils, les « bonnes pratiques » qui lui permettront de procéder à cette métamorphose : la transformation de productions verbales en un corpus oral, susceptible d'être étudié et conservé, et par conséquent de prendre place dans le patrimoine culturel de la nation.

Sans doute les productions langagières dans leur forme *écrite*, fixe et définitive, d'œuvres littéraires ou de documents d'histoire, n'ont-elles jamais cessé d'être au cœur des politiques mises en œuvre par le Ministère de la Culture, qu'il s'agisse du livre ou des archives. Mais ce n'est que tout récemment qu'on s'est avisé de porter intérêt à l'aspect vivant du langage dans son jaillissement spontané, dans son énonciation quotidienne, ordinaire, et dans l'extraordinaire variété de ses parlers… Pour la première fois, s'esquisse ainsi la possibilité de constituer, sur des bases assurées, de véritables archives de la parole. Le progrès des technologies devrait y contribuer.

Un corpus oral, en effet, n'est pas une simple collection d'enregistrements de la parole humaine, c'est un objet « construit » : le traitement des données (numérisation, transcription, indexation) permet non seulement de les conserver, mais les fait passer à un statut nouveau, matière de recherche et de valorisation. Encore faut-il pouvoir s'appuyer sur des prescriptions de méthode, cohérentes et faciles à mettre en œuvre.

Grâce au « Guide des bonnes pratiques », c'est un nouveau et vaste domaine qui s'offre désormais à la curiosité des chercheurs. Par l'intermédiaire de son *Observatoire des pratiques linguistiques,* la Délégation générale à la langue française et aux langues de France a donné l'impulsion de départ, puis s'est attachée à regrouper et à coordonner les énergies et les ressources diverses qui ont produit ce travail, qu'elles proviennent du monde de la recherche, ou des différents horizons du Ministère de la culture concernés par cette initiative.

Assurer le développement des corpus oraux, leur diffusion et leur conservation, c'est aussi rendre accessible, donner à entendre le patrimoine linguistique français dans sa diversité, sa richesse et sa vérité. C'est encore se donner un outil précieux de connaissance des pratiques langagières nécessaires à la définition des politiques de la langue, mais aussi des politiques éducatives et sociales.

Pendant plusieurs mois, cette démarche a rassemblé juristes, linguistes, conservateurs et informaticiens dans la volonté de concilier les chemins nouveaux de la culture et de la recherche avec le respect du droit. C'est le résultat d'un effort commun de pensée que nous présentons aujourd'hui, avec l'espoir qu'il féconde à son tour de nombreux travaux.

PREFACE DE BERNARD MEUNIER, PRESIDENT DU CNRS

L'oral et l'écrit. Ces deux mots possèdent une force évocatrice puissante. Nous pensons à la manière dont les civilisations se sont structurées par les pratiques orales et ensuite par la création d'écritures permettant de mieux transmettre dans l'espace et dans le temps les paroles des uns et des autres.

Mon regard de chercheur sur le rôle respectif de l'oral et de l'écrit dans la diffusion des connaissances scientifiques ne me fait pas oublier que, bien au-delà du rôle primordial de l'écrit, la présentation orale devant ses pairs, ou un large public, est toujours essentielle pour diffuser, convaincre, faire partager des idées. L'oral garde une force de conviction, permettant d'atteindre le plus grand nombre dès lors qu'il peut être enregistré et transmis à l'aide des moyens audiovisuels actuels.

La collecte et l'utilisation des corpus oraux doivent se faire selon le respect de « bonnes pratiques », comme cela se fait pour celles des corpus écrits. Nous savons tous combien une phrase, sortie de son contexte et diffusée sans retenue, peut devenir dangereuse pour son auteur, un groupe de personnes ou une communauté.

Les auteurs de ce remarquable travail ont abordé en profondeur tous les aspects juridiques de la collecte et de l'usage des corpus écrits. Je souhaite que cet ouvrage bénéficie de la meilleure diffusion auprès des acteurs et des utilisateurs des corpus oraux que nous sommes tous, à un moment ou à un autre.

PREFACE DE JEAN-NOËL JEANNENEY, PRESIDENT DE LA BIBLIOTHEQUE NATIONALE DE FRANCE

La Bibliothèque nationale de France est heureuse d'avoir contribué à l'élaboration de ce *Guide*. Elle entretient, en effet, un rapport ancien et étroit avec les langues parlées, leur préservation et leur diffusion. Son département de l'Audiovisuel est l'héritier des *Archives de la Parole* de Ferdinand Brunot, créées dès 1911. Depuis cette date, notre établissement s'est constamment préoccupé d'assurer les meilleures conditions de captation et de conservation des expressions orales de toute sorte, comme de leur diffusion auprès du public le plus large.

Aujourd'hui, les technologies numériques renforcent ce lien historique et scientifique. En matière de conservation, un plan ambitieux de numérisation de nos collections a été engagé dont les documents sonores et audiovisuels bénéficient en particulier. D'autre part la diffusion de ces richesses dans nos murs et à distance est servie par l'essor spectaculaire de notre bibliothèque numérique en ligne, « Gallica », qui permet à chaque internaute, où qu'il se trouve et quel que soit l'objet de sa recherche ou de sa curiosité, d'accéder à ces sources fondamentales de la connaissance.

Fruit d'une confiante collaboration, ce *Guide* témoigne de la complémentarité des savoirs entre linguistes, juristes, conservateurs, informaticiens, techniciens du son et de l'image : je me réjouis que la Bibliothèque nationale de France ait contribué à cette entreprise novatrice et féconde.

Cet ouvrage applique les rectifications de l'orthographe, étudiées par le Conseil supérieur de la langue française (1990), et approuvées par l'Académie française et les instances francophones compétentes.

Les mots signalés par un astérisque renvoient au glossaire juridique situé en fin d'ouvrage.

1 PRESENTATION

1.1 LES OBJECTIFS

Il existe actuellement quantité de recherches fondamentales ou appliquées, qui se fondent sur l'exploitation de « corpus oraux » (collections ordonnées d'enregistrements de productions linguistiques orales et multimodales). Issu de la prise de conscience de linguistes soucieux d'assurer la pérennité des sources et un accès diversifié aux documents oraux qu'ils produisent, ce *Guide des bonnes pratiques* aborde en priorité les « corpus oraux », créés et utilisés par et pour des linguistes. Mais les questions soulevées par la création et l'exploitation documentaire de ces corpus se retrouvent dans de nombreuses disciplines, l'ethnologie, l'anthropologie, la sociologie, la psychologie, la démographie, l'histoire orale notamment utilisent l'enquête orale, le témoignage, l'interview, le récit de vie. Fondé sur la démarche des linguistes, ce *Guide* recoupe toutefois les préoccupations d'autres chercheurs qui utilisent des corpus oraux (par exemple en synthèse et reconnaissance de la parole), même si les besoins spécifiques de ceux-ci ne sont pas systématiquement abordés dans le présent document.

Le *Guide* que nous vous proposons s'est fixé pour premier objectif de fournir les *informations* nécessaires à la constitution de corpus de données orales ou multimodales, et d'offrir des *propositions* utiles concernant non seulement les aspects juridiques, mais aussi les aspects matériels touchant aussi bien à la collecte, qu'à la structuration et la mise en forme des données, qu'à l'exploitation, la communication et la conservation de ces données.

Le second objectif de ce *Guide* est d'aider les chercheurs qui constituent ou enrichissent des corpus oraux à *anticiper* certaines « difficultés à retardement » qui risquent de grever lourdement l'exploitation puis le devenir de leur corpus. Certains choix initiaux, certains manques peuvent révéler leur importance à des étapes ultérieures du processus, alors qu'il est trop tard pour modifier quoi que ce soit.

Le troisième objectif est de favoriser l'émergence de *pratiques communes*, afin de satisfaire aux exigences actuelles de conservation et d'interopérabilité des corpus, d'évaluation, et d'éthique tant dans la constitution que dans l'usage des données.

1.2 LES CONDITIONS D'ELABORATION

Le conseil scientifique de l'Observatoire des pratiques linguistiques (Délégation générale à la langue française et aux langues de France) a souhaité encourager fortement les actions de conservation, de constitution et de valorisation des corpus oraux et multimodaux pour les raisons suivantes :

- permettre la sauvegarde d'un riche patrimoine sur les pratiques linguistiques en France ;
- aider à la constitution de grands corpus de référence, pour la recherche, l'enseignement, les industries de la langue mais aussi le patrimoine ;
- aider au développement des outils informatiques de traitement, d'enrichissement et de valorisation des corpus ;
- favoriser la mise à disposition de ces corpus.

1.3 LES ASPECTS JURIDIQUES

Très vite il est apparu que les aspects juridiques liés à la constitution et à l'utilisation des corpus oraux constituaient un obstacle récurrent et capital.

Ces aspects juridiques concernent principalement les questions de droits moraux et patrimoniaux et de propriété des données, que l'on retrouve à chacune des quatre grandes étapes du travail sur corpus :

- le recueil des données et l'enregistrement (droit à l'image, à la voix, situation d'enquête, autorisations…) ;
- l'utilisation et l'exploitation informatisée des données (archivage, base de données à des fins de recherche, d'enseignement, d'ingénierie…) ;
- la diffusion et la mise en circulation des données (droits, droit de citation, diffusion en ligne…) ;
- la conservation des données.

Au vu du grand nombre de domaines concernés, la DGLFLF a suscité la création d'un comité composé d'experts de diverses disciplines. Ce comité a instauré un *groupe de travail* ayant pour objectif d'aider les équipes de recherche à normaliser les pratiques de recueil et d'exploitation de corpus au regard de la législation en tenant compte de l'ensemble des contraintes liées à la recherche. Le guide que nous présentons ici est le résultat d'une quinzaine de mois de travail de ce groupe.

Ce groupe de travail devait évidemment comprendre des juristes spécialistes du droit de la recherche, mais pas seulement : la nécessité de compétences en termes de constitution des corpus, d'utilisation et de conservation ont conduit à adjoindre aux juristes des linguistes pratiquant de la « linguistique de corpus » et travaillant sur des données orales, des représentants des grandes institutions de conservation patrimoniale (INA, INSI, BnF) et des informaticiens spécialistes en gestion de corpus.

Pour remplir sa mission, ce groupe de travail s'est donné pour objectifs notamment de :

- recenser les pratiques actuelles et définir en priorité les contraintes méthodologiques et théoriques liées à la recherche ;
- diffuser une synthèse sur la législation existante ;
- établir des recommandations ;
- et, le cas échéant, en cas de vide ou de flou, formuler des propositions pour l'élaboration de normes et règles juridiques (notamment européennes).

Il fallait pour cela tout d'abord :

- recenser les domaines juridiques concernés ;
- identifier et quantifier les risques ;
- repérer les réponses existantes ;
- et ensuite construire ces réponses sous la forme d'une série de recommandations de bonnes pratiques (juridiques et éthiques).

Pour cela le groupe a décidé de travailler en étroite relation avec plusieurs équipes témoins pratiquant ou ayant pratiqué le recueil de données orales ou audio-visuelles.

Le but était de parvenir ainsi à une « typologie des situations », et de faire le tour de toutes les pratiques et solutions déjà utilisées, tant en France qu'ailleurs.

1.4 LES AUTRES ASPECTS

Chemin faisant le groupe de travail s'est aperçu que proposer uniquement une série de recommandations ou de solutions de nature juridique ne permettrait pas de répondre de façon satisfaisante aux difficultés rencontrées.

Il est en effet apparu que bien souvent la difficulté ou la solution étaient liées au type de pratique de collecte ou d'utilisation ; que certaines solutions passaient par des voies techniques qui avaient un retentissement sur les données elles-mêmes (anonymisation ou floutage) ; qu'il n'était pas indifférent de résoudre tel ou tel problème juridique à tel moment plutôt qu'à tel autre. Bref, proposer des solutions à des questions juridiques revenait à évoquer le processus même de collecte ou de mise en forme, de transmission ou d'utilisation de ce type de données.

Enfin, au-delà du respect dû aux droits des personnes enregistrées, s'est posée la question du « droit d'auteur » de ce type de données : quels sont les droits des collecteurs de ces données ? Qui en est juridiquement responsable, qui a le droit de les transmettre ? Sous quelles formes ? Comme on le voit, les aspects juridiques liés à la propriété scientifique ou à la responsabilité pénale étaient, eux aussi, indissociables de la pratique de recueil et d'utilisation des données.

Dès lors, ne valait-il pas mieux élargir la compétence du « Guide » projeté, et évoquer non seulement les pratiques juridiques, mais aussi l'ensemble des pratiques mises en jeu dans ce type de corpus ? C'est le choix qui a été fait, car cela permettait de maintenir liés tous les aspects, tels qu'ils le sont dans la réalité.

1.5 LA METHODE

La méthode à laquelle s'est rallié le groupe de travail se caractérise par les traits suivants :

- la conviction qu'il ne faut pas laisser croire qu'il existe des réponses toutes faites à tout type de situation ;
- la volonté de ne pas « brider » les chercheurs (en interdisant certaines pratiques par exemple) ;
- le respect de la méthodologie du chercheur et des contraintes liées à l'observation (les chercheurs souhaitent enregistrer des situations sans que les contraintes, notamment techniques et juridiques, les modifient).
- la nécessité d'élaborer et de rédiger ce guide en mettant en commun les compétences requises aux différentes étapes (linguistes, juristes, conservateurs) ;
- l'affichage d'une démarche fondée sur le respect de la loi et de l'éthique ;
- la nécessité de fournir à travers ce *Guide* un outil d'expertise des risques (repérage, mais aussi évaluation).

1.6 LE CADRE JURIDIQUE FRANÇAIS

Un bon nombre de questions et de solutions tournent autour de la notion de *consentement* des enquêtés mais aussi de la responsabilité des instances *propriétaires*. C'est certes un point nodal. Mais il est loin d'être le seul en cause, et par ailleurs les réponses à une telle question se sont révélées complexes.

Les pratiques actuelles de recueil de consentement et d'autorisation sont très variées. Il n'existe pas de normes reconnues, et les difficultés sont multiples.

Tout d'abord, le consentement doit être *éclairé* (cadre, finalités, « risques » pour l'enquêté).

Mais le recueil de consentement a priori peut parfois gêner l'enquête (paradoxe de l'observateur) en formalisant une situation alors qu'on souhaite obtenir des données « naturelles » proches de la conversation familière.

Ainsi, par exemple, une pratique qui s'est révélée intéressante et efficace consiste (en plus du recueil de l'autorisation) à laisser aux enquêtés un document expliquant le cadre, les finalités, les risques, l'accessibilité, et les coordonnées permettant de retrouver ultérieurement les références des publications et des résultats.

La difficulté provient également d'une *contradiction* entre l'obligation d'indiquer les finalités de l'enquête pour éclairer le consentement, et l'impossibilité de prévoir à l'avance l'ensemble des finalités *et les possibilités futures d'utilisation des données, étant donné le souci actuel de parvenir à* une interopérabilité *maximale*.

Il faut noter enfin que certaines cultures orales (et pas seulement à l'autre bout du monde) n'offrent pas la possibilité de proposer et de garder une trace écrite du consentement.

Et toutes les autres questions de nature juridique offrent la même complexité : anonymat, cryptage, floutage, définition des responsabilités, dépôt, communications, etc., toutes pratiques nécessairement liées à la constitution et à l'existence d'un corpus oral. Aucun de ces aspects ne repose sur une pratique unique, définie clairement et partout reconnue.

Et chacune de ces étapes se retrouve intimement liée à des choix techniques, à des pratiques sociales ou scientifiques, tout cela étant difficilement dissociable.

D'où le choix du groupe de travail, d'offrir un Guide qui ne soit pas seulement un « mémento juridique », mais aussi une aide pratique et fiable envisageant tous les aspects du processus.

1.7 UN « GUIDE DES BONNES PRATIQUES » ?

Prenant en compte les cadres juridiques existant en France (et plus généralement dans un certain nombre de points en Europe), ce guide s'appuie sur les questionnements des chercheurs qui ont participé à son élaboration. Ceux-ci ont cherché à comprendre les fondements des règles juridiques applicables et les enjeux liés à leur respect et à leur mise en œuvre. C'est donc une *vision dynamique de la régulation juridique* qui sert de trame à ce guide, à travers la démarche que suivent les chercheurs. Les auteurs du guide, eux-mêmes impliqués sur les terrains de recherche dont il est question ici, ont eu le souci de proposer des pratiques et usages

respectueux des droits existants. Pour cela, la démarche du chercheur doit consister à connaitre l'existence de ces droits et des contraintes qui en découlent. Il s'agira ensuite de tirer les conséquences de ces contraintes tant dans la phase du recueil des données que dans celle de leur valorisation.

Pour présenter de façon rigoureuse et crédible une telle démarche, il faut tout d'abord la situer dans son contexte, que celui-ci soit scientifique, politique, juridique ou institutionnel. Les usages et pratiques proposés seront tout au long « éclairés » par ce contexte, de façon à mieux comprendre quels sont les enjeux du respect ou du non respect de ces usages ou pratiques.

1.8 QUELQUES QUESTIONS FREQUENTES

Le premier objectif de ce guide est d'apporter des informations et des éléments de réponse aux questions qui se posent à tous chercheurs ou responsables de la constitution, de l'exploitation, de la conservation et de la diffusion de corpus.

Pour répondre à cet objectif, le guide a été conçu avec de nombreux renvois qui forment autant de parcours de lecture possibles. Les questions suivantes représentent les interrogations qui se posent traditionnellement au commencement d'un projet de recherche et proposent ainsi un premier exemple de parcours.

FOIRE AUX QUESTIONS

1. *Quelles autorisations dois-je faire signer aux locuteurs que j'enregistre pour pouvoir ensuite exploiter ce corpus et pouvoir :*
   a. le citer dans un travail universitaire ;
   b. le citer dans un article publié dans une revue scientifique ;
   c. le citer dans un ouvrage à diffusion commerciale ;
   d. le mettre à disposition sur un site ;
   e. le diffuser sur CD.

   Ces différents types d'exploitation sont-ils soumis aux mêmes règles ?

   Réponse : Les questions a, b et c relèvent du droit de citation (voir fiche *Droit de citation*). Les éléments de réponses aux questions d et e sont notamment présentés dans les chap. 2.1.5, 2.3 et 3.5. (voir fiche *Consentement* et *les exemples d'autorisations*).
2. *J'ai fait un enregistrement de personnes que je connais bien.*
   a. A quelles conditions puis-je l'exploiter ? (exploiter est pris au sens de la question 1)
   b. Peuvent-elles revenir sur leur autorisation ?

   Réponse : Tout le chapitre 3.4. est une réflexion sur les conditions de recueil des données, qui veut sensibiliser aux problèmes nombreux qui peuvent se poser au cours du recueil. Le fait de bien connaitre les personnes concernées ne diminue pas les exigences juridiques (qu'il faut a à leur égard), au contraire (il pose des questions de confiance qui peuvent donner lieu à des situations assez complexes). Voir fiche *Consentement*.
3. *Lorsque j'enregistre des enfants,*
   a. qui peut donner son consentement ?
   b. lorsque l'enfant sera majeur peut-il revenir sur ce consentement ?
   c. si l'enregistrement a lieu dans le cadre scolaire, faut-il des autorisations particulières ?

   Réponse : Ce cas rejoint le cas plus général des personnes pour lesquelles il faut demander une autorisation supplémentaire des responsables et tuteurs (parents et institution scolaire, dans ce cas) (voir chap. 3.3.2 catégorie des participants).

4. *Dans le cadre d'un travail au sein d'un laboratoire,*
   a. Qui est considéré comme l'auteur du corpus ?
   b. Quel(s) droit(s) ce travail donne-t-il au chercheur ?
   Réponse : Voir chap. 2.3 (droit d'auteur) et la fiche Droit d'auteur.
5. *Qui est considéré comme « responsable » de la diffusion et du traitement d'un corpus ?*
   Réponse : Voir chap. 2.3 et la fiche Responsable du traitement.
6. *Si je masque les noms propres de personnes, cela suffit-il pour que je puisse utiliser librement une transcription ?*
   Réponse : L'anonymisation ne consiste pas simplement en un effacement des noms propres. Voir chap. 3.5 Anonymisation et la fiche *Données personnelles et anonymisation*.
7. *Sous quelles conditions puis-je archiver mon corpus sous la forme de fichiers informatiques ?*
   Réponse : Il faut prendre en compte les aspects juridiques (protection de la vie privée, loi informatique et liberté, demande d'autorisation, voir les fiches *Données personnelles et anonymisation*, *Responsable du traitement* et les aspects techniques de conservation (voir fiches techniques).
8. *Si les personnes que j'ai enregistrées (dans les médias ou en privé) sont décédées, ai-je une liberté d'exploitation de ces enregistrements ?*
   Réponse : Les droits des auteurs survivent 70 ans après leur mort ! Quant à la protection au titre de la vie privée, elle ne peut être invoquée après la mort de la personne sauf si de son vivant la personne a interdit la diffusion. Par ailleurs, les membres de la famille du défunt peuvent invoquer leur droit personnel à la protection de la vie privée. Voir chap. 2.3.1 et fiche *Données personnelles et anonymisation*.
9. *Je découvre dans une armoire des enregistrements. Je voudrais pouvoir les exploiter. Je n'ai plus la trace de qui a enregistré ou qui a été enregistré.*
   a. Puis-je me servir de ces documents ?
   b. Quelles précautions (quelles garanties) dois-je prendre ?
   Réponse : On ne saurait trop inciter à la prudence et il est nécessaire de faire des recherches pour identifier les documents, y compris pour des raisons de rigueur scientifique. Voir chap. 2.3 et chap. 3.5.
10. *J'enregistre une émission à la radio (ou à la télévision).*
    a. Puis-je utiliser librement la transcription ?
    b. Puis-je utiliser la version sonore ?
    c. Du point de vue des autorisations, y a-t-il une différence entre émissions des radios publiques et des radios privées ?
    d. Y a-t-il une différence entre enregistrer des personnalités connues et enregistrer des « anonymes » (personnes qui témoignent, s'expriment en libre antenne, auditeurs qui posent des questions, etc.) ?
    e. les droits d'exploitation sont-ils différents si j'achète une cassette, un dévédé ou un cédé de l'émission ou si j'enregistre moi-même l'émission lorsqu'elle est diffusée ?
    Réponse : Les émissions radio sont protégées, qu'elles soient publiques ou privées. Voir 3.3.1 sur la reprise d'enregistrements médiatiques, et plus particulièrement la notion de documents d'actualité.
11. *J'aimerais constituer un corpus de données authentiques. Quelles sont les précautions que je dois prendre ?*
    Réponse : Voir chap. 3. où est proposée une réflexion articulant méthodologie de recherche sur le terrain et problèmes éthico-juridiques rencontrés au fil de la démarche.

    Beaucoup d'autres questions encore…

2 LE CONTEXTE

SCIENTIFIQUE, POLITIQUE, JURIDIQUE ET INSTITUTIONNEL

Qui dit contexte dit « *mise en perspective* ». Telle est la finalité de ce chapitre qui présente ce qu'est le travail scientifique du linguiste sur l'oral. La mise en perspective se devait d'être aussi *politique* et *juridique*. Le contexte institutionnel a une importance grandissante compte tenu des besoins d'assurer, sur la durée, la « traçabilité » et la poursuite des recherches. En garantissant la pérennisation tant des données qui ont permis à un chercheur de travailler que des résultats obtenus, le chercheur comme l'institution participent au développement des connaissances dans un avenir proche ou plus lointain.

2.1 LA LINGUISTIQUE ET LES CORPUS ORAUX

Depuis une vingtaine d'années, les études sur les corpus de langues parlées ont complètement renouvelé les sciences du langage. Il suffit, pour s'en convaincre, de consulter les bibliographies récentes, en France et hors de France (par exemple la *Revue Française de Linguistique Appliquée* ou les *Recherches sur le Français Parlé)*. Ces études ont permis de formuler de nouvelles hypothèses sur le fonctionnement normal et pathologique du langage et elles sont devenues une composante essentielle du dialogue entre les linguistes et les informaticiens. En France, jusqu'à cette période encore récente, l'intérêt pour les langues parlées était essentiellement réservé aux domaines où il s'exerçait « par défaut » : en premier lieu les études sur les aspects proprement sonores de la langue (phonétique, phonologie et prosodie), le parler des jeunes enfants, ou tout ce qu'on classait parmi les « langues sans traditions écrites », en France les langues régionales et parlers locaux et, hors de France, tout ce qu'on nommait « langues exotiques ». A cela s'ajoutaient quelques essais isolés, dans les années 1950-1960, pour rassembler des modèles de français parlé afin d'enseigner le français en tant que langue étrangère, notamment le *Français Fondamental* et le *Corpus d'Orléans*.

Les représentations de la langue française, en particulier dans les grammaires, restaient fondées sur des données de langue écrite, littéraire ou non, les « grapholectes », comme les nommait Ong (1988), ou sur des données fournies par l'intuition. Cette mise à l'écart des données de langue parlée a entrainé deux conséquences majeures, d'une part l'image très négative que les Français ont de leur propre langue et d'autre part une influence considérable sur les théories linguistiques les plus courantes. Les nouvelles données révélées par les corpus de langue parlée n'ont sans doute pas encore fait évoluer l'image de la langue dans le grand public, mais elles ont déjà beaucoup fait évoluer les théories parmi les spécialistes.

De nouveaux domaines, abordés dès les années 1970 en Grande-Bretagne (Sinclair & Coulthard, 1975 pour l'École de Birmingham), ont émergé en France, comme l'essor des modèles de l'interaction et l'analyse conversationnelle (article fondateur de Sacks, Schegloff, Jefferson aux États-Unis en 1974, articles de Bange et de Quéré en France, en 1983 et 1984).

Les données de langue parlée collectées avant l'ère de l'informatique ne peuvent pas être comparées à ce qu'on appelle aujourd'hui « corpus de langue parlée ». Chacune

des collections anciennes, dispersées au gré des recherches, suivait ses propres règles de choix, d'enregistrement, de transcription et de conservation, de sorte qu'il est difficile maintenant d'y accéder et de les mettre en commun (les enregistrements du *Français Fondamental* ont été effacés, ceux du *Corpus d'Orléans* doivent être aujourd'hui retranscrits). Aucune ne pouvait atteindre de très grandes dimensions (il s'agissait généralement de quelques heures d'enregistrements seulement) et, dans ces données, la recherche d'informations ne pouvait se faire que manuellement. A partir des années 1980-90, le développement de l'informatique a permis de créer des corpus modernes de langue parlée dans le monde entier, en premier lieu dans les pays anglo-saxons. Une nouvelle discipline est née, celle des linguistiques de corpus (G. Kennedy en a donné une description en 1998 pour l'anglais et Habert et ses collaborateurs pour le français en 1997), qui intéressent les universitaires et les industries de la langue et qui, au titre de *Language Resources*, font maintenant partie des patrimoines nationaux. La France, qui était en avance pour la mise au point des corpus de langue écrite (en particulier pour FRANTEXT qui est à la source du *Trésor de la Langue Française*), a pris un grand retard dans la constitution des corpus de langue parlée.

Il existe de nombreux types de corpus de langue parlée, prévus pour divers objectifs, dans plusieurs disciplines. Il s'agit toujours d'enregistrements de données sonores, éventuellement accompagnées de données visuelles (prises en vidéo, ou à la télévision), presque toujours accompagnées de transcriptions et de traitements informatisés. Sans prétendre tout exposer ici, on en présentera quatre aspects : les types de données et de locuteurs, la dimension des corpus, les transcriptions, et un bref panorama des exploitations et des résultats.

2.1.1 TYPE DE DONNEES ET DE LOCUTEUR

Certaines données sont « sollicitées ». On fait par exemple venir dans des laboratoires de phonétique des locuteurs qui, agissant en tant que « cobayes », fournissent des types de prononciations et d'intonations, dans de très bonnes conditions d'enregistrement. On leur fait prononcer des mots et des listes de mots, des nombres et des listes de nombres, ou on leur fait lire des textes ou fragments de textes. Ces documents servent à différentes exploitations, soit pour consigner et étudier les prononciations en tant que telles, comme le font J. Durand, B. Laks et Ch. Lyche pour étudier la prononciation du français contemporain (projet PFC), soit pour tester un comportement langagier (comme on le fait dans des services hospitaliers qui étudient des phénomènes d'aphasie), soit pour établir des analyses qui servent à la synthèse de la parole ou à la lecture automatique de textes écrits (*Text-to-Speech data)* ou aux dialogues homme-machine (c'est l'objectif de *SpeechDat Exchange*, qui stocke de 500 à 5 000 enregistrements téléphoniques pour 28 langues). Dans toutes ces situations, les locuteurs savent généralement qu'ils sont enregistrés et ils ont une idée, précise ou approximative, de la finalité de leur prestation.

D'autres données sont dites « de parole continue », avec divers degrés de spontanéité (la notion a été spécialement étudiée dans un numéro de la *Revue Française de Linguistique Appliquée*). Certaines sont recueillies dans des situations qui n'ont pas été provoquées par le chercheur et qui auraient eu lieu de toute façon sans lui. D'autres,

plus ou moins « sollicitées », sont orchestrées et organisées par le chercheur. L'idéal du spontané total serait d'enregistrer les locuteurs sans qu'ils s'en doutent (micros cachés, enregistrements pirates), en le leur disant ensuite ou sans le leur dire, l'objectif étant de saisir leur langage « en toute liberté », avec un minimum de contrôle. Les dispositions juridiques limitent cette possibilité. La présence de l'enquêteur et des appareils apporte de toute façon un frein à cette liberté (c'est la question du « paradoxe de l'observateur » popularisée par W. Labov). Dans la pratique, divers degrés de contrainte peuvent être identifiés, selon qu'il s'agit de parole privée ou de parole publique, devant des familiers ou des étrangers, avec diverses formes de complicité ou non, selon qu'il s'agit de parole en face-à-face ou de parole transmise par un canal comme le téléphone, le répondeur, la radio, la télévision ou d'autres dispositifs techniques. Une bonne approche ethnographique (enregistrements répétés) permet de résoudre le problème de la sensibilité au micro. Mais cela demande qu'on y consacre beaucoup de temps pendant la phase de recueil des données.

Il est rare que les corpus modernes soient composés de paroles « de tout venant ». Le choix des locuteurs et des situations d'enregistrement est généralement fixé en fonction des objectifs donnés au départ. Les chercheurs proposent de collecter des conversations entre adultes, des négociations professionnelles, des entrevues (préparées ou non), des prises de parole dans des organismes publics, des discours électoraux, des explications entre services publics et utilisateurs, des cours publics, des sermons, des discours politiques, des conférences (spécialisées ou de vulgarisation), des témoignages historiques, des récits de faits-divers, des récits de vie (produits par des individus, des groupes, des représentants de groupes, des porte-paroles), des dialogues entre mères et jeunes enfants, des enfants enregistrés dans un contexte scolaire ou en dehors (dans leurs jeux ou dans leurs récits, en réponse à des tests ou en dehors, dans des situations scolaires ou non, dans des jeux libres ou contraints, avec parodie et jeux de rôles), des malades dans les hôpitaux, etc. Un exemple : une banque de données CLAPI (Corpus de Langue Parlée en Interaction) est constituée actuellement à Lyon (laboratoire ICAR) afin de réunir des corpus de « parole en interaction » les plus diversifiés possibles, dans des situations non provoquées par les chercheurs : conversations à table, concertations entre notaires, appels à des centres d'aide sociale d'urgence et à des consultations thérapeutiques, etc. Cette banque de données comporte 300 h. d'enregistrements audio et vidéo, des transcriptions et des « métadonnées » décrivant les caractéristiques des locuteurs.

De nombreuses disciplines cherchent à étudier les corrélations entre les productions de langue parlée et d'autres phénomènes. Les corrélations entre langage et paramètres socio-économiques ont été à la base des recherches de sociolinguistique. Aux Etats-Unis, W. Labov avait produit de célèbres études sur les Noirs des grandes villes américaines de l'Est, en enquêtant dans les domiciles, dans les rues ou dans les grands magasins (avec des conditions d'enregistrement souvent défectueuses). Les études sur le développement du langage se font en fonction de l'âge des enfants, des activités observées, des consignes fournies et des données familiales. La prise en compte des « genres » (tels que les conçoit D. Biber pour l'anglais) amène à faire des corrélations avec les lieux de prise de parole, les sujets dont il est question, les types

d'interlocuteurs et le type d'échanges (monologues, dialogues, conversations à plusieurs). Pour pouvoir mesurer ces corrélations, le contenu et la taille des corpus sont généralement définis à l'avance : tant de types de situations et de locuteurs (comme l'avait fait l'équipe Sankoff-Cedergren dans les années 1970 pour étudier la variation sociale dans la ville de Montréal). Dans d'autres cas, les chercheurs découpent, à l'intérieur de corpus existants, des sous-corpus représentatifs adaptés à leur étude (c'est ce qu'a proposé D. Biber pour faire des échantillonnages dans le grand *British National Corpus*). Il s'agit en ce cas de corpus « fermés » et « échantillonnés ».

Les linguistes, de leur côté, ont souvent collecté des corpus « ouverts », qu'ils modifient au gré de l'avancement de leur travail, sans délimiter à l'avance un objet de recherche pré-déterminé, parce qu'ils sont certains de découvrir des phénomènes nouveaux, impossibles à prévoir au départ : répartition du langage formel et informel, relations entre grammaire et lexique, liens entre degrés de complexité de la syntaxe et type de situations de parole, utilisation de la morphologie orale, rôle des contextes dans la construction du sens des énoncés, rôle de la prosodie dans la structuration des textes, etc.

La qualité technique des enregistrements dépend bien évidemment des équipements techniques utilisés, mais aussi des types de situations et de locuteurs choisis (lieux bruyants, locuteurs trop nombreux, locuteurs affectés d'un défaut de parole). Ces situations diverses influent également sur le consentement des locuteurs : il est plus facile d'obtenir l'autorisation d'enregistrer la parole publique que la parole privée, les propos d'un locuteur sûr de lui-même plutôt que d'un locuteur inquiet et sensible à ce que l'on a pu appeler « l'insécurité linguistique ».

Dans tous les cas, il est bien difficile de justifier les enregistrements par l'étude de la langue. Si on explique cette finalité, les locuteurs français ont immanquablement l'impression qu'ils parlent mal et que l'étude va les ridiculiser. Peu d'entre eux sont détendus sur cette question. Presque tous les chercheurs ont mis au point des stratégies pour aborder le problème de biais : en disant qu'ils s'intéressent au contenu, aux témoignages, aux explications, au savoir particulier des locuteurs (qui peut être un savoir de langage, dans le cas des recherches sur les régionalismes). Dans les travaux sur la parole en interaction, les choses sont un peu différentes : les chercheurs peuvent dire qu'ils s'intéressent précisément à la manière dont les participants interagissent entre eux, à leur coordination, aux ajustements remarquablement précis auxquels ils recourent, par la parole, les gestes, les mimiques, les regards et l'ensemble des attitudes (ressources multimodales, difficilement contrôlables dans leur ensemble même par des locuteurs qui se surveillent).

2.1.2 DIMENSIONS

La dimension utile des corpus et des unités qui les constituent varie selon l'étude prévue. Les études de phonétique, de phonologie et de prosodie peuvent donner de bons résultats avec des unités sonores de durée assez limitée. Mais, si l'on veut étudier des corrélations entre le langage et d'autres phénomènes, ou si l'on veut étudier le lexique, il y faut des unités beaucoup plus développées, en quantité plus importante et dans des domaines d'activité plus diversifiés. La dimension des corpus

de langue parlée et des éléments dont ils sont composés se mesure avec deux sortes d'unités. On utilise des unités de temps lorsqu'on s'intéresse prioritairement à l'enregistrement sonore, en faisant abstraction de la transcription. On classe par exemple comme très petits éléments de corpus ceux qui durent entre quatorze et trente secondes (quatorze secondes étant la moyenne pour une information à la radio). Mais on tient compte de sous-unités encore plus petites quand on observe les chevauchements de parole entre les locuteurs ou quand on mesure les pauses (jusqu'au dixième de seconde). Les petites unités sont utilisées par exemple par les compagnies de téléphone qui construisent actuellement des services européens de renseignements par téléphone dans toutes les langues de l'Europe (EuroSpeech 2003). On classe comme petits éléments ceux qui durent dix minutes et comme très grands éléments ceux qui ont une durée de soixante ou quatre-vingt-dix minutes. En totalisant l'ensemble de ces éléments, on dira par exemple qu'on dispose de réserves de 100 ou 500 heures d'enregistrements.

Mais ces mesures sont peu fiables pour les grands composants de corpus, parce que la densité des enregistrements dépend du débit des locuteurs. En français, on estime que les locuteurs qui parlent lentement prononcent 110 mots par minute et que ceux qui parlent très vite en prononcent 350 par minute (dans certains types d'aphasie, et sous l'influence des neuroleptiques, le débit tombe au-dessous de 100 mots par minute, ce qui est pénible à écouter. Au-dessus de 350 mots par minute, l'écoute et la transcription deviennent très difficiles). La densité varie donc de un à trois, ce qui est considérable. Selon les deux débits extrêmes qui viennent d'être cités, une heure d'enregistrement peut correspondre à 6 600 ou à 21 000 mots. On a donc intérêt à évaluer les grands corpus en fonction du nombre de mots graphiques que comporte la transcription. Les grands corpus de langue parlée collectés aujourd'hui dans le monde sont de l'ordre de grandeur de la dizaine de millions de mots transcrits pour l'anglais, américain ou britannique. Malheureusement, les corpus actuels de français parlé sont de l'ordre de grandeur du million de mots. Avec une taille aussi limitée, il n'est guère possible de faire des recherches lexicales, ni d'établir des statistiques fiables sur les usages.

2.1.3 TRANSCRIPTIONS

Les transcriptions de langue parlée qui ont cours aujourd'hui sont tellement différentes les unes des autres qu'il est difficile de les rassembler sous une même étiquette. Dans certains cas, lorsqu'on ne retient que le contenu des enregistrements, en en changeant librement la forme, les termes de *transposition* ou d'*adaptation* conviendraient mieux. C'est ce que font souvent les journalistes, lorsqu'ils rapportent les propos de personnes interviewées, en résumant ces propos et en leur donnant généralement une tournure plus normative (là où un homme politique important dit *ça, je sais pas, pour pas que…,* ils rétablissent *cela, je ne sais pas, pour que …ne pas…*). Les historiens et les sociologues ont parfois des pratiques voisines, lorsqu'ils s'intéressent avant tout au contenu informatif : ils font un tri dans les données, coupent les passages qui ne les intéressent pas et suppriment les particularités de la production orale qui leur paraissent gênantes, répétitions, hésitations ou retouches.

Certains secteurs d'activité, comme les transcriptions de débat parlementaire, ont même codifié ces tâches, en établissant plusieurs degrés d'adaptation.

Lorsqu'il s'agit de s'intéresser au langage lui-même, le choix d'un type de transcription dépend des finalités de l'étude (des projets européens et internationaux se sont donné des consignes d'édition de corpus) et, comme le signalait déjà E. Ochs en 1976, la transcription engage toujours une théorie. Certaines études nécessitent de disposer de transcriptions phonétiques ou phonologiques. Le standard Unicode, synchronisé sur la norme ISO-10646, comporte déjà dans sa version 4.0 plus de 96 000 caractères dont, en particulier, ceux de *l'Alphabet Phonétique International*. C'est une nécessité pour tous les travaux qui concernent la prononciation, mais aussi pour tous les cas où il est difficile de dégager des morphèmes stables qu'on pourrait écrire en orthographe standard : langage des très jeunes enfants (modèle international CHILDES), langage des étrangers en cours d'acquisition de la langue, notation de certains régionalismes, notation de certaines formes d'aphasie comme les jargons (Abou-Haidar 2002). Ces transcriptions, qui ne peuvent se faire que pour de petites quantités de corpus, sont souvent accompagnées de traductions juxtalinéaires. La représentation de la prosodie exige des modèles spécifiques, très développés dans les techniques récentes (Martin 1987). Les enregistrements vidéo demandent des notations spéciales, qu'on peut pousser plus ou moins loin (Van der Straten 1998, Mondada 2006).

En ce qui concerne les grands corpus de langue parlée, ils sont transcrits en orthographe standard, de façon à en rendre la lecture facilement accessible. A partir de ce choix, plusieurs options sont possibles : orthographe standard avec ou sans adaptations, avec ou sans ponctuation, avec ou sans indications de pauses, allongements, rythmes, accentuations, hésitations, toux, rires, gestuelle, etc. De grands débats ont eu lieu sur tous ces points, pour dégager les conditions optimales de transcription, adaptées aux objectifs de la recherche. Un exemple : les linguistes qui s'intéressent aux unités syntaxiques de la langue parlée se méfient généralement de la ponctuation, qui impose des délimitations propres à la langue écrite et qui s'avère souvent trompeuse quand on la met avant d'avoir suffisamment bien analysé les textes. Mais les textes non-ponctués indisposent les informaticiens, dont les analyses automatisées réclament des repères de ponctuation. Des négociations sont parfois menées entre les linguistes et les informaticiens (ICOR au laboratoire ICAR) afin d'établir des conventions de transcription qui tiennent compte de ces problèmes et des standards internationaux (GAT, TEI, Du Bois, Jefferson).

Les transcriptions qu'utilisent les linguistes conservent soigneusement toutes les particularités des productions orales : répétitions, hésitations, amorces de mots, retouches. Elles exigent que le transcripteur veille à ne pas projeter sur la transcription ses propres interprétations (ajouter ou ôter des *ne* de négation, par exemple, ou reconstruire une portion de texte selon les stéréotypes attendus). Ce souci du détail exige un entrainement et une formation spécifique des transcripteurs. La tâche, longue et couteuse, est pleine de pièges (Leech 1991). Selon les estimations courantes, un minimum de trente minutes de travail est nécessaire pour transcrire une minute d'enregistrement (les concepteurs du corpus néerlandais estiment que cela revient à un euro par mot graphique !). En raison même de leur fidélité, les

transcriptions de la langue parlée déplaisent aux profanes : ils y voient quantité de « fautes de français », de répétitions, de dissolutions de l'information. Montrer à un informateur profane une transcription de sa parole provoque souvent le rejet. Ce n'est pas un très bon moyen pour obtenir son autorisation de transcrire et publier le résultat de la recherche.

L'outillage informatique a transformé le travail de transcription, d'une part par les aides qu'il a apportées, d'autre part par les exigences nouvelles qu'il a introduites. Les aides à la transcription (Anvil, Clan, Elan, Ite, Praat, Transcriber…) facilitent les manipulations et permettent de réécouter facilement les portions d'enregistrement sous étude. La technique des *corpus synchronisés* permet de lire sur écran des portions de texte écrit en même temps qu'on écoute les mêmes portions dans leur déroulement sonore (*Speech Communication* 33, numéro spécial sur les annotations et les outils d'analyse des corpus). Les exigences nouvelles concernent les annotations informatisées : étiquetage morpho-syntaxique de tous les éléments du texte, arborescences, métadonnées (concernant les circonstances d'enregistrement, les situations et les locuteurs). Divers classements et codages permettent de faire les lemmatisations et les concordanciers nécessaires pour pouvoir formuler des requêtes sur l'ensemble du corpus. Une polémique s'est engagée, dans les années 2000, autour du degré de sophistication des annotations qui semblait nécessaire (Sinclair, Teubert). La standardisation se fait maintenant au plan européen (SpeechDat Exchange Format).

2.1.4 TRAITEMENT AUTOMATIQUE DE LA PAROLE

Contrairement à bien d'autres domaines de recherche autour de la parole, la transcription automatique de la parole, qui s'effectue sur un flux acoustique continu, nécessite une modélisation de l'ensemble des phénomènes observés dans le signal sonore. Il faut donc modéliser, au-delà des mots auxquels est associée une représentation phonologique dans le dictionnaire de prononciation, des phénomènes extra-lexicaux : respirations, hésitations, fragments de mots, etc.

Suivant les genres de document traités, les systèmes de reconnaissance automatique obtiennent des taux d'erreurs très variables[1]. Cependant, s'il y a un décalage entre les modèles (en gros les connaissances) du système et les corpus à transcrire, ces taux d'erreurs peuvent augmenter rapidement. Afin d'arriver aux meilleurs résultats possibles, les systèmes de transcription doivent être adaptés en fonction des corpus à transcrire.

Les recherches actuelles montrent que la transcription automatique est en train de devenir un instrument précieux pour aider la transcription et l'annotation de corpus. Par exemple dans (Barras *et al.* 2004) est montrée l'utilité de la transcription automatique pour la génération semi-manuelle de transcriptions acoustiques fines (c'est-à-dire comprenant non seulement tous les mots orthographiques mais également les « disfluences » et autres évènements extra-lexicaux). Les recherches en cours montrent également que la transcription automatique de la parole peut devenir un

[1] Les travaux du LIMSI (Barras 2004) présentent des résultats allant de 10 à 30 % d'erreur de mots avec des systèmes optimisés pour une tâche donnée.

instrument précis pour explorer, analyser des corpus, quantifier des phénomènes linguistiques. Plus généralement, on peut penser qu'à l'avenir il faudra de moins en moins opposer les visions des linguistes et des informaticiens. À cet égard, l'émergence de la linguistique des corpus oraux comme domaine de recherche doit reposer sur la formation de linguistes informaticiens et d'informaticiens linguistes.

2.1.5 *Exploitations et resultats*

Les grands corpus actuels de langue parlée sont chers. Certains corpus, notamment dans l'ingénierie, sont exploités en association avec les industriels : dialogue homme-machine, reconnaissance et synthèse de la parole, communications téléphoniques, etc. (des organismes comme ELRA/ ELDA se sont spécialisés dans la diffusion des corpus et des ressources disponibles dans ce domaine).

Les grands corpus servent en premier lieu de documentation générale sur la langue nationale. Les grands *corpus de référence*, échantillonnés en tenant compte des régions et des données socio-économiques et culturelles, permettent de guider les politiques linguistiques à grande échelle. Par exemple, le corpus de référence du portugais parlé, qui comporte des enregistrements réalisés au Portugal, en Afrique, au Brésil et en Asie, permet d'évaluer les différences selon la géographie mondiale, et de fonder sur cet examen, certains usages de pratiques scolaires et même des décisions gouvernementales. Le *British National Corpus* a servi de base à la fabrication d'une grande grammaire, la *Longman Grammar of Spoken and Written English*, conçue sur des bases très nouvelles. Une grande activité éditoriale s'est développée en langue anglaise, en utilisant ces matériaux. C'est ainsi que l'éditeur Collins a utilisé les corpus anglais pour la publication de nombreux ouvrages didactiques servant à l'enseignement de l'anglais comme langue maternelle et comme langue étrangère. Une documentation sur la langue parlée est parfois le point de départ pour lancer des activités nouvelles : des corpus de langue parlée ont servi de base pour diffuser des langues peu (ou pas du tout) écrites, comme on l'a fait pour la langue maori, qui a servi de modèle pour développer des émissions de radio et de télévision (Kennedy 1998 : 72).

La comparaison entre langues parlées appartenant à un même groupe linguistique permet d'évaluer *in vivo* les ressemblances et différences à l'intérieur d'une grande aire linguistique.

Une exploitation importante est celle qu'offrent les corpus multilingues (appelés aussi corpus parallèles ou alignés), qui servent aux traducteurs, à l'enseignement des langues et à l'étude contrastive. Il en existe pour la langue écrite :

- anglais/français à l'université de Lancaster, à l'université d'Oslo, à Mannheim, à l'université de Gand en Belgique (Contragram, bank.ugent.be/contragram/newslet.html), à l'université de Montréal,
- français/ anglais/ néerlandais, à l'université de Courtrai,
- français/anglais/espagnol à l'université de Pennsylvanie.

Une étude récente, fondée sur des enregistrements et transcriptions de quatre langues romanes (italien, français, portugais, espagnol) permet de comparer la prosodie (intonations, accentuations, rythmes), en tenant compte de différentes situations et différents médias (C-ORAL-ROM, Cresti & Moneglia).

C'est ainsi que les grands corpus de langue parlée ont renouvelé quantité de problèmes linguistiques. Sur les données livrées par ces grandes collectes, de nouvelles disciplines se sont fondées, comme l'analyse conversationnelle et l'analyse des interactions, des négociations et des codes de politesse. Les recherches en pragmatique s'appuient massivement sur ces données. Certaines connaissances ont été nettement modifiées, comme par exemple les études portant sur la production et sur la perception du langage parlé et, par voie de conséquence, sur la fragilité de l'intuition linguistique (Blanche-Benveniste 1997). On a pu montrer quel est le degré d'organisation ordonnée et systématique dans les interactions. On s'en est servi pour remettre en cause certaines unités de base comme la *phrase*, et pour en introduire d'autres comme les unités de *macro-syntaxe*, utilisée maintenant par plusieurs équipes de linguistes (Blanche-Benveniste *et al.*, 1999, Scarano 2003, Nolke 2002). L'étude de l'intonation a été prise en charge très sérieusement dans la délimitation des unités de macro-syntaxe (Cresti & Moneglia 2005, Couper-Kuhlen & Selting, 1996). Dans les interactions, on a montré qu'intervenaient plusieurs niveaux d'organisation imbriqués (Turn-Constructional Units ou « Unités de Construction du Tour », Selting 1995, 1998, 2000, Auer *et al.* 1999, Ochs, Thompson & Schegloff, 1996). Dans différentes langues, on a pu montrer quel était le rôle des caractéristiques de productions orales que sont les particules discursives, les répétitions, les hésitations ou les « réparations », qui intéressent actuellement les neurosciences. Les perspectives sur l'histoire des langues en ont même été modifiées, dans la mesure où l'on peut maintenant étudier l'influence qu'exercent les différentes situations de parole sur le type de grammaire adopté (Biber 1987). On peut montrer par exemple, pour le français, que les récits d'explication et les argumentations révèlent des pratiques de syntaxe à haut degré d'enchâssement, alors qu'il y en a rarement dans les conversations, ou que les récits d'accidents contiennent des organisations chronologiques complexes. On sait que les thèmes réputés « sublimes » (discours sur la morale, la religion, la mort) déclenchent des caractéristiques de « langue de cérémonie », par exemple, en français, un grand nombre de liaisons, des emplois massifs du *ne* de négation et même parfois des emplois inattendus de passé simple. Les grands corpus permettent de suivre certains processus de grammaticalisation en cours. Ils montrent l'importance numérique des énoncés parenthétiques, des focalisations et des thématisations. Ils obligent à considérer que les locutions figées occupent une place très importante par rapport à la libre composition des énoncés, de sorte que le lien entre la grammaire et le vocabulaire apparait maintenant plus nettement qu'auparavant, beaucoup de tournures grammaticales n'étant utilisées par les locuteurs que pour une petite liste de mots du lexique. Il faut en conclure que, lorsqu'on parle, on ne choisit pas « un mot » mais un ensemble préconstruit (Sinclair, 1991).

Cela remet en cause, évidemment, les théories linguistiques qui visaient à isoler la syntaxe comme une composante du langage indépendante.

Ces grands corpus, lorsqu'ils existent, rendent un service primordial : ils servent de base de données pour toutes les comparaisons concernant le langage : pour évaluer le langage des enfants à divers stades d'acquisition, pour soutenir les diagnostics dans les pathologies de langage, pour évaluer le degré d'accomplissement dans l'acquisition des langues maternelle et étrangère, pour calculer les effets des langages

de groupes et des langages professionnels (Gadet), pour étudier les modes de coordination dans une équipe ou dans un groupe, pour comprendre les spécificités des types d'activités et des contributions qui y sont adéquates dans des contextes institutionnels différents ou pour connaitre l'effet des influences régionales. Un exemple : avant de juger qu'une tournure est caractéristique du parler des enfants de tel âge ou de telle origine, il est indispensable de recourir à une base de données de comparaison pour savoir si la tournure est spécifique ou non (les fautes les plus courantes sur les relatifs *dont* et *lequel*, premier degré, se retrouvent chez les adultes les plus scolarisés, et depuis assez longtemps, pour autant qu'on puisse en juger).

CORPUS DE LANGUES A TRADITION ORALE

Les problèmes rencontrés lors de la constitution, l'exploitation, la diffusion et la conservation de corpus oraux dans les sociétés dites « à tradition orale », ou « ethniques », ou « exotiques », recoupent partiellement ceux rencontrés lors de l'établissement des grands corpus de langues occidentales. Les précautions à prendre (telles que préconisées dans le guide) pour respecter les personnes sont alors à adapter au contexte dans lequel se déroule le travail de terrain.

Dans une société à tradition orale, l'autorisation après information (sur le modèle du « consentement éclairé » décrit dans le guide, là encore adapté à la situation) peut, dans certains cas, n'avoir de valeur que si elle est orale, et accordée par la personne qui en a le pouvoir (tout comme dans une situation d'enquête en milieu médical, l'autorisation n'aura de valeur que si elle est accordée par l'Ordre des médecins). Par ailleurs, l'information du locuteur n'est pas sans poser problème dans des sociétés où l'activité de recherche, les objectifs de la constitution du corpus et ses réseaux de diffusion (publications, internet) ne correspondent à rien de concret.

Le chercheur doit par ailleurs s'informer du droit en vigueur dans le pays dans lequel il va travailler. Par exemple, le droit français ne reconnait pas la propriété intellectuelle ni les droits d'auteur dans le cas de recueil de contes ou de mythes, considérés comme faisant partie du patrimoine et relevant du domaine public. Dans plusieurs pays d'Afrique en revanche (voir annexe), il a été créé des bureaux de droit d'auteur pour protéger ce type de productions et leurs auteurs. Par ailleurs, certaines communautés ne reconnaissent pas le droit national de l'État dans lequel elles vivent. C'est par exemple le cas dans certaines communautés amérindiennes en Guyane qui fonctionnent selon un droit collectif et non pas privé (Tiouka 2005, pour une réflexion sur l'intégration du droit coutumier dans le droit français et européen). Certaines autorisations n'auront de valeur pour ces communautés que si elles respectent le droit coutumier, et bien que le chercheur se sente protégé en respectant le droit national, il peut se retrouver en conflit avec les autorités coutumières et se voir refuser l'accès au terrain.

L'exploitation du corpus nécessite dans la plupart des cas l'intervention de plusieurs personnes : le transcripteur, qui peut être le locuteur de l'enregistrement, mais pas toujours ; le traducteur (id). Les droits de ces personnes sur le corpus sont là encore à définir selon plusieurs paramètres : le droit national s'il a un sens pour la communauté, ou bien le droit coutumier.

Il est deux points sur lesquels la constitution de corpus oraux dans certaines sociétés diffère de celle des grands corpus de langues nationales :

1. la taille du corpus

Il est difficilement envisageable d'arriver à recueillir des corpus sur certaines langues qui pourraient atteindre la taille des grands corpus de langues européennes, comprenant plusieurs millions de mots. Le problème de la représentativité des corpus se pose alors de façon différente.

2. le retour à la communauté

La pratique du terrain (anthropologie du début du 20e siècle, linguistique missionnaire, etc.) jusqu'à il y a encore une cinquantaine d'années a laissé des

traces dans les communautés qui se sont senties pillées et exploitées sans avoir jamais eu accès aux résultats de la recherche. Celles-ci réclament maintenant que les recherches aient des retombées directes sous diverses formes, et ces revendications sont reprises par tous les organismes qui financent ou organisent des recherches sur les langues en danger (UNESCO, projet DOBES du Max-Planck Institute[2], etc.). Les revendications qui émanent des communautés n'ont rien à voir avec le dédommagement du travail individuel des différents locuteurs impliqués dans la constitution et l'exploitation du corpus, et demandent une implication du chercheur (participation à des programmes d'éducation, restitution des enregistrements et des matériaux collectés, constitutions d'archives accessibles aux communautés, etc.). La nécessité de rendre à la communauté les matériaux collectés devrait d'ailleurs motiver les chercheurs à constituer des bases de données et à archiver leurs corpus.

Cependant, il semblerait que ces corpus restent encore souvent des « outils personnels » uniquement destinés à servir de base à l'analyse linguistique du seul chercheur. Les raisons sont multiples : comme pour la constitution de n'importe quel corpus de langue orale, l'aspect technique et le temps nécessaire à la mise en forme (numérisation, dans certains cas synchronisation, etc.) rebutent le chercheur, et ce d'autant plus que ce travail n'est pas valorisé par les institutions scientifiques. Par ailleurs, sur certains terrains, la constitution du corpus est le résultat d'une relation de confiance qui s'est établie entre le chercheur en tant que personne (et non pas en tant que représentant d'une communauté scientifique) et la communauté ou certaines personnes de la communauté, dans des contextes difficiles. La décision de diffuser le corpus au sein de la communauté scientifique ou plus largement (accessibilité par Internet) interroge l'éthique du chercheur, et ne relève plus du cadre juridique. Dans tous les cas, il est souhaitable que cette diffusion se fasse après la restitution du corpus à la société concernée.

Le chercheur se trouve donc en porte-à-faux entre la volonté de préserver une relation privilégiée avec son terrain, et avec la croissante nécessité de mettre à la disposition de la communauté scientifique les ressources qui sont à la base de l'analyse et des résultats de la recherche.

Au fur et à mesure que se développent les grands corpus de langue parlée actuels, la standardisation progresse (depuis les consignes diffusées par EAGLES en 1993) et les champs de recherche deviennent de plus en plus intéressants. Dans cette perspective, rendre accessibles les corpus de français parlé existants ou ceux de toute autre langue et en créer de nouveaux est une tâche importante du « patrimoine immatériel ». Les problèmes juridiques de protection de la parole, qui ont longtemps été considérés, à tort, comme secondaires, sont actuellement des freins très importants : beaucoup de chercheurs refusent de faire circuler leurs corpus parce qu'ils ne sont pas sûrs d'avoir « les bonnes autorisations ». Beaucoup hésitent à en lancer de nouveaux, parce que la demande d'autorisations leur parait fondamentale mais difficile à accomplir. C'est pourquoi une réflexion collective sur cette question est maintenant indispensable.

[2] Voir dans la bibliographie les liens vers les sites de ces organisations.

2.2 CADRES POLITIQUES DE LA DIFFUSION DE LA RECHERCHE

LA DIFFUSION DES RESULTATS DE LA RECHERCHE FAIT PARTIE DES MISSIONS DES CHERCHEURS

> *« Les organismes publics doivent avoir le souci constant de faire bénéficier au mieux la collectivité nationale des fruits de leurs travaux... ».*
>
> *« La politique de la recherche et du développement technologique vise à l'accroissement des connaissances, à la valorisation des résultats de la recherche, à la diffusion de l'information scientifique et technique et à la promotion du français comme langue scientifique »*[3].

C'est en ces termes que le rapport annexé à la loi d'orientation du 15 juillet 1982 définit les contours de la valorisation. Il ne fait aucun doute que ces principes généraux trouvent à s'appliquer aux chercheurs dont les travaux aboutissent à la constitution de corpus oraux. Toutefois les conditions de la valorisation et de la diffusion dépendront aussi des possibles droits existants sur les contenus collectés et sur les résultats du traitement de ceux-ci par les chercheurs.

LA DYNAMIQUE DE L'ECHANGE ET LES OCCASIONS OFFERTES AUX TITULAIRES DE DROITS POUR FACILITER LA LIBERTE D'ACCES DANS LA SOCIETE DE L'INFORMATION

Sans doute peut-on parler aujourd'hui d'une nouvelle manière de voir le rapport de chacun dans l'échange de l'information. Cette dynamique de l'échange engendre, de fait, de nouveaux comportements. La liberté d'accès, la gratuité et le droit de réutilisation semblent aller de soi quand ils s'inscrivent dans la réciprocité.

Le 22 octobre 2003, à Berlin, la plupart des Directeurs Généraux des Établissements Publics à caractère Scientifique et Technologique (EPST) ont signé la *Déclaration de Berlin sur le Libre Accès à la Connaissance en Sciences exactes, Sciences de la vie, Sciences humaines et sociales*, dont l'objectif est de promouvoir Internet « comme instrument fonctionnel au service d'une base de connaissance globale de la pensée humaine ».

En signant cette déclaration, les responsables politiques chargés de la science, les institutions de recherche, les agences de financement, les bibliothèques, les archives et les musées se sont engagés à envisager un certain nombre de mesures. Ces mesures doivent permettre de « trouver des solutions aptes à soutenir le développement des cadres actuels, juridique et financier, en vue de faciliter un accès et un usage optimaux » d'Internet. Le texte reconnait aussi l'existence d'une possible contradiction entre les demandes de protection et de libre accès. Enfin, de cette déclaration, il ressort que le libre accès requiert l'engagement de chacun en tant que producteur de connaissances scientifiques ou détenteur du patrimoine culturel, ce

[3]Art 5 de la Loi n°82-610 du 15 juillet 1982 modifiée d'orientation et de programmation pour la recherche et le développement technologique de la France, aujourd'hui art. L 111-1 du code de la recherche. *JO* du 16-07-1982, p. 2273 et ss.

libre accès se faisant « dans le respect des droits des auteurs ou des titulaires ». Le libre accès doit donc être réglementé et modulé par les titulaires de droits. Les auteurs (ou l'institution) peuvent concéder un « droit gratuit, irrévocable et mondial d'accéder à l'œuvre » ou bien « une licence autorisant à copier, utiliser, distribuer, transmettre, montrer en public, réaliser et diffuser des œuvres dérivées, sur quelque support numérique que ce soit et dans quelque but responsable que ce soit, sous réserve de mentionner comme il se doit son auteur ». Peut être uniquement concédé « le droit d'en faire des copies imprimées en petit nombre pour un usage personnel ». La formalisation de ces autorisations peut se faire sous forme de licences de type **creatives commons* (autorisations d'utilisation données directement par les auteurs, sans contrepartie financière. Les auteurs peuvent en revanche, le cas échéant, poser des limites à cette utilisation en la réservant exclusivement à des usages à but éducatif).

Appliquées aux corpus oraux, ces licences peuvent être un moyen de mettre à la charge des futurs utilisateurs le respect des engagements souscrits par le chercheur créateur du corpus à l'égard de tous ceux qui ont contribué à son élaboration.

Les programmes de numerisation patrimoniale

Le contexte de la société de l'information a suscité de nombreuses initiatives publiques dans le dessein d'assurer la pérennisation de la mémoire culturelle. En 2001 à Lund, en Suède, un groupe de représentants nationaux des États membres de l'Union européenne, intéressés par les problèmes de numérisation, a élaboré un texte qui prône notamment : la mise en place de standards d'interopérabilité ; la diffusion de bonnes pratiques dont la gestion des *droits de propriété intellectuelle ; l'organisation de centres de compétences sur la numérisation dont les professionnels de l'information ont la responsabilité.

La question de la conservation des résultats de la recherche se pose aujourd'hui avec d'autant plus d'acuité que les résultats, mais aussi les matériaux mêmes qui ont servi à ces recherches sont sur des supports numériques. Comment assurer la « traçabilité » des différentes étapes du travail de recherche ? Que faut-il conserver ? Qui assurera cette conservation ? Dans quelles conditions ? Ces questions doivent aujourd'hui être posées et trouver des éléments de réponse pour chaque opération de recherche. Si des recommandations générales peuvent être données, cela n'exonère en rien les responsabilités de ceux qui initient une recherche dont l'un des objectifs, ou l'une des étapes consiste en l'élaboration d'un corpus oral.

2.3 Cadres juridiques

Le propos de ce guide qui s'adresse à des chercheurs n'est pas de traiter de toutes les techniques juridiques à appréhender (on renverra pour une présentation plus détaillée de certains des sujets abordés à des fiches spécifiques en annexe). Il s'agit de sensibiliser le lecteur et de l'inviter à se poser les questions nécessaires pour comprendre ses obligations mais aussi ses droits.

Quel peut être le statut juridique de chacun des corpus oraux constitués par les chercheurs ? Cette question peut *a priori* sembler théorique, mais nous ne pouvons pas l'occulter car c'est en fonction des réponses apportées qu'il sera possible de déterminer les conditions d'exploitation et de diffusion des corpus. Pour répondre à

cette question, il faut tout d'abord connaitre les conditions d'élaboration du corpus et de ses différentes composantes. Le corpus est-il constitué d'informations du *domaine public ? Est-il le produit d'une ou plusieurs créations intellectuelles susceptibles d'être protégées par le *droit d'auteur ? Les contenus du corpus sont-ils des *données personnelles ? Quels sont alors les droits des locuteurs ou des personnes concernées ?

Ces statuts juridiques déterminés et les droits qui en découlent une fois connus, il convient de s'enquérir des modalités de la gestion contractuelle de ces droits. Les titulaires des droits se sont-ils prononcés sur les conditions de mise à disposition et de réutilisation des corpus ?

Enfin, ce sont les questions de la responsabilité de tous ceux qui auront à intervenir dans la « vie du corpus » qui méritent attention : responsabilité des créateurs, responsabilités des hébergeurs, des diffuseurs, des archiveurs… (voir annexe).

Pour faciliter la démarche du chercheur, on donnera ici un aperçu sur quatre grandes questions qui reviennent de façon récurrente dans la constitution et la vie des corpus : qu'est-ce que le domaine public, c'est à dire « l'inappropriable » ? Quand est-il question de droit d'auteur à propos des corpus ? Comment assurer la protection des données personnelles au regard du traitement des informations constituant les corpus oraux ? Quelles sont les responsabilités des personnes en charge de la diffusion des corpus sur Internet ?

2.3.1 LE DOMAINE PUBLIC ET LE DROIT D'AUTEUR

QU'EST-CE QUE LE DOMAINE PUBLIC ?

Si l'expression « domaine public » est généralement connue de tous, l'acception juridique du terme peut être entendue dans des sens différents qu'il est important de préciser pour éviter des ambiguïtés ou des incompréhensions lors de la constitution des corpus oraux. Au sens juridique, le domaine public est un concept multiforme qui peut renvoyer autant à un lieu, qu'à un régime ou à des contenus.

Le domaine public peut, ainsi, être « l'endroit où la société civile s'efforce d'influer sur la manière dont les biens collectifs sont gérés et distribués ». C'est dans ce sens que l'UNESCO est à l'origine d'une véritable politique des contenus et développe une stratégie de promotion d'un domaine public fort, accessible en ligne et hors-ligne. Le domaine public recouvre non seulement les idées de liberté d'accès et de gratuité d'utilisation des données, mais aussi la possibilité pour chacun de les exploiter. Il se caractérise, en outre, par l'absence de monopole, puisque les informations qui tombent dans le domaine public deviennent *de facto* des « choses communes ».

En revanche, deux types d'informations peuvent être distingués : celles qui sont nées dans le domaine public et celles qui y sont « tombées ». Les idées, la langue, les textes de loi et tous les éléments qui fondent le patrimoine commun d'une communauté donnée, constituent, de par leur nature, le « fonds commun » du domaine public. Ce fonds commun reste pourtant difficile à délimiter. Les enregistrements linguistiques suscitent ainsi de nombreuses hésitations. Mis à part les droits de celui qui a enregistré, le contenu d'une langue, son expression phonique, font-ils ou non partie du domaine public ? La question peut aussi se poser à l'égard des traditions et des

coutumes. En outre, ce fonds commun est-il universel ou bien seulement commun à une petite communauté ? Aujourd'hui, il fait de plus en plus l'objet de revendications identitaires qui soulèvent de nouvelles interrogations.

Au-delà d'un certain délai, les œuvres protégées par le *droit de la propriété intellectuelle, notamment par le droit d'auteur ou les brevets, finissent par entrer dans le domaine public. Le droit d'auteur, par exemple, protège les œuvres soixante-dix ans après la mort de leur auteur. En droit français, à l'expiration de ce délai, d'autres types de protection peuvent subsister sur les œuvres de l'esprit : les droits patrimoniaux d'une part ; les attributs imprescriptibles du *droit moral d'autre part. Par conséquent, certains éléments du domaine public peuvent encore bénéficier de la protection du droit moral.

Ces distinctions font apparaitre deux types de situations apparemment opposés : soit les corpus sont constitués d'œuvres du domaine public ne pouvant faire l'objet d'une appropriation (de par leur nature ou du fait de l'expiration du délai de protection) et de ce fait sont libres de droit, soit les corpus sont soumis au droit d'auteur et donc soumis aux autorisations requises. En réalité, nous l'avons vu, il existe une possibilité intermédiaire où les corpus protégés par le droit d'auteur peuvent être mis en libre accès dans le cadre d'une licence accordée par les titulaires de droits autorisant l'utilisation et l'exploitation des résultats. Sans être dans le domaine public, ces corpus sont – de par la volonté de leurs créateurs – libres d'accès et d'utilisation. Néanmoins, si les créateurs peuvent renoncer à exercer leurs *droits patrimoniaux, il ne leur est pas possible de renoncer à leur droit moral, qui reste imprescriptible.

Le droit d'auteur et les corpus

Quelles sont les conditions pour qu'un corpus soit protégé ? Il y en a trois.

Il faut en premier lieu qu'il corresponde à l'exigence d'une *activité créatrice* : un travail de compilation d'informations n'est pas protégé en soi.

Pour être protégé, il est par ailleurs indispensable que le corpus ait une *forme définie*. Ce qui est protégé, ce n'est pas le contenu du corpus mais son enveloppe, son architecture.

Enfin, la forme du corpus doit répondre à la condition d'être *originale*. Que signifie l'originalité d'un corpus ? L'originalité de nombreuses créations de l'ère du numérique, comme les logiciels ou les bases de données, ne peut être appréciée que d'après des critères objectifs. Il semble qu'il en soit de même des corpus oraux, ceux-ci pouvant le plus souvent être assimilés à une base de données. C'est alors, le plus souvent, le fait que le corpus soit ou non copié et révèle un minimum d'activité créative qui servira de critère pour déterminer s'il est ou non original (et non pas uniquement la prise en compte de l'empreinte de la personnalité de son auteur).

« Il n'y a pas de place pour les droits des auteurs quand il n'y a pas d'auteur »

L'auteur est en principe la (ou les) personne(s) physique(s) sous le nom de laquelle (ou desquelles) l'œuvre est divulguée. Le travail scientifique suppose l'intervention de nombreux acteurs dont bon nombre sont susceptibles de revendiquer la qualité d'auteur sur les résultats de la recherche.

Certains corpus oraux, comme les autres produits de la recherche, peuvent rester l'œuvre d'un auteur unique, alors que d'autres peuvent être l'œuvre de plusieurs auteurs. Dans le cas de pluralité d'auteurs, le droit distingue les œuvres de collaboration des œuvres collectives. Pour les premières, chaque co-auteur dispose des mêmes prérogatives. D'autres œuvres – telles que les bases de données ou les dictionnaires – peuvent être qualifiées d'œuvre collective lorsqu'elles sont créées

> *« sur l'initiative d'une personne physique ou morale qui l'édite, la publie et la divulgue sous sa direction et sous son nom, et dans laquelle la contribution personnelle des divers auteurs se fond dans l'ensemble »*[4].

Dans ce dernier cas, c'est la personne physique ou morale qui a pris l'initiative de l'œuvre qui dispose des droits d'auteur. Par ailleurs, le contexte de la création ou le statut de l'auteur peuvent avoir des incidences sur la détermination du titulaire des droits d'auteur. L'œuvre a-t-elle été créée dans le cadre d'une mission de service par un employé ou un fonctionnaire ? Quels sont les droits respectifs de l'auteur et de son employeur ? Si la question est résolue le plus souvent par le contrat de travail, elle reste plus délicate quand le créateur est un fonctionnaire. En effet, depuis plusieurs années, deux logiques s'affrontent, celle de la reconnaissance d'un droit de la personne créatrice d'une part, et d'autre part celle de la reconnaissance uniquement d'un droit de l'État sur les créations de fonctionnaires. La transposition de la directive sur les droits des auteurs dans la société de l'information a incité les pouvoirs publics à proposer une voie médiane qui reconnait à la fois le droit des auteurs et les droits de l'employeur « État » quand la création de l'œuvre s'inscrit dans l'exécution de la mission de service public. Si ce texte est voté par le Parlement, les droits des auteurs pourraient naitre sur la tête du fonctionnaire.

En contrepartie, tous les droits d'exploitation de l'œuvre pour les besoins de sa mission seraient cédés à son employeur État (droit de communiquer ou de diffuser pour la mission). Toutefois, dans le cas d'une exploitation commerciale, l'auteur personne physique recouvrera ses droits avec l'obligation d'accorder un droit de préférence à son employeur et la possibilité d'être intéressé à l'exploitation commerciale. Ce texte n'est pas sans soulever des débats et des interrogations. Comment sera déterminé le périmètre de la mission de service des chercheurs qui interviennent dans l'établissement du corpus ? Comment distinguer l'exploitation pour la mission du service et l'exploitation commerciale quand – nous l'avons vu précédemment – le chercheur a pour mission de communiquer les résultats de sa recherche et de les valoriser par la publication ?

QUELS DROITS POUR LES AUTEURS SUR LES CORPUS ORAUX ASSIMILABLES A DES ŒUVRES ?

Il convient de distinguer les droits patrimoniaux des prérogatives du droit moral. On rappellera aussi que la loi pose quelques limites aux droits exclusifs des auteurs.

Les droits patrimoniaux se résument en un droit exclusif au profit de l'auteur (ou des titulaires) ou des ayants droit (bénéficiaire d'une cession, héritiers...) d'autoriser ou

[4] Art. L. 113-2 du CPI.

interdire la reproduction ou la communication au public de l'œuvre protégée. Si le corpus oral est une œuvre, toute reproduction (la numérisation est pour le droit une reproduction) et toute mise à disposition du public (sur un site Internet comme sur tout autre support) nécessitent l'autorisation expresse de l'auteur ou du titulaire de droit.

Quant aux prérogatives du droit moral, toujours attachées à la personne physique créatrice de l'œuvre protégée, elles sont au nombre de quatre : *le *droit de divulgation, le *droit de repentir et de retrait, le *droit à la paternité* et *le *droit au respect de l'œuvre*. Chacun de ces droits est applicable aux corpus oraux. L'auteur du corpus (au titre de son droit de divulgation) peut décider du moment ou des modalités de la mise à disposition du corpus au public, le dépôt aux archives ne valant pas nécessairement divulgation. Un corpus inédit ne peut donc être mis à la disposition du public sans l'autorisation de son auteur. Le chercheur auteur qui refuse de divulguer le corpus qu'il a créé est dans son droit (au titre du droit d'auteur), même si par ailleurs il peut être sanctionné administrativement pour ne pas avoir exécuté sa mission de service public qui est de communiquer les résultats de sa recherche. Le droit de repentir ou de retrait peut s'exercer aussi sur un corpus oral, ces regrets ne pouvant porter que sur le contenu intellectuel de l'œuvre et non pas sur les conditions matérielles de sa diffusion. Si le droit à la paternité est en soi facile à comprendre, on peut se demander ce que signifie le droit au respect de l'œuvre appliqué à un corpus oral. Ce droit correspond autant au respect de la forme de l'œuvre (pas de suppression, d'adjonction ou de modification…) qu'au droit au respect de l'esprit de l'œuvre (altération de la finalité du corpus).

Comme tout monopole, les droits exclusifs des auteurs souffrent des limites. On peut en premier lieu rappeler qu'ils sont limités dans le temps et qu'au-delà de cette limite, les œuvres tombent dans le domaine public (cf. *supra*). Ces limites peuvent aussi trouver leurs justifications dans le type d'usage qui est fait des œuvres. On parlera alors d'exceptions au droit d'auteur qui sont justifiées par les finalités ou le contexte ou encore l'intérêt général.

Enfin, *le droit à la copie privée* ou le droit de citation concernent directement les corpus oraux (voir fiche *Droit de citation*).

2.3.2 LE RESPECT DE LA VIE PRIVEE

LE RESPECT DE LA VIE PRIVEE DANS LA CONSTITUTION, L'EXPLOITATION, LA DIFFUSION ET LA CONSERVATION DES CORPUS

La création d'un corpus passe le plus souvent par la collecte de données. Celles-ci pouvant être des données personnelles, cette collecte doit être faite dans le respect de la loi *Informatique et libertés* : licéité et loyauté, information préalable, obtention du consentement des personnes concernées (voir fiche *Consentement*), respect des finalités annoncées[5]… Quand il s'agit de finalités de recherche, faut-il entendre de façon restrictive une recherche spécifique identifiée comme telle, ou peut-on entendre de façon plus large l'expression « finalités de recherche » ? Le problème se

[5] http://www.cnil.fr/fileadmin/documents/approfondir/textes/CNIL-78-17_definitive-annotee.pdf

pose quand, une fois le corpus constitué et exploité scientifiquement par les chercheurs qui ont été à l'origine de sa création, on envisage une réutilisation et de nouvelles exploitations scientifiques. La recherche scientifique bénéficie aujourd'hui d'une exception au principe général avec l'application de ce que l'on appelle *l'extension de finalité*. Toutefois, toute nouvelle exploitation scientifique devra se faire en respectant les formalités préalables à tout traitement (nouvelle procédure de déclaration ou d'autorisation) et les principes posés par la loi (information, consentement et/ou autres garanties appropriées…).

Même si la diffusion des corpus et leurs nouvelles exploitations sont faites dans les conditions requises, se pose le problème de la conservation des données personnelles.

Si les données sont « anonymisées » de manière irréversible, elles sortent du champ de la loi et peuvent être conservées (voir fiche *Données personnelles et anonymisation*). Toutefois dans la recherche, le besoin de « traçabilité » nécessite souvent de sauvegarder les données personnelles.

Et pourtant, en principe, sur le fondement du **droit à l'oubli*, les données personnelles ne doivent pas être conservées au-delà de la durée initialement prévue, et quand la finalité initiale annoncée lors de la collecte de ces informations n'a plus de raison d'être, ces données doivent être détruites. Cela veut-il dire qu'il n'est pas possible de conserver certains corpus contenant des données personnelles si celles-ci n'ont pu être anonymisées ? Non, mais il ne peut s'agir que de cas exceptionnels où le maintien des données personnelles se justifie pour des raisons scientifiques. Dans ces cas, les corpus oraux pourraient bénéficier – en tant qu'archives publiques – d'une dérogation au droit à l'oubli permettant leur conservation au-delà de la durée prévue, en vue d'un traitement à des fins de recherche, historique ou scientifique. C'est alors la loi sur les archives qui fixera les conditions de leur mise à disposition en libre accès (délais plus ou moins longs – 60 à 150 ans[6] – suivant le degré de sensibilité des données contenues dans le corpus).

QUELLES SONT LES RESPONSABILITES DES PERSONNES CHARGEES DE LA DIFFUSION DES CORPUS SUR INTERNET ?

La diffusion des corpus oraux sur Internet peut être assimilée à « l'édition d'un service de communication au public en ligne ». Il est donc important d'apprécier les obligations et responsabilités des éditeurs d'un service de communication au public en ligne (voir fiche *Responsable du traitement*).

[6] Voir art. 213-2 du code du patrimoine (ancien article 7 loi de 1979).

3 LA DEMARCHE
CONSTITUTION, EXPLOITATION, CONSERVATION, DIFFUSION

3.1 EXPLICITER LA DEMARCHE

Les objectifs, notamment scientifiques, liés à la constitution, à l'exploitation, à la conservation et à la diffusion des corpus oraux sont très diversifiés, et le respect de ceux-ci, ainsi que leur hétérogénéité, impliquent que soit reconnue la diversité des *démarches* qui peuvent être adoptées par les chercheurs et par les responsables de la diffusion et de la conservation de ces corpus.

Le *Guide des bonnes pratiques* n'a pas vocation à contraindre cette démarche en prescrivant une méthodologie type, mais souhaite fournir toutes les informations nécessaires au repérage des points juridiques et éthiques « sensibles ». Seule l'identification précise et détaillée des éléments de la situation en jeu et notamment de la forme des données et de leurs supports, des pratiques de terrain, mais aussi des différentes étapes de leurs traitements, permet d'apporter à la fois des éléments de réponses juridiques correspondant à la situation, et une évaluation des « risques » éventuels. Enfin, une analyse réflexive sur la démarche liée à la constitution et aux traitements des corpus oraux est le premier élément de l'élaboration d'une éthique reconnue par l'ensemble d'une communauté scientifique.

3.2 ÉLEMENTS DE LA SITUATION EN JEU

Les enregistrements qui constituent les données primaires de l'enquête linguistique sont loin de former un objet uniforme. Ainsi, un conte enregistré sur une bande magnétique lors d'une cérémonie traditionnelle sur la place d'un village est un objet scientifique et patrimonial fort différent de l'enregistrement numérique d'un texte lu par un « informateur rémunéré » dans les locaux d'un laboratoire universitaire, des réponses à un questionnaire enregistrées sur minidisque par un chercheur au domicile de la personne interrogée ou bien encore d'une conversation spontanée non sollicitée par les chercheurs, se déroulant dans un café et filmée par une ou plusieurs caméras.

Il convient donc, dans un premier temps, d'identifier les éléments qui caractérisent les données récoltées en situation :

- le *type de données* qui constitue le corpus et leurs supports (d'enregistrement, mais aussi de stockage pour exploitation, et de conservation),
- les *différentes techniques* employées par les chercheurs pour récolter les données,
- la définition des *participants* et de leur rôle,
- la catégorisation des *lieux* de la collecte.

3.2.1 CORPUS ET TYPE DE DONNEES

Si la volonté de « capturer » la parole est fort ancienne, c'est récemment que les avancées technologiques et la recherche (notamment en linguistique) ont permis de concevoir les enregistrements comme de véritables « données ». Ainsi, l'Alphabet Phonétique International est un exemple de système « d'enregistrement »

alphabétique inventé par des linguistes afin de normaliser le codage de la transcription phonétique et/ou phonologique de la parole. L'histoire moderne des enregistrements audio et vidéo se déroule au fil des transformations des modes d'enregistrement comme à travers celui des supports d'inscription utilisés.

Les modes d'enregistrement

Le mode d'enregistrement analogique a été le premier à être utilisé pour l'enregistrement et la conservation du son. Il code les variations mesurées sous forme de signaux obéissant à la même loi de variation que celle qui régit leur propagation dans un milieu naturel. Depuis quelques décennies, c'est plutôt un mode d'enregistrement numérique qui est privilégié. Dans ce mode, les mesures ponctuelles de la pression de l'air sont régulièrement effectuées (échantillonnage). Ces mesures sont ensuite codées sous forme d'une valeur numérique exprimée dans une échelle de référence puis sont représentées sur le support de stockage sous la forme d'une suite organisée d'unités binaires.

Les supports d'enregistrement

- *Supports physiques*

Les premiers supports modernes permettant la conservation de la parole ont été les supports physiques. Ce terme est dû au fait que les variations de pression mesurées par un appareil (microphone) sont inscrites physiquement dans la matière du support. On compte parmi eux les anciens cylindres, les disques vinyles, etc. Ces supports conservent dans la matière qui les compose (vinyle, cire, etc.), sous la forme d'un sillon ondulé, une image analogique des variations de pression mesurées. Ces supports utilisés au siècle dernier sont pratiquement abandonnés. Ils posent aujourd'hui des problèmes d'accès et de conservation.

- *Supports magnétiques*

Les supports magnétiques sont apparus plus tardivement, dans la deuxième moitié du 20e siècle. Différents supports de stockage ont été et sont encore utilisés de nos jours (fil, bande, disque) dans différents conditionnements (bobine, cassette, cartouche, etc.). Le principe ici repose sur la rémanence des particules magnétiques réparties tout le long du support (i.e. la propriété qu'ont ces particules de conserver durablement leur aimantation). Cette aimantation des particules pourra, suivant les modes d'enregistrement, coder des informations sous forme binaire (comme dans les supports disque-dur, cassette DAT, disquette informatique, etc.) ou bien des informations sous forme analogique (comme dans les mini-cassettes audio, les cassettes VHS, etc.). Une partie de ces supports est destinée à être utilisée sur du matériel informatique, une autre sur du matériel audio/analogique. Ici encore, comme pour tout support, ceux-ci se dégradent inexorablement au cours du temps. Ces supports demeurant encore très populaires, l'accès aux outils qui en permettent la lecture et l'écriture reste aisé.

- *Supports optiques*

Les supports optiques sont les derniers apparus ; ils sont connus principalement dans leur forme de Compact-Disc (CD-audio, CD-ROM, etc.). La technologie repose sur les propriétés optiques des composants, à savoir par exemple pour les

CD-audio, la capacité des alvéoles qui les composent de réfléchir la lumière d'un faisceau laser. Ces supports sont principalement utilisés pour stocker des données numériques (exception faite de certains disques laser peu populaires, et des films argentiques peu utilisés pour l'enregistrement sonore). Une grande partie de ces supports est destinée à une utilisation sur des équipements informatiques, ce qui facilite l'accès, le transfert et le traitement des données. Les problèmes de conservation sont les mêmes que pour tout type de support, même s'ils ne sont pas sensibles aux mêmes agressions (lumière, chaleur, champs magnétique, humidité, etc.). Comme les supports magnétiques, ils ont l'avantage d'être récents et populaires, ce qui rend leur utilisation facile aujourd'hui.

Il existe d'autres types de supports mélangeant par exemple les techniques optiques et magnétiques. (voir fiche *Supports pour enregistrer et archiver le son*).

Les criteres de choix

La conservation des supports posant de toute façon des problèmes similaires quel que soit le type de support choisi, les critères de choix du bon support d'enregistrement puis de conservation reposeront plutôt sur la qualité du codage, la facilité d'accès et de traitement ainsi que sur la possibilité de reproduire son contenu sans perte d'information. On privilégiera donc les supports numériques par rapport aux supports analogiques, car ils peuvent être dupliqués à l'identique et à l'infini. On privilégiera aussi les supports informatiques en raison de la panoplie des outils que l'informatique offre pour la gestion, l'accès à du matériel de lecture, la diffusion et le traitement des données (cryptage, techniques d'anonymisation, etc.) tout en considérant que ces outils posent encore de nombreux problèmes de standardisation (par exemple en ce qui concerne le choix des logiciels, des formats, des codecs de compression). Enfin, pour la conservation, un support qui ne peut pas être effacé est aussi, peut être, une bonne garantie pour éviter les accidents malencontreux.

Le choix d'un format qui permet la reproduction à l'identique garantit une forme de pérennité aux données. Elle met en cause la notion même d'« original » qui se réfère alors moins au support qu'aux données elles-mêmes.

Standardisation des annotations

Les corpus oraux sont en général composés d'enregistrements audio ou vidéo et d'annotations de ces derniers.

- *Données primaires vs. données secondaires*

On distingue généralement entre données primaires et données secondaires :

- les *données primaires* sont constituées par les *enregistrements*, ayant un lien le plus proche possible avec l'évènement documenté. Elles comprennent aussi les autres objets recueillis dans le contexte de l'action, comme les documents lus ou écrits durant l'action enregistrée, les objets manipulés, les images consultées, etc. Elles comprennent aussi les traces informatiques laissées par l'activité.
- les *données secondaires* sont constituées par la série de descriptions, transcriptions, annotations qui viennent enrichir les données primaires et qui sont souvent fournies après coup et sur la base des données

primaires. Elles comprennent aussi les métadonnées, les conventions de transcription, les autorisations des participants, etc.

La distinction entre données primaires et données secondaires est utile notamment pour différencier des niveaux d'interprétation et souligner l'importance du retour aux données primaires et de leur disponibilité. Ainsi une analyse porte sur la bande audio ou vidéo et non exclusivement sur la transcription, même si celle-ci est un adjuvant important sans lequel l'analyse serait probablement impossible. C'est dans ce sens que sont développés les outils d'alignement entre la source sonore/visuelle et le texte de la transcription. Toutefois, cette distinction entre données primaires et données secondaires a ses limites : elle ne doit pas faire oublier le fait que tout enregistrement est le fruit de décisions à la fois techniques et théoriques – concernant par ex. le choix du moment à enregistrer et la délimitation du segment enregistré, le choix du cadrage et de l'optique pour la vidéo, du positionnement et de l'orientation du micro pour l'audio – qui reposent sur une connaissance préalable de l'activité enregistrée. Les « données » ne sont jamais « offertes » ni « (re)cueillies » mais elles sont activement produites par les chercheurs (Mondada, 2006).

- *Explicitation de la structure des données*

Pour l'écriture des annotations, on utilise des formalismes d'expression qui permettent à la fois de coder le contenu des commentaires ainsi que d'expliciter de quel type de commentaire il s'agit. Par exemple, dans les bases de données relationnelles, on va utiliser des tables comportant des champs avec des noms (i.e. *pos* pour « *part of speech* ») qui vont servir à stocker des valeurs (par ex. « verbe ») exprimées dans des types particuliers de structure (chaîne de caractères, nombre, etc.).

Un formalisme alternatif et très largement utilisé dans le domaine de l'annotation textuelle est celui apporté par la grande famille des langages de balisage de textes. Ce formalisme délimite les commentaires par des marques formelles (i.e. balises) indiquant de quel type de commentaire il s'agit. Il existe aujourd'hui un consensus assez vaste, toutes disciplines confondues, sur l'adoption du récent langage de balisage de texte XML comme formalisme de structuration et d'échange de documents (voir fiche *Codages et formats*).

- *Standardisation/Normalisation*

Alors que le choix d'un formalisme permettant d'exprimer l'ensemble des annotations ainsi que d'expliciter leur structure est indispensable, il n'est pas pour autant suffisant pour permettre l'échange ou la conservation d'un document. Pour échanger ou conserver un document, il faut que le langage utilisé pour coder sa structure ainsi que le contenu de ses commentaires soit commun entre les participants (dans le cadre d'un échange) ou qu'il puisse rester connu au cours du temps (dans le cadre d'une conservation à long terme). Dans le contexte d'un document utilisant un langage de balisage de textes, les noms des éléments de structure (balises, attributs...) doivent être connus et leur définition acceptée et partagée, ainsi que l'ensemble des contraintes (enchainement de balises, vocabulaires contrôlés, caractère optionnel ou obligatoire de certaines structures…).

Quand un grand nombre de personnes ou toute une communauté parviennent à s'entendre sur un langage commun, on parle de standardisation. C'est ce qui s'est passé par exemple avec l'Alphabet Phonétique International (API). Alors que la standardisation est nécessaire pour l'échange, la conservation à long terme réclame des garanties sur la transmission et sur l'accès à la documentation des langages communs mis en place. A ce titre, les organismes de normalisation doivent pouvoir apporter une certaine pérennité aux normes qu'ils mettent en place, ainsi qu'une indépendance vis-à-vis des intérêts privés. Ils doivent aussi être représentatifs de l'intérêt général. A ces conditions, il sera avantageux, partout où elles existent, d'utiliser des normes pour le codage et le formatage des données. On peut citer à titre d'exemple le codage des caractères ISO-10646, plus connu sous le nom d'Unicode, qui est un code-caractère qui se veut universel et prend en compte la plupart des écritures du monde, y compris l'Alphabet Phonétique International. Pour le codage de l'analyse linguistique, il sera intéressant de lire les recommandations de la Text Encoding Initiative (TEI) qui propose des analyses pour des structures de données telles que les dictionnaires, les poèmes ou la transcription de la parole. Il sera aussi très utile de suivre les progrès du groupe de travail de l'ISO sur la gestion des ressources linguistiques TC37 SC4 (voir fiche *Codages et formats*).

Ainsi, les principes qui doivent guider le choix d'une technologie plutôt que d'une autre pour l'annotation peuvent être résumés en quatre questions :

- Cette technologie permet-elle de *coder de manière explicite* toutes les annotations ?
- Cette technologie présente-t-elle un *caractère propriétaire* ou *une limite légale* qui empêcheraient de partager les annotations avec d'autres (formats propriétaires, techniques basées sur des brevets, etc.) ?
- Cette technologie est-elle *acceptée par la communauté* avec laquelle l'échange des données est envisagé ?
- Cette technologie a-t-elle fait l'objet d'une *normalisation* ?

3.2.2 TECHNIQUES D'ENQUETE

RECUEIL ET PRODUCTION DE DONNEES

Les enquêtes linguistiques n'ont pas toujours donné lieu à des enregistrements pour des raisons techniques (les premiers outils d'enregistrement de la parole ont à peine plus d'un siècle) mais aussi méthodologiques et théoriques. Ainsi, les questionnaires écrits, la prise de notes, le recours à l'intuition et/ou à l'observation du chercheur ont et sont encore des outils de description utilisés par les linguistes. La possibilité d'enregistrer la parole et l'évolution des techniques (miniaturisation des appareils, qualité du signal enregistré, numérisation et traitements informatiques des données sonores et vidéo), ont néanmoins permis aux enquêtes de terrain de développer des méthodologies qui restent toutefois très différentes ne serait-ce que par la diversité des domaines scientifiques concernés (dialectologie, sociolinguistique, analyse conversationnelle, psycholinguistique, linguistique de l'oral, traitement automatique de la parole, ethnolinguistique…). Cependant les recherches sur la méthodologie de l'enquête ont conduit les chercheurs à considérer les données enregistrées comme

étant le produit de la situation d'enquête par opposition à une conception de données préexistantes simplement (re)cueillies (Cameron *et al.*, 1991).

Enfin, les techniques d'enquête ont un rôle important dans la possibilité qu'elles offrent (ou qu'elles n'offrent pas) de contrôler les données fournies aux chercheurs par la personne interrogée. La suite de ce chapitre est consacrée à un inventaire des différentes techniques d'enquête utilisées lors de la constitution de corpus oraux.

Le questionnaire

Le questionnaire oral enregistré peut revêtir différentes formes ; il est le plus souvent composé de questions fermées ou semi-ouvertes et de listes de termes lexicaux ou de textes préparés par le chercheur. Seul le cas des textes préexistant au questionnaire peut poser éventuellement la question de la propriété intellectuelle (comme par exemple la lecture d'un texte protégé par le droit d'auteur, ou une production dont l'originalité du contenu serait protégée). Dans les autres cas il s'agit de capter, notamment, les variations, les régularités et les perceptions de ces régularités par le questionné, en référence à un système linguistique commun.

Le degré de sensibilité des informations collectées est le plus souvent prévisible, puisque c'est le chercheur qui élabore le questionnaire et qui peut donc évaluer les risques selon la nature des questions. Toutefois, des questions apparemment anodines peuvent aussi receler des enjeux, insoupçonnés par le chercheur, surgissant du contexte particulier de l'enquête. Il convient en outre de souligner que le questionnaire contient plus que toute autre technique la marque de l'acte de questionnement et de la *prise* du chercheur (Encrevé, 1983) et donc potentiellement un sentiment d'évaluation, même si celui-ci est souvent atténué par la possibilité explicitement offerte de ne pas répondre à tout ou partie des questions (pour une analyse des situations de questionnaire cf. Achard 1991). Enfin, soulignons ici un point qui concerne de nombreuses situations d'enquête, mais qui est particulièrement lié au questionnaire : celui-ci contient souvent une partie consacrée au recueil de données personnelles (âge, catégorie socio-professionnelle…) dans le but de dresser le profil sociologique de l'enquêté.

L'entretien

L'entretien est composé de questions ouvertes, l'objectif étant principalement de recueillir une quantité importante de données linguistiques. L'entretien suppose toujours un guidage de la part de l'enquêteur, qui peut être plus ou moins fort (de l'entretien directif au semi-directif, voire au non-directif ; du plus standardisé au moins standardisé), le rapprochant ainsi du questionnaire oral ou de l'interaction moins contrainte (Maynard *et al.*, 2002, Houtcoop-Streenstra, 2000). Bien que dans l'entretien le chercheur introduise souvent les catégories et les thèmes qu'il souhaite voir traités par les informateurs, la méthodologie des chercheurs peut aussi requérir, par souci de collecter les productions les plus naturelles possibles, que l'objet de la recherche ne soit pas précisé en détail avant l'entretien, et pose donc le problème du choix du moment et du contenu des informations fournies aux interviewés (cf. Mondada 2001).

Du point de vue juridique, les entretiens sont le plus souvent des sources de données et d'informations concernant la vie privée de l'interviewé ou de personnes mentionnées dans le cours de l'entretien et sont donc à protéger en tant que tels.

Le recueil de contes, chants…

Le recueil de contes, chants et productions orales de cultures traditionnelles est une pratique fréquente dans les domaines de la description des langues à tradition orale et de l'ethnolinguistique notamment. Outre l'importance de contextualiser ces chants, contes et récits (des significations implicites dans un contexte culturel peuvent échapper ou paraitre anodines dans un autre), deux éléments sont principalement à prendre en compte : la propriété intellectuelle de productions traditionnelles d'une communauté, et les conditions de recueil, souvent liées à des activités sociales dans un cadre public ou privé.

Les recits de vie

Les récits de vie sont couramment sollicités lors de recherches en anthropologie, en histoire, en ethnolinguistique, mais aussi en dialectologie, et dans de nombreux autres domaines (Guillaumou *et al.* 1997). Ces types d'enregistrement représentent nécessairement une source importante de données personnelles concernant l'auteur du récit et de tierces personnes, qui peuvent éventuellement être associées à un contexte social ou historique particulièrement sensible, notamment quand le récit personnel fait écho à un évènement vécu par une ou plusieurs communautés.

Ainsi, même dans le cas de recherche sur des phénomènes exclusivement linguistiques, les propos contenus dans des récits de vie et la question de l'impact de leur diffusion dans l'espace public ne peuvent échapper à la responsabilité du chercheur qui les sollicite et les exploite.

De plus, les conditions d'exploitation et de diffusion de ces récits peuvent se faire dans un contexte social très différent de celui, très particulier, qui a marqué le recueil et qui a souvent lieu dans un cadre précis et grâce à une relation privilégiée entre le chercheur et le témoin.

Enfin la question de la propriété intellectuelle d'un récit de vie et du droit moral inaliénable peut s'avérer particulièrement pertinente dans le cas de récits originaux.

L'enregistrement en laboratoire

Les enregistrements en laboratoire selon un protocole expérimental sont utilisés en sciences du langage notamment dans les domaines de la psycholinguistique, de la phonétique et du traitement automatique de la parole. Ainsi certains corpus intéressent directement la recherche appliquée et les entreprises concernées par l'ingénierie linguistique, et font donc parfois l'objet de financement partiel ou total sur des fonds privés.

De même que pour les questionnaires, sauf dans les cas particuliers de textes soumis aux droits d'auteur, les productions des participants selon un protocole expérimental élaboré par les chercheurs ne semblent pas devoir être concernées par le droit de la propriété intellectuelle (sauf les cas particuliers). La situation particulière de la personne enregistrée reste toutefois à rapprocher de tous les cas de recherches expérimentales sur personne humaine.

L'enregistrement d'activites provoquees

Il s'agit principalement d'enregistrements d'activités dans le contexte ordinaire des acteurs sociaux concernés, même si les consignes proviennent du chercheur (activités proposées à des enfants en milieu scolaire, tâches simulées en situation professionnelle, etc.). Cette situation combine à la fois les caractéristiques d'enregistrements selon un protocole expérimental (qui est de la responsabilité du chercheur) et les caractéristiques du contexte ordinaire en milieu écologique ; elle offre donc un double cadre contrôlable par le chercheur. Cette intervention explicite du chercheur (dont le rôle peut être clairement identifié par les participants) facilite les conditions d'obtention d'un consentement éclairé ; toutefois une attention particulière doit être apportée au milieu professionnel qui peut contraindre le consentement (confidentialité…).

L'enregistrement d'activites dans leur contexte ordinaire

Les recherches en sociolinguistique, analyse conversationnelle et analyse des usages des technologies (Computer Supported Cooperative Work ; Dialogue Homme Machine), s'intéressent au recueil de données en situation d'activité non orchestrée par le chercheur et non provoquée par ses consignes. Il s'agit ici d'activités telles qu'elles ont lieu de manière ordinaire, même en l'absence du chercheur. Ces activités peuvent être fort variées : réunions, activités professionnelles, demandes de renseignements, interactions téléphoniques, etc. Les techniques de collecte sont également très différentes. Elles vont de l'observation participante à l'enregistrement autorisé, en passant par l'utilisation « de personnes ressources » choisies au sein du groupe de pairs observés et en particulier chargées de porter le dispositif d'enregistrement (micro, éventuellement caméra).

L'objectif partagé par ces techniques est la recherche de données en situation naturelle et suppose donc une méthodologie s'efforçant de minimiser les effets produits par les dispositifs d'enregistrement (Heath, 1997 ; Jordan & Henderson, 1995). Il y a donc de fortes probabilités pour que les données contiennent des informations sensibles au regard de la protection de la vie privée. Les modalités du recueil du consentement doivent en tenir compte et s'y adapter.

La reprise d'enregistrements

Certains corpus constitués d'enregistrements produits par des acteurs différents des enquêteurs pour des finalités autres que scientifiques ou autres que les finalités évoquées lors du recueil de consentement peuvent donner lieu à une reprise dans un but de recherches linguistiques ou à une mise à disposition du public à des fins patrimoniales, mémoriales ou politiques (c'est ainsi que par exemple la Brigade des Pompiers de New York a mis à disposition en aout 2005 les enregistrements des communications radio durant l'attentat du 11 septembre 2001). Ces corpus sont donc caractérisés par l'absence de consentement pour la nouvelle finalité et par le fait que les propos archivés n'ont pas été produits en connaissance de cette finalité, mais dans un autre cadre et avec d'autres objectifs. Ainsi, lors d'interviews ou de séminaires – enregistrés par exemple dans le but d'une diffusion des contenus transcrits – l'autorisation de diffusion peut concerner les propos transcrits et validés, et non une reprise ultérieure des enregistrements.

La reprise d'enregistrements mediatiques

La reprise d'enregistrements médiatiques est un cas particulier de la catégorie précédente, qui offre la particularité de concerner des données produites dans un cadre de diffusion publique.

Là encore, si le contenu des enregistrements est protégé par le droit d'auteur (par exemple dans le cas d'une production originale), le recueil du consentement est un préalable à toute exploitation. Une exception existe toutefois pour un laps de temps déterminé, lorsqu'il s'agit de discours destinés au public et prononcés en public, tels que spécifiés dans les lignes suivantes :

> *Code de Propriété Intellectuelle, art 122.5 :*
>
> *La diffusion, même intégrale, par la voie de presse ou de télédiffusion, à titre d'information d'actualité, des discours destinés au public prononcés dans les assemblées politiques, administratives, judiciaires ou académiques, ainsi que dans les réunions publiques d'ordre politique et les cérémonies officielles*[7].

Rappelons que l'enregistrement personnel d'une émission correspond à une licence légale pour cette copie strictement réservée à l'usage privé. La représentation de celle-ci ne peut se faire que dans le cadre du « cercle de famille ». De même pour une cassette ou un dévédé du commerce, le droit de copie (avec ses limites) ne donne aucun droit d'exploitation.

Enfin précisons que le caractère public du contexte de diffusion médiatique ne signifie pas une restriction de la protection des données personnelles.

La diversité des techniques utilisées pour la collecte de données, défini autant de *situations* qui mettent en évidence des *participants* dont le rôle est le premier élément de catégorisation.

3.2.3 Role des participants

Les participants à l'enquête et aux activités enregistrées sont catégorisables de différentes manières, qui toutes éclairent de façon spécifique ce qu'ils font et ce qu'ils disent (Sacks, 1972). Ainsi les participants à une situation d'enregistrement peuvent-ils être à la fois considérés comme des enquêtés (si l'on rapporte la situation au fait qu'elle est un objet d'enquête) et comme des acteurs sociaux – dont la caractérisation précise dépend du contexte, de l'activité, des formes d'engagement et de participation, impliquant à la fois l'histoire sociale des personnes et l'accomplissement local de leur rôle, mais aussi de leur identité durant la rencontre. Selon la manière dont les chercheurs eux-mêmes traitent ces multiples catégories, différentes conséquences peuvent apparaitre à la fois pour l'objet de l'enquête et pour l'évaluation du caractère plus ou moins sensible de l'activité.

Categories de participants

La terminologie très variée utilisée dans la littérature pour définir les catégories de participants à une enquête révèle des implications éthiques et théoriques diverses

[7] Art. 122.5 du code de la propriété intellectuelle.

(Cameron *et al.*, 1991). Voici une liste non exhaustive des termes utilisés dans différents contextes de recherche pour caractériser les participants du point de vue de leur engagement dans l'enquête :

– informateurs,
– locuteurs,
– sujets,
– « cobayes »,
– natifs,
– acteurs sociaux,
– participants,
– collaborateurs,
– partenaires,
– enquêtés,
– témoins.

Ces choix terminologiques sont le plus souvent le produit de considérations théoriques et politiques qui révèlent le type de relations préexistantes, construites, ou développées entre l'enquêté et l'enquêteur.

Si nous ne pouvons développer ici les enjeux de ces considérations théoriques, il est néanmoins important de repérer les marques d'une *relation* particulière qui fonde différentes réalisations de la *paire* enquêté/enquêteur, impliquant différents droits et obligations selon les caractéristiques de cette relation (Sacks, 1972).

Deux éléments définissent notamment cette relation : la proximité/distance des participants et les rôles en action et en situation.

- *Proximité/distance*

La question de l'accessibilité des situations enquêtées pour le chercheur s'est posée depuis toujours et a motivé différentes formes de *fieldwork* (travail de terrain), allant de l'immersion dans une communauté totalement étrangère au chercheur à l'exploitation de ses liens d'appartenance à sa communauté.

Ces problèmes ont été traités en termes de paradoxe de l'observateur - selon lequel le phénomène enquêté se dissout dès qu'il est observé (tel le vernaculaire pour Labov, 1972) - aussi bien qu'en termes de violence symbolique entre l'enquêté et l'enquêteur (Bourdieu, 1993). Ils ont aussi été traités en termes de *réflexivité* - par des chercheurs intégrant leur présence et celle du dispositif d'enquête dans l'analyse de l'objet enquêté (en anthropologie notamment, Clifford & Marcus, 1986 ; Mondada, 1998).

Les enquêtes chez les « proches » du chercheur, lorsque celui-ci exploite ses propres réseaux pour un travail d'enquête, facilitent les prises de contact et l'accès au terrain, tout en posant souvent des problèmes d'indistinction entre les relations dictées par l'enquête et les relations personnelles. Ces questions ne se posent pas dans le cas des enquêtes chez les « lointains » (communauté observée, panel échantillonné, témoins non sélectionnés par le chercheur,…) où les difficultés d'accès peuvent être supérieures mais où une fois gagnée la confiance et établie une relation, l'enquêteur a un statut souvent plus clair et mieux reconnu en tant que tel (Beaud & Weber, 1977).

La recherche en sciences sociales et humaines a en outre souvent utilisé des « populations captives », dans le sens où l'enquêteur dispose d'un accès facilité par des institutions (l'école, l'hôpital...) et où ces populations ont des possibilités limitées de refuser de collaborer (enfants, élèves…). Dès lors, une attention particulière doit être consacrée à l'approche de populations telles que :

– les personnes défavorisées,
– les personnes handicapées,
– les enfants,
– les élèves et étudiants,
– les employés d'entreprises ou d'institutions contactés par le biais de leur hiérarchie,
– etc.

L'usage est alors de doubler les autorisations pour les personnes par l'autorisation d'un responsable légal (enfants relayés par les adultes).

Ce cas particulier montre que l'autorisation signée ne peut pas toujours être considérée comme un acte suffisant et qu'il convient de protéger certains enquêtés au-delà de ce qu'ils ont signé (responsabilité de l'enquêteur).

- *Rôles en situation*

Lors de la situation d'enquête, et selon les techniques utilisées, la relation enquêteur–enquêté peut prendre des formes très différentes et impliquer des engagements plus ou moins directs.

- *Rôles de l'enquêteur*

– *observateur extérieur,*
– *observateur participant,*
– *observateur engagé* (défendant la communauté),
– *membre de la communauté* participant à une recherche-action (projet émanant de la communauté ou tenant compte de ses problèmes et objectifs, et y intervenant par une démarche spécifique sous la forme d'une « recherche-action »),
– *observateur déguisé* (*cross-dressing* dans la tradition ethnographique) s'intégrant dans la communauté par le biais de relations, d'un travail ou d'une fonction, mais ne déclarant pas son identité d'enquêteur,
– « *magicien d'Oz* » : enquêteur qui se dissimule derrière un dispositif technologique qui est censé répondre à l'informateur.

- *Rôles des enquêtés*

– *enquêté*/informateur/locuteur focalisé
– « *périphériques* » : les techniciens, les passants, les spectateurs...
– *associés* aux participants ratifiés à l'enquête (ex. clients appelant un centre d'appels, ou bien époux de la femme interviewée)
– le « *compère* » : informateur privilégié qui porte l'outil enregistreur et qui permet à l'enquêteur de pénétrer un groupe dont le compère fait partie ou auquel il a accès.

Ces rôles rendent compte notamment des variations possibles entre participation et observation, dans la tension entre les deux reflétée par le terme d'« observation participante » (Becker, 1960 ; Platt, 1983 ; Spradley, 1980). Selon ces rôles (Adler, 1987), l'engagement par rapport à l'enquête et aux enregistrements sera très différent, ainsi que les modalités de contact pour l'obtention d'un consentement éclairé.

3.2.4 LIEUX

L'information sur le lieu de la collecte conditionne des éléments de réponses juridiques particuliers de par ses propres caractéristiques et le rôle qu'il tient dans la situation d'enquête.

Ainsi on peut tout d'abord différencier les *lieux publics,* au sein desquels l'activité scientifique d'enregistrement audio-vidéo ne requiert pas d'autorisation autre que celle de la personne enregistrée, et les *lieux privés,* soumis à l'autorisation préalable du propriétaire/responsable qui est distincte du recueil du consentement de l'enquêté.

Le lieu peut également être défini selon la relation que les participants établissent. S'agit-il d'un lieu où la présence de la personne enregistrée est du fait de l'enquêteur (laboratoire, salle d'enregistrement…) ou est-ce celui-ci qui se déplace sur le terrain et investit donc l'espace propre de l'enquêté ?

Enfin, le lieu d'enregistrement peut être intégré aux données (caractéristiques audios ou visuelles présentes dans les données) ou ne relever que d'une information éventuellement présente parmi les métadonnées.

3.3 PRATIQUES DE TERRAIN

Ce chapitre a pour but de montrer l'omniprésence des enjeux éthiques et juridiques dans les étapes qui constituent la démarche de terrain ayant pour fin la constitution de corpus de données orales, interactives et multimodales. Nous insisterons notamment sur les phases *préparatoires* de l'enquête, préalables à l'enregistrement des données, où il s'agit notamment d'établir une relation avec les personnes concernées : ces modes d'approche sont étroitement liés non seulement aux méthodologies d'enquête (cf. *supra* 3.3.2) mais aussi aux possibilités et limitations techniques du dispositif d'enregistrement choisi, dont dépendent les contraintes spécifiques pour les autorisations à effectuer un enregistrement. Une fois terminée l'enquête et analysées les données, il s'agit d'organiser le *retour sur le terrain* pour différentes formes de « rendu » des résultats et des expériences – retour qu'il vaut mieux anticiper et qui configure le type d'engagement pris envers les personnes concernées.

3.3.1 MODES D'APPROCHE

Les enquêtes dont la finalité est le recueil de données enregistrées dépendent nécessairement de la qualité de la relation avec les personnes ressources – qu'on les appelle des informateurs ou des partenaires (cf. *supra* 3.2.3). La mobilisation de ces personnes varie selon la méthode d'enquête choisie : nous insisterons ci-dessous sur la temporalité des différentes approches auprès des personnes directement

concernées ou de leur hiérarchie, et sur la question de savoir comment organiser le retour, le contre-don, éventuellement la rémunération de ces personnes.

Typologie des relations et modes d'approche

On peut considérer que la façon dont les personnes sont approchées sur le terrain – la façon dont une relation personnelle et sociale est établie – est un acte ayant immédiatement des implications éthiques et juridiques. L'établissement de la relation avec les informateurs a d'une part des effets sur la qualité de leur collaboration et donc, en définitive, sur la qualité des données ainsi constituées ; d'autre part, elle a des effets sur les relations de confiance, d'acceptation, voire d'intérêt ou de curiosité scientifique que les informateurs nourriront envers les enquêteurs.

On peut esquisser une typologie des relations établies avec les informateurs en l'articulant au moment où ils sont approchés dans le processus de l'enquête :

- Quand l'enquête procède par *convocation nominale* des informateurs en laboratoire, les modalités de leur engagement sont généralement explicitées *préalablement*, au moment où les personnes acceptent de collaborer à l'enregistrement, effectué dans des lieux et à des moments convenus à l'avance. Les personnes sont alors soit sélectionnées et contactées par le chercheur (ou par une institution travaillant pour lui), soit elles répondent à un « appel à volontaires ». L'appel ou l'annonce de recrutement est le premier acte de communication qui manifeste (ou suscite des attentes quant à) la forme du contact, voire du contrat qui s'établit avec le chercheur.
- Quand l'enquête procède sous la forme d'un *fieldwork* (travail de terrain) impliquant une présence plus ou moins longue de l'enquêteur sur le terrain et des formes d'*observation participante* – classiquement discutées au sein des méthodes ethnographiques empruntées par les linguistes comme par d'autres chercheurs en SHS (Depperman 2000, Duranti, 1997, Hammersley, Atkinson 1995 & Moerman 1988) – la relation aux informateurs s'établit dans la *durée* de cette présence et est souvent associée à la construction de relations personnelles impliquant entre autres une confiance réciproque. Sur certains terrains, le chercheur n'est pas le premier à intervenir et d'autres l'ont peut-être précédé. Selon le comportement de ses prédécesseurs, l'accueil sera plus ou moins facile de la part de la communauté et, en particulier, les exigences en matière de retour (cf. 3.3.4) seront plus ou moins grandes.
- Quand l'enquête procède par *des entretiens, des « micro-trottoirs », des enregistrements d'activités* réalisés de manière *aléatoire* dans des espaces publics, sans viser des témoins particuliers mais des passants choisis simplement à cause de leur présence sur les lieux au moment de l'enregistrement, une rencontre préalable avec les informateurs est par définition impossible. C'est donc *juste avant, pendant* ou *juste après* la réalisation de l'enregistrement qu'ont lieu l'explication des finalités et la demande d'autorisation.

– Dans certains cas, il est possible d'envisager un contact *postérieur* à l'enregistrement : tel est le cas d'enregistrements réalisés à l'insu d'une partie des participants dont l'entrée sur la scène enregistrée n'était pas prévisible (c'est le cas des conversations téléphoniques par exemple, où une partie collabore à l'enquête et l'autre n'est pas toujours au courant de l'enregistrement ; elle est contactée ensuite pour donner son accord).

La forme du contact, de l'engagement, de la crédibilité, de la confiance varie énormément selon que la relation d'enquêteur à enquêté est établie au préalable ou durant le travail sur le terrain, de manière durable ou au moment même de l'enregistrement, ou encore après celui-ci.

Les personnes contactees

Dans la présentation que nous venons de faire, nous avons considéré, pour des raisons de simplicité, que le contact s'établissait avec la ou les personnes directement concernées par l'enregistrement ; or il s'agit souvent de personnes appartenant à un groupe ou à une institution – ce qui implique des prises de contacts multiples. Il s'agit ainsi de distinguer :

– Le cas où l'informateur agit *en son propre nom,* de manière individuelle.
– Le cas où l'informateur *est contacté* dans le cadre de ses activités professionnelles ou institutionnelles, et intervient donc en tant qu'appartenant à une organisation. La hiérarchie des personnes visées par l'enquête est aussi contactée au préalable : tel peut être le cas de la direction d'une entreprise, ou du chef d'une tribu, ou des parents d'élèves. Il convient de remarquer que la relation entre la personne et sa hiérarchie ne va souvent pas de soi et invite à différencier ce qui sera promis, expliqué, montré, etc. aux personnes et à leur hiérarchie.

Remunerations

Lors de l'approche des personnes concernées par l'enquête, des promesses peuvent être faites, de véritables contrats peuvent être proposés, des contreparties, rémunérations, remboursements peuvent être proposés. Ces engagements peuvent être à la fois éthiques et juridiques, sociaux, matériels voire financiers. De toute façon, la question se pose d'une forme de « dédommagement » des informateurs – qui est très différente si on la catégorise comme « contre-don », « rémunération », « dédommagement », « service rendu »…

Plusieurs cas de figure sont envisageables :

– *durant, voire avant l'enquête :* rémunérations financières promises dès l'établissement du premier contrat, contre-dons en nature, contre-dons symboliques, prestations pour la communauté concernée ;
– *après l'enquête :* reconnaissance de l'informateur sous des formes allant du remerciement ou de la citation à la mention comme co-auteur ou comme collaborateur, voire comme partenaire de la recherche, restitution des résultats, restitution des données/corpus sous forme d'archives, diffusion de savoir-faire, retours bénéfiques attendus pour la

communauté au sens large et sur le long terme (sur le modèle des bénéfices attendus d'une recherche médicale).

Pour une discussion de ces formes de « rendu » nous renvoyons (*infra*, 3.3.4) à la discussion du « retour » sur le terrain. La question reste de savoir ce qu'on peut/doit promettre aux informateurs lors de l'établissement de la relation, en tenant compte du fait que :

- Cette relation se modifie dans le *temps* (notamment si l'enquête de terrain implique une durée).
- Cette relation peut plus ou moins reconnaitre l'« informateur » comme un *partenaire* du projet de recherche (et non seulement comme un « objet »), dans des projets participatifs où le « natif » apporte plus que ses propres performances (par exemple en collaborant aux transcriptions, aux traductions, aux gloses des données).
- La *rémunération financière* peut être moins problématique pour des informateurs recrutés (parfois par des organismes spécialisés) dans le cadre d'un contrat formel ; elle peut être plus problématique sur le terrain, où elle implique une mise en concurrence non seulement entre les informateurs possibles, mais aussi entre les chercheurs pouvant y avoir accès (tel est le problème par exemple pour des linguistes d'universités moins dotées de moyens face à des chercheurs venant d'universités mieux dotées – et pouvant de ce fait être privilégiés par les informateurs ou générer chez eux des demandes difficiles à satisfaire). Les pratiques des anthropologues et des linguistes diffèrent sur ce point. Dans le cas d'une observation participante, il peut être délicat pour un anthropologue de rémunérer les personnes qui lui délivrent les informations, au risque d'entrainer une surenchère du cout de l'information. En revanche, la rémunération du locuteur et/ou traducteur qui passe plusieurs heures par jour avec le linguiste est un juste dédommagement pour un véritable travail, et n'entrave pas forcément la relation de confiance qui a pu s'instaurer entre les deux personnes.
- La rémunération financière n'est qu'un cas parmi d'autres de « *retour* » (ou de dédommagement, de salaire…), qui pour les enquêtes de terrain se fait toujours de manière plus ou moins implicite, au fil de la vie quotidienne et de la négociation des relations mutuelles.

3.3.2 DISPOSITIF D'ENREGISTREMENT

Le choix du dispositif d'enregistrement des corpus a des effets sur la manière dont les personnes concernées vont être traitées, dont leur consentement va être obtenu, dont l'acceptation ou l'acceptabilité de l'enregistrement vont se négocier.

Nous allons ici discuter quelques aspects qui peuvent se révéler pertinents, allant du choix des contextes dans lesquels effectuer l'enregistrement aux modalités de l'enregistrement.

Contexte de l'enregistrement

Par définition, il n'est pas possible de *tout* enregistrer et les chercheurs sont obligés de faire des choix. Ceux-ci dépendent de l'objet de recherche visé, des contraintes techniques (par exemple, difficulté à enregistrer en vidéo la nuit ou en audio dans des lieux très bruyants), et aussi du respect des personnes enregistrées.

Interviennent notamment :

- le choix du *moment* à enregistrer : il s'agit de trouver un équilibre entre les moments intéressants pour l'enquêteur et le respect de la vie privée de l'enquêté ;
- le choix des *activités* à enregistrer : celles-ci peuvent être davantage publiques et sociales ou bien intimes et privées ;
- le choix du *lieu* où enregistrer : là aussi il y a une tension entre des lieux publics détachés de la vie privée ordinaire des personnes et des lieux intimes : le *laboratoire* est un lieu totalement détaché de l'espace de vie des informateurs – et c'est d'ailleurs ce qui fait que les chercheurs voulant travailler sur les pratiques sociales situées l'évitent ; le *domicile* des personnes est leur lieu de vie, lui-même articulé en lieux plus « publics » ou plus « intimes » (un repas pris à la salle à manger, à la cuisine ou au lit n'a pas la même teneur, ainsi qu'un entretien effectué au salon ou autour de la table de la cuisine) ; les *espaces de travail* sont eux aussi, quoique de manière différente, structurés par des questions de confidentialité qu'il s'agit de respecter ; leur non-respect peut risquer d'impliquer pour les données recueillies un devoir de confidentialité qui signifie l'impossibilité de leur exploitation (voir 3.4) ; les *espaces religieux*, sacrés et/ou soumis à des tabous doivent également être respectés. De manière générale, une bonne connaissance du lieu et de son organisation géographique et sociale est nécessaire avant d'envisager tout enregistrement (image ou son).

L'équilibre à trouver se situe donc entre contextualité et naturalité des données enregistrées et voyeurisme – le choix des moments à enregistrer pouvant avoir des conséquences importantes sur la suite de l'enquête (sur les autorisations pour exploiter les données et sur le droit de rétractation *post hoc* des sujets).

Modalites d'enregistrement

Les modalités d'enregistrement interviennent souvent dans le choix des contextes à enregistrer (cf. *supra*), des activités visées ainsi que dans les modalités d'acceptation ou de résistance des personnes concernées. Différentes dimensions techniques peuvent intervenir concernant l'acceptabilité de l'enregistrement par les personnes enregistrées :

- Le fait que l'enregistrement soit réalisé en *audio* ou en *vidéo* : pour certaines activités, les personnes concernées peuvent préférer l'audio à la vidéo – jugée plus invasive -, quittes à passer de l'audio à la vidéo dans un deuxième temps, une fois constatés les modalités et les effets de l'enregistrement sur l'activité.

– Le fait que l'enregistrement soit réalisé par *l'enquêteur présent*, par des *techniciens* ou par un *dispositif pré-installé* et fonctionnant en l'absence du chercheur a des effets sur son acceptation : même si la caméra ou le micro sont souvent traités comme des « prothèses » ou des prolongements du chercheur (par exemple, quand les participants s'adressent directement à eux), l'absence du chercheur peut être préférée par certains participants.
– Le fait que l'enregistrement soit réalisé par le *chercheur* ou par *les participants eux-mêmes* : d'une part, la délégation de l'enregistrement aux participants peut être vue comme une forme de contrôle de leur part sur ce qui est enregistré ; d'autre part, cette délégation peut être refusée comme une forme trop poussée de collaboration détournant le participant de son activité.
– Le fait que l'enregistrement soit réalisé par un *dispositif voyant ou discret*, voire caché : il existe de nombreux débats sur le fait de recourir à un micro caché et sur les conséquences de ce choix sur les relations possibles avec les participants (Mitchell, 1991, Mondada, à paraitre, Welland & Pugsley, 2002) ; par ailleurs, même lorsque les participants sont au courant de l'enregistrement, le fait de recourir à un dispositif voyant peut aussi bien être perçu comme un gage de transparence que comme une gêne. Souvent, la miniaturisation des dispositifs permet de les installer d'une manière qui, sans du tout les dissimuler, en fait rapidement des éléments intégrés dans le décor.
– Le fait que l'enregistrement dépende de *moyens techniques nécessitant une intervention à brève échéance* (relative par exemple à la durée de la batterie ou à la durée de la cassette) implique des perturbations de l'activité par le chercheur (ou par les participants qui effectueraient le remplacement de la cassette) qu'évitent d'autres dispositifs dotés d'une plus grande autonomie (enregistrant par exemple directement sur des disques durs). Cela peut avoir des répercutions sur le comportement des témoins en raison du dérangement occasionné, en particulier pour certaines activités (comme opérer un patient, effectuer une consultation en thérapie, discuter d'un contrat délicat, être engagé dans un processus de création).
– Le fait que l'enregistrement offre ou non des *angles morts* aux participants qui voudraient lui échapper un instant : par exemple, le cadre et le champ délimités par une seule caméra permettent d'inférer des zones qu'elle ne couvre pas alors que la puissance imaginée d'un micro ou le fait de recourir à plusieurs caméras sur la même scène peuvent donner l'impression d'un dispositif de surveillance auquel on ne peut se soustraire.
– Le fait de pouvoir arrêter ou imposer des *coupures à l'enregistrement* peut intervenir comme une matérialisation de la possibilité de rétractation ; le fait que l'effacement ou la coupure de l'enregistrement puissent être effectués par les participants quand ils le désirent ou bien doivent être

effectués plus tard, ou par des tiers, peut donner l'impression d'une plus ou moins grande latitude pour intervenir sur les données et suppose des relations de confiance différentes. Cette question – comme bien d'autres – est, là aussi, liée aux contraintes techniques de l'enregistrement et à la sophistication du dispositif. On pourra en tenir compte dans le choix de supports permettant ou non un effacement immédiat des données ou bien permettant ou non un visionnement sur place de ce qui a été enregistré.

Ces considérations (Mondada 2006) montrent bien l'imbrication des questions techniques et des questions juridiques, le respect à la fois personnel, éthique et juridique des participants étant matérialisé dans les choix techniques mis en œuvre.

3.3.3 DEMANDE D'AUTORISATION ET CONSENTEMENT ECLAIRE

La définition du « consentement éclairé » et sa traduction dans des formes de relation sociale (le contact avec les informateurs) et des formes matérialisées (les documents échangés et signés) sont sensibles au contexte et aux objets de l'enquête, ainsi qu'aux conditions socio-culturelles du groupe dans lequel cette enquête se déroule. Nous esquissons ici quelques pistes de réflexion, en partant de la définition même du « consentement éclairé », en reprenant la question du moment auquel ces questions se posent, ainsi que la question des personnes que l'on informe et à qui on demande l'autorisation, des formes que prend cette information, des objets à propos desquels on choisit d'informer, et des formes du consentement lui-même.

DEFINITION DU « CONSENTEMENT ECLAIRE »

On parle souvent de formulaires d'autorisation à soumettre aux informateurs ; il est cependant important de faire dépendre cette autorisation de l'information préalable donnée aux personnes concernées : sans *information*, la *demande d'autorisation* n'a pas d'objet ni de sens. C'est pourquoi on parle de *consentement éclairé (informed consent*), dans le sens où l'acceptation de l'enregistrement est étroitement dépendante de la compréhension des finalités pour lesquelles il est effectué. Sur certains terrains, la difficulté de faire comprendre les finalités de la recherche ne doit cependant pas inciter le chercheur à passer outre la demande de consentement, et celle-ci doit alors être formulée en accord avec le type de société dans laquelle se déroule le terrain (par exemple, comment concevoir un consentement individuel signé dans une société à tradition orale dans laquelle le droit privé n'a aucun sens ?).

MOMENT DE L'INFORMATION ET DE LA DEMANDE

La demande d'autorisation dépend du mode d'approche des personnes enregistrées. Elle peut différer selon le moment où elle a lieu :

- information et demande *préparée à l'avance* durant une permanence sur le terrain et dépendant de la relation d'interconnaissance et de confiance avec l'enquêteur,
- information et demande faite *juste avant* l'enregistrement,
- information et demande faite *juste après* l'enregistrement,
- information et demande orale effectuée *avant* et demande écrite effectuée *après* l'enregistrement (avec possibilité de rétractation).

L'information est plus abondante lorsqu'elle bénéficie de la présence prolongée de l'enquêteur sur le terrain ; elle est plus restreinte lorsque la demande d'autorisation se fait rapidement avant ou après l'enregistrement, sans autre forme de contact entre les enquêteurs et les enquêtés.

Le moment où se situent l'information et la demande d'autorisation peut être choisi en relation avec ses effets envisagés sur la structuration de l'activité enregistrée : souvent le moment de l'information et de la demande d'autorisation est choisi de manière à ne pas perturber l'activité du point de vue des participants (par ex. une demande d'autorisation à un client au moment de la vente peut provoquer un risque de perturbation de la vente pour le vendeur et donc être refusée à l'enquêteur qui désirerait documenter cette activité), ou du point de vue des enquêteurs (par ex. une demande d'autorisation en ouverture de conversation modifie l'organisation du déroulement séquentiel de cette ouverture).

Si l'information et la demande interviennent *après* l'enregistrement, l'information peut apparaitre comme un « dévoilement », une « révélation » qui *a posteriori* qualifie l'enregistrement de « dissimulation » : cela peut faire intervenir des recatégorisations des participants et des activités (celui qui s'était présenté comme un touriste perdu dans la ville demandant son chemin devient un enquêteur travaillant sur les descriptions spatiales dans les demandes d'itinéraire) (Mondada à paraitre). En outre, cette technique n'est pas envisageable pour de nombreux terrains de recherche. Ainsi ces cas de dissimulation sont particulièrement mal venus dans certaines communautés et font alors beaucoup de tort à la communauté scientifique dans son ensemble et aux chercheurs qui suivront.

Statut du demandeur

Même si le chercheur est celui qui informe et demande habituellement l'autorisation d'enregistrer, différents cas de figure sont envisageables :

- Le cas le plus classique est celui de *l'enquêteur* se chargeant de l'information et de la demande d'autorisation.
- Souvent toutefois le chercheur envoie sur le terrain des *étudiants* ou *des collaborateurs* qui sont autant de porte-paroles du projet.
- Dans certains cas, il est envisageable que les participants deviennent eux-mêmes *les porte-paroles du projet* : cela est le cas lorsque le chercheur demande à un participant d'informer d'autres participants (par ex. l'hôte qui invite chez lui des amis à un repas qui sera enregistré ; le commerçant qui demande à ses clients d'accepter de se laisser enregistrer ; l'enseignant qui demande l'autorisation à ses élèves ou étudiants, etc.). Cette délégation fait partie des collaborations sur le terrain entre enquêtés et enquêteurs ; elle peut toutefois être la source de malentendus et de difficultés.

De même, l'autorisation peut concerner les signataires eux-mêmes ou des personnes qui dépendent d'eux (subalternes, enfants, étudiants, etc.). Dans ce dernier cas, il est important de tenir compte du fait que *autorisation* ne se confond pas toujours avec *acceptation*. Dans les sociétés où le droit individuel n'existe pas, l'avis et l'autorisation

du groupe dans son ensemble ou de certains de ses responsables (politiques ou religieux) sont souvent indispensables.

QU'EST-CE QU'*INFORMER* ?

Au cœur du consentement éclairé, il y a l'exigence d'informer les participants enregistrés. Toutefois, dès que l'on interroge cette exigence, les questions surgissent. Qu'est-ce qu'« informer » ? Informer « à propos de quoi » ? A quelles conditions peut-on dire que cette information produit le statut « éclairé » de son destinataire ?

La notion même d'« information » peut laisser penser à un simple transfert de messages et de contenus ; elle tend à gommer les processus, les contextes et les contingences qui caractérisent cette activité communicationnelle par laquelle un enquêteur explique l'objet de son enquête à ses partenaires sur le terrain. Dès que l'on réfléchit en termes de type d'activité, l'« information » aux enquêtés pose une série de problèmes à résoudre :

- *l'adéquation au destinataire* : l'explication du projet de recherche, pour être comprise et partagée, demande à être ajustée aux compétences, au niveau de langue et de compréhension du destinataire, cet ajustement concerne aussi le contexte et les modalités de l'enquête, prenant en compte l'adéquation entre ce que les partenaires voient faire sur le terrain et les explications qu'on en donne ;
- l'explicitation des *finalités de l'enquête* doit se faire sans *nuire* à celle-ci : cela pose la question de l'équilibre à trouver entre la transparence de l'enquête et les transformations qu'elle peut induire sur les conduites des participants ;
- l'explication du projet de recherche peut se faire à des *niveaux de généralité différents* (de « c'est une enquête sur les façons de parler des gens » à « c'est une enquête sur la fréquence et les contextes de la liaison non obligatoire en français »).

L'information aux enquêtés comprend non seulement des explications du projet scientifique mais aussi des informations précises concernant par exemple :

- les *responsables* de l'enquête et leur affiliation institutionnelle, ainsi que les financeurs,
- une *adresse* de contact,
- les *personnes qui auront accès aux données* et qui travailleront sur elles,
- *la façon* dont les enquêtés ont été choisis et la population dont ils font partie,
- la façon dont les données seront *anonymisées*,
- le fait que les données seront transcrites selon des *conventions particulières* (possibilité de donner un exemple),
- la façon dont les données seront *archivées* une fois l'enquête terminée (conservation ou destruction à la fin de l'enquête, conservation auprès de quel garant, modalités de réutilisation éventuelle, transmission à d'autres chercheurs),
- les *modalités d'accès* aux informations relatives au projet et concernant tout particulièrement les données/analyses faisant référence à la

personne (possibilité d'accès aux fichiers et informations concernant tout particulièrement la personne),
- les droits de la personne, notamment le droit de *rétractation*,
- les *risques* éventuels ainsi que les retombées positives, morales ou matérielles, de l'étude.

Les modalités d'information peuvent, elles aussi, varier selon la culture des destinataires, en particulier :

- l'information peut se faire de manière orale : individuellement dans des *conversations familières*, collectivement dans des *réunions d'information*…
- elle peut se faire de manière écrite (par une brochure, un dépliant…) ou par courriel.

Dans le contexte d'une culture écrite, il est recommandé de laisser un texte ; de même, l'indication d'un site Internet où suivre l'évolution du projet (éventuellement avec des modes d'accès particuliers) peut être utile.

L'OBJET DE LA DEMANDE D'AUTORISATION

Ce n'est qu'après cette phase d'information que la demande d'autorisation de collecter les données peut intervenir. La question qui se pose est de savoir comment circonscrire l'objet de cette autorisation.

L'autorisation concerne en effet les dimensions suivantes, qui peuvent interagir et se superposer :

- les *actions* effectuées par les chercheurs dans le cadre du projet : l'enregistrement, la préparation du corpus (transcription, traduction, annotation, etc.), les conditions d'archivage (lieu de dépôt, durée prévue de la conservation, institutions garantes...), l'analyse dans le cadre des objectifs annoncés, les usages des données de manière intégrale ou non, la diffusion des résultats de l'analyse, la conservation/destruction des données une fois terminée l'enquête ;
- les *formats* et les conditions de l'enregistrement : audio/vidéo, avec plusieurs caméras/micros, à des moments connus ou non des enquêtés, bien circonscrits ou couvrant de longues durées, tout choix technique intervenant dans la façon dont la personne figurera dans les données peut être explicité voire négocié ;
- les conditions *de diffusion* des données et des résultats : sous forme intégrale ou partielle (courts extraits dont la longueur maximale peut être prévue), sous forme uniquement textuelle (transcriptions) ou audio-visuelle (dans des documents Powerpoint par exemple) ;
- les contextes de diffusion des données et des résultats : des contextes de recherche (*workshops* [ateliers], colloques, congrès), des contextes d'enseignement universitaire, des contextes de formation et de vulgarisation plus larges, des contextes liés au terrain (par exemple il faut demander explicitement l'autorisation de réutiliser les données dans le contexte d'une formation dans la même institution où elles ont été recueillies – où elles peuvent se révéler très sensibles) ;

– des contextes larges de diffusion : sous la forme d'un cédé ou sur un site internet.

L'explicitation de ces contextes se superpose avec celle des activités dans lesquelles les données seront utilisées ; l'enjeu dans les deux cas est celui des personnes qui auront accès aux données dans le cadre de ces activités. On peut différencier les contextes de diffusion soumis à un certain contrôle de la part du chercheur (par des conventions, par exemple) et les contextes de diffusion incontrôlables par définition (site Internet par exemple).

Il est envisageable de laisser la possibilité à l'enquêté d'ajouter des contraintes qui lui seraient personnelles ; toutefois cette éventualité pose le double problème de sa légalité ainsi que celui de son interprétabilité. Un des problèmes majeurs qui se posent dans la demande d'autorisation – comme d'ailleurs pour l'information – concerne l'évolution toujours possible des finalités de l'enquête, qui peuvent ne pas être totalement fixées à son début et surtout se transformer au fil du travail sur le terrain et sur les corpus. Pour cela, il est important de formuler les finalités de manière suffisamment générale pour intégrer d'éventuelles évolutions des finalités pouvant émerger au cours du travail de recherche. Par contre, tout changement de finalité devra faire l'objet d'une nouvelle demande (cf. *infra*).

Les formes de l'autorisation

La demande d'autorisation peut prendre différentes formes, qui dépendent elles aussi du contexte socio-culturel dans lequel se déroule l'enquête : ainsi par exemple exiger la signature de l'enquêté n'a de sens que dans les cultures de l'écrit, de la *littéracie* où cette procédure a un sens, n'effraie pas et n'est pas liée à d'autres pratiques avec lesquelles elle pourrait être confondue (comme la signature de chèques).

On peut donc différencier les formes de la demande selon le support sur lequel elles sont consignées :

– demande écrite et signée,
– demande orale,
– il est possible et utile de prévoir que l'autorisation orale soit elle-même enregistrée, sous forme audio ou vidéo, ce qui permet d'en assurer la traçabilité ; c'est la solution à favoriser lors du travail dans des sociétés à tradition orale, en respectant, en fonction des besoins, le degré de formalité requis par les pratiques langagières de la communauté concernée et le choix de la langue (par ex. enregistrement individuel avec le locuteur pour une autorisation ponctuelle, ou autorisation enregistrée lors d'une réunion plus formelle avec les autorités).

Dans le cas de la demande écrite, celle-ci peut se présenter sous différentes formes – dans un texte préformé (formulaire) :

– Un texte *compact* qui synthétise les différents aspects de la demande d'autorisation et qui demande un accord (ou un refus) global.
– Un texte présentant des cases à cocher et donc des *choix* : cette forme a l'avantage sur la première de matérialiser des choix véritables pour l'enquêté et donc de lui laisser la possibilité de refus partiels (par ex. il

peut accepter l'enregistrement audio mais refuser l'enregistrement vidéo) voire d'ajouts de contraintes (par ex. il peut demander l'anonymisation de la vidéo en plus de l'audio). La question qui se pose alors est celle de la formulation des alternatives, de manière à ce qu'elles ne soient pas redondantes et qu'elles ne soient ni trop compliquées ni trop longues à traiter pour l'enquêté.

Un problème peut se poser lors des demandes collectives, lorsque des groupes sont concernés (par exemple dans le cadre d'enregistrements de réunions) : si de trop nombreuses alternatives sont laissées au choix des participants, il est possible que les réponses mènent à des résultats contradictoires où n'émerge aucun dénominateur commun ; dans ce sens, les demandes à des groupes présentent des problèmes et des contraintes qui ne sont pas les mêmes que pour les individus.

Pour aller plus loin :
voir Exemples de demande d'autorisation en annexe.

3.3.4 APRES L'ENQUETE : RETOURS, DEBRIEFINGS

On insiste souvent sur la préparation du terrain, mais il est également important de prévoir le départ et le retour sur le terrain. Cela présente une importance à la fois scientifique et éthico-juridique : le retour sur le terrain peut se révéler nécessaire à tout moment pour une vérification, un complément d'enquête, une reprise de contact avec les informateurs. Si le départ du terrain s'est mal passé, le retour sera impossible. Par ailleurs, la présence sur le terrain produit non seulement des relations de confiance, mais aussi des attentes qui engagent dans la durée : quitter le terrain en disparaissant tout simplement, après avoir pratiqué une immersion qui souvent établit des relations étroites avec les participants et leur demande de l'aide et des prestations, peut produire de grosses déceptions. Une fois « prélevé » du savoir, des réponses, des corpus sur le terrain, il s'agit donc de savoir comment « rendre » quelque chose aux personnes sans lesquelles l'enquête aurait été impossible (voir aussi les questions de rémunération traitées *supra*). Il est par exemple désormais impossible de travailler sur certains terrains (dans le cas des langues en danger) sans envisager une restitution au locuteur et à la communauté, voire un engagement du chercheur, sous quelque forme que ce soit (implication dans des projets éducationnels, de littéracie, etc.)[8].

Il convient en outre de signaler que les « feedbacks », les « debriefings », les retours d'expérience peuvent se faire déjà pendant le travail sur le terrain, sous la forme de comptesrendus de résultats partiels par exemple. La distinction entre le « pendant » et l'« après » du terrain peut ainsi être relativisée.

---

[8] Rapport de l'UNESCO (2001), *Language vitality and Endangerment* : « Any research in endangered language communities must be reciprocal and collaborative. Reciprocity here entails researchers not only offering their services as a quid pro quo for what they receive from the speech community, but being more actively involved with the community in designing, implementing, and evaluating their research projects ».

Plusieurs types de pratiques sont envisageables pour assurer un « retour » auprès des populations enquêtées. Nous en énumérons quelques-unes, allant de la présentation de résultats la plus proche du contexte académique à la formulation de savoirs et savoir-faires la plus proche du terrain. C'est sans doute dans l'évaluation de la distance entre le « retour » et l'académie ou le terrain que se situent les choix de « politique du terrain » :

– Présentation des résultats à la fin du projet : la formulation des résultats peut être plus ou moins vulgarisée, plus ou moins proche des préoccupations des enquêtés, la présentation des résultats peut comporter notamment des exemples de *transcription* et d'analyse de transcriptions : les participants réagissent de manière très différente (parfois surpris, parfois choqués) à la représentation de leur voix.
– Démarche *d'empowerment* (restitution) : elle consiste à ne pas simplement penser le « retour » en termes d'« information » mais aussi en termes d'apport en savoirs et savoir-faires à la communauté des enquêtés (Cameron *et al.*, 1991) : on peut ainsi songer non seulement à présenter des analyses mais à permettre aux participants de continuer à *collecter* des données et d'analyser leurs propres données pour leurs propres fins, on peut formuler les *retombées de l'analyse* dans les termes de l'agenda, des thèmes, des préoccupations des acteurs, on peut répondre, dans la mesure des compétences du chercheur, aux *demandes d'expertises* souvent exprimées par les communautés (par ex., ateliers de réflexion sur le passage à l'écrit, ou sur la traduction de documents officiels, implication dans des programmes d'éducation bilingue), on peut mettre au service de la communauté les *savoirs produits* par l'enquête en les matérialisant dans d'autres formes que les écrits universitaires traditionnels (par ex.. sous forme d'expositions, ou d'autres produits culturels dérivés), on peut offrir une *formation* basée sur les résultats/les méthodes de l'enquête ; de manière plus générale, on peut songer à transmettre des outils d'analyse, à transférer des compétences qui pourraient être utiles sur le terrain.
– La question du « retour » des données elles-mêmes sous forme de corpus ou d'archives peut se révéler délicate : elle peut s'imposer dans certains cas (ainsi, pour les langues en danger, il devrait[9] être constitué des archives patrimoniales léguées à la communauté) mais aussi devoir être évitée pour protéger les informateurs (ainsi dans le cas d'enquêtes dans des entreprises ou des institutions, les données collectées pourraient intéresser certains niveaux de la hiérarchie mais nuire à des subalternes). Le retour des archives, s'il est pertinent, pose donc souvent des questions : d'accès limité des personnes pouvant consulter ces archives, en tenant compte des risques et des avantages que produit la mise à disposition sur le terrain, de modes et de technologies d'accès

[9] C'est même un devoir selon les recommandations de l'UNESCO en la matière (voir fiche *Unesco*).

aux archives : si les archives sont formatées pour que la population concernée puisse y avoir accès, les technologies doivent être adaptées aux usages et aux possibilités de ces populations (il ne sert à rien de faire un DVD si personne n'a de lecteur de DVD, ou de faire un site Internet si personne n'a d'accès à l'informatique, se pose ici la question de la gestion de l'asymétrie entre « l'académie » et le « terrain »), la garantie *d'accès aux publications* pose des questions analogues à celle de l'accès aux données, quoique de manière souvent moins difficile.

3.4 ANONYMISATION

La possibilité ou la garantie (que nous relativiserons plus bas) de rendre les données recueillies anonymes est importante pour la protection de la vie privée des personnes concernées par l'enquête et pour la légalité des corpus recueillis par les chercheurs. L'anonymisation des données n'est toutefois ni un processus simple ni une garantie non-problématique, car elle fait surgir de nombreux problèmes à la fois techniques, scientifiques et sociologiques.

L'anonymisation des données est une garantie importante en matière de légalité des données et de leur usage ; dans certains cas, si elle garantit véritablement la non-identification des personnes concernées, et si par ailleurs les données ne sont pas protégées par le droit d'auteur, elle peut permettre d'utiliser des données même en l'absence de demande d'autorisation préalable. Il convient toutefois d'être prudent sur ce point – en considérant toutes les limitations et les difficultés auxquelles on se heurte dans l'anonymisation (cf. *infra*).

3.4.1 DEFINITION

Bien qu'on parle souvent d'anonymisation, la question légale qui se pose est celle de *l'impossibilité d'identifier des personnes* : l'enjeu est que, sur la base des données recueillies et de leurs modes de représentation (transcription par exemple), on ne puisse pas identifier les personnes concernées. Les procédures d'identification sont bouleversées par les technologies actuelles qui offrent des facilités de stockage et de diffusion des données, mais aussi de puissants outils de traitement des informations (tri, recoupement, requêtes croisées…).

Ces considérations concernent :

- tout ce qui permet d'identifier *directement* une personne : par référence au locuteur ou à un tiers et à sa sphère privée, sur la base des manifestations du locuteur, comme sa voix ou son apparence physique ;
- tout ce qui peut lui porter préjudice ;
- tout ce qui peut indirectement permettre, par recoupement d'informations, de remonter au locuteur concerné.

Les opérations qui suppriment ces références ou ces manifestations sont appelées des procédures d'« anonymisation » des données.

3.4.2 DONNEES CONCERNEES

L'anonymisation ne concerne pas uniquement les enregistrements ou les transcriptions, mais un ensemble de données qui sont contenues dans les corpus et qui se différencient selon divers supports et formats – dont dépendront les techniques d'anonymisation :

- les données primaires vidéo,
- les données primaires audio,
- les données primaires textuelles : documents, officiels ou non recueillis sur le terrain,
- les données secondaires : transcription, notes de terrain, métadonnées, analyses, descriptions ethnographiques,
- les données secondaires visuelles : copies d'écran (*screen shots*), voire représentations de la voix (oscillogrammes, spectrogrammes…).

On remarquera que certaines données personnelles échappent à l'anonymisation : tel est le cas des hommes et des femmes publics, dans des interventions à caractère public (par exemple des hommes politiques à la télévision), où ils interviennent en connaissance de cause en ce qui concerne la diffusion de leur image et où leurs propos sont eux-mêmes considérés comme un discours public. Dans ce cas, les propos, s'ils sont considérés comme « originaux », seront soumis aux contraintes de diffusion régissant le droit d'auteur avec une tolérance pour un laps de temps déterminé par « l'actualité ». Dès lors que ces interventions ne sont plus considérées comme liées à l'actualité, elles échappent à cette qualification[10].

3.4.3 QUAND ANONYMISER ?

On peut distinguer différents moments auxquels peut intervenir l'anonymisation. Selon les finalités de l'étude et les contextes de l'enquête, on peut considérer que l'anonymisation doit se faire le plus *tôt* ou le plus *tard* possible. La première solution augmente les garanties de confidentialité pour la personne, la seconde maximise les possibilités d'analyse pour le chercheur. Les temporalités peuvent varier selon les types de données aussi :

- on évite l'anonymisation sur les données primaires originales de référence car elle pourrait endommager les données elles-mêmes ; par contre les données ainsi non anonymisées doivent être conservées dans un lieu sûr,
- les données peuvent/doivent/ne doivent pas (selon les politiques adoptées) être anonymisées lors de leur dépôt pour conservation ; le rôle de garant des institutions assurant la conservation est ici concerné,
- on peut travailler (dans un groupe de recherche bien délimité et qui garantit la non circulation externe des données) sur des données non anonymisées et garantir en revanche une anonymisation de tout extrait figurant dans un écrit ou une présentation orale,

[10] Cf. l'art. 122.5 du code de la propriété intellectuelle.

– on effectue toujours l'anonymisation sur les copies destinées à circuler entre chercheurs extérieurs au projet et parfois entre chercheurs internes au projet (c'est le cas notamment pour de grands consortiums de recherche ou des projets articulant des réseaux d'équipes importants).

3.4.4 COMMENT ANONYMISER ?

Les modes d'anonymisation touchent à la fois les supports et les formats des données et mettent ainsi en jeu des possibilités et des contraintes technologiques ; ils concernent aussi des formes et des manifestations symboliques de l'identité des personnes et mettent ainsi en jeu des questions d'analyse.

FORMES OU ELEMENTS CONCERNES PAR L'ANONYMISATION

Comme nous allons le voir, il est difficile – voire impossible – de constituer une liste finie des formes concernées par l'anonymisation. On peut toutefois souligner les formes principales :

– formes nominatives (nom, prénom, surnom ou petit nom, sigle d'entreprise…),
– données personnelles (adresse, numéro de téléphone, numéro de passeport, numéro de compte, âge, lieu de naissance…),
– profession, statut, titres,
– activités sociales,
– parenté, réseaux,
– référence à des lieux (toponymes, institutions, services…),
– référence à des caractéristiques de la personne (physiques, culturelles, médicales…) uniques ou rares dans son milieu identifié,
– caractéristiques physiques : voix, visage, caractéristiques corporelles,
– etc.

L'« etc. » clôturant cette liste souligne le fait que tout élément, selon les contextes d'enregistrement et de réception de cet enregistrement, peut devenir un porteur d'informations sur l'identité des personnes. L'identification des formes concernées par l'anonymisation suppose donc une compétence sociologique et culturelle qui rende le chercheur capable d'imaginer les usages, les connaissances et les associations qui pourraient permettre l'identification d'une personne sur la base d'une forme donnée.

FORMES DE REMPLACEMENT

Une fois identifiées les formes pouvant porter à l'identification des personnes, il s'agit de les transformer pour effectuer les opérations d'anonymisation.

On fera remarquer que la forme la plus radicale d'anonymisation est la *suppression* pure et simple des données – bien que l'on cherche souvent d'autres moyens d'assurer l'anonymisation qui puissent mieux les préserver. On notera cependant que la suppression peut être partielle (on peut envisager de détruire des extraits qui seraient porteurs de trop d'éléments problématiques et confidentiels pour qu'ils soient utilisables en l'état).

La forme d'anonymisation généralement adoptée procède par *remplacement* d'éléments confidentiels par des formes neutres. Ces formes varient selon les supports techniques concernés : nous distinguerons ici entre le texte, l'audio et la vidéo.

- *Texte*

Les textes concernés sont d'abord la transcription et toutes ses mentions dans des articles, exempliers, cours, conférences… D'autres textes devant être anonymisés sont les données primaires textuelles (documents recueillis sur le terrain). Celles-ci peuvent se présenter d'ailleurs sous une forme textuelle ou sous la forme d'image (tel est le cas d'une lettre, d'un document administratif, d'un manuscrit qui est conservé sous forme photocopiée ou numérisée).

Le principe de la substitution consiste à rendre visible la portion de texte qui a été remplacée, et ainsi à donner des informations générales sur elle (concernant au moins sa durée).

– *Remplacement par un « blanc »* : c'est la solution la moins informative et surtout la moins visible.
– *Remplacement par un hyperonyme* ou une abréviation, tel que NN ou NVILLE ou NHOPITAL pour nom, nom de ville, nom d'hôpital, etc. Cette solution peut rester informative (on précise le type de référence de la forme anonymisée). Elle est utile dans les cas où la substitution par pseudonyme (cf. *infra* ici-même) est impossible, difficile ou non vraisemblable. Cette solution implique le développement de conventions spécifiques pour la notation de ces hyperonymes, qui ne sont pas de même nature que le texte qu'ils remplacent (c'est pourquoi l'emploi des majuscules est parfois proposé, quand il n'entre pas en contradiction avec d'autres emplois de majuscules prévus dans les conventions de transcription).
– *Remplacement par un pseudonyme* : c'est la solution la plus souvent utilisée, du moins pour les noms de personnes car elle permet une bonne intégration de la forme de remplacement dans le fil du discours, n'attire pas l'attention sur elle, est vraisemblable et garde un certain nombre d'indications contenues dans la forme initiale. Cela n'est toutefois possible que si le choix des pseudonymes est réfléchi et répond aux problèmes suivants : le pseudonyme est choisi dans le même champ paradigmatique que la forme qu'il remplace (par exemple « Ahmed » sera remplacé par « Moustapha » plutôt que par « Albert », le pseudonyme tentant de conserver des traits d'ethnicité), dans certains cas, notamment si l'interaction enregistrée le rend pertinent, on veillera à conserver les connotations possibles du nom (par ex. s'il est à la base de plaisanteries ou de jeux de mots) et le nombre de syllabes et certaines caractéristiques phonétiques et prosodiques (si elles sont exploitées dans l'interaction) ; le pseudonyme est choisi de manière à éviter de pouvoir reconstituer le nom initial (dans ce sens, le choix d'un pseudonyme commençant par les mêmes lettres que l'original est à éviter, même s'il présente des avantages pour sa mémorisation) ; le pseudonyme est choisi de manière à éviter de ridiculiser la personne (dans ce sens, sont à

éviter les pseudonymes qui renverraient à des caractéristiques de la personne – par ex. « Monsieur Gros ») ; les noms des rues, les numéros de téléphone, etc. peuvent être remplacés de la même manière que les noms de personne.

On remarquera qu'il est plus facile de choisir un pseudonyme pour les personnes que pour les noms de villes (on peut imaginer un nom de petite ville ou de quartier ou encore de rue mais beaucoup moins un nom de grande ville ou de capitale) ; il est parfois envisageable mais pas toujours possible de penser à des pseudonymes pour des noms de services institutionnels (cela n'a pas de sens de remplacer « département de chirurgie » par « département de dermatologie » dans le cas d'un hôpital). Dans le cas où le choix d'un pseudonyme est difficile ou invraisemblable, on recourra à la solution de l'hyperonyme.

- *Audio*

– *remplacement par du silence* ; cette solution a comme désavantage le fait que le remplacement peut être confondu avec une pause,
– *remplacement par un bip ou un autre bruit* qui ne se confond avec aucun signal pouvant intervenir dans l'enregistrement,
– *remplacement par le signal original filtré et déformé* ; cette technique est surtout utilisée dans les médias pour rendre la voix non identifiable ; quand elle est pratiquée par des non spécialistes, elle peut poser des problèmes quant à son irréversibilité (possibilité de rétablir le signal original).

- *Image*

L'image concernée est surtout celle, dynamique, des enregistrements vidéo. Mais on peut penser aussi aux images fixes, par exemple à des photographies sur des documents et à des captures d'écran dans les transcriptions. De même, on peut songer à l'anonymisation d'une représentation visuelle du flux sonore (dans un spectrogramme par exemple) lorsqu'elle pourrait rendre reconnaissable la prononciation d'un nom ou d'un numéro.

– pour ces données, la *suppression* est envisageable sous forme de coupures lors du montage ; dans ce cas, il est conseillé de marquer la durée du segment coupé sur la bande et de ne pas donner l'impression d'une continuité ;
– *remplacement par un brouillage du signal* : par floutage, par pixélisation ou par contourage de l'image ou par application d'autres types de filtres (ce traitement peut concerner *toute l'image ou un détail uniquement*) ; dans ce dernier cas, elle est d'une technique plus complexe à réaliser quand ce détail est en mouvement ;
– placement d'un bandeau noir sur les yeux de la personne.

3.4.5 LES LIMITES DE L'ANONYMISATION

Même si l'anonymisation est une opération fondamentale pour assurer la circulation légale des données, il convient d'être prudent par rapport aux promesses et garanties faites aux enquêtés concernant l'anonymisation des données.

Les limitations sont essentiellement de deux ordres très différents, le premier concernant les contextes qui augmentent ou diminuent la reconnaissabilité des personnes, le second concernant les contraintes que l'anonymisation fait peser sur les objets mêmes de la recherche.

Contextes de production et de circulation

L'anonymisation est relativisée par différents facteurs intervenant soit lors de la production des données – et selon les spécificités de ce qui se passe durant l'enregistrement –, soit lors de la réception de ces données :

- L'anonymisation opère d'abord sur une série de formes censées contenir les indications principales permettant l'identification de la personne ; néanmoins n'importe quelle référence ou forme peut, selon les contextes, conduire à l'identification de la personne, et souvent d'une manière qui passe au premier abord inaperçue pour l'enquêteur. Ainsi, par exemple, la mention d'un détail rare dans l'interaction (une pathologie rare de la personne, un attribut extraordinaire, une caractéristique unique et connue dans la région de la personne…) peut se révéler significative pour certains (dans certains cas sans que l'enquêteur ne s'en aperçoive).
- Le caractère reconnaissable de ces détails dépend de manière cruciale du contexte de réception et plus spécifiquement du public qui consultera ou prendra connaissance des corpus. Ainsi les membres d'un département d'anesthésie reconnaitront facilement un de leurs collègues sur la base d'expressions typiques, d'expertises spécifiques ou de façons propres de parler ou d'agir ; en revanche les mêmes détails passeront inaperçus chez les professionnels d'un autre hôpital ou a fortiori chez des étudiants en linguistique. Mais, là encore, la reconnaissabilité ne dépend pas simplement de l'éloignement géographique ou social du contexte dans lequel ont été enregistrées les données : les personnes sont mobiles dans l'espace et dans les milieux sociaux et il n'est pas impossible que le fils d'un patient puisse reconnaitre son père dans un cours universitaire portant sur des consultations thérapeutiques. La valeur identifiante d'un détail dépend donc du contexte de réception des données.
- Selon les cas, la référence à une institution ou à un organisme peut rendre nécessaire ou non l'anonymisation : par exemple la référence à une grande enseigne doit être anonymisée s'il s'agit du lieu de travail d'un employé, mais n'a pas besoin de l'être si elle intervient comme élément du paysage dans une indication d'itinéraire, et doit à nouveau être anonymisée si elle est citée dans des propos diffamatoires.
- D'autres aspects sont liés au *recoupement* d'informations venant de plusieurs sources (cela peut concerner par exemple la relation entre données anonymisées et métadonnées).

Pratiques d'analyse

Les limitations de l'anonymisation peuvent venir d'un autre type de considérations, davantage liées aux pratiques d'analyse des chercheurs.

Le problème fondamental est posé par la contradiction éventuelle entre anonymisation et disponibilité des détails pour l'analyse (sur le principe de disponibilité Mondada, 2003). En effet, les enregistrements et les transcriptions visent à produire la disponibilité des détails observables pour qu'ils puissent être exploités par l'analyse ; l'anonymisation au contraire peut rendre indisponibles certains de ces détails en les effaçant ou en les transformant.

Cela peut être le cas par exemple de l'anonymisation par bipage d'un nom qui est prononcé en chevauchement avec un autre tour de parole et qui rend impossible l'analyse de ce chevauchement.

Cela peut être le cas de l'anonymisation de numéros de téléphone lors d'appels d'urgence qui rend indisponible la manière dont l'appelant donne son numéro de téléphone dans une situation de stress et d'émotion et qui peut donc affecter de manière cruciale cette information.

Cela peut être le cas de l'anonymisation des visages sur une bande vidéo qui rend impossible une analyse des regards.

De manière analogue, le filtrage de la voix (tel que pratiqué par les médias) n'est pas envisageable pour la plupart des études linguistiques qui se basent sur les qualités intrinsèques du signal sonore.

C'est pourquoi les chercheurs affirment souvent la nécessité et revendiquent le droit de travailler – en garantissant la sécurité et l'inaccessibilité des données – sur des données non anonymisées, de les conserver sous cette forme et de faire intervenir l'anonymisation le plus tardivement possible et d'une manière qui tienne compte de ce qui est pertinent pour l'analyse.

3.5 TRANSCRIPTION

La transcription est une pratique qui, loin de se limiter à un exercice technique de reproduction, intègre de nombreux enjeux théoriques et interprétatifs (déjà Ochs, 1979). Dans le passage de l'oral à l'écrit graphico-visuel, de nombreuses opérations de catégorisation sont effectuées, soit quant aux formes linguistiques, segmentées visuellement en unités (Blanche-Benveniste & Jeanjean, 1987 ; Mondada, 2000), soit quant à l'identité des locuteurs eux-mêmes (Mondada, 2003). Du point de vue de la protection de l'image et de l'identité des personnes enquêtées et enregistrées, il convient d'apprécier ces effets pour éviter la surinterprétation, la stéréotypisation (Jefferson 1996) et la stigmatisation des locuteurs et de leurs façons de parler. Nous nous limiterons ici à ces enjeux de la transcription ; dans la section suivante, nous prendrons en compte un tout autre aspect, celui des questions de standardisation des transcriptions et de leurs conventions.

3.5.1 LES DESCRIPTIONS ETHNOGRAPHIQUES

La transcription est souvent accompagnée d'une brève description ethnographique qui esquisse le contexte dans lequel elle a été recueillie ainsi que le type d'activité et l'identité des participants. Cette description, qui intègre des éléments issus des métadonnées du corpus, peut avoir plusieurs effets sur la lecture (ou sur la réception d'un exposé oral) :

– Elle peut contenir des informations, permettant l'identification des personnes, qui entrent en contradiction avec les principes de l'anonymisation.
– Elle peut contenir des indications qui forcent la lecture ou l'interprétation des données. En restituant l'appartenance à telle catégorie ou à telle autre dimension pertinente de l'enquêté, ces indications peuvent donner une image particulière de l'activité et des locuteurs.
– En particulier, elle peut contenir des allusions, permettre des inférences qui renforcent certains stéréotypes (voire qui les utilisent pour provoquer des effets comiques pour conquérir le public – cela n'étant pas rare dans les exposés oraux).

Ces remarques ne concernent pas uniquement la description des données mais aussi les noms des corpus, qui peuvent parfois intégrer des éléments confidentiels. Dans ce sens, même si cela a souvent une fonction mémorielle, il convient d'éviter d'intégrer le nom des acteurs concernés dans le nom du corpus.

3.5.2 L'IDENTIFICATION DES LOCUTEURS

La transcription doit intégrer les résultats de l'anonymisation. Là où l'annotation prévoit un codage des tours de parole, des parties de transcription peuvent être attribuées à des locuteurs distincts et identifiés de diverses manières. L'usage des pseudonymes est assez répandu, mais d'autres possibilités sont envisageables, qui ont néanmoins des effets variables sur l'interprétation du texte qui les suit. Tout choix effectué en la matière pose le problème de la manière dont est traité le locuteur. Par exemple :

– A, B, C… : solution qui est la moins connotée mais qui en adoptant l'ordre alphabétique ordonne les locuteurs en premier, deuxième, troisième....
– E1, E2, E3... (pour des élèves) : choix qui homogénéise les personnes au sein d'une même classe, désignée par une catégorie unique. La même chose vaut pour L1, L2, L3 où L renvoie au Locuteur : si le linguiste peut considérer que tous les locuteurs sont égaux et que les acteurs sociaux l'intéressent avant tout en tant qu'êtres parlants, du point de vue de l'activité en cours, ceux-ci participent d'abord sous d'autres catégories, que ce soit enquêteur/enquêté, père/fils, médecin/patient, etc.
– H, F (pour homme et femme) : là encore, le choix privilégie la catégorie du sexe/genre sur toute autre catégorie, en postulant ainsi la pertinence généralisée de cette catégorie pour la compréhension des activités en cours.

Ces remarques invitent à se demander quels effets interprétatifs produisent les choix des identifiants. Il convient de ce point de vue de se demander quels sont les identifiants pertinents pour les participants – surtout dans des démarches analytiques qui se préoccupent de la perspective des participants (comme l'analyse conversationnelle). C'est pourquoi les solutions alternatives peuvent être les suivantes :

– EVA, MAR, ROB, AND… : indication des 3 premières lettres des pseudonymes, que ce soit des prénoms ou des noms propres – selon la tonalité de la conversation,

– APP/OPE pour appelant/opérateur ou DOC/PAT pour docteur/patient, ou encore INTE/IEUR pour interviewé/intervieweur lorsque l'activité institutionnelle est régie par des paires catégorielles de ce type. Sur ces questions, on peut renvoyer aux réflexions de H. Sacks sur les catégorisations des personnes et sur la pertinence des catégories selon l'activité et le contexte en cours (une personne qui est médecin dans un contexte peut très bien être père de famille dans un autre ; la manière de l'identifier dépend donc de l'activité en cours) (Sacks, 1972, 1992).

La notion de vie privée et d'intimité n'ayant pas la même valeur dans toutes les sociétés, il conviendra que le chercheur se renseigne sur les souhaits des locuteurs concernant l'anonymisation des données. Dans certaines communautés, le fait de ne pas mentionner les noms des personnes est considéré comme un manque de respect pour l'auteur du récit ou les personnes qui y participent, alors que dans d'autres, les mentionner est une atteinte à la vie privée. Sur ce point, il semble y avoir par exemple de grandes différences entre certains terrains en Afrique (où les locuteurs souhaitent être cités) et des terrains comme ceux de l'Amazonie, en particulier en Guyane française.

3.5.3 ENJEUX

Lorsqu'on transcrit, on prend sans cesse des décisions quant à la manière de représenter les locuteurs et leurs manières de parler. Ainsi, l'analyse – et parfois le jugement – se glissent immédiatement dans la pratique de la transcription. Nous soulignerons quelques enjeux des choix effectués dans la transcription elle-même.

ENJEUX (ORTHO)GRAPHIQUES

Depuis plus de vingt ans, de nombreuses discussions ont eu lieu sur l'emploi de l'orthographe standard, de l'orthographe adaptée et de l'API dans les transcriptions (voir 2.1.3). Les transcriptions phonétiques (API ou autres) ne sont lisibles que par les spécialistes et seulement pour des textes courts. Ainsi, pour lire de grands corpus, de nombreux linguistes européens ont choisi l'orthographe standard, mais proposent aussi de pouvoir superposer d'autres notations, lorsqu'il s'agit d'observer plus en détail certains phénomènes.

A l'inverse, dans certaines disciplines comme la phonétique, une transcription orthographique peut dans certains cas être non pertinente (par exemple pour la transcription de logatomes, de pseudo-mots, etc.).

Toutefois, la représentation écrite de la langue surprend souvent les locuteurs, et peut même leur déplaire considérablement. Il arrive qu'ils refusent l'image de leur langue transmise par la transcription, qu'ils désavouent le chercheur et qu'ils refusent son travail.

LA REPRESENTATION DU PARLER EXOLINGUE

Le choix de transcrire en API certains passages ou uniquement ceux de certains locuteurs plutôt que d'autres permet certes une plus grande précision dans la représentation des détails de leur parler mais risque aussi de provoquer des effets d'asymétrie non maitrisés.

Ainsi le recours à l'API et à l'orthographe adaptée peut produire des effets de stigmatisation et d'asymétrie à l'encontre de locuteurs « non-natifs » – lorsque ces derniers sont représentés de manière différente par rapport aux locuteurs « natifs » (ceux-ci par des notations standard, les « non-natifs » par des orthographes spéciales qui en mettent en relief non seulement la différence mais aussi l'« anormalité », l'« anormativité »).

De manière comparable, la notation explicite, par convention, de la variété de langue du locuteur (différenciation grâce à des polices, styles, alphabets spécifiques aux différentes langues utilisées dans une conversation bilingue, ou spécifique à l'interlangue de l'apprenant dans une conversation exolingue) opère une précatégorisation de cette variété : or cette variété se trouve être souvent un élément négocié par les participants et changeant au fil de la conversation (où par moment certaines formes sont marquées comme « étrangères » ou « étranges » et où à d'autres moments leur différence n'est pas du tout prise en considération).

Les mêmes questions se posent pour la traduction de la transcription :

- le fait de traduire les paroles de certains locuteurs plutôt que d'autres peut être considéré comme un jugement de valeur ;
- la façon dont on traduit, plus ou moins littéralement, peut amener à produire une version appauvrie de la parole du locuteur, et à en effacer ou au contraire à en souligner la différence ;
- différents formats existent pour la traduction (fournie en note, à la suite de l'original, ligne par ligne ; ou bien de manière à proposer un équivalent à la forme originale, de manière à respecter un lien quasi littéral à l'original, de manière à en fournir une glose grammaticale) qui produisent chacun une image différente de la culture et de la langue de l'autre (Traverso, 2003).

Précisons qu'il s'agit ici de traduction dans le cadre spécifique des corpus oraux. Cette traduction est indispensable pour le travail sur des langues autres que le français, mais reste souvent un outil pour le chercheur, et dans ce cas il ne doit pas chercher à être le reflet de la parole du locuteur. Elle doit s'accompagner de renseignements métalinguistiques qui permettent de mieux retranscrire les nuances nécessaires à une analyse approfondie de la langue. Ainsi, si une publication bilingue du corpus est prévue, un véritable travail de traduction devra alors être envisagé, dans une optique totalement différente de celle du recueil des données en vue de l'analyse de la langue.

Enjeux du multimodal et du detail de la transcription

Le fait de ne noter que les activités verbales et d'ignorer d'autres indications communicationnelles – comme c'est actuellement le cas dans la plupart des transcriptions – peut produire une image aberrante de certains comportements des locuteurs. Cela peut être le cas notamment de locuteurs aphasiques ou d'enfants s'exprimant par d'autres moyens que les moyens linguistiques standards : ne pas tenir compte de la totalité des ressources mobilisées par ces locuteurs signifie en donner une image réduite, qui pathologise ou anormalise leur comportement.

De même, différents degrés de granularité de la transcription (Jefferson, 1985) peuvent nuire à la représentation de conduites non-standards (par ex. la vocalisation prononcée par un patient aphasique peut être significative et demander une transcription adéquate ; mais elle peut aussi être réduite à un simple « bruit » sans aucun sens dans une transcription superficielle).

Le caractère plus ou moins approfondi ou détaillé de la transcription ne répond donc pas uniquement à des exigences scientifiques ; elle répond aussi à des exigences éthiques et juridiques, qui permettent de nuancer et de complexifier l'image que l'on donne des locuteurs – en s'éloignant d'autant plus du risque de le caricaturer et de le stigmatiser à travers des comportements stéréotypés.

4 LES CORPUS ORAUX, OBJETS DE PATRIMOINE ? UNE SOLUTION POUR LA PRESERVATION ET L'ACCES AUX CORPUS ORAUX ?

4.1 RAPPEL DE LA SITUATION

LES CORPUS ORAUX, PRODUITS PAR DES CHERCHEURS, AU SEIN DES INSTITUTIONS

L'enregistrement de corpus oraux s'inscrit dans une histoire déjà longue d'un siècle, à laquelle la possibilité de fixer la voix a conféré une dimension nouvelle et singulière. Dès 1896, érudits, chercheurs (anthropologues, ethnomusicologues, linguistes) fixent leurs collectes sur des cylindres, puis des disques. Les chercheurs étant conscients de créer des collections nouvelles à transmettre aux générations futures, les productions enregistrées lors des « missions ethnographiques » trouvent naturellement place dans des instituts sous l'égide de l'État. Les Archives de la Parole, conservatoire des langues et dialectes de France, naissent au sein de l'Université de Paris en 1911, la phonothèque du Musée de l'Homme en 1932, la Phonothèque Nationale en 1938, et elle sera en 1977 intégrée au sein du Département de l'Audiovisuel de la BnF. Les grandes collectes ethnographiques menées par le Musée National des Arts et Traditions Populaires[11], également Centre d'ethnologie de la France, concernent par exemple la Bretagne en 1939 et l'importante enquête pluridisciplinaire sur l'Aubrac qui, entre 1964 et 1968, produisit notamment près de quatre mille phonogrammes et une douzaine de films. Ce sont les linguistes puis les ethnologues qui se soucient de façon prioritaire de l'avenir de leurs enregistrements, y compris de leur utilisation par d'autres chercheurs. Dans les années 70, certains sociologues comme Daniel Berteaux[12] introduisent le « récit de vie » dans leurs méthodes. Cette piste ouvre la voie à des recherches pluridisciplinaires dont les « Ethnotextes » ont constitué une voie, expérimentée par Jean-Claude Bouvier et Philippe Joutard.

MAIS LA FRANCE EST UN PAYS DE TRADITION ECRITE ET L'ORAL NE BENEFICIE PAS DE VALEUR CULTURELLE, ENCORE MOINS DE STATUT PATRIMONIAL.

L'Université n'a donc pas développé de méthodologie critique spécifique et adaptée à sa problématique. L'absence de vocabulaire normalisé pour définir les différentes formes de corpus oraux est révélatrice de l'absence de statut scientifique et patrimonial des corpus oraux. Chaque discipline utilise sa terminologie en lui conférant un sens précis. Claude Martel[13] rappelle la variété des définitions connues

[11] Le MNATP est devenu en juin 2005 le MCEM, Musée national des Civilisations de l'Europe et de la Méditerranée.

[12] Daniel Berteaux, « L'approche biographique. Sa validité méthodologique, ses potentialités » *Cahiers internationaux de sociologie*, 1980.

[13] Voir l'article de Claude Martel « la recherche et les sources orales, les mots pour le dire » in : *Bulletin de liaison des adhérents de l'AFAS* 10, 1998.

pour des termes comme récits de vie, témoignages, entretiens selon le domaine disciplinaire de celui qui les utilise.

Les historiens ont éprouvé pendant fort longtemps des réticences à considérer le témoignage oral comme une source fiable et digne de considération. Philippe Joutard, un des promoteurs de l'histoire orale, rappelle l'isolement de la France face aux autres pays européens comme par exemple la Grande-Bretagne, l'Italie, l'Espagne, l'Argentine qui connaissent, au sein même de l'Université, un développement dynamique et foisonnant de cette discipline. De nombreuses revues attestent de cette vitalité (voir bibliographie).

L'excellente enquête[14] menée entre 2001 et 2003 à la demande du Ministère de la Recherche par Françoise Cribier et Elise Feller, a prouvé que, dans les trente dernières années, les chercheurs français, dans toutes les disciplines des sciences humaines et sociales à l'exception de l'histoire, ont énormément enregistré. Mais leurs enregistrements, sans reconnaissance officielle ni lieu pour les accueillir, sont restés dans les laboratoires. Surtout ils n'ont été ni décrits ni documentés, et les autorisations des témoins, lorsqu'elles existent, sont limitées, dans le meilleur des cas, à l'usage des chercheurs qui les ont réalisés.

Les fonds sont souvent conservés en mains privées car, la plupart du temps, les collectes orales réalisées lors des campagnes d'enregistrement officielles embarrassent les pouvoirs publics. A cet égard, la grande entreprise coordonnée par la DGRST au début des années 1960 autour de Plozévet, village bigouden, est tout à fait exemplaire. L'enquête très importante, menée par le Musée de l'Homme, qui a duré pendant près de cinq années, a mobilisé des historiens, des géographes, des sociologues, des économistes, des ethnologues. Nombre d'entre eux étaient équipés de magnétophones. Mais cette enquête, au lieu de fournir un travail pluridisciplinaire, n'a produit qu'un ensemble de monographies et personne ne s'est soucié des enregistrements réalisés, à l'exception de ceux produits par l'ethnologue Donatien Laurent. Il est l'un des rares chercheurs qui, non seulement a documenté l'ensemble de sa collecte, mais l'a déposée au Centre de recherche et de culture celte et bretonne de l'Université de Brest. Aujourd'hui ses enregistrements sont numérisés et consultables dans le cadre universitaire. Les autres enregistrements ont été perdus, ou, par manque de crédits, les bandes ont été réenregistrées.

L'ERE DU NUMERIQUE ET DU TRAVAIL EN RESEAU : LES ANNEES 80

Aussi, ces collections sans statut scientifique posent, pour certaines encore, du point de vue de leur préservation et de leur consultation, des questions juridiques toujours non résolues.

---

[14] Cribier F. & Feller E. (2003) *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »* rapport présenté à Madame la Ministre déléguée à la Recherche et aux nouvelles technologies. dactylogr. 2 vol.
http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf Une autre enquête succincte a été réalisée par Dubar C. à la demande du CNRS (voir bibliographie).

Enregistrés en analogique, les documents sonores ne peuvent être consultés qu'en temps réel. Leur indexation ne suffit pas toujours à en prendre rapidement connaissance. Ce travail rebute la plupart des chercheurs.

Dans les années quatre-vingt, les techniques de numérisation[15] marquent un nouvel intérêt pour l'oral, donnée sensible et contenu souvent unique. En effet les enregistrements produits numériquement, indexés par le chercheur lui-même au moment de sa réalisation, permettent de « feuilleter » rapidement le son comme on peut le faire avec de l'écrit.

Mais, si les techniques numériques ont, comme pour l'écrit et l'image, révolutionné l'accès aux corpus oraux, elles ont introduit, par le caractère parfait des copies réalisées, une autre révolution intellectuelle beaucoup plus importante, notamment pour les usages ultérieurs. *En gommant la notion d'original, elles ont oblitéré les repères* qui jusqu'alors jalonnaient le domaine des collections. Versés par leur producteur au sein d'une institution patrimoniale, les corpus oraux deviennent objets de collection mais il devient alors impossible de distinguer entre le premier enregistrement réputé « original » et les copies successives d'un corpus oral.

Le support ne permettant plus d'identifier les différents éléments, qui décidera de sélectionner et de figer l'instant T de la version qui, en entrant dans une institution, témoignera de la recherche de son producteur ? Quel type de *métadonnées* seront simultanément intégrées aux collections ?

COLLECTIONS ORALES SANS STATUT

Les corpus oraux ne figurent pas dans le Code de la Propriété au titre des œuvres protégées, sauf si elles ont une forme identifiée et, comme telle, protégeable : *les témoignages, les interviews, les entretiens, les émissions radiophoniques*.

D'une façon générale, les collections orales, mais également la dimension sonore en général, ne sont pas prises en compte dans cette grande entreprise culturelle lancée en 1964 par André Malraux : *l'Inventaire général des monuments et richesses artistiques de la France.* Aucun des dispositifs qui fondent un patrimoine[16] ne leur est attribuable. Pas de classement ni d'inscription et, par voie de conséquence, aucune commission spécialisée « du patrimoine » ne s'en préoccupe. Seule, l'UNESCO a pris des initiatives dans ce sens (voir fiche *UNESCO*). Plus modestement, la Mission du Patrimoine ethnologique créée dans les années 80 au sein du Ministère de la culture et de la communication va placer les corpus oraux au rang d'objets. Cette préoccupation a très vite disparu des programmes.

---

[15] « Musique et son : les enjeux de l'ère numérique. Création musicale, recherche, archivage, transmission », *Culture et Recherche* 91-92, 2002.

[16] Sur le terme très galvaudé de « patrimoine » on lira Jean-Pierre Babelon et André Chastel, *La notion de patrimoine*, Liana Levi, 1994 et l'analyse historique très complète que lui a consacrée André Desvallées, « Emergence et cheminement du mot Patrimoine » dans *Musées et collections publiques de France* 208 : 6-29,1995.

4.1.1 LES COLLECTIONS DE CORPUS ORAUX

PRATIQUES ET USAGES DES INSTITUTIONS PATRIMONIALES

Les corpus oraux produits de façon unique par des producteurs individuels ou institutionnels ne constituant pas une catégorie particulière au regard du patrimoine et du Code de la Propriété intellectuelle, le législateur n'a pas prévu de dispositif particulier pour les collecter et organiser leur préservation.

L'Université s'est désintéressée de cet ensemble riche et foisonnant qui ressortissait de domaines disciplinaires trop diversifiés. *Il n'existe donc pas de dépôt légal des corpus oraux*.

Les corpus oraux ne peuvent être protégés dans une institution patrimoniale qu'au travers d'une *initiative volontaire* (don ou dépôt) de celui qui les a collectés ou par *décision de l'institution* soucieuse de constituer des collections orales sur des thématiques qui lui sont propres. Les institutions patrimoniales peuvent donc être, à la fois ou successivement, productrices de corpus oraux et conservatrices de documents oraux produits par d'autres. Les institutions responsables de ce type de collections engagent des recherches sur la conservation des documents sonores.[17] Elles mettent également en œuvre des critères sélectifs de constitution des fonds.

- De façon générale, c'est *le principe de cohérence des fonds* qui préside à la constitution des collections au sein des institutions patrimoniales (archives, bibliothèques patrimoniales, musées). Un enregistrement isolé ne signifiera que pour lui-même. L'enregistrement unique de la voix d'un écrivain dans le musée qui lui est consacré demeure anecdotique.
- Cela signifie que la constitution d'un fonds cohérent est le résultat d'une *politique de tri et de sélection exigeante* selon les axes prioritaires définis par l'institution (fonds parlé pour la BnF, fonds sur la déportation pour les Archives Nationales) mais suffisamment larges et complets pour qu'ils constituent pour demain des sources de référence significatives. Dans les musées de société, héritiers des écomusées définis dans les années 1970 à l'initiative de Georges-Henri Rivière, la collecte d'enquêtes orales vise à combler l'absence d'objets ou leur difficulté à témoigner de la dimension humaine à l'intérieur d'une collectivité. A Fécamp, l'enregistrement des ouvrières des anciennes pêcheries révèle une forme d'organisation sociale de la cité dans la première moitié du 20e siècle dont aucun objet ni aucun écrit ne peut rendre compte[18]. Il en est de même au Musée de la manufacture des tabacs à Morlaix, à l'Ecomusée de la communauté urbaine du Creusot-Montceau-les-Mines (Saône-et-Loire).

---

[17] Calas, M.-F. Fontaine, J-M. (1996) *La conservation des documents sonores*, Paris : CNRS-Editions.

[18] Cette série d'entretiens réalisés en collaboration entre le Musée et le service d'archives municipales a donné lieu à un disque avec livret *Femmes de marins, compagnes de pêche*, Fécamp, Musée des Terre-Neuvas, 2003.

- La collecte n'est pas toujours considérée comme un objet de collection ou comme une œuvre. A la BnF, aux Archives nationales, le traitement documentaire n'est pas déterminé par le support de la collection. Rien de semblable dans les musées. A l'exception du Musée National des Civilisations de l'Europe et de la Méditerranée (ancien Musée national des Arts et Traditions populaires), du Musée Dauphinois qui, très tôt, a intégré au même titre que les objets, les enquêtes de Charles Joïsten sur l'inventaire du musée, la plupart des musées comme le Musée-conservatoire de Salagon par exemple, portent les corpus oraux sur des inventaires de type bibliothèque. De même, l'Ecomusée de Saint-Quentin-en-Yvelines, a choisi d'inscrire les entretiens qu'il mène avec les acteurs politiques et les habitants sur un registre à part qui répertorie les collections d'études. A la fin des années 90, on a assisté à un intérêt très fort, voire excessif, pour la quête identitaire et le devoir de mémoire. Ces archives orales ne bénéficient pas encore d'une reconnaissance bien établie.
- Les collectes orales ne sont pas réductibles à l'enregistrement des voix. Elles ne prennent sens que dans la mise à disposition des données temporelles, techniques, scientifiques de leur production. L'ensemble de ces éléments de contextualisation (métadonnées), spécifiques du corpus enregistré, constitue avec lui un tout indissociable, sans lequel l'enregistrement serait privé de temporalité et de sens. Et on pourrait alors lui faire signifier tout et son contraire.
- Comme tout objet patrimonial, le document oral, bien que daté, identifié, n'est pas, comme nombre de chercheurs l'ont cru très longtemps, réductible au seul usage de son producteur. Les enquêtes orales dépassent souvent le projet dans lequel elles ont été menées. Elles peuvent être utilisées dans le cadre d'autres disciplines

  > *« Une nouvelle lecture conduit à porter un autre regard sur ce qui a été dit, parce que le temps a passé, et que les questions qu'on se pose se sont déplacées »*[19].

- Elles doivent pouvoir être analysées, au cours des temps, par différents chercheurs à travers leur grille d'analyse personnelle. Mais le Plan de numérisation des documents sonores mis en place fin 1999 par le Ministère de la Culture et de la Communication a révélé le déficit d'informations relatif à ces collections orales. Certains fonds considérés comme historiques ne pouvaient témoigner convenablement de leur intérêt en l'absence de documents indispensables de contextualisation. En outre, aucune des collections ayant répondu à l'appel à numérisation ne détenait les droits d'exploitation permettant d'organiser la consultation du public, notamment via internet.

[19] Françoise Cribier & Elise Feller, *op. cit.*

- De quel type de protection les corpus oraux bénéficient-ils dans les institutions publiques et privées ? Le versement d'une collecte au sein d'une institution n'a pas de **valeur probatoire*. La date de versement peut-elle indiquer une preuve éventuelle d'antériorité par rapport à un enregistrement qui se révèlerait être une contrefaçon du premier ? A l'exception des dépôts qui, par nature, sont toujours révocables, les collections entrent (sous la forme de supports ou de données numériques) de façon définitive et imprescriptible dans les fonds de l'institution. Cette cession, nous venons de le rappeler, n'emporte pas, sauf accord spécifique, cession des droits d'exploitation. Les institutions s'engagent a priori à assurer la pérennité physique et à organiser la consultation des corpus oraux dans le respect des droits de ceux qui ont participé à la création, mais il est indispensable que les cessionnaires cèdent au moins les autorisations de consultation. Depuis les années 80, la consultation à distance des collections a, en quelque sorte, « réveillé » l'intérêt pour les corpus oraux, et laissé entrevoir des possibilités d'accès autrefois inimaginables L'accessibilité aux corpus analogiques pose le problème en termes de conservation et d'identification des sources préalables à la numérisation. Il se chiffre également en moyens financiers et en personnel.

4.1.2 LA BIBLIOTHEQUE NATIONALE DE FRANCE

CONSERVATOIRE DE L'ORALITE

Héritier des Archives de la Parole fondées en 1911 par Ferdinand Brunot, du Musée de la Parole et du Geste qui leur succède en 1928, puis de la Phonothèque nationale créée en 1938, le département de l'Audiovisuel de la Bibliothèque nationale de France inscrit son action dans la continuité de ces institutions. Aujourd'hui, c'est donc plus d'un siècle d'une mémoire de l'oralité qui est ainsi conservée et mise à la disposition du public.

Mais, parallèlement, le département de l'Audiovisuel mène une politique active de développement de ses collections, notamment dans le domaine de l'oralité. En effet, outre la collecte du dépôt légal (voir fiche *Bibliothèque nationale de France*), la Bibliothèque nationale de France a vocation et mission d'enrichir ses collections par acquisitions, donations, dons, legs, dations, etc. C'est donc le cas du département de l'Audiovisuel qui, de manière complémentaire au dépôt légal des documents sonores, vidéographiques, multimédia et informatiques dont il a la charge, a défini les grands axes d'une politique d'enrichissement de ses collections en matière d'enregistrements sonores inédits. On trouvera à la fin de cette présentation quelques-uns des fonds entrés récemment au département de l'Audiovisuel, représentatifs de la place de l'oral dans ses collections.

LES GRANDES LIGNES D'ENRICHISSEMENT DES COLLECTIONS DU DEPARTEMENT DE L'AUDIOVISUEL

Le département de l'Audiovisuel définit comme documents « inédits », des documents « source » à l'état « unique », non diffusés en nombre, et qui ne sont pas

déterminés par une forme éditoriale précise. Cela posé, face à l'extension indéfinie du champ et à la multiplicité des contenus (linguistique, ethnologie, histoire orale…), à la multiplicité des sources possibles (institutionnelles, chercheurs indépendants…), à la nécessaire complémentarité avec d'autres institutions en même temps que face aux vides à combler en matière de conservation, de diffusion et de valorisation, le département de l'Audiovisuel a déterminé un certain nombre de principes forts à même de guider sa politique d'enrichissement en la matière.

Le critere documentaire et patrimonial

La politique du département repose tout d'abord sur un principe de sélection. Le critère fondamental qui amène à accepter ou à refuser un don d'inédits est avant tout l'intérêt documentaire et/ou patrimonial du fonds proposé. Ce critère peut être assimilé à celui de « mémoire nationale ». En d'autres termes, quels sont les enregistrements inédits que l'on peut considérer comme relevant d'une mémoire, d'un patrimoine national ? Ce critère ne limite pas le champ de la politique documentaire au « terrain » français, mais donne priorité aux fonds ayant – soit en termes de source (le collecteur, l'institution…), soit en termes de contenu – un rapport avec la France. Le don du fonds de Deben Bhattacharya, ethnomusicologue indien ayant enregistré à travers le monde, mais ayant vécu à Paris de 1954 à 2001, ou celui des collectes pygmées de Simha Arom (Lacito-CNRS), en sont l'illustration.

En étroite articulation avec ce critère d'intérêt documentaire et/ou patrimonial, et étroitement délimité par lui, le département de l'Audiovisuel accorde une attention privilégiée à des documents ou à des fonds pour lesquels n'existe a priori aucun lieu de conservation et/ou de consultation déterminé. C'est le cas, par exemple, de certaines archives personnelles ou de fonds en déshérence dans certains laboratoires, faute de structures appropriées.

L'acceptabilite du fonds et le principe documentaire

Ce principe de sélection et les critères documentaire et patrimonial établis, des conditions d'acceptabilité sont posées quant à la réception d'un fonds. Il s'agit tout d'abord de conditions documentaires. Ainsi, pour être reçues ou acquises, les sources inédites doivent être documentées et/ou exploitables d'un point de vue documentaire. On pourra envisager, soit que le traitement documentaire soit fourni en même temps que l'archive sous forme de métadonnées ; soit, éventuellement, que toutes les informations soient fournies à la BnF sous une forme ou une autre pour lui en permettre le traitement documentaire.

L'acceptabilite du fonds et le principe juridique

Les conditions juridiques forment une autre composante des conditions d'acceptabilité. La personne – physique ou morale – qui réalise le don doit notamment s'assurer :

– Qu'il est le propriétaire des supports physiques sur lesquels ont été réalisés les enregistrements, et que ces enregistrements sont susceptibles d'être donnés à la Bibliothèque ;

– Qu'il est titulaire ou qu'il peut garantir, les droits d'auteur sur les œuvres réalisées et les droits voisins du producteur de phonogrammes et éventuellement des interprètes musicaux.

Pour la BnF, recevoir les supports nécessite également de disposer des droits d'auteur et droits voisins requis pour leur reproduction et leur communication aux lecteurs, les documents sonores devant faire l'objet d'actes de reproduction et de représentation pour être conservés et consultés. Or, la personne – physique ou morale – qui réalise le don n'a pas toujours la capacité juridique de délivrer ces autorisations de reproduction et de communication.

Doivent pouvoir être cédés à la BnF :

– le droit de reproduction du document, c'est-à-dire la possibilité de transférer son contenu sur un support idoine (numérique) pour des raisons de conservation du signal ;
– le droit de représentation. Ce droit se comprend comme étant, au minimum, la possibilité d'une consultation par le public de chercheurs en salle P (au niveau « Recherche » de la Bibliothèque). On pourra admettre le principe d'une autorisation de communication au cas par cas. De même, pour certains documents, on acceptera qu'un délai de réserve de communication puisse être exigé pour des raisons autres que celles tenant au droit d'auteur (confidentialité de données relatives à la vie privée…).

Quelques exemples parmi les derniers fonds inedits reçus en don par le departement de l'Audiovisuel

(classés par ordre d'arrivée dans les collections) :

– fonds des atlas linguistiques régionaux (1979 et suivantes) ;
– fonds du Centre de Recherche Historique, EHESS/CNRS (1979) : histoire orale, récits de vie, années 1970-1980 ;
– fonds Félix Quilici (1981) : musiques corses de tradition orale, 1959-1963 ;
– fonds Geneviève Massignon (1985) : collectes ethno-linguistiques, Acadie, Ouest de la France, Corse…, 1946-1963 ;
– fonds Nicole Revel (1995) : épopées Palawan, Philippines, années 1980 ;
– fonds Gilles Deleuze (1997) : cours, Université Paris VIII, 1979-1984 ;
– fonds Deben Bhattacharya (2003) : collectes ethnomusicologiques, Asie, Europe…, 1954-2000 ;
– programme « Archivage », LACITO/CNRS (2005) : langues rares, transcriptions, annotations, http://lacito.vjf.cnrs.fr/archivage/

4.1.3 LES ARCHIVES DE FRANCE

Dans le Livre II du Code du Patrimoine, les archives sont définies à l'article L 211-1 comme suit :

« Les archives sont l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité. La conservation de ces documents est organisée dans l'intérêt public tant pour les besoins de la gestion et de la justification des droits des personnes physiques ou morales, publiques ou privées, que pour la documentation historique de la recherche. ». Les archives constituent deux catégories : les archives publiques qui procèdent de l'activité de l'État, des collectivités locales et des entreprises publiques et les archives privées (voir fiche *Archives : législation*).

C'est le mode de production et non le type de support ou le sujet qui définit l'appartenance à l'une ou l'autre catégorie. L'enregistrement d'une séance du Conseil général est un document d'archive public alors que l'enregistrement d'un personnage politique à la radio est un document d'archive privée.

La consultation des fonds sonores varie selon qu'il s'agit d'archives publiques ou privées. Si les premières sont clairement réglementées, c'est la volonté du déposant qui fixe les règles en matière d'archives privées.

QUELQUES EXEMPLES DE CORPUS ORAUX DANS DES FONDS D'ARCHIVES

- *Les Archives nationales*

Placées sous la responsabilité de la direction des Archives de France, elles regroupent cinq centres.

- Le Centre Historique des Archives Nationales (CHAN) à Paris. C'est au sein de la section XX^e^ qu'a été créée dans les années 80 une cellule d'archives orales. Cette cellule reçoit des versements, par exemple ceux réalisés par la Fondation pour la mémoire des déportés, mais elle produit des témoignages en complémentarité des archives écrites « en disant ce qui ne s'écrit pas, en redimensionnant l'évènementiel à l'échelle humaine, et en venant le cas échéant, par la narration de détails occultés, combler les lacunes historiques existantes » [20]. Il en va de même pour les enregistrements vidéo des archives judiciaires (procès de Klaus Barbie, de Paul Touvier ou du « sang contaminé ») et les Archives de la Présidence de la République : discours et conférences de presse des présidents de la République Georges Pompidou et Valéry Giscard d'Estaing.
- Les sources créées par les conservateurs correspondent à deux approches : le récit autobiographique sert l'écriture de l'histoire des élites, et les corpus thématiques peuvent permettre de croiser

[20] Agnès Callu, « Aux Archives nationales, une politique raisonnée en faveur des témoignages oraux » *Colonnes : archives d'architecture du XX^e^ siècle*, 20, décembre 2002.

plusieurs récits sur un même fait (par exemple la fonction d'instituteur dans les années 50).

- Le Centre des Archives contemporaines (CAC) à Fontainebleau. C'est là, par exemple, que sont versées les 400 heures d'enregistrement réalisées dans le cadre du programme lancé par le Comité d'histoire de la Sécurité sociale par Dominique Aron-Schnapper (voir point : Statut des collections d'archives...).
- Le Centre des Archives du Monde du Travail (CAMT) à Roubaix qui collecte tout type d'archive sur son domaine, dont des enregistrements.
- Des deux autres centres, celui d'Esperran ne conserve que des microfilms et celui des archives d'outre-mer conserve surtout un fonds imprimé clos.

- *Les Services d'archives départementales*

Décentralisés bien avant d'autres, ces services collectent souvent des copies d'émissions de radio, des films d'amateurs, des documentaires, conduisent des programmes d'enquêtes orales seuls ou avec des concours associatifs et universitaires. Leur situation est très diversifiée et l'importance des fonds oraux tient aux thématiques couvertes et surtout à la motivation et à l'intérêt du directeur.

Les services d'archives municipales, dans le courant de la *patrimonialisation* de la mémoire, ont souvent confié la réalisation d'archives orales à des emplois-jeunes recrutés sur des postes de « gardiens de la mémoire » (exemples : Martigues, Lille).

4.1.4 PLACE DES CORPUS ORAUX DANS LES MUSEES

Est considéré comme musée, dans son acception la plus large, toute collection permanente composée de biens dont la conservation et la présentation revêtent un intérêt public, et organisée en vue de la connaissance, de l'éducation et du plaisir du public.

Les collections sont constituées de tout type d'objet et d'œuvre dont la matérialité est tangible.

Les enregistrements oraux, par définition, constituent, pour le musée, de l'immatériel. Pourtant l'ICOM, l'association internationale des collections de musées, ONG qui au sein de l'UNESCO, préside au développement de toutes les formes de musées, a lancé le débat en 2004 sur la dimension immatérielle du patrimoine intangible. Le malaise ressenti par les musées occidentaux en général, face à l'intégration de la dimension sonore, audiovisuelle, paysagère au sein des musées, révèle parfaitement cette forme de contradiction, pour un musée, entre objets et oralité.

Par contre, les musées d'histoire, les écomusées, les musées de société utilisent parfois depuis de très longues années (exemple, le Musée dauphinois de Grenoble) l'enregistrement de la mémoire orale comme un des éléments essentiels du projet culturel et scientifique autour duquel le musée va s'organiser. Les collections sonores sont inscrites sur le Registre d'inventaire du Musée comme les autres collections, mais le cas est loin d'être généralisé et nombre d'enregistrements sonores et vidéo sont au mieux inscrits sur le Registre des collections d'étude ou documentaires.

Si les corpus oraux étaient, comme au Musée dauphinois, reconnus comme des œuvres inscrites sur le Registre d'inventaire dont les modalités de rédaction sont définies par les textes législatifs, ils seraient inaliénables et imprescriptibles.

4.1.5 *LES « CORPUS ORAUX » A L'INA*

Au travers de la consultation du dépôt légal de la radio télévision, l'Ina, de fait, donne accès à une grande variété de corpus oraux constitués par divers témoignages, paroles, allocutions, discours enregistrés dans une perspective de diffusion.

Les chercheurs, usagers du Centre de consultation de l'Inathèque, constituent pour leurs besoins spécifiques des corpus à partir des sources de la radio télévision, qui s'inscrivent dans différentes logiques disciplinaires d'exploitation de corpus oraux : linguistique, sociologie, histoire…

L'étude de ces corpus peut porter sur les procédés discursifs dans tel ou tel genre d'émission (l'interview télévisée, les commentaires radiophoniques…), sur différents types d'analyse du discours (politique, journalistique…), sur la création de répertoires lexicographiques, sur des analyses sociolinguistiques (la parole du danseur, paroles d'ouvrières) etc.

Certaines collections d'émissions archivées à l'Ina constituent d'emblée des « corpus fermés » de productions orales.

Pour n'en citer que quelques unes : « *Les archives du vingtième siècle* » produites par Jean-José Marchand, recueils d'entretiens avec des personnalités du monde littéraire et artistique, « *Les conteurs* », une collection réalisée par André Voisin, produite par le service de la recherche de l'ORTF, recueil d'histoires personnelles, régionales (*Ceux de La Hague, Au cœur de l'Aubrac*…).

Par ailleurs l'Ina est engagé depuis sa création dans la production de collections d'enregistrements patrimoniaux et de recueil de témoignages.

Ces entretiens, de durée variable (jusqu'à 15 heures d'entretien), sont accessibles par le biais d'une interface de consultation interactive, «@propos», facilitant la navigation dans le programme.

- Ainsi, la collection « Musique Mémoires » est fondée sur une campagne d'archivage visant à recueillir le témoignage de compositeurs, interprètes, chefs d'orchestres et personnalités dont les créations et l'action ont marqué la vie musicale des soixante dernières années. Ces entretiens, menés par Bruno Serrou, explorent le parcours propre à chacun des artistes : origine, formation, influences, rencontres, exercice du métier... Entretiens déjà réalisés : François Bayle, Claude Helffer, Betsy Jolas, Claude Ballif, Pierre Boulez, Marius Constant, Antoine Duhamel, Luis de Pablo, Yvonne Loriod, Michel Fano, Ivo Malec.
- « Histoires d'historiens » offre une collection d'autoportraits d'historiens contemporains ; l'histoire de leur vie ainsi racontée permet une meilleure intelligence de leur œuvre. Entretiens déjà réalisés : Maurice Agulhon, Pierre Chaunu, Emmanuel Le Roy

Ladurie, Claude Nicolet, Pierre Nora, Robert Paxton, Madeleine Rebérioux, René Rémond, Zeev Sternhell, Jean Tulard.

- « Télé notre histoire » est une collection de longs entretiens offrant une véritable mémoire de la télévision racontée par ceux dont l'itinéraire personnel et la pratique professionnelle éclairent l'histoire de ce média : auteurs, artistes, producteurs, programmateurs, ingénieurs, techniciens, décideurs, pionniers ou praticiens plus récent. Entretiens déjà réalisés : Igor Barrère, Marcel Bluwal, Yves Jaigu, Jacques Krier, Claude Santelli, Pierre Tchernia…
- D'autres entretiens qui ne s'inscrivent pas dans une logique de collection offrent néanmoins les témoignages d'acteurs essentiels de la vie culturelle, scientifique et artistique contemporaine. Entretiens déjà réalisés : Françoise Gilot, K.S. Karol, Claude Lévi-Strauss.
- « Mémoires de la Shoah » en cours de production est une collection de 110 entretiens de 3 heures environ de témoins de la Shoah : déportés, orphelins, « justes ».

Toutes ces collections seront, à terme, accessibles au centre de consultation de l'Inathèque de France.

4.2 LES INITIATIVES PRIVEES

L'enregistrement de témoignages oraux connait depuis 1972 (date de la création de la Commission permanente d'histoire de l'Éducation) un développement notable au sein de programmes mis en place par les *Comités d'Histoire orale* créés par les institutions publiques soucieuses de valoriser la mémoire de leurs institutions.

Aujourd'hui on dénombre 67 comités et services[21] intégrés à une institution (Comité d'histoire du Ministère de la Culture et de la Communication, Comité d'histoire de la BnF).

L'*AHICF***,** Association pour l'histoire des chemins de fer en France, occupe une place à part. Elle se met au service des institutions dont elle se propose de faire l'histoire. L'AHICF, créée en 1987, a deux missions : recherche et sauvegarde du Patrimoine. Elle favorise la préservation des sources, mais n'a pas vocation à l'assurer elle-même. Il existe des services à la carte (historiens) pour aider à la création de la mémoire dans le domaine industriel.

D'une façon générale, ces comités considèrent les enregistrements réalisés comme des archives privées couvertes par le droit d'auteur La clause de dévolution des corpus oraux produits, au bénéfice d'un service d'archives en cas de dissolution des associations, est une règle assez répandue.

---

[21] Guide des Comités d'histoire et des services historiques. Paris, Comité pour l'Histoire économique et financière de la France

On peut citer parmi les partenaires actifs d'un réseau « archives orales » les Pôles associés de la BnF comme la FAMDT, DASTUM, la MMSH Maison Méditerranéenne des Sciences de l'Homme à Aix-en-Provence (cf. BnF, Pôles associés). Ces centres ne disposent que très rarement des droits complets des corpus qu'ils conservent.

4.3 L'ACCES AUX COLLECTIONS

Il n'existe pas de catalogue collectif des corpus oraux. Plusieurs initiatives ont permis d'identifier des structures institutionnelles et associatives qui produisent ou collectent des corpus oraux à des fins de préservation et de consultation. Encore font-elles davantage du signalement global que du détail des contenus, la plupart de ces corpus étant très peu décrits par leurs producteurs. La publication[22] qui vient de paraitre cette année résulte du dépouillement d'une vaste enquête sur les sources orales en sciences sociales conservées en France. Elle marquera peut-être, si le catalogue informatisé permet une actualisation par le réseau des producteurs, le début de la constitution d'une source collective pour l'oralité.

Les conditions de consultation sont définies par le contrat. Or, il n'existe pas de contrat-type.

Dans les institutions, les enregistrements oraux, dans leur majorité, sont traités à travers le Code de la Propriété intellectuelle. D'une façon générale, le témoin a un droit de regard sur l'utilisation de sa voix (loi du 17 juillet 1970). Nul ne peut fixer, conserver, divulguer sans son accord les propos et l'image d'une personne privée se trouvant dans un lieu privé. Le Code civil article 9 et le Code pénal article 226-1 obligent à obtenir le consentement écrit de la personne. Le témoin, s'il fait preuve d'originalité dans ses propos, peut être considéré comme auteur, et bénéficier à ce titre d'un droit moral et des droits moraux afférents. L'utilisation de son enregistrement peut passer par l'obligation d'une rémunération définie dans le cadre d'un contrat. Le collecteur devra obtenir l'autorisation de consultation la plus large.

> *L'accessibilité pose des questions de droit et de déontologie (respect de la vie privée, droit à sa voix pour un témoin, histoires de vie, témoignages délicats, propos qui risquent de devenir diffamatoires...). Or, pour des raisons qui tiennent à la nature du contenu (récits de vie et témoignages mettant en cause d'autres personnes, entretiens en milieu psychiatrique), ces corpus oraux ne peuvent être donnés en consultation sur place, encore moins être diffusés sur Internet.*

Chaque cas est donc particulier et la reconnaissance des droits des uns et des autres, relève d'une analyse fine et périlleuse au cours de laquelle les questions suivantes devront avoir reçu une réponse : qui détient des droits ? Le détenteur accepte-t-il de les céder, dans quelles conditions et pour quel usage ? Pour quelle durée ? De façon immédiate ? Différée ?

---

[22] Callu, A., Lemoine, H. (2004) *Patrimoine sonore et audiovisuel français : entre archive et témoignage : guide de recherche en sciences sociales*, Paris, Belin, 7 vol., 1 CD-Rom, 1 DVD-Rom.

Le collecteur-chercheur pour qui l'enregistrement des corpus constitue un moment dans une recherche approfondie devrait pouvoir être protégé en tant qu'auteur. Il est dans la plupart des cas appelé *collecteur*. Pour reconnaitre un droit d'auteur à l'intervieweur, il faudrait pouvoir mettre en évidence la forme originale de son propos.

Les institutions, en conséquence, elles ne peuvent souvent que donner en consultation dans leurs propres locaux, et les travaux de numérisation dont ils peuvent prendre l'initiative sont souvent faits sans autorisation des véritables détenteurs de droits (Plan national de numérisation).

Il subsiste bien des difficultés.

La question du collecteur salarié agissant dans le cadre de ses missions publiques, censé faire l'abandon de ses droits au bénéfice de l'État souligne un problème sur les droits des salariés « auteurs » qui bute dans la fonction publique sur des questions financières non résolues.

Que dire des droits que pourraient revendiquer des étudiants bien peu aguerris à la technique de l'interview et qui sont payés pour poser les questions dans l'ordre d'un questionnaire préétabli ?

Le statut des collections d'archives orales n'est pas indifférent

Le cas de la première grande enquête sur l'histoire de la sécurité sociale, conduite entre 1973 et 1975 par Dominique Schnapper à la demande du Comité d'histoire de cette institution créée en 1973, a permis l'enregistrement de 200 témoins qui ont donné lieu à 400 heures d'interviews et de témoignages. Il s'agissait, par définition, d'archives privées. Or, avant que ne débute la campagne, il a été décidé que l'ensemble de l'enquête serait classée comme une archive publique et, comme telle, consultable au bout de soixante ans. Cette décision a eu des conséquences importantes. Philippe Joutard à plusieurs reprises a évoqué cet exemple, dans lequel il voit une des raisons possibles du manque de dynamisme du développement de l'histoire orale en France.

De même, Florence Descamps partage cette analyse en stigmatisant ces archives orales novatrices qui ont été, dès le début « gelées ».

Les chercheurs, peu enclins à voir institutionnalisés leurs corpus, les ont gardés par-devers eux, peu encouragés par les organismes comme le CNRS et l'Université (exception : la convention signée entre le CNRS et la BN en 1979 pour la sauvegarde des Atlas linguistiques) qui, jusqu'à une date récente, n'ont jamais pris d'initiatives constructives pour préserver des corpus oraux qui échappaient à toute définition académique, alors que l'histoire orale connait en Grande-Bretagne où elle est née, tout comme dans les pays latins autres que la France, un grand foisonnement.

4.3.1 QUEL RESEAU POUR DEMAIN ?

UN RESEAU DE GESTION, DE PROTECTION DES COLLECTIONS DE CORPUS ORAUX ORGANISE PAR LES UNIVERSITES ET LES INSTITUTIONS DE RECHERCHE OU DES INSTITUTIONS PATRIMONIALES ?

En dehors des institutions patrimoniales, les universités, les organismes de recherche, à l'instar de ce qui existe dans de nombreux pays européens, pourraient avoir la capacité et la volonté de créer un grand réseau des sciences humaines et sociales, à travers lequel les corpus mis à la disposition des autres chercheurs pourraient être protégés et rendus accessibles à d'autres.

Le Rapport[23] rédigé par Françoise Cribier avec la collaboration d'Elise Feller a étudié la situation et les réseaux existant pour la sauvegarde et l'accès aux données qualitatives des sciences sociales dans six pays européens. Deux initiatives sont présentées comme d'éventuels modèles pour les chercheurs français : Qualidata (Grande-Bretagne) et SIDOS (Suisse).

Qualidata[24] en Grande-Bretagne a été créé en 1994. Il est implanté à Colchester dans le département de sociologie de l'Université d'Essex. Cette initiative s'est inscrite dans un contexte universitaire largement sensibilisé à la préservation des données orales notamment par l'enquête menée par Paul Thomson à l'initiative de l'ESRC (Conseil de la Recherche Économique et Sociale du Royaume Uni). Elle pourrait servir d'exemple. Le service est très sélectif pour les fonds produits après 1995 (parmi les critères : thèmes bien identifiés, corpus documentés, documents sonores numérisés et en excellent état et dont les caractéristiques sont juridiquement établies).

Le service retient des critères utiles dans la perspective d'une analyse secondaire à venir. Il est intéressant de noter l'investissement de la structure dans la formation des chercheurs futurs producteurs de données.

Elle peut ainsi constituer un moyen de mieux maitriser la recherche dans certains secteurs en évitant les redondances.

[23] *op.cit.*

[24] Qualidata, UK Data Archive, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK. www.qualidata.ac.uk.Voir aussi l'Annexe 3 (Cribier, 2005).

Le SIDOS, Service suisse d'information et d'archivage des données pour les sciences sociales, créé en 1992 par l'Académie suisse des sciences humaines et sociales, constitue lui aussi une sorte d'agence de gestion des données qualitatives ou quantitatives produites par les chercheurs[25].

Le SIDOS considère le producteur de données comme un auteur et tout travail d'archivage comme un travail d'édition des données et de la documentation.

L'archivage est orienté vers l'échange de données entre chercheurs. Il constitue un instrument d'enrichissement de l'activité scientifique, à la condition que ces données soient ensuite convenablement diffusées.

La mise en place de tels réseaux aurait un intérêt incontestable pour la recherche. Nous ne sommes pas persuadés que le statut patrimonial et la pérennité de ces collections orales seraient mieux garantis.

Quelles sources orales pour demain ?

Depuis le début de l'enregistrement numérique, la question de la pérennité à long terme fait encore problème, notamment de par l'obsolescence rapide des standards et de la compatibilité des systèmes. Mais la cohérence future des collections est bousculée par les modalités d'archivage des données. Verser ses fonds représente pour le chercheur un véritable *travail d'édition* des corpus et de leur documentation, afin de toujours rendre accessibles des documents compréhensifs et cohérents. Ce travail devrait toujours être réalisé par le chercheur. Quand en prendra-t-il le temps ? Quelle image de ses travaux souhaitera-t-il verser ? Quelle forme conserver ? Quel intérêt pour le chercheur de demain ? Il n'y a pas de réponse unique.

Le chercheur qui souhaite utiliser des corpus oraux créés par d'autres a besoin d'une médiation, c'est-à-dire d'une documentation qui décrit les variables mais aussi la collecte des données et le contexte du projet.

Dans ce dernier cas, le chercheur producteur n'est pas le mieux à même de décrire ses données, dont l'usage sera fait par des personnes non familières de son domaine. Il appartient au professionnel du traitement documentaire, bibliothécaire, documentaliste, archiviste de décrire *grâce à des outils normalisés et compréhensibles par tous les corpus destinés à des tiers.*

La description trop précise, *témoignages* « *ultérieurs* », « *rétrospectifs* » « *récits de vie a posteriori* », fondée sur la notion de temporalité, certes utile pour les besoins d'analyse du chercheur, n'est pas opérante pour la gestion de ces collections au sein d'une institution de conservation. Ces critères font certes partie de la description objective

[25] Voir enquête réalisée par F. Cribier & E. Feller , *op. cit.* Annexe 3 : 14-20.

du document oral, mais il n'appartient pas à l'institution de les *classer dans des catégories trop étroites qui procèdent déjà de l'analyse et limitent la liberté des futurs usagers en contraignant leur point de vue.*

En résumé, le producteur de corpus est certainement le seul à pouvoir documenter ses corpus oraux. Leur utilisation par des tiers ne pourra se faire que si le signalement est rédigé par des professionnels de la documentation.

4.3.2 VERS LA RECONNAISSANCE D'UN STATUT DU PATRIMOINE ORAL

L'avenir des sources orales n'est pas une question exclusivement juridique. Cette dimension peut être résolue par des solutions contractuelles pragmatiques. Ce *Guide* n'a d'autre ambition que de le montrer.

Mais le véritable enjeu de la question des sources orales est d'ordre culturel et politique. Leur reconnaissance nécessite à la fois l'élaboration de critères de tri exigeants, sans lesquels aucun patrimoine digne de ce nom ne peut exister, et dans le même temps une prise de conscience de la société, qui consiste à conférer, à ces documents produits scientifiquement, *un statut d'objet du patrimoine.*

Leur intégration au sein du dispositif qui régit les objets du patrimoine sera alors chose naturelle.

La France, il faut le noter, accuse à l'égard du patrimoine immatériel un retard singulier.

5

6 ANNEXES

Fiches juridiques

L'Œuvre Orale
Les œuvres protégées
Données personnelles et anonymisation
Le droit de citation
Le Consentement
Exemples d'autorisations
Bases de données, objet d'un droit « sui generis »
Responsable du traitement
Le Patrimoine immatériel et l'UNESCO

Fiches techniques

Prise de son et enregistrement sur le terrain
Supports pour enregistrer et archiver le son
Supports pour enregistrer et archiver la vidéo
Codages et formats

Institutions

Bibliothèque nationale de France
Les Archives : législation
Musées de France : législation
Inathèque de France

Travaux

Programme « ARCHIVAGE » du LACITO
CLAPI
PFC
DELIC
ESLO
Inventaire des corpus

L'ŒUVRE ORALE

L'élaboration des corpus oraux peut se faire à partir de différents types de productions orales. Un certain nombre de ces productions peuvent être des œuvres de l'esprit protégées par le droit d'auteur. Nous présenterons ici quelques-uns des exemples les plus courants en analysant leur statut juridique et les conséquences qui en découlent.

L'INTERVIEW, RECITS DE VIE

Le fait de répondre à des questions ou de livrer un témoignage, voire d'enregistrer une personne en situation peut constituer un élément important dans la réalisation d'un certain nombre de corpus oraux. Du fait de cette importance, il est nécessaire de cerner le cadre juridique dans lequel s'inscrivent la réalisation et l'exploitation des interviews. Des jurisprudences récentes concernant des personnes publiques permettent aujourd'hui de faire le point et d'inciter à une certaine prudence dans la prise en compte des droits des interviewés et des obligations du responsable du corpus. Les conséquences du non-respect de ces obligations peuvent être un obstacle à l'utilisation, voire à la diffusion ou la publication du corpus. Il faut cependant garder à l'esprit que tout enregistrement ne constitue pas nécessairement une interview. Lorsqu'il n'y a pas communication d'une pensée, l'enregistrement ne sera pas considéré comme une interview. Ainsi, lorsqu'il s'agit de faire réciter une liste de nombres, lire un texte imposé ou répondre à des questions qui n'impliquent aucun élément personnel (ex. le temps qu'il fait), le régime de l'interview ne trouve pas à s'appliquer.

Premier principe :

Toute interview ne peut se faire qu'avec l'ACCORD de la personne interrogée. La jurisprudence reste constante sur ce point. Par exemple, un journaliste s'est vu condamner car il avait enregistré une personne clandestinement en dissimulant le magnétophone sous sa serviette[26]. Dans le cas où il n'est pas souhaitable pour des raisons liées à l'intérêt scientifique, l'accord peut être demandé après l'interview. Par exemple, pour des enregistrements en situation, pour ne pas influencer le sujet, on peut lui demander son accord lorsqu'il sort du lieu choisi. En principe, il est préférable pour des raisons de garantie des droits des personnes que cet accord soit formalisé dans un écrit. Mais ce n'est pas toujours possible ni souhaitable compte tenu du contexte. Dans tous les cas, il faut d'une façon ou d'une autre garder une trace de cet accord (un enregistrement oral, un écrit…). Voir fiche *Consentement*.

Deuxième principe :

Une fois l'enregistrement terminé, si nous sommes en présence d'une œuvre de l'esprit, ce qui sera bien souvent le cas, il faut déterminer quels sont les droits de l'interviewé. Est-il auteur et peut-il bénéficier à ce titre des droits correspondants (voir fiche *Œuvres protégées*). Partage-t-il ces droits avec l'intervieweur ?

C'est en fonction de l'apport de l'un et de l'autre que l'on pourra déterminer si l'on est ou non en présence d'une œuvre et qui en est le (ou les) titulaire(s) : banalités (ex. : questions courantes, propos impersonnels) ou originalité (travail créatif de clarification et structuration des questions ou des réponses)

[26] Cour d'appel de Versailles première chambre, 29 novembre 2001.

des questions posées et des propos enregistrés. La jurisprudence a récemment admis que lors des longues séries d'interviews filmées du président Mitterrand, ses propos constituaient en eux-mêmes une œuvre de l'esprit protégée par le droit d'auteur[27]. Le cas échéant, l'intervieweur et l'interviewé peuvent être co-auteurs du résultat de l'interview (voir fiche *Œuvres protégées*)[28]. Quand le corpus oral est aussi une œuvre audiovisuelle, la loi énumère les différentes catégories de personnes pouvant revendiquer, aussi, la qualité de co-auteurs (voir fiche *Œuvres protégées*)[29]. Quand cela est possible, et si la personnalité de l'interviewé ou le contexte le justifient, un contrat peut être une bonne solution pour ménager les droits de chacun.

Il reste un dernier aspect à ne pas négliger, celui de la relecture. En effet, toute personne ayant un droit d'accès et de rectification sur ses propos (voir fiche *Consentement*), mieux vaut offrir à la personne interrogée la possibilité de corriger ses dires.

LES DISCOURS POLITIQUES

Les actualités télévisées utilisent abondamment les discours politiques, d'où la naissance d'une confusion. Les actes officiels, tels les textes de lois, les décrets ou encore les rapports gouvernementaux sont dépourvus de toute protection et appartiennent au domaine public. Les discours et autres allocutions sont d'une nature bien différente ; ils ne fixent pas de règle pour les citoyens. On ne peut donc pas les définir comme des actes officiels, même s'ils sont prononcés dans le cadre de fonctions officielles ; ce sont des créations intellectuelles qui appartiennent à leurs auteurs. Ainsi le texte d'une loi votée au Parlement relève du domaine public, mais le discours de présentation bénéficie de la protection du droit d'auteur[30]. L'auteur jouit donc de droits moraux et patrimoniaux.

Il existe deux exceptions au droit de reproduction :

- L'article L122-5 3° CPI autorise une reproduction, même intégrale, pour la presse ou la télévision à des fins d'informations d'actualité[31]. Cela signifie que ce type de reproduction est réservé aux organes de presse ou de télévision (pas aux chercheurs). Il faut en plus qu'il y ait un rapport avec l'actualité immédiate. L'autorisation est donc limitée dans le temps. La notion d'actualité s'apprécie en fonction de la périodicité du média (un hebdomadaire aura plus de temps qu'un journal télévisé quotidien). L'œuvre devra avoir un rapport direct avec l'actualité immédiate. Lors d'une prise d'otages, la photo du preneur d'otage fait partie de l'actualité, pas celle de gens passant à proximité du lieu. Dès que le temps de

[27] Note P. Sirinelli à propos de TGI Paris, 16 Septembre 2003, C.Sosnowski c/ France 2 *et al. Propriétés Intellectuelles*, 9, Octobre 2003, : 380-382. On notera que le Président Mitterrand avait aménagé par contrat les droits qu'il revendiquait sur le résultat de l'interview.

[28] CA Paris, 4e chambre, 5 décembre 1997, SA Les Belles Lettres et autres c/ Éditions Albin Michel et autres, *Recueil Dalloz* 1999, 65.

[29] Voir TGI Paris 16 septembre 2003, précité.

[30] TGI Paris 3e chambre, 25 Octobre 1995, François Mitterrand c/ ID Éditions et autres in *Revue Internationale du Droit d'Auteur*, 167, Juillet 1995 : 294-298.

[31] Art. cité en note 3.

l'actualité passe, toute utilisation d'une œuvre de l'esprit ne peut se faire sans autorisation; et versement de droits le cas échéant.

- Le droit de citation permet de s'exonérer du droit de reproduction à condition de citer ses sources, que l'extrait soit court et qu'il n'ait qu'un rôle d'illustration. Si l'on soustrait les citations, l'œuvre doit conserver son caractère original. (voir fiche *Citation*).

LES EMISSIONS DE RADIO ET DE TELEVISION

Ce sont des œuvres qui sont destinées au public, ce qui ne veut pas dire qu'elles soient libres de droits. Une chaine, qu'elle soit radiophonique ou de télévision, reçoit la qualification juridique d'entreprise de communication audiovisuelle selon l'article L216-1 al. 2 du CPI.

L'entreprise dispose de droits voisins sur les programmes (voir fiche *Œuvres protégées)*. Toute utilisation dans un cadre professionnel devra donc se faire uniquement après l'obtention d'un accord de l'entreprise. Un corpus oral ne pourra utiliser des émissions de radio ou de télévision que si les droits ont été acquis préalablement. Le plus souvent, c'est le producteur qui doit donner son autorisation. Il faut donc s'adresser à la station de radio, à la chaine de télévision ou encore à l'INA.

Le corpus ne va pas, la plupart du temps, utiliser l'intégralité d'une œuvre radiophonique ou audiovisuelle ; il se contentera d'extraits. Le droit de citation existe à condition d'indiquer clairement la source et que l'œuvre à laquelle on l'intègre ait un caractère « critique, polémique, pédagogique, scientifique ou d'information... » selon l'article L213 3° CPI. L'indication de la source doit apparaitre explicitement. Un extrait d'émission de télévision peut se faire si le logo de la chaine est visible[32]. Toutes les règles énoncées dans la fiche *Citation* restent valables.

[32] TGI Paris 31 mars 1999.

LES ŒUVRES PROTEGEES

LES ŒUVRES PROTEGEES PAR LE DROIT D'AUTEUR ET LES DROITS VOISINS

Un corpus oral s'obtient au terme de nombreuses étapes. Une ou plusieurs personnes réaliseront une collecte, puis d'autres (ou les mêmes) effectueront la transcription ; et ainsi de suite jusqu'à l'obtention du corpus. La question qui se pose alors est de savoir qui sont le ou les auteurs ? Quels droits cela leur confère-t-il ?

La réponse s'obtient en s'intéressant d'abord aux notions d'auteur et d'œuvre. Il faut ensuite décrire les droits liés à la qualité d'auteur. Enfin, nous ne devons pas oublier les auxiliaires de la création que sont les producteurs et les artistes-interprètes.

1. L'AUTEUR ET SON ŒUVRE

Aucune définition légale ne vient déterminer ce qu'est un auteur si ce n'est son lien avec sa création : « la qualité d'auteur appartient, sauf cas contraire, à celui ou à ceux sous le nom de qui l'œuvre est divulguée »[33]. C'est donc l'œuvre qui est l'élément déterminant du droit d'auteur.

L'ŒUVRE

La qualification d'œuvre ne prend en compte ni le genre, ni la forme d'expression, ni le mérite, ni la destination.

Sans liste limitative, peuvent être œuvres de l'esprit des œuvres écrites, orales, audiovisuelles, produits de la recherche littéraire, artistique, musicale, quel que soit le mode d'exploitation (vidéo, cinéma...), les destinataires ou les sens auxquels elles s'adressent (l'ouïe, la vue, l'odorat...).

Toutefois, pour qu'une œuvre soit protégée par le droit d'auteur, il faut :

- qu'elle soit concrétisée dans une forme ;
- qu'elle soit originale.

Le droit d'auteur ne protège que la forme de l'œuvre et non les idées contenues dans celle-ci. Un cours d'université donné oralement, une conférence sont protégés. En revanche, les hypothèses scientifiques traitées différemment ne sont pas protégées. À propos des logiciels, seule la forme du programme, c'est-à-dire l'enchainement des instructions, peut être protégée.

L'originalité : l'œuvre ne doit pas avoir été copiée ou être le résultat d'un plagiat. Elle doit présenter une certaine créativité marquant soit l'empreinte de la personnalité de son auteur, soit un apport intellectuel, un effort personnalisé allant au-delà de la simple logique automatique.

Une œuvre est protégée dès sa création, aucune formalité n'est nécessaire pour jouir des droits qui lui sont attachés. Cette protection dure toute la vie du créateur et se prolonge après sa mort, sans limite pour le droit moral durant soixante-dix ans pour les droits pécuniaires.

[33] Art. 113-1 du CPI.

UNICITE OU PLURALITE D'AUTEURS

Suivant les conditions dans lesquelles l'œuvre a été créée, il peut y avoir un ou plusieurs auteurs :

- **Œuvre dérivée ou seconde :** lorsqu'une œuvre existante est incorporée, sans l'intervention de son auteur, à une autre œuvre. Une nouvelle œuvre est ainsi créée, donnant des droits à son auteur, malgré son originalité relative. Mais celui dont l'œuvre a été utilisée conserve sur celle-ci toutes ses prérogatives. Ainsi, une traduction ou une adaptation d'un texte ouvrent des droits à celui qui la réalise ainsi qu'à l'auteur du texte traduit ou adapté.
- **Œuvre collective** : quand une personne physique ou morale dirige ou coordonne plusieurs contributions qui se retrouvent fondues pour former une œuvre unique[34]. Il n'est alors pas possible de distinguer les apports de chacun. L'auteur sera donc celui qui en aura eu l'initiative ou qui aura joué le rôle de coordination. Il est primordial d'étudier le processus qui a conduit à sa création.
- **Œuvre de collaboration :** œuvre créée par plusieurs auteurs. La contribution de chaque auteur est clairement identifiable[35]. Les œuvres cinématographiques ou audiovisuelles comportent le plus souvent une pluralité d'auteurs. Sont présumés coauteurs, sauf preuve contraire, le réalisateur, le scénariste, le dialoguiste, le compositeur, l'adaptateur (ainsi que l'auteur de l'œuvre adaptée). Pourra également être considérée comme coauteur toute autre personne qui fera la preuve d'un apport original (création spéciale ou indépendance vis-à-vis du réalisateur).

Un corpus oral peut être qualifié, selon les cas, d'œuvre dérivée, d'œuvre collective ou d'œuvre de collaboration.

La frontière entre ces notions n'est pas toujours évidente à définir. Il faut pourtant bien prendre garde à ne pas vouloir les faire entrer trop hâtivement dans une catégorie afin d'éviter des procédures devant les juridictions civiles. Voyons maintenant quels sont les droits que la création ouvre à son ou ses auteurs.

2. LES DROITS DE L'AUTEUR

Le droit d'auteur se décompose en deux prérogatives bien distinctes :

LE DROIT MORAL

Le droit moral est *perpétuel*. Il survit à la mort de l'auteur. Il est *inaliénable*, l'auteur ne peut y renoncer, ni même le transmettre. Il est juridiquement *imprescriptible*. L'exercice de ce droit est *absolu*; l'auteur peut en user à discrétion sauf lorsqu'il y a plusieurs auteurs ou abus pour nuire à autrui, ou encore détournement.

La première prérogative est le droit de *divulgation* : l'auteur décide seul de rendre ou non publique sa création. Il peut choisir ensuite d'y inscrire son nom ou ses qualités ; c'est le droit de *paternité*. S'il désire rester anonyme, cela n'autorise en rien l'appropriation par d'autres personnes. L'œuvre ne peut être altérée sans un accord exprès de l'auteur; il faut respecter son *intégrité*. Enfin, un auteur peut exprimer des regrets face à sa création et demander le retrait

[34] Art. L113-2 al. 3 CPI.
[35] Art. L113-2 al. 1 CPI.

de son œuvre. Il fait alors valoir son *droit de retrait* ou *repentir*. L'auteur doit alors indemniser l'éditeur de l'œuvre.

À la mort de l'auteur les droits de retrait et de repentir disparaissent. Restent le droit au nom et à la paternité. À la mort de l'auteur, le droit de divulgation se voit placé sous le contrôle du juge afin de faire respecter la volonté du créateur.

LES DROITS PATRIMONIAUX

Ils sont *limités* dans le temps : soixante-dix ans après la mort du créateur. Ils peuvent faire l'objet de contrats car ils sont *cessibles* ; à condition toutefois de ne pas violer la liberté de décision de l'auteur.

L'auteur dispose du droit exclusif d'autoriser ou d'interdire la reproduction ou la représentation de son œuvre.

Le droit de reproduction se définit comme « la fixation matérielle de l'œuvre par tous procédés qui permettent de la communiquer au public de manière indirecte ». Ainsi toute copie de l'œuvre, quel que soit le support, ne peut être faite sans l'autorisation de l'auteur.

Le droit de représentation s'entend, lui, comme la mise en contact direct de l'œuvre avec le public. La communication de l'œuvre au public est une représentation, le public pouvant être les chercheurs du laboratoire, comme toute autre personne destinataire (amphi d'étudiants, colloque...).

Ces droits exclusifs souffrent des exceptions étroitement limitées par le législateur. Elles sont principalement :

- l'exception de copie privée qui se limite à l'usage personnel et privé du copiste ;
- le droit de citation (voir fiche *Droit de citation*).

Il existe un droit à la copie privée, mais il se limite à l'usage personnel et privé du copiste. Les usages professionnels ou collectifs (mêmes internes) sont donc proscrits.

Le droit de suite se trouve un peu en marge, car il s'applique pour les œuvres d'art plastique. Lors de la revente de l'œuvre, 3 % du prix va au créateur s'il est vivant, sinon à ses héritiers.

3. LES DROITS VOISINS

Souvent, l'artiste ne peut développer seul sa création. Pour la communiquer au public, il a besoin d'auxiliaires pour assurer l'effort financier, le producteur de phonogrammes ou de vidéogrammes, ou pour donner vie à son œuvre, l'artiste-interprète. Des droits leur sont accordés pour une durée de cinquante ans, à partir de la première communication.

DROITS DU PRODUCTEUR

Le producteur a le droit d'autoriser ou d'interdire la reproduction directe ou indirecte de l'œuvre qu'il a produite. Il peut en contrôler aussi la forme de communication (diffusion, vente, échange, location).

Dans le cas d'un phonogramme produit à des fins commerciales, il ne peut s'opposer à la communication directe de l'œuvre dans un lieu public (sauf sonorisation de spectacle) ou sur une radio. En contrepartie, il reçoit une rémunération fixée par la loi. Ce sont généralement les sociétés de gestion

collective (SACEM...) qui collectent les sommes correspondantes auprès des utilisateurs puis qui les reversent.

DROITS DE L'ARTISTE-INTERPRETE

L'artiste-interprète a droit au respect de son nom, de sa qualité, de son interprétation[36].

Il a le droit d'autoriser ou d'interdire la fixation, la reproduction et la communication au public de sa prestation. Il est fréquent qu'il apporte ses droits à une société de gestion collective (ADAMI, SPEDIDAM...) qui délivre l'accord pour toute utilisation de la prestation et qui perçoit les sommes dues à l'artiste en cotrepartie.

[36] Art. L212-2 CPI.

DONNEES PERSONNELLES ET ANONYMISATION

La collecte d'informations personnelles devenant chaque jour plus simple, la vie privée doit être vraiment protégée. Les méthodes de profil de personnalité sont utilisées par bon nombre d'entreprises pour mieux connaître, soit leurs clients, soit aussi parfois leurs employés. La source d'informations disponible sur Internet ne cesse de croitre. L'anonymisation constitue un mode important de protection. Les corpus oraux sont souvent grands consommateurs de données ; il faut donc concilier les impératifs légaux avec les exigences de la recherche.

Les textes fondamentaux encadrant la protection des données personnelles ne définissent pas la notion d'anonymisation car toute donnée peut devenir sensible, selon la finalité de son traitement. Il nous faut donc reprendre les concepts-clés des différents textes afin d'avoir une vision plus précise de l'anonymisation.

LES DONNEES A CARACTERE PERSONNEL

La loi du 6 janvier 1978 s'articule autour de la notion de donnée nominative. La Convention 108 du Conseil de l'Europe de 1981 lui préfère celle de données personnelles et la directive 95/46/CE choisit l'expression « *donnée à caractère personnel* ». Le projet de transposition de la directive reprend d'ailleurs ce dernier terme. Au-delà des différences de termes, il faut souligner la généralité des formules. Ainsi le considérant 26 de la directive dispose :

> *« …pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne ».*

Les traitements ne sont à déclarer à la CNIL que lorsqu'ils utilisent des données à caractère personnel. Mais il n'y a *pas de critère précis* qui puisse être dégagé. Les décisions de la CNIL s'appuient sur le type d'informations utilisées, mais surtout sur la logique qui va présider à leur traitement. L'article 2 de la directive définit la notion de données à caractère personnel et donne quelques exemples, sans pour autant délimiter le champ d'application. Pourtant, à la lecture du rapport du Sénat sur le projet de transposition, le lecteur pourrait avoir l'impression inverse, puisque l'auteur écrit :

> *« La directive prévoit des critères permettant de délimiter le champ des données concernant une personne identifiable… ».*

Toutes les données, quelle que soit leur forme ou leur support, peuvent tomber dans le champ d'application du cadre légal « informatique et libertés ». Le considérant 14 mentionne explicitement le son, l'image et la voix. On peut aujourd'hui y ajouter, en raison des progrès de la génétique l'ADN, ou l'iris de l'œil.

L'identification d'une personne peut se faire de manière directe ou indirecte. Le caractère personnel d'une donnée dépend des moyens de tri, de rapprochement qui pourraient être mis en œuvre. Cela conduit donc à une évolution constante du champ des données personnelles, la technique mettant à la disposition du plus grand nombre des outils de plus en plus performants.

Certains auteurs avancent qu'il suffit qu'il y ait une *probabilité suffisante de rapprochement* avec une personne pour qu'une donnée acquière un caractère personnel indirect. Les analyses ne sont pas toujours explicites mais il n'est pas possible de négliger cet argument[37]. Dans le domaine statistique, la CNIL a imposé des seuils au-delà desquels des rapprochements d'agrégats de données – pourtant individuellement anonymes – sont interdits.

Le caractère personnel d'une information dépend de l'objet qu'elle décrit, du contexte dont elle provient, mais aussi de la personne qui la reçoit. Pour pouvoir identifier un individu ou un groupe, nous avons besoin des informations, mais nous n'y arriverons pas sans un élément de connaissance propre qui déclenchera le mécanisme d'association. *Le récepteur constitue un élément important* de l'équation. Le phénomène croissant de marchandisation des données attire les convoitises de toutes sortes d'individus, pas toujours à même de les exploiter en dehors d'une transaction commerciale. Il nous semble donc important d'inclure dans une réflexion les capacités des personnels qui traitent les données. Pour un informaticien qui gère un système traitant des données génétiques, les données auxquelles il accède, sont-elles pour lui des données à caractère personnel ?

LA NOTION DE TRAITEMENT DES DONNEES A CARACTERE PERSONNEL

L'article 2b de la directive européenne définit ainsi le traitement des données :

« toute opération ou ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé, et notamment la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction. »

La longueur de la définition illustre avant tout le champ des possibilités ouvert par l'outil informatique, tout en allant plus loin car c'est à tous les types de traitements qu'il est ici fait allusion. La distinction entre traitement automatisé et non automatisé n'a plus cours. Il en va de même pour la notion de fichier, le législateur européen met sur le même plan les fichiers informatiques et manuels : *il suffit que les données soient organisées suivant une structure définie*[38].

Le traitement n'implique pas forcément une manipulation du fichier, un simple stockage suffit à le faire entrer dans le champ d'application. La difficulté vient une nouvelle fois de la très grande portée de la définition.

Nous retrouvons l'interrogation que nous avions soulevée : la difficulté qu'éprouve le législateur à éviter la systématisation trop grande des notions qu'il veut défendre. Une donnée n'acquiert pas forcément de caractère personnel par sa nature, tout dépend de celui qui l'utilise. Nombreux sont ceux qui soulignent à juste titre qu'il est impossible d'admettre des définitions trop vastes, sous peine de les rendre inapplicables, et que mieux vaudrait se concentrer sur des types définis qui remettent en cause des valeurs fondamentales[39].

---

[37] Lamy *Droit de l'informatique et des Réseaux* 508 et suivants.
[38] Considérant 27, de la directive 95/46
[39] Frayssinnet, J. dir. (2001) *Droit de l'Informatique et de l'Internet*, Paris, PUF, §127 : 85-86.

L'ANONYMISATION

L'*anonymisation* sert à qualifier l'opération par laquelle se trouve supprimé dans un ensemble de données, recueilli auprès d'un individu ou d'un groupe, tout élément qui permettrait l'identification de ces derniers. Le nom propre n'est donc pas le seul élément qu'il faille prendre en compte. On pourrait parler de « *dépersonnalisation* » des données comme dans la loi fédérale allemande sur la protection des données à caractère personnel du 23 mai 2001.

Lorsqu'on réfléchit à l'anonymisation, il convient de connaitre les éléments à traiter, mais aussi les opérations que vont subir les données.

La difficulté soulevée par la question de l'anonymisation apparait plus clairement après avoir rapidement passé en revue quelques principes clés des lois encadrant l'informatique. Il ne s'agit pas tant de savoir comment effectuer le travail d'anonymisation, mais plutôt de définir quelles données doivent être anonymisées, pour qui, et dans quel contexte.

L'exemple des pratiques autorisées pour la recherche médicale fournit quelques pistes de réflexion.

La condition première est d'avoir un responsable ainsi qu'une ou plusieurs finalités précises. Les données transmises ne peuvent l'être que si elles sont destinées à des membres du même milieu professionnel, soumis aux mêmes règles déontologiques. Le plus souvent celui qui reçoit les données doit pouvoir travailler sur des données anonymes.

L'anonymat doit être irréversible, et la CNIL est seule habilitée à autoriser la fourniture de données non anonymisées après examen du projet scientifique. La publication ou un autre mode d'exploitation des résultats ne peut donner lieu *en aucune manière* à une possible identification des personnes.

L'obligation d'obtenir un consentement préalable peut être levée si retrouver les personnes concernées s'avère difficile. S'il n'a pas été recueilli immédiatement, le consentement doit être obtenu avant le premier traitement. Les demandes de dérogation sont du ressort exclusif de la CNIL.

Voici les divers procédés qu'elle préconise :

- Le codage : les données personnelles sont cryptées par des clés cryptographiques générées par des logiciels informatiques.
- Les bases de données séparées : Le réseau SESAME-VITALE utilise bien entendu le cryptage des données. Mais pour garantir un maximum de confidentialité, deux types de bases de données ont été distingués. Des bases primaires contiennent toutes les données mais elles ne sont pas connectées au réseau, elles servent de sécurité et disposent de tables de concordance pour lever l'anonymat après autorisation. D'autres bases de données assurent le fonctionnement quotidien du réseau, mais seules les données nécessaires sont présentes.

Une autre voie existe : Les limitations techniques. La loi québécoise « concernant le cadre juridique des technologies de l'information » propose de protéger l'anonymat non pas en modifiant les données, mais en limitant les possibilités de recherche, voire en les adaptant à la personne qui consulte la base selon des critères bien précis (sa profession, une autorisation, sa présence dans le fichier, etc.)

Cette dernière perspective offre pour la constitution et l'exploitation de corpus oraux la possibilité de faire coïncider les obligations légales avec les nécessités

du travail de recherche. *Toute donnée étant potentiellement sensible, une anonymisation systématique s'avère de plus en plus complexe ; elle peut même mettre en danger l'intérêt de certaines recherches.* En effet, des détails concernant les personnes comme par exemple le nom, ou le lieu d'habitation peuvent constituer un élément important du corpus, et des résultats qui peuvent en être obtenus. C'est pourquoi la possibilité de ménager des niveaux d'accès selon des critères stricts (ex : chercheur ou non, présence d'autorisation, but de la consultation, etc.) semble une alternative efficace.

Il existe d'autres procédés à inventer. En effet, l'article 11-2 de la nouvelle loi ouvre la possibilité de faire certifier des techniques nouvelles par la CNIL. Ce n'est pas au chercheur de présenter son procédé, mais à l'institution à laquelle il appartient.

Il faut bien sûr que le type de données collectées ait fait l'objet d'une réflexion quant à son intérêt pour l'étude entreprise ; sous peine de mettre à mal les garanties mises en place et au risque de ne pas obtenir l'accord des autorités compétentes.

Pour le droit à l'image se référer à :

Pierrat, E. (2002) *Reproduction interdite : le droit à l'image expliqué aux professionnels et à ceux qui souhaitent se protéger*, Paris, Maxima Laurent du Mesnil.

Isgour, M. & Vinçotte, B. (1998) *Le droit à l'image,* Bruxelles, Larcier.

Serna, M. (1997) *L'image des personnes physiques et des biens*, Paris, Economica.

Bécourt, D. (2004) *Image et vie privée*, Paris, L'Harmattan.

Bloch, P. dir. (2002) *Image et Droit,* Paris, L'Harmattan.

LE DROIT DE CITATION

Citer des œuvres est un acte important qui s'impose particulièrement dans tout travail scientifique et la constitution et l'utilisation de corpus n'échappent pas à la mise en œuvre du droit de citation. Même si le droit d'auteur prend en compte la citation, l'interprétation qui est faite de ce droit suscite bien souvent des interrogations. Beaucoup d'idées reçues entourent le droit de citation ; comme par exemple l'existence d'un pourcentage défini entre l'extrait choisi et l'œuvre dont il est extrait. Il convient donc de traiter ici le cadre général du droit de citation et les applications particulières à certaines catégories d'œuvres ou certains modes de citation.

LE CADRE GENERAL DU DROIT DE CITATION

Le droit de citation est une exception au droit de reproduction. En effet, tout ou partie d'une œuvre originale ne peut être reproduit par quelque moyen que ce soit sans autorisation de l'auteur (voir fiche *Œuvres protégées*). Ainsi l'article L122-5 3° permet les analyses et courtes citations. Trois conditions doivent alors être observées :

- **La citation doit être justifiée** : il faut un but (critique, scientifique…) et elle doit être incorporée à un développement lié à ce but (démonstration, exposé). Sinon on en revient au cas du recueil qui constitue une œuvre en lui-même ; sauf pour ce qui appartient au domaine public.
- **La citation doit être courte** : les extraits ne peuvent reprendre l'essentiel de l'œuvre dont ils sont issus. L'œuvre qui incorpore des citations doit pouvoir « survivre » à leur suppression. L'appréciation de la brièveté est fonction du rapport entre les citations et l'œuvre citante dans laquelle elles sont incorporées.
- **La citation doit respecter le droit moral de l'auteur** : cela signifie la mention explicite de l'auteur de l'œuvre citée (*droit de paternité*) ; mais aussi la préservation de l'intégrité de l'œuvre, tant dans la forme que dans l'esprit. Enfin, si une œuvre n'a pas été divulguée, la citation est interdite.

Toutes les règles qui viennent d'être énoncées correspondent parfaitement aux œuvres écrites. Il faut aussi s'intéresser aux questions particulières des autres types d'œuvres et voir dans quelles conditions le code de la propriété intellectuelle prévoit des aménagements.

LES CAS PARTICULIERS

- **Les œuvres graphiques ou appartenant aux arts plastiques** : « une jurisprudence constante exclut la reproduction intégrale d'une œuvre au titre de la citation. Quant à la reproduction partielle de l'œuvre (partie du dessin, tableau ou photo) elle porte atteinte à l'intégrité de l'œuvre et ne peut se faire sans l'autorisation de l'auteur. En conséquence, l'exception (le droit) de citation ne peut être invoquée. »
- **Les œuvres musicales** : il n'y a pas d'exclusion de principe de la citation dans le domaine musical. Les règles encadrant leur citation sont les mêmes que pour les autres types d'œuvres.

- **Les bases de données** : pour le droit, une base de données constitue un objet spécifique. Elle bénéficie d'ailleurs d'un droit spécifique (voir fiche *Base de données*). Concernant la citation, la jurisprudence lui a aussi reconnu un régime spécifique. Depuis l'arrêt *Microfor* de 1988, il est admis qu'une base de données peut être constituée exclusivement d'extraits d'œuvres sans qu'il y ait d'autres apports. Dans ce seul cas, les deux exigences posées dans le cadre général (citation justifiée, brièveté) disparaissent. Toutefois subsiste l'obligation de mentionner de façon explicite l'auteur de l'œuvre et l'origine de celle-ci.
- **Le lien hypertexte** : l'usage du lien hypertexte constitue la base de la navigation Internet. Insérer un lien vers un autre document peut s'assimiler à une citation si les trois règles de bases sont respectées, et surtout si l'objet cité n'est pas illicite ou si sa reprise n'est pas interdite par son auteur. La fourniture de lien vers un objet contrefait devient, par exemple, de la complicité. Si la liberté de citation n'apparait pas clairement, il est là aussi préférable d'entrer en contact avec l'auteur.

LE CONSENTEMENT

Le consentement de la personne concernée, dans la plupart des situations de traitement de données personnelles, constitue le plus souvent une manière d'assurer sa protection. Toutefois, il n'est pas toujours possible d'obtenir ce consentement lors de la collecte des données. De plus, il est important d'indiquer que l'obtention du consentement n'exonère pas le responsable du traitement de ses obligations à l'égard des personnes concernées (voir fiche *Responsable du traitement*).

LE PRINCIPE DU CONSENTEMENT

Le consentement doit être éclairé. Pour ce faire, le responsable du traitement doit, en principe, procéder ou faire procéder à une *information préalable* de la personne avant de recueillir son consentement. Le nouvel article 32 de la loi de 1978 énonce clairement les informations qui doivent être fournies :

- l'identité du responsable du traitement et, le cas échéant, celle de son représentant ;
- la finalité poursuivie par le traitement auquel les données seront soumises ;
- le caractère obligatoire ou facultatif des réponses ;
- les conséquences éventuelles, à cet égard, d'un défaut de réponse ;
- les destinataires ou catégories de destinataires des données ;
- l'existence d'un droit d'accès, de rectification voire d'opposition à la collecte ;
- les transferts de données à caractère personnel envisagés à destination d'un État non membre de la Communauté Européenne.

En principe, le consentement doit être exprès. La personne affirme clairement qu'elle accepte que les données personnelles la concernant fassent l'objet d'un traitement. Même si le législateur ne l'impose pas, le consentement écrit est considéré comme une bonne pratique. Dans des situations particulières, d'autres formes peuvent être choisies, l'important est de pouvoir faire la preuve de la volonté de la personne concernée (ex. : enregistrement d'un accord verbal).

Qui consent ? Toute personne physique. Lorsqu'il s'agit d'une personne déclarée incapable (qu'elle soit majeure ou mineure), l'information doit parvenir au représentant légal. Pour le cas d'enfants, il faut l'autorisation des parents ou du dépositaire de l'autorité parentale.

Il est vrai que l'écrit tient une place importante dans le formalisme du consentement. Quand, dans certaines situations, cela se révèle impossible à mettre en œuvre sous peine de fausser les résultats de la recherche engagée, des solutions alternatives existent.

Même si elle s'est exprimée de façon expresse, toute personne ayant consenti au recueil des données et à leur traitement dispose, pour des motifs légitimes, d'un droit de rétractation ou d'opposition à ce que des données à caractère personnel la concernant continuent à faire l'objet d'un traitement.

DES ALTERNATIVES AU CONSENTEMENT

Le consentement de la personne concernée ne suffit pas toujours à la protéger contre des utilisations abusives des données qui la concernent. Il faut aussi veiller à ne pas faire de l'exigence du consentement une obligation administrative qui ferait perdre de vue ce qui compte : la protection de la personne. C'est pourquoi les textes récents ont mis en place des alternatives au consentement ou même des garanties spécifiques pour le traitement de certains types de données (voir fiche *Responsable du traitement*). Parmi les alternatives pouvant s'appliquer au traitement des corpus oraux on citera :

> *«la réalisation de l'intérêt légitime poursuivi par le responsable du traitement ou par le destinataire, sous réserve de ne pas méconnaître l'intérêt ou les droits et libertés fondamentaux de la personne concernée. » (art. 7, 5°).*

Toutefois, l'intérêt légitime de la recherche et du traitement aura à être démontré. En pratique l'application de cette alternative a pour conséquence de dispenser le responsable du traitement de demander un consentement exprès à condition, bien entendu, de ne pas méconnaitre l'intérêt ou les droits et libertés fondamentaux de la personne concernée.

EXEMPLES D'AUTORISATIONS

Voici à titre *d'exemples* deux formulaires d'autorisation, tirés d'expériences de chercheurs en France (ICAR) et aux États-Unis (Ervin-Tripp).

Les formulaires d'autorisation diffèrent notamment en ce qui concerne la présentation des options proposées à l'informateur. Celles-ci concernent essentiellement les contextes d'exploitation des données et la forme des données montrables en public (différents supports, anonymisés ou non).

Ces exemples ne peuvent constituer un modèle à reprendre tel quel, seul un travail sur l'explicitation de la démarche et les objectifs d'exploitation de chaque projet permet de construire un formulaire d'autorisation adéquat.

De manière générale, il est fortement conseillé d'adapter le formulaire aux visées particulières de l'enquête, notamment aux objets que l'on désire recueillir et étudier, aux types d'acteurs sociaux concernées par l'enquête et aux conditions d'exploitation et de diffusion du corpus.

1) Exemple de formulaire type de demande d'autorisation mis au point au laboratoire ICAR (UMR 5191 CNRS)

[papier avec entête officiel]

Autorisation

pour l'enregistrement audio/vidéo et l'exploitation des données enregistrées

Présentation de l'enquête

[Peut se présenter sous forme de brochure séparée laissée aux enquêtés]

[Préciser l'institution d'où émane la recherche, la personne qui dirige/qui est responsable du projet, les chercheurs concernés sur le terrain.

Préciser le thème général du projet, le type de corpus qui est recueilli de manière générale, le type d'enregistrement qui est recueilli auprès de ces informateurs en particulier, son traitement et utilisation prévus.

Souligner les apports du projet, valoriser la collaboration de l'informateur, expliciter les bénéfices éventuels qu'il peut en tirer et les risques éventuels qu'il peut courir.]

Ces recherches ne sont possibles que grâce au consentement des personnes qui acceptent d'être enregistrées. Nous vous demandons par conséquent votre autorisation à procéder aux enregistrements.

Autorisation (biffer les paragraphes qui ne conviennent pas)

Je soussigné(e)

- autorise par la présente NN et NN à enregistrer en audio/vidéo le [préciser le type d'événement enregistré].

- autorise l'utilisation de ces données, sous leur forme enregistrée aussi bien que sous leur forme transcrite et anonymisée (cf. *infra*) :

a) à des fins de recherche scientifique (mémoires ou thèses, articles scientifiques, exposés à des congrès, séminaires) ;

b) à des fins d'enseignement universitaire (cours et séminaires donnés à des étudiants avancés, à partir du niveau maitrise, en sciences du langage et en sciences sociales) ;

c) pour une diffusion large dans la communauté des chercheurs, sous la forme d'éventuels échanges et prêts de corpus à des chercheurs, moyennant la signature d'une convention de recherche ;

d) pour une diffusion sur un site Internet dédié à la recherche.

- prends acte que pour toutes ces utilisations scientifiques les données ainsi enregistrées seront anonymisées, cela signifie :

a) que les transcriptions de ces données utiliseront des pseudonymes et remplaceront toute information pouvant porter à l'identification des participants ;

b) que les bandes audio qui seront présentées à des conférences ou des cours (généralement sous forme de très courts extraits ne dépassant pas la minute) seront « bipées » lors de la mention d'un nom, d'une adresse ou d'un numéro de téléphone identifiables (qui seront donc remplacés par un « bruit » qui les effacera) ;

c) en revanche, pour des raisons techniques, le projet ne peut pas s'engager à anonymiser les images vidéo mais s'engage à ne pas diffuser d'extraits compromettant les personnes filmées.

- souhaite que la contrainte supplémentaire suivante soit respectée :

Lieu et date:

Signature :

[Prévoir un double ou un autre document qui sera laissé à la personne, comportant une adresse de contact et éventuellement une adresse Internet où consulter les résultats publiés du projet].

2) Exemple de demande d'autorisation de Susan Ervin-Tripp, Univ. de Californie, Berkeley

Researcher name

LETTER OF CONSENT

PHOTOGRAPHIC, AUDIO, AND/OR VIDEO RECORDS RELEASE CONSENT FORM

As part of this project we have made a photographic, audio, and/or video recording of you while you participated in the research. We would like you to indicate below what uses of these records you are willing to consent to. This is completely up to you. We will only use the records in ways that you agree to. In any use of these records, names will not be identified.

1. The records can be studied by the research team for use in the research project.

Photo ..... Audio ..... Video .....

[Please use initials]

2. The records can be shown to subjects in other experiments.

Photo ..... Audio ..... Video .....

[Please use initials]

3. The records can be used for scientific publications.
Photo ..... Audio ..... Video .....
[Please use initials]
4. The written transcript can be kept in an archive for other researchers.
Photo ..... Audio ..... Video .....
[Please use initials]
5. The records can be used by other researchers.
Photo ..... Audio ..... Video .....
[Please use initials]
6. The records can be shown at meetings of scientists interested in the study of.....
Photo ..... Audio ..... Video .....
[Please use initials]
7. The records can be shown in classrooms to students.
Photo ..... Audio ..... Video .....
[Please use initials]
8. The records can be shown in public presentations to nonscientific groups.
Photo ..... Audio ..... Video .....
[Please use initials]
9. The records can be used on television and radio.
Photo ..... Audio ..... Video .....
[Please use initials]
I have read the above description and give my consent for the use of the records as indicated above.
Date .....
Signature .....
Signature of Guardian, if Applicable .....
Native language(s) .....
Where native language learned (city or region) .....
Languages used on the tape
Where language(s) used on tape were learned .....
Age at which each language used on tape was learned .....
Education Occupation .....
Name Age ..... Sex.....

BASES DE DONNÉES, OBJET D'UN DROIT « SUI GENERIS »

Avec l'utilisation des nouvelles technologies, la création d'un corpus oral aboutit, le plus souvent, à la création d'une « base de données » renfermant toutes les informations recueillies, transformées et produites au cours des différentes phases du travail de recherche. Peu d'activités n'utilisent pas de bases de données.

L'alinéa 2 de l'article L112-3 du Code de Propriété Intellectuelle (CPI) définit la notion de base de données :

> *« On entend par base de données un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique et individuellement accessibles par des moyens électroniques ou par tout autre moyen ».*

Ainsi, pour constituer une base de données, il faut des données mais aussi une structure pour les ordonner. Les données ont un statut propre, indépendant de celui de la base de données. Elles peuvent être des œuvres protégées par le droit d'auteur, elles peuvent être des données personnelles. En fonction de chaque statut, elles relèveront des cadres juridiques correspondants.

La base de données, quant à elle, relève des œuvres susceptibles d'être protégées par le droit d'auteur. A côté de ce régime de protection dont bénéficie l'auteur de la base de données, l'investisseur (ou producteur) de la base dispose d'un droit dit « sui generis » qui le protège contre des exploitations et utilisations abusives des données de la base. Ainsi, l'auteur et le producteur bénéficient de droits différents et ces régimes ne s'excluent pas. Ces deux régimes de protection peuvent bénéficier à une seule et même personne qui serait à la fois auteur et producteur. Les bénéficiaires peuvent être aussi des personnes différentes.

LE DROIT « SUI GENERIS » DES BASES DE DONNEES, PROTECTION DE L'INVESTISSEMENT

Qui bénéficie de ce droit ? En quoi consiste le régime de protection mis en place ?

LA TITULARITE DES DROITS

Dans le Code de la propriété intellectuelle le bénéficiaire du « droit sui generis »[40] est appelé le « producteur ». Selon l'article L341-1 du CPI le producteur est celui « *qui prend l'initiative et le risque des investissements correspondants… »*. Il ne s'agit donc *pas* forcément du *concepteur* de la base, mais plutôt de celui, ou de ceux qui ont pris *l'initiative*, les décisions clés et qui, avec les investissements requis, ont permis la réalisation de la base.

L'article L341 du CPI précise qu'il faut un *investissement* financier, matériel ou humain *substantiel* pour bénéficier du droit sui generis. Ce dernier critère permet de distinguer une base de données d'une simple compilation (simple reprise d'éléments contenus dans une autre base). Ainsi, une simple reprise des annuaires de France-Télécom ne peut faire l'objet d'une protection par ce régime particulier. Il faut aussi mettre en évidence le volume de travail, le cout des interventions lors de la création de la base comme dans les mises à jour.

[40] Art. L 341-1, 342-2, 342-1 à 342-5 du CPI.

La jurisprudence refuse la protection du droit « sui generis » à une revue d'annonces légales au motif que la revue ne justifie pas d'investissements substantiels dans leur obtention et dans leur traitement.

Les « producteurs » de corpus oraux ne pourront bénéficier de la protection du droit « sui generis » que s'ils font la démonstration de la réalité de l'investissement substantiel réalisé. Dans ce cas, ils bénéficieront des droits correspondants.

LES DROITS DU PRODUCTEUR

Le droit principal concerne l'*interdiction* pour l'utilisateur légitime d'extraire de façon substantielle les données contenues dans la base. Par *substantielle*, le code vise autant la qualité que la quantité des données extraites (art. L 341-1 et 341-2 du CPI). Ainsi, le terme *substantiel* s'apprécie au cas par cas. Des informations rares – bien qu'en petit nombre – peuvent tomber sous le coup de l'interdiction. Une extraction, même non substantielle, peut se voir interdite si elle a un caractère *répété* ou *systématique*. Le but avoué est ici d'empêcher le pillage de bases de données par des concurrents mal intentionnés.

Il faut garder à l'esprit que ces possibilités d'interdiction sont un droit et non pas une obligation. Le producteur peut autoriser – moyennant contrepartie – ces extractions. Le droit sui generis du producteur prend effet à compter de l'achèvement de la fabrication de la base de données et expire quinze ans après le 1er janvier de l'année civile qui suit celle de cet achèvement (art. L 342-5 du CPI). Si à la fin de la période, il y a un nouvel investissement substantiel, la protection se voit renouvelée. Les atteintes au droit sui generis sont sanctionnées pénalement. En cas d'infraction, les peines sont de deux ans de prison et de 150 000 euros d'amende. Pour les personnes morales, l'emprisonnement se transforme en interdiction d'exercice.

LA COEXISTENCE AVEC LES AUTRES DROITS ET LES LIMITES

Les droits du producteur trouvent leur première limite :

- Dans le droit des utilisateurs légitimes à extraire ou réutiliser une partie non substantielle du contenu de la base.
- Le législateur a prévu, dans certains cas[41], un statut dérogatoire à l'extraction à des fins privées d'une partie de la base, que cette extraction soit qualitativement ou quantitativement substantielle. Il faut noter que la directive européenne de 1996 prévoyait, derrière cette dérogation, le cas de l'enseignement ou de la recherche scientifique à but non commercial. Cette dernière disposition n'a pas été retenue par le législateur. En fait, les deux conditions imposées limitent fortement ce statut dérogatoire : seuls sont visés les contenus de bases de données non électroniques, et cette extraction doit respecter les droits d'auteurs ou les droits voisins sur les œuvres ou éléments incorporés dans la base.
- Le producteur doit aussi, le plus souvent, assurer l'*accès* aux informations tout en garantissant leur *licéité* ainsi que leur *fiabilité*. Les informations doivent donc être mises à jour et avoir été obtenues de manière légale (les droits éventuels liés à ces informations ne peuvent pas être ignorées).

[41] Art. 342-3 2°.

RESPONSABLE DU TRAITEMENT

Tout traitement de données doit avoir un responsable. Sa mission est d'éviter ou de circonvenir les risques inhérents à la gestion et à l'utilisation des données recueillies. La loi définit qui est le responsable et lui fixe donc des obligations.

L'article 3-1 de la loi du 6 janvier 1978 dispose :

> *« Le responsable d'un traitement de données à caractère personnel est, sauf désignation expresse par les dispositions législatives ou réglementaires relatives à ce traitement, la personne, l'autorité publique, le service ou l'organisme qui détermine ses finalités et ses moyens. »*

Qui est-il ? C'est une personne physique qui détient le pouvoir de décision sur les finalités et les moyens à mettre en œuvre.

LES PRINCIPES GENERAUX

Le responsable du traitement se doit donc de veiller à la qualité des données, au respect des finalités indiquées, au respect du principe de licéité et aux conditions de conservation.

LA QUALITE DES DONNEES

Pour pouvoir être traitées, les données doivent avoir été recueillies selon un ensemble de principes qui garantissent la protection des personnes. Une donnée doit être :

- **adéquate, pertinente et non excessive**. Toutes les données faisant l'objet d'un traitement doivent être en lien avec la finalité poursuivie. La CNIL se montre particulièrement vigilante sur ce point. L'INSEE s'est souvent vu refuser ses questionnaires ou être obligé de les revoir car les données collectées étaient jugées trop nombreuses ou inutiles par rapport à la finalité annoncée. Plus on acquiert de données sur un même individu, plus le risque est grand de voir ce traitement surveillé étroitement, voire refusé par la CNIL.
- **exacte**. Les données doivent être exactes et mises à jour. Ceci renvoie au droit d'accès, d'opposition et de rectification ouvert à chaque personne concernée par le traitement.

LE RESPECT DES FINALITES INDIQUEES

La finalité du traitement sert à justifier celui-ci. Il s'agit de répondre à la question du but de la mise en œuvre d'un ou plusieurs traitements. De même il peut y avoir plusieurs finalités. Le responsable du traitement - selon la définition - détermine la finalité. Il doit donc annoncer *par avance* le but du traitement qu'il s'apprête à réaliser. Le responsable du traitement qui justifierait après le traitement les finalités poursuivies manquerait à ses obligations légales et serait susceptible d'être sanctionné pénalement.

LE RESPECT DU PRINCIPE DE LICEITE

Toute donnée collectée doit avoir été recueillie loyalement. Cela suppose une *information préalable*, une *demande écrite de consentement* (voir fiche *Consentement*), l'explication quant à la *finalité du traitement*, le *nom du responsable du traitement*, ainsi que les *conséquences en cas de refus*. La notion de loyauté renvoie au contexte dans lequel s'est effectuée la collecte.

LES CONDITIONS DE CONSERVATION

Confidentialité et conservation limitées dans le temps. Il appartient au responsable du traitement d'assurer la confidentialité et le respect des règles de communication de ces données hors du cadre défini pendant toute la durée de conservation de ces données. La durée de conservation varie selon la finalité du traitement effectué. Les données peuvent être conservées au-delà de la durée prévue initialement quand elles présentent un intérêt pour des fins historiques, statistiques ou scientifiques (art. 36). Cette possibilité de conservation n'entraine pas la possibilité d'exploitation ni de diffusion, les conditions d'accès aux données étant réglées par la loi sur les archives.

LES FORMALITES PREALABLES : DECLARATION ET AUTORISATION

DECLARATION

En principe, pour les catégories les plus courantes de traitement dont la mise en œuvre n'est pas susceptible de porter atteinte à la vie privée ou aux libertés, la formalité requise est une déclaration à la CNIL. Dans les situations répétitives, cette déclaration peut être simplifiée. La CNIL délivre sans délai un récépissé et, dès réception de celui-ci, le responsable peut mettre en œuvre le traitement. La déclaration comporte l'engagement que ce traitement satisfait aux exigences de la loi (respect du principe de licéité, voir *supra*).

AUTORISATION

Si les corpus contiennent des données sensibles, le responsable du traitement devra demander une autorisation à la CNIL qui dispose d'un délai de deux mois pour se prononcer (délai susceptible d'être renouvelé une fois). L'absence de réponse de la CNIL dans les délais doit être interprétée comme un rejet de la demande d'autorisation.

LE PATRIMOINE IMMATERIEL ET L'UNESCO

Ces fiches ont pour objet de présenter les questions posées par les documents (déclarations, conventions, autres…) de l'UNESCO. En effet, la constitution de corpus oraux et la recherche sur les langues participent à la protection du patrimoine culturel de l'humanité qui est une des grandes missions de cette organisation internationale. La constitution de grands corpus oraux peut servir de documentation générale sur des langues et contribuer à l'élaboration des outils de diffusion de langues peu (ou pas) écrites[42].

L'UNESCO s'est intéressée aux différentes formes de régulation des recherches portant sur le patrimoine culturel : régulation éthique, déontologique et juridique.

ÉTHIQUE ET DEONTOLOGIE DE LA RECHERCHE

Dans la recommandation de 1989 (article E.g) la communauté scientifique internationale est encouragée « *à se doter d'une éthique appropriée à l'approche et au respect des cultures traditionnelles* ». Le chercheur doit être animé d'un souci de respect à l'égard de ses collaborateurs occasionnels (sujets de recherche), dont il devra rechercher la confiance, et à l'égard des traditions de ceux qu'il étudie.

Par ailleurs, l'exigence pour les chercheurs de se doter d'un code de déontologie a été posée notamment lors de la Conférence de Washington sur l'évaluation globale de la recommandation de 1989 relative à la sauvegarde de la culture traditionnelle et du folklore[43]. La cinquième recommandation faite à l'UNESCO l'invite à : *« encourager les groupements internationaux (chercheurs, professionnels de la culture…) à créer et à adopter des codes déontologiques qui assurent que des démarches appropriées et respectueuses sont suivies vis-à-vis de la culture traditionnelle et du folklore.* »

LES RECOMMANDATIONS RELATIVES A L'ENCADREMENT JURIDIQUE DES TRAVAUX SUR LES LANGUES

Invitant le lecteur à s'intéresser aux cadres normatifs dans lesquels les recherches sur les langues sont menées et souhaitant ouvrir sur des exemples de législations nationales, ces fiches proposent des pistes de réflexion sur les questions posées par la recherche sur les langues en voie de disparition. Bien entendu, il n'était pas possible de traiter de toutes les situations locales et ces exemples sont là pour inciter les chercheurs à se renseigner sur les droits nationaux susceptibles de s'appliquer dans les pays où sont menées les recherches.

Une des questions qui se posent lorsqu'on appréhende les travaux sur la langue est celle de la détermination de son statut juridique en tant qu'élément du patrimoine culturel : fait-elle partie du domaine public et, partant, libre de tout droit ou, au contraire, s'agit-il d'un bien appropriable, et, dans ce cas, quelles sont les conséquences pour le travail des chercheurs ?

Ces différentes interrogations nous conduisent à analyser dans les documents de l'UNESCO et les législations de quelques pays africains ce qui est dit sur le statut des langues (fiche I). De ce statut découleront les conditions dans

---

[42] Voir *supra* 2-1.

[43] Conférence précitée, voir note 26.
http://www.folklife.si.edu/resources/Unesco/actionplan_french.htm.

lesquelles peuvent être menées les recherches et la constitution des corpus oraux (fiche II). La question d'un droit à la protection de la vie privée à travers les recommandations de l'UNESCO et dans quelques pays africains complètera cette présentation (fiche III).

LA DETERMINATION DU STATUT JURIDIQUE DE LA LANGUE

L'UNESCO et la reconnaissance explicite de la langue comme élément du patrimoine culturel immatériel de l'Humanité

La caractéristique des textes de l'UNESCO est de cerner la langue d'un point de vue collectif en ceci qu'elle fait partie du patrimoine culturel de l'Humanité. On peut à cet égard se référer à la description qui en est donnée par sa section du patrimoine immatériel. Il en ressort que :

> *« Les langues sont la plus grande création et expression du génie de l'humain. Elles ne sont pas uniquement des outils complexes et raffinés de communication. Elles constituent un élément déterminant de l'identité humaine et, à ce titre, représentent un noyau primordial du patrimoine culturel de l'Humanité. »*

Trois textes principaux contribuent à cette appréhension de la langue comme élément du patrimoine culturel :

> *Tout d'abord, la Recommandation sur la sauvegarde de la culture traditionnelle et populaire*[44]*. Selon l'article A de cette Recommandation, la culture traditionnelle et populaire est : « l'ensemble des créations émanant d'une communauté culturelle fondées sur la tradition, exprimées par un groupe ou des individus et reconnues comme répondant aux attentes de la communauté en tant qu'expression de l'identité culturelle et sociale de celle-ci, les normes et les valeurs se transmettant oralement, par imitation ou par d'autres manières. Ses formes comprennent, entre autres, la langue, la littérature, la musique, la danse, les jeux, la mythologie, les rites, les coutumes, l'artisanat, l'architecture et d'autres arts. »*
>
> *Ensuite, la Déclaration universelle de l'UNESCO sur la diversité culturelle*[45]*, qui considère que la culture est : « l'ensemble des traits distinctifs spirituels et matériels, intellectuels et affectifs qui caractérisent une société ou un groupe social et qu'elle englobe, en outre les arts et les lettres, les modes de vie, les façons de vivre ensemble, les systèmes de valeurs, les traditions et les croyances. » Le cinquième point de son Plan d'Action vise à : « sauvegarder le patrimoine linguistique de l'Humanité et soutenir l'expression, la création et la diffusion dans le plus grand nombre possible de langues. »*

[44] Recommandation sur la sauvegarde de la culture traditionnelle et populaire, 15 novembre 1989.
[45] Déclaration universelle sur la diversité culturelle, 17 octobre 2001.

> *Enfin, la Convention pour la sauvegarde du patrimoine culturel immatériel*[46]*, qui est la consécration du patrimoine immatériel. Par patrimoine culturel immatériel, il faut entendre selon l'article 2 de ladite Convention : « les pratiques, représentations, expressions, connaissances et savoir-faires - ainsi que les instruments, objets, artefacts et espaces culturels qui leur sont associés - que les communautés, les groupes et, le cas échéant, les individus reconnaissent comme faisant partie de leur patrimoine culturel. Ce patrimoine culturel, transmis de génération en génération, est recréé en permanence par les communautés et les groupes en fonction de leur milieu, de leur interaction avec la nature et de leur histoire, et leur procure un sentiment d'identité et de continuité, contribuant ainsi à promouvoir le respect de la diversité culturelle et de la créativité humaine. »*

Il se manifeste dans les domaines suivants :

- les traditions et expressions orales, y compris la langue comme vecteur du patrimoine culturel immatériel ;
- les arts du spectacle ;
- les pratiques sociales, rituels et événements festifs ;
- les connaissances et pratiques concernant la nature et l'univers;
- les savoir-faires liés à l'artisanat traditionnel.

LANGUES ET FOLKLORE

Les composantes sus-citées du patrimoine culturel (y compris la langue) qui sont l'objet des recherches en linguistique, reçoivent aussi la qualification de folklore ou d'expressions du folklore. C'est d'ailleurs cette terminologie qui a été proposée comme modèle pour les législations nationales. Pour s'en convaincre, on peut se référer aux Dispositions types de législation nationale sur la protection des expressions du folklore contre leur exploitation illicite et autres actions dommageables[47]. Mais cette catégorie ne doit pas faire illusion car elle semble se fondre - même si les logiques ne sont pas les mêmes - dans les catégories précitées. Ainsi, un auteur[48], pour définir le folklore, se réfère entièrement à la convention de 1989 sur la culture traditionnelle et populaire. Par expressions du folklore, il faut entendre, au sens de l'article 2 des Dispositions types :

> *« les productions se composant d'éléments caractéristiques du patrimoine artistique traditionnel développé et perpétué par une communauté ou par des individus reconnus comme répondant aux aspirations artistiques traditionnelles de cette communauté, en particulier*
>
> *les expressions verbales telles que les contes populaires, la poésie populaire et les énigmes ;*

[46] Convention pour la sauvegarde du patrimoine culturel immatériel, 17 octobre 2003.

[47] Élaborées conjointement par l'Unesco et l'OMPI et approuvées par un comité d'experts gouvernementaux en 1985.

[48] Folarin, S. (2002) « Conservation, préservation et protection juridique du folklore en Afrique », *Bulletin du droit d'auteur*, XXXII/4 :41.

> *les expressions musicales telles que les chansons et la musique instrumentale populaires ;*
>
> *les expressions corporelles telles que les danses et les spectacles populaires ainsi que les expressions artistiques des rituels. »*

C'est de la conjonction des trois textes principaux mentionnés ci-dessus que résulte la définition de la langue comme élément du patrimoine culturel immatériel de l'Humanité. Mais que recouvre cette notion ?

LA NOTION DE PATRIMOINE CULTUREL IMMATERIEL DE L'HUMANITE

Nous ne pouvons ici développer les notions de patrimoine culturel et de patrimoine culturel immatériel. Nous renvoyons, en note, à des études sur le sujet[49].

Toutefois nous dirons quelques mots sur la notion de patrimoine culturel de l'Humanité. Sans entrer dans le débat sur ce que recouvre la notion d'Humanité, nous mentionnerons l'importance d'une coopération entre tous les acteurs concernés pour la sauvegarde du patrimoine :

> *« …les États parties reconnaissent que la sauvegarde du patrimoine culturel immatériel est dans l'intérêt général de l'Humanité et s'engagent, à cette fin, à coopérer aux niveaux bilatéral, sous-régional, régional et international.»*[50]

On remarque aussi, par ailleurs, que les expressions du folklore, qui font pourtant partie du patrimoine culturel de l'Humanité, sont l'objet d'une « appropriation nationale ». Les langues ne peuvent-elles pas, alors, relever à la fois du patrimoine de l'Humanité et par ailleurs être prises en compte dans la protection des patrimoines culturels nationaux ?

LES ÉTATS AFRICAINS ET LA RECONNAISSANCE IMPLICITE DE LA LANGUE COMME ELEMENT DU PATRIMOINE CULTUREL NATIONAL A TRAVERS LA NOTION DE FOLKLORE

On retrouve la notion de patrimoine culturel immatériel au niveau régional dans l'Accord de Bangui sur la propriété intellectuelle[51]. Comme dans les législations nationales, la mention se fait par référence au folklore.

L'OAPI ET LE PATRIMOINE CULTUREL IMMATERIEL

Le titre II de l'annexe VII de l'Accord de Bangui porte sur la protection et la promotion du patrimoine culturel. Aux termes de l'article 67 al. 1, « ***le patrimoine culturel est l'ensemble des productions humaines matérielles et immatérielles caractéristiques d'un peuple dans le temps et dans l'espace***.»

Cette définition a le mérite d'être complète en ce qu'elle cerne les deux aspects du patrimoine culturel, l'aspect matériel et l'aspect immatériel qui inclut indubitablement la langue. Même si la langue n'est pas expressément visée dans le texte, on peut la retrouver implicitement dans la référence à l'oralité.

---

[49] Cornu, M. (2003) « Droit des biens culturels et des archives » : 1, sur le patrimoine culturel immatériel.
http://www.unesco.org/culture/heritage/intangible/html_fr/index_fr.shtml

[50] Convention sur le patrimoine immatériel, article 19.2.

[51] Elaboré par l'OAPI (Organisation Africaine de la Propriété Intellectuelle), adopté en 1977 et révisé le 22 février 1999.

En effet, le folklore qui est une production du patrimoine culturel comprend : *« les productions littéraires de tout genre et de toute catégorie orale ou écrite, contes, légendes, proverbes, épopées, gestes, mythes, devinettes*[52].»

LES LEGISLATIONS NATIONALES

La langue, et partant le patrimoine immatériel, ne sont pas expressément et directement envisagés par la plupart des législations africaines[53].Toutefois, ils y sont intégrés par référence aux lois sur le droit d'auteur, qui soulignent dans une formule assez générique que : *« le folklore appartient à titre originaire au patrimoine culturel national.*[54] »

Les langues ne sont prises en compte que comme des « vecteurs » de contenus spécialisés (chant, conte, proverbe, énigme, etc.) qui sont des expressions du folklore et sont à ce titre protégés au titre du droit d'auteur.

En conséquence, lorsque les corpus réunissent ces contenus du folklore, il faut demander les autorisations requises aux titulaires des droits d'auteur.

LA RECHERCHE SUR LES LANGUES SUSCEPTIBLES DE RENTRER DANS LE CHAMP DE PROTECTION MIS EN PLACE PAR L'UNESCO

La recherche scientifique est un moyen de sauvegarde de la langue. Ainsi, l'identification de la culture traditionnelle et populaire participe à la protection du patrimoine culturel immatériel permettant de *« créer des systèmes d'identification et d'enregistrement (collecte, indexation, transcription) ou développer des systèmes déjà existants au moyen de guides, guides de collecte, de catalogues types, etc., eu égard à la nécessité de coordonner les systèmes de classement utilisés par différentes institutions*[55] ».

LA SAUVEGARDE DU PATRIMOINE IMMATERIEL

L'UNESCO fait ainsi figure de pionnière en matière de promotion de la recherche pour la sauvegarde du patrimoine immatériel. En plus de son programme spécifique aux langues en danger, elle déclare, dans le cadre de la musique traditionnelle du monde, devoir adapter son action aux besoins des chercheurs. Elle soutient la préservation des archives sonores et des centres de documentation et elle encourage la recherche.

L'article D.e de la Recommandation de 1989 demande aux États de « *promouvoir la recherche scientifique se rapportant à la préservation de la culture traditionnelle et populaire* ». De même l'article 13.c de la Convention de 2003 dispose que les États doivent s'efforcer « *d'encourager les études scientifiques, techniques et artistiques ainsi que les méthodologies de recherche pour une sauvegarde efficace du patrimoine culturel immatériel, en particulier le patrimoine culturel immatériel en danger* ». Cette promotion passe par l'adoption de diverses mesures juridiques, techniques, administratives et financières appropriées.

---

[52] Article 68 al. 2.a.

[53] Sauf la loi ivoirienne du 28 juillet 1987, qui vise dans son article 3 les œuvres du folklore.

[54] Par exemple l'article 8 de la loi ivoirienne du 25 juillet 1996 ; l'article 5-1 de la loi camerounaise du 19 décembre 2000 ; l'article 82 de la loi tchadienne du 2 mai 2003. La loi de la République démocratique du Congo, dans son article 6, dispose que le folklore est « l'un des éléments fondamentaux du patrimoine culturel traditionnel ».

[55] Recommandation sur la culture traditionnelle et populaire, article B.b.

Orchestrées par l'UNESCO, diverses actions visent plus spécialement à la sauvegarde du patrimoine linguistique de l'Humanité. Elles concernent, en général, l'encouragement de la diversité linguistique – dans le respect de la langue maternelle – la promotion de la diversité linguistique[56] avec un programme spécifique concernant les langues en péril. « *Une langue est en péril lorsque ses locuteurs commencent à la délaisser, réservant son utilisation à des contextes de moins en moins nombreux, et ne la transmettant plus de génération en génération*[57]. »

Les chiffres et estimations sur ce phénomène sont quelque peu alarmants. Selon l'UNESCO, il existe environ 6 000 langues dans le monde, dont plus de la moitié sont en danger. On estime par ailleurs qu'une langue meurt tous les quinze jours. Il est estimé de même « *qu'une langue ne peut survivre qu'à la condition de compter au moins 100 000 locuteurs. Or, sur les 6 700 langues actuelles, la moitié compte moins de 10 000 locuteurs*[58]. » Tel est le cas de la langue zapara qui est parlée couramment par seulement cinq personnes[59].

Les projets de recherche concernant les langues en péril doivent viser deux objectifs principaux.

Il s'agit d'une part d' «épargner l'humanité de la perte qui peut découler de l'extinction d'une langue en danger. Cette approche met en exergue l'archivage, qui consistera à collecter autant de documentation que possible sur la langue, et à procéder à une description linguistique aussi complète que le temps le permettra. »

Il s'agit, d'autre part, de « revitaliser la langue en encourageant son utilisation dans l'alphabétisation et dans l'enseignement primaire[60].»

En suivant cette logique des textes de l'UNESCO, la finalité de la recherche doit s'inscrire dans une optique de sauvegarde du patrimoine culturel, y compris les langues. À cet égard, il est important de rappeler que la recommandation de 1989 porte sur la sauvegarde de la culture traditionnelle et populaire. Il en va de même pour la convention de 2003 concernant le patrimoine culturel immatériel. Cet objectif affiché de sauvegarde du patrimoine immatériel induit diverses mesures. Ainsi, selon l'article 2 al 3 de la convention de 2003 : « *on entend par 'sauvegarde' les mesures visant à assurer la viabilité du patrimoine culturel immatériel, y compris l'identification, la documentation, la recherche, la préservation, la protection, la promotion, la mise en valeur, la transmission, essentiellement par l'éducation formelle et non formelle, ainsi que la revitalisation des différents aspects de ce patrimoine.* »

LE DROIT D'ACCES AUX MATERIAUX DU PATRIMOINE IMMATERIEL

La promotion du patrimoine impose de laisser aux chercheurs un droit d'accès à ce patrimoine. Ce droit d'accès implique pour le chercheur de pouvoir collecter les matériaux nécessaires à son travail mais aussi de pouvoir travailler à partir de ceux déjà collectés et conservés. Il ressort de la lecture de

---

[56] Respectivement les points 5 ; 6 et 10 du plan d'action de la Déclaration universelle sur la diversité culturelle.

[57] Unesco, Kit d'information de la Section du patrimoine immatériel, Division du patrimoine culturel, Secteur de la culture.

[58] La mort des langues,
http://www.tlfq.ulaval.ca/axl/Langues/2vital_mortdeslangues.htm.

[59] Sources, No 106 – juillet août 2001, p.6, Unesco.

[60] http://www.acalan.org, mission et vision de l'Acalan.

l'article 13d.ii de la Convention de 2003 que l'État doit adopter des mesures juridiques, techniques, administratives et financières appropriées visant à « *garantir l'accès au patrimoine* ». Dans la Recommandation de 1989, ce droit est affirmé de façon particulière. En effet, cette Recommandation fait de la recherche, la finalité de la conservation des matériaux du patrimoine culturel. Autrement dit, les matériaux doivent être conservés pour que des recherches puissent être menées. Aux termes de l'article B : « *la conservation concerne la documentation relative aux traditions se rapportant à la culture traditionnelle et populaire et a pour objectif, en cas de non-utilisation ou d'évolution de ces traditions, que les chercheurs et les porteurs de la tradition puissent disposer de données leur permettant de comprendre le processus de changement de la tradition* »

LE DROIT A LA PROTECTION DES MATERIAUX COLLECTES

La Recommandation de 1989 (article F) prévoit, de façon explicite, la protection des « intérêts des collecteurs en veillant à ce que les matériaux recueillis soient conservés dans les archives, en bon état et de manière rationnelle ». Les chercheurs peuvent se prévaloir d'un tel droit dans la mesure où leur travail consiste en partie dans la collecte de données sur le terrain. Toutefois, l'exercice de ce droit implique aussi que les données recueillies soient déposées aux archives prévues à cet effet.

L'ARCHIVAGE DES MATERIAUX COLLECTES

L'accès aux matériaux du patrimoine immatériel peut se faire de plusieurs façons. Il peut se réaliser dans le cadre d'institutions de documentation (ou musées) que les États devront tout d'abord mettre en place. Dans le texte de l'article 13b de la Convention de 2003, il est question d'organismes compétents pour la sauvegarde du patrimoine culturel immatériel sur un territoire. La recommandation de 1989 établit une liste de mesures (sept au total[61]). Parmi lesquelles, on peut citer :

- la mise en place de services nationaux d'archives où les matériaux de la culture traditionnelle et populaire collectés puissent être stockés dans des conditions appropriées ;
- la création de musées ;
- l'octroi de moyens en vue d'établir des copies d'archives et de travail de tous les matériaux de la culture traditionnelle et populaire.

Une fois de tels organismes institués, l'État doit en faciliter l'accès et les matériaux doivent être réellement mis à disposition.

La responsabilité du service d'archives peut se trouver engagée. La Recommandation de 1989 (article F.iv) invite à « *reconnaître que les services d'archives ont la responsabilité de veiller à l'utilisation des matériaux recueillis* ».

LE DROIT A L'INFORMATION

Ce droit peut être perçu comme une modalité particulière du droit d'accès comme il appert dans la diffusion de l'information. Ainsi, l'Académie Africaine des Langues (Acalan)[62] se fixe pour objectif de « *faciliter la documentation et l'échange d'information par la mise en place d'une base de données, la collecte*

[61] Article B alinéas a à g.

[62] http://www.acalan.org, mission et vision de l'Acalan (Académie africaine des langues).

et l'archivage des documents, la publication et consacrer une bonne partie de ses ressources à l'impulsion de la recherche et à la coordination des activités de recherche. » L'UNESCO a recommandé, en 1989, la mise en place d'une unité centrale d'archives aux fins de la prestation de certains services (indexation centrale, *diffusion de l'information* relative aux matériaux de la culture traditionnelle et populaire et aux normes applicables aux activités la concernant, y compris l'aspect préservation)[63].

LA RETRIBUTION DES PERSONNES SOLLICITEES

En occultant les aspects de propriété intellectuelle qui peuvent résulter de l'utilisation des expressions du folklore ou de toute expression langagière qualifiable d'œuvre de l'esprit, la rétribution se pose pour toutes les personnes auditionnées pour la recherche, surtout lorsqu'on se trouve dans une optique de protocole expérimental. Les textes ne prévoient pas cette éventualité. L'exigence sur ce point est d'ordre éthique. Selon G. Durnon, le montant doit être fixé en commun accord avec les intéressés et doit être équitable et s'il existe des obstacles (milieu traditionnel opposé à une telle pratique), le chercheur peut participer à une œuvre d'intérêt collectif.

LE DROIT DE DIFFUSION DES RESULTATS DE LA RECHERCHE

L'UNESCO ne vise pas directement la diffusion des résultats de la recherche. Cependant, les publications scientifiques participent à la *large diffusion* des éléments constituant le patrimoine qu'encourage l'UNESCO, la seule limite consistant dans le fait *d'éviter, lors de cette diffusion, toute déformation* pouvant porter atteinte à son intégrité.

LE PARTAGE DES RESULTATS DE LA RECHERCHE

Quand il s'agit des travaux de recherche effectués par les spécialistes d'un État membre dans un autre État membre, la Recommandation de l'UNESCO de1989 dispose dans son article Gd que les États devraient : « *garantir aux États membres sur le territoire desquels ont été effectués des travaux de recherches le droit d'obtenir de l'État membre concerné copie de tous documents, enregistrements vidéo, films et autres matériels*. »

L'idée, ici, est de participer, à l'issue de la recherche, à l'enrichissement des archives des institutions locales : « *en fournissant des copies des documents sonores ou audiovisuels recueillis au cours de la recherche*[64]». Ce qui semble recommandé, c'est de faire bénéficier les communautés sollicitées des retombées positives de la recherche. Cela se justifie si on part du postulat que chaque peuple a un droit sur sa propre culture. C'est ce qu'affirme en substance la Recommandation de 1989 dans son article D. On remarque par ailleurs l'existence de fortes revendications émanant des communautés dans lesquelles sont effectuées les recherches. Comme le note un rapport de l'UNESCO[65] : « *speakers increasingly demand control over the terms and*

[63] Article C.b de la Recommandation Unesco de 1989 sur la sauvegarde de la culture traditionnelle et populaire.

[64] Durnon, G. (1981) *op. cit.* :51 ouvrage précité, p.51. Cela peut se faire par le biais d'une « procédure de restitution des biens culturels aux pays d'origine qui est appliquée depuis plusieurs décennies au département d'ethnomusicologie du Musée de l'Homme à Paris. »

[65] Language Vitality and Endangerment,
http://portal/unesco.org/culture/en/ev.php-URL_ID=9105&URL_DO=DO_TOPIC&URL_SECTION=201.html

conditions that govern research; furthermore, they claim rights to the outcomes and future uses of the research". Une prise en compte de ces droits met à la charge du chercheur une obligation d'implication des populations dans la mise en œuvre de la recherche. Aux termes du rapport précité : « *any research in endangered language communities must be reciprocal and collaborative. Reciprocity here entails researchers not only offering their services as a quid pro quo for what they receive from the speech community, but being more actively involved with the community in designing, implement and evaluating their research project*. »

LA PROTECTION DES PERSONNES CONCERNEES PAR LA RECHERCHE

La constitution des corpus oraux donne lieu à la collecte de nombreuses informations sur les personnes : nom, prénom, âge, appartenance ethnique, sexe, statut social, lieu de résidence et de naissance, image et voix (enregistrement audio et vidéo, photographie), etc.

LES PERSONNES CONCERNEES

Ces données concernent diverses personnes. Il pourra s'agir selon les cas d'un chanteur, d'un compositeur, d'un traducteur, d'un interprète, d'un conteur, ou d'un locuteur ordinaire. La Recommandation de 1989 traite de l'informateur qu'il faut protéger en tant que porteur de la tradition. Une autre qualification existe en ce qui concerne les personnes : celle des « trésors humains vivants ». Les « trésors humains vivants » sont « *des personnes qui possèdent à un très haut niveau les connaissances et les savoir-faires nécessaires pour interpréter ou créer des éléments spécifiques du patrimoine culturel immatériel que les États membres ont choisi comme témoignages de leurs traditions culturelles vivantes et du génie créateur des groupes, des communautés et d'individus présents sur leur territoire.* » C'est le cas par exemple des membres de la famille Dökala en Guinée, qui assurent l'enseignement de l'histoire familiale, locale, régionale conformément à l'héritage légué par les anciens. « *Ce sont eux qui détiennent au plus haut niveau les valeurs authentiques de la civilisation mandingue* »[66]. Le système des trésors humains vivants a été adopté en 1993 et est censé être une partie essentielle de la mise en œuvre de la Convention sur le patrimoine immatériel.

LE CHAMP DE LA PROTECTION

Le champ de la protection des personnes concerne, selon l'article F.i de la Recommandation de l'UNESCO de 1989, la vie privée et la confidentialité. Le droit à la confidentialité d'une information dont bénéficie une personne peut être perçu d'un point de vue négatif en tant qu'*obligation de secret* qui pèse sur une autre (généralement un professionnel). Visant à mettre le bénéficiaire à l'abri de divulgations, le droit à la confidentialité se situe dans le prolongement du droit au respect de la vie privée.

En ce qui concerne ce droit, au-delà de la première difficulté à laquelle on se heurte - résidant dans le défaut de définition légale de la notion même de vie privée, en droit français comme dans les législations africaines - une seconde

[66] Namankoumba Kouyaté, Méthodes traditionnelles de transmission de l'oralité : l'exemple du Sosso-Bala. Conférence : « Evaluation globale de la recommandation de 1989 sur la sauvegarde de la culture traditionnelle et populaire ; pleine participation locale et coopération internationale », Washington, 1997
http://www.folklife.si.edu/resources/Unesco/kouyate.htm

difficulté résulte d'une certaine divergence de conception. Cl. Ouoba écrit à ce sujet que *« si l'on estime que le secret et le privé sont indispensables à l'épanouissement individuel de chaque citoyen dans les sociétés occidentales, il serait un peu trop idéaliste de transposer une telle appréciation dans les sociétés où le partage et la communion sont les bases de l'existence même*[67]*.»* Abondant dans le même sens, A. Sow Sidibé ajoute que « *alors que la conception occidentale est fondée sur l'individualisme, celle négro-africaine repose sur des valeurs communautaires qui privilégient le groupe*[68] ».

Le droit à la vie privée est garanti, directement ou indirectement, dans la grande majorité des États africains tant par le biais de l'adhésion à des conventions internationales que par des législations spécifiques[69]. Au titre du droit au respect de la vie privée se trouve aussi consacré un droit à l'image. « *Représentation physique ou morale, elle (l'image) appartient en propre à l'individu, et sa reproduction ou sa divulgation, en somme sa violation, ne peut se faire sans son accord*[70].» Toute reproduction de l'image d'une personne doit se faire avec son consentement, sauf exception (notamment en matière d'information). Ainsi que le recommandait G. Durnon, dans le cadre de la collecte des musiques : *« la prise en vue nécessite l'assentiment des intéressés*[71]. »

UN REGIME PARTICULIER POUR LES DONNEES TRAITEES A DES FINS DE RECHERCHE

Ainsi l'article 68[72], relatif à la recherche scientifique, dispose que l'utilisation des données est soumise à une autorisation préalable du concerné ou de ses ayantdroits et de la commission nationale. Et selon l'article 49 : *« les données à caractère personnel, traitées pour des finalités particulières, peuvent être communiquées en vue d'être traitées une autre fois pour des fins historiques et scientifiques, à condition d'obtenir le consentement de la personne concernée, de ses héritiers ou de son tuteur ainsi que l'autorisation de l'Instance Nationale de Protection des Données à Caractère Personnel.* » Il s'agit ici de la reconnaissance du principe de finalité compatible pour le traitement des données à caractère personnel à des fins historiques et scientifiques lorsque les données avaient été collectées pour une autre finalité.

Même si le contexte et les enjeux ne sont pas les mêmes dans la recherche biomédicale et en recherche linguistique, on peut citer la déclaration du Réseau africain sur l'éthique, le droit et le VIH : *« (…) la recherche doit être effectuée*

---

[67] Ouoba, Cl. (2002) *Le droit à la vie privée au Burkina Faso, Conception, réalité juridique et socioculturelle*, Thèse de l'Université de Grenoble 2 : 50.

[68] Sow Sidibé, A. « Le secret médical aujourd'hui », revue électronique *Afrilex* 2 : 25. http://www.afrilex.u-bordeaux4.fr/

[69] A titre d'exemple : ONU, Déclaration universelle des droits de l'homme, 10 décembre 1948, Charte africaine des droits de l'homme et des peuples, 26 juin 1986, Burkina Faso (Loi No 101-2004/an du 20 avril 2004 portant protection des données à caractère personnel), Tunisie (Loi organique No 2004-63 du 24 juillet 2004 portant protection des données à caractère personnel).

[70] Ouoba, Cl. *op.cit.*: 37.

[71] Durnon, G. (1981) *Guide pour la collecte des musiques et instruments traditionnels*, Editions de l'Unesco : 95.

[72] Tunisie : Loi organique n° 2004-63 du 24 juillet 2004 portant protection des données à caractère personnel.

sur la base d'un consentement libre et éclairé de la personne sans intrusion dans sa vie privée et sans coercition (…) »[73].

L'INFORMATION DES PERSONNES

Au-delà du fait qu'il s'agit d'une obligation légale et éthique du chercheur et qu'y satisfaire peut s'avérer difficile voire inapproprié dans le cadre de certaines recherches, l'information peut présenter un avantage. Elle peut être un moyen pour solliciter la collaboration des personnes concernées. G. Durnon écrit à ce sujet que « *on se rend compte de l'intérêt qu'elles portent aux recherches menées quand, au cours de l'enquête, on s'attache à expliquer qui nous sommes, les raisons de notre présence, l'objet et le but de nos recherches, ce qu'il adviendra des objets et informations recueillis sur place, de l'aide que nous sollicitons et ce qui peut être proposé en retour*[74]. »

Quelques exemples d'organismes africains :

- Acalan, Académie africaine des langues,
- Celhto, Centre d'étude linguistique et historique pour la tradition orale (Niamey, Niger),
- Cerdotola, Centre régional de documentation sur les traditions orales et les langues africaines (Yaoundé, Cameroun).

Principales sources bibliographiques

Ouoba, C. (2002), *Le droit à la vie privée au Burkina Faso, Conception, réalité juridique et socioculturelle*, Thèse, Université de Grenoble 2.

Durnon, G. (1981) *Guide pour la collecte des musiques et instruments traditionnels*, Paris, Éditions UNESCO.

Recommandation UNESCO de 1989.

Recommandation UNESCO de 2003.

Cornu, M. (2005) « A propos de l'adoption du code du patrimoine, Quelques réflexions sur les notions partagées », *Recueil Dalloz* 22 : 1452-1458

[73] Réseau africain sur l'éthique, le droit et le VIH, Déclaration de Dakar, juillet 1994, principe de l'éthique dans la recherche.
[74] G. Durnon, *op. cit.* : 41.

PRISE DE SON ET ENREGISTREMENT SUR LE TERRAIN[75]

PRINCIPES

En matière de prise de son, il n'existe pas de solution unique répondant à tous les besoins. Les principaux critères à prendre en compte sont :

- la nature de la source à enregistrer (plusieurs sources d'émission du son ou une seule : un ou plusieurs locuteurs) ;
- le contexte, et les perturbations sonores ou le parasitage qu'il peut produire ;
- la durée des entretiens et le besoin d'autonomie du matériel qui peut en découler.

Les moyens financiers dont on dispose restreignent souvent le choix en équipement. Cependant, à moyens égaux, la part consacrée à ce chapitre tend à être négligée. Au cours des quarante dernières années, la démocratisation du matériel d'enregistrement a généralement conduit à une désaffection envers le matériel haut de gamme et de fait à une moindre exigence de qualité (exemple : substitution de la cassette audio à la bande ¼ de pouce et au Nagra). Aujourd'hui comme hier, acquérir le matériel adéquat peut représenter un investissement conséquent dans un projet de collecte sonore.

Par ailleurs, tout matériel nécessite un temps d'appropriation. La prise de son requiert un certain apprentissage. Il existe des formations pratiques de quelques jours aux bases du métier de preneur de son, qui peuvent permettre d'une part de tirer tout le parti des ressources dont on dispose, et d'autre part de libérer le collecteur de soucis techniques lors de l'entretien.

Les qualités couramment attendues d'un matériel d'enregistrement sont les suivantes :

- facilité d'utilisation (pour éviter les erreurs de manipulation en cours d'enregistrement) ;
- autonomie (batterie suffisante) ;
- robustesse ;
- légèreté ;
- ergonomique (poids, taille, bouton, lisibilité des vumètres...) ;
- capacité du support d'enregistrement (pour éviter les interruptions dues aux changements de face ou de support) ;
- niveau d'entrée audio réglable (de manière à éviter sous-modulation – autrement dit un son trop faible – et surmodulation – autrement dit une saturation) ;
- sortie casque réglable ;
- en numérique, possibilité d'enregistrer dans le format recherché, et en particulier dans un format linéaire (non compressé) et pérennisable (voir fiche *Supports pour enregistrer et archiver le son*) ;
- interopérabilité et rapidité de transfert vers une station informatique (qui servira de plate-forme d'édition et de gravure).

D'autres caractéristiques moins évidentes sont à rechercher pour une réalisation de qualité :

- bruit de fonctionnement de l'appareil le plus faible possible ;

[75] Fiche rédigée par Luc Verrier et Alain Carou (BnF).

- qualité des circuits analogiques (notamment préamplification micro) ;
- câblage et connectique professionnels (symétriques : les 2 fils conducteurs sont entourés d'une tresse métallique qui les protège des parasites) ;
- alimentation fantôme 48 volts pour micro statique ;
- qualité des convertisseurs analogique/numérique (bande passante, dynamique, bruit).

Des conditions spécifiques peuvent nécessiter la prise en compte d'autres éléments :

- discrétion de l'équipement (petite taille) ;
- matériel dit « tropicalisé » (adapté aux conditions climatiques extrêmes : chaleur, froid, humidité) ;
- traitement du signal, notamment pour le travail en conditions difficiles (filtre coupe-bas pour le vent, limiteur pour l'enregistrement de sources au niveau sonore aléatoire) ;
- système d'édition intégré, pour permettre un dérushage immédiat ;
- système de gravure intégré, pour gagner en autonomie et en sécurité sans perdre en portabilité.

TYPES D'ENREGISTREURS

ENREGISTREUR SPECIALISE SUR MEMOIRE FLASH, MICRO DRIVE OU DISQUE DUR

Cette technique très fiable, de haute capacité et ouverte à tout type de format numérique, se démocratise actuellement :

- mémoire flash en baisse (1 Go coute moins de 100 euros début 2006) ;
- émergence des Micro Drive (4 Go), plus chers et plus gourmands en énergie ;
- apparition de disques durs 1.8'' (80 Go), issus de la technologie des ordinateurs portables.

Le support de stockage peut être amovible. Le transfert vers un ordinateur ou un système de stockage externe (recommandé) est très rapide. Le média de stockage, s'il est amovible, peut être directement raccordé à un ordinateur ou à une autre unité de stockage (voire à un graveur autonome) via un adaptateur et/ou une connexion informatique incorporée (USB, FireWire, SCSI).

Ces appareils possèdent généralement des entrées et sorties audio numériques (SPDIF, AES) permettant un raccordement et un transfert des données sans passer par le domaine analogique (donc sans perte).

Les composants analogiques sont similaires à ceux des DAT Pro, la partie mécanique (fragile) en moins. On peut disposer de plus de 2 canaux sur certains modèles professionnels.

Certains modèles disposent d'un système d'édition intégré.

Au sommet de cette catégorie se classent les enregistreurs sur disque dur, tels que le Nagra V, considéré comme la « Rolls » de l'enregistrement de terrain et digne remplaçant des Nagra analogiques (pour un cout équivalent). Ce type de modèle se décline également en multi-pistes (HHB, Cantar).

BALADEUR « MP3 » (VARIANTE BON MARCHE DU PRECEDENT)

Maintenant très répandu (iPod, iRiver…), ce type d'appareil est d'une utilisation très simple pour la lecture, moins pour l'enregistrement. L'autonomie est élevée (25 h), ainsi que la capacité de la mémoire (100 Go).

Le microphone intégré est d'une qualité médiocre, acceptable comme dictaphone. La qualité des circuits analogiques et l'ergonomie sont les éléments qui laissent le plus à désirer. On peut adjoindre au baladeur un pré-ampli micro/convertisseur pour pallier ses carences.

L'ergonomie est très limitée (enchainement de menus). Attention à bien disposer d'un modèle offrant un format d'enregistrement ouvert, mais aussi à paramétrer correctement le format désiré.

Ce matériel bon marché est pour le moment du « gadget » promis à une durée de vie commerciale courte : il n'y a pas d'entretien durable assuré. Avant de miser sur cette technologie, il est donc recommandé d'attendre l'arrivée prochaine d'appareils semi-professionnels plus spécialisés.

Autre innovation récente en développement : l'enregistrement sur PDA, offrant une interface couleur et les fonctionnalités d'une micro-station d'édition.

ORDINATEUR « PORTABLE »

Simple d'utilisation et générant peu de frais si on dispose déjà d'un portable pour d'autres usages, cette solution offre d'intéressantes facilités. Cependant, elle se classe plutôt dans les solutions « transportables » que « portables ». L'enregistrement se fait directement sur le disque dur de l'ordinateur ou autre solution de stockage externe, ce qui simplifie le transfert et permet de le faire dans tout type de format ouvert. Le logiciel d'édition permet l'enregistrement et le montage, voire la gravure. La stabilité de la configuration logicielle doit avoir été testée avant utilisation.

Le bruit et le rayonnement électromagnétiques générés par l'ordinateur peuvent pénaliser la qualité du signal. Les cartes-son intégrées sont souvent de qualité médiocre, et mieux vaut éviter d'utiliser l'entrée micro intégrée. Il est préférable de faire l'acquisition d'un module externe pré-ampli/micro/convertisseur, branché via une interface informatique USB ou FireWire (le module peut être simplement stéréo, mais aussi multi-canaux si besoin).

DAT

Cette technologie est en fin de vie, mais reste souvent utilisée. Elle contraint à un transfert rapide des bandes numériques, qui ne doivent en aucun cas être archivées telles quelles vu leur fragilité. Attention : la récupération d'index lors du transfert est quasi-impossible.

Techniquement, c'est une solution semi-professionnelle à professionnelle tout à fait éprouvée. Les préamplis intégrés sont de qualité. Un des talons d'Achille est la fiabilité de la mécanique d'entrainement de la bande (« machine tournante », par différence avec les modèles présentés ci-dessus), et désormais les couts d'entretien, de plus en plus onéreux.

MINIDISC

C'est une solution très bon marché, mais en passe d'être détrônée par les baladeurs-enregistreurs MP3. Le format numérique d'enregistrement a été longtemps exclusivement compressé et propriétaire (ATRAC). Le Hi-MiniDisc, lancé en 2004, permet dorénavant d'enregistrer dans un format linéaire mais

toujours propriétaire (OpenMG). Avec le logiciel SonyStage, il est possible de transférer rapidement des données en OpenMG vers un PC. Le format OpenMG n'est cependant en aucun cas à considérer comme un format d'archivage, du fait de la dépendance par rapport à l'offre technologique Sony.

Sony a récemment mis à la disposition de ses utilisateurs un utilitaire de conversion d'OpenMG vers WAV. On ignore à ce jour si la transformation est entièrement transparente.

L'ergonomie est limitée (enchaînement de menus, affichage de taille réduite). Les niveaux d'entrée ne sont pas réglables sur tous les modèles. La connectique est non professionnelle, vulnérable. Le choix de micros est restreint, sans adaptateur dédié.

CASSETTE AUDIO

Autre solution très bon marché, la cassette audio dispose toujours d'un marché (pays africains notamment) et a donc encore plusieurs années assurées. Sa robustesse et sa fiabilité sont éprouvées. Cependant, le matériel de niveau professionnel se raréfie (reste notamment Marantz PMD222).

Les limites de la cassette sont connues : durée par face, qualité moyenne, navigation difficile dans le document…

Mieux vaut éviter d'utiliser les réducteurs de bruit (Dolby), l'appareil servant par la suite de lecteur ayant en général des réglages différents de l'enregistreur.

La microcassette (utilisée dans les dictaphones) présente une qualité et une espérance de vie insuffisantes pour être encore utilisée.

Essai de synthèse comparative :

	Enregistreur dédié Flash, MicroDrive, disque dur	Baladeur MP3	Ordinateur portable
Technologie	actuelle	actuelle	actuelle
Prix	1 000/10 000 €	150-450 €	Ordinateur +200 € d'équipement spécifique
Facilité d'utilisation	semi-pro et pro	grand public	grand public ou semi-pro
Ergonomie	+	-	+
Capacité	selon disque dur	selon disque dur	selon disque dur
Formats d'enregistrement	wav, BWF, MP2, MP3	wav, MP2, MP3…	potentiellement tous
Interopérabilité	+	-	+
Qualité des convertisseurs A/D	++	-	+

	Nagra analogique	DAT	Cassette audio	MiniDisc
Technologie	fin de vie	fin de vie	fin de vie	fin de vie
Prix	1 000 € (occasion)	1 500 €	700 € (si neuf)	150 €
Facilité d'utilisation	pro	semi-pro	grand public	grand public
Ergonomie	+	+	+	-
Capacité	30 min	2 h	1-2 h	80 min
Formats d'enregistrement	analogique	PCM 16/32 à 48	analogique	ATRAC, Open MG (Hi-MD)
Interopérabilité	sans objet	-	sans objet	-
Qualité des convertisseurs A/D	sans objet	+	sans objet	-

CONSEILS PRATIQUES

MEDIAS DE COLLECTE VIERGES ET DISQUES DURS

Acheter des médias de qualité et déjà éprouvés.

Veiller à stocker les médias vierges dans un environnement de même qualité que l'archive (les dégradations physico-chimiques intervenant que le média soit enregistré ou non). Éloigner des sources de magnétisme et de chaleur.

Éviter un stockage excessif.

Protéger les cassettes, MD, DAT contre le réenregistrement (ergot de protection). Les médias magnétiques (cassettes, DAT, bandes) perdent en fiabilité au fil des cycles effacements/réenregistrements.

Les médias magnéto-optiques (MD) et mémoires flash sont réenregistrables quasiment sans limite dans la pratique (100 000 cycles écriture/lecture). La vulnérabilité réside dans la partie sensible des mémoires flash (connecteur), à manipuler avec précaution.

Les disques durs sont garantis par les fabricants pour fonctionner régulièrement et n'offrent pas les mêmes garanties en cas de stockage dormant.

CHOIX DES MICROS

Deux critères fondamentaux pour le choix d'un micro, selon l'usage auquel on le destine, sont sa sensibilité et sa directivité :

- un manque de sensibilité devra être compensé par un gain de préampli plus important, ce qui augmentera d'autant le bruit de fond (souffle) ;
- un micro couvre un espace sonore plus ou moins large, de 360° (omnidirectionnel) à 30° (micros « canons »), en passant par les micros directionnels (hypercardioïde, semi-canon).

Exemple : on choisira un micro canon pour des prises de sons précises (chant d'un oiseau), et un couple de micros cardioïdes pour des ambiances et l'enregistrement de musiques.

A côté des micros dynamiques (utilisés par les journalistes et pour la scène, permettant d'encaisser de forts niveaux) et des micros statiques (les plus respectueux des timbres, mais plus fragiles et délicats sur le terrain) existe une gamme de micros grand public : ainsi les électret (MiniDisc, alimentés par pile) et autres micros avec préampli intégré (pour usage spécifique : cravate, perche), généralement pourvus de connectique grand public (mini jack).

Conseils :

Bon rapport qualité prix : le micro Sennheiser K6 avec une capsule ME64 ou 66 (semi canon directif) assez polyvalent (ajouter une bonnette Rycotte pour les prises de son en extérieur).

Micro main type Shure SM58 ou LEM D021 : excellent micro de reportage pour les interviews en milieu bruyant, mais nécessitant d'être placé le plus près possible de la bouche...

Prévoir les accessoires nécessaires :

- pied de micro (encombrant mais adapté) ;
- pied de table (plus transportable mais peut poser des problèmes) ;
- perche, permet une optimisation rapide de la distance (solution cinéma, nécessite de la pratique et une certaine acuité) ;
- système HF ou micros sans fil: chers mais très pratiques (spectacle, cinéma, TV…) ;
- bonnette (anti-vent, plop voix…) ;
- suspension élastiques pour micro (isolation mécanique des vibrations) ;
- filtre adaptateur coupe-bas (vent, plop).

REGLAGES

Faire des essais avant de lancer l'enregistrement. Bien se préparer (longueur de câble, batterie chargée, bloc secteur, cassettes vierges de durée suffisante…)

En numérique, vérifier que l'appareil est bien paramétré : bon format et bonne résolution.

Attention à éviter la saturation (dépassement du niveau maximum) : écouter, observer les indicateurs de niveau. En cas de saturation audible, si les vumètres n'indiquent pas le maximum, c'est que le préampli est saturé en entrée : enclencher l'atténuateur d'entrée micro. Régler le niveau d'enregistrement afin qu'il n'y ait pas de dépassement du 0 dBFs seuil critique (moyenne –10 dBFs).

Attention au problème de « Larsen » (sifflement suraigu) qui peut être lié au réglage du volume du casque. Le microphone re-capte le son émis par le casque : baisser, voire couper le signal qui va au casque lorsque celui-ci n'est pas sur la tête.

Il est essentiel d'écouter ce que l'on enregistre de façon régulière pendant l'enregistrement !

Attention : certains systèmes « ne conservent pas » l'enregistrement s'il est interrompu accidentellement. Les solutions les plus évoluées (Nagra V) permettent d'avoir l'équivalent d'une lecture analogique « après enregis-

trement » (i.e. possibilité de contrôler ce qui est effectivement enre-gistré sur le disque dur).

QUELQUES CONSEILS TECHNIQUES

Un micro à la main (dynamique) doit être tenu à environ 20 cm de la bouche. Si possible, laisser quelques secondes de silence entre chaque question.

Idéal : micro-cravate couplé avec micro d'ambiance.

Si on dispose de deux entrées micros : 1. interviewé 2. les questions (on peut utiliser une petite mixette pour plus de sources).

Faire attention aux bruits de manipulation du micro et des câbles.

Ne pas coller le micro près de l'appareil (bruits mécaniques).

Penser à prendre quelques minutes en fin d'interview pour capter un « silence plateau » ou une ambiance (geste de souplesse au niveau du montage).

Une annonce en début d'enregistrement (sujet, interviewer, interviewé, date, lieu…) est un moyen simple et pérenne de garantir l'identification du contenu de l'enregistrement.

Le support d'enregistrement n'est pas forcément un support qualifié pour l'archivage (voir fiche *Supports pour enregistre et archiver le son*).

TRANSFERT ET EDITION

Cette étape est cruciale même si elle parait simple au premier abord. On utilisera une carte son avec entrée et sortie numérique optique (ADAT), SPDIF ou AES (permet un transfert sans perte et immune aux bruits parasites). La carte son ou l'interface audio et le logiciel d'édition doivent fournir un large choix de fréquences d'échantillonnage et de résolution (44.1 à 96 kHz, 16 et 24 bits), et permettre l'acquisition et la conversion de différents formats (wave, bwf, mp2, mp3…). Attention aux niveaux en numérique ou en analogique (norme d'alignement analogique-numérique : 0dBvu = -18dBFs).

Le logiciel d'édition effectue sur le son les mêmes opérations qu'un éditeur de texte (Word par exemple) : couper/copier/coller, enregistrer sous différents formats, accélérer ou ralentir le son… Il peut permettre aussi de réaliser certains filtrages (coupe-bas, ronflette, correction…) afin de fournir un document de qualité facilement écoutable. Une indexation pourra être faite afin de permettre une meilleure navigation au sein du document.

Pour finir, on réalise avant gravure une « image » du support d'archivage, incluant les données audio et les métadonnées.

SUPPORTS POUR ENREGISTRER ET ARCHIVER LE SON[76]

Assurer la conservation à long terme du son numérique

SUPPORTS D'ENREGISTREMENT, SUPPORTS D'ARCHIVAGE

L'arrêt progressif de la fabrication des matériels professionnels analogiques (à bandes et à cassettes) conduit à exclure l'archivage sous une forme autre que numérique. Tout enregistrement réalisé aujourd'hui sur support analogique impliquera à court terme un investissement non négligeable en temps et en argent pour le convertir et l'archiver dans un format numérique.

D'autre part, tous les supports d'enregistrement numérique ne réunissent pas les qualités attendues d'un support d'archivage.

Les qualités généralement attendues d'un support d'enregistrement sont sa capacité, sa maniabilité, éventuellement la possibilité d'indexation. Souvent, l'autonomie et la robustesse du matériel d'enregistrement associé, ainsi que son prix, sont les critères décisifs du choix.

Un support d'archivage numérique doit quant à lui réunir de tout autres qualités :

- Garantie de pouvoir trouver du matériel de lecture à moyen terme : large diffusion de la technologie, fabrication par plusieurs constructeurs différents.
- Possibilité de coder l'audio dans un format « ouvert » de qualité satisfaisante : la relecture du fichier ne peut être garantie à moyen et long terme si la syntaxe du format est secrète, ou en d'autres termes si la relecture de l'archive est dépendante de l'offre commerciale d'un industriel.
- Existence d'outils pour contrôler l'état de l'enregistrement sur le support : en effet, la lecture d'un support numérique ne nous apprend rien de son état de conservation, sauf lorsque l'information qu'il contenait devient illisible. La perte n'est pas proportionnelle à la dégradation comme dans le domaine analogique, mais obéit à un effet de seuil (« tout ou rien »). Il est indispensable de disposer d'outils d'évaluation de l'état du support pour engager les copies d'information en temps utile.
- Robustesse (capacité de conserver l'information dans son intégrité pendant plusieurs années). Outre ces quatre garanties fondamentales, sans lesquelles il n'est pas d'archivage numérique viable, deux autres sont également à rechercher, particulièrement dans l'optique d'une gestion de masse : simplicité des opérations de copie ; protection contre l'effacement accidentel.

Un support d'enregistrement commode et bon marché, si on considère uniquement la phase de collecte, peut se révéler bien plus couteux si l'on intègre dans le calcul la dimension archivage à long terme (en particulier s'il doit y avoir conversion du format natif à un autre format). Voici quelques exemples :

- L'usage du **MiniDisc** implique l'enregistrement dans un format propriétaire Sony. Même si les supports magnéto-optiques sont réputés d'une bonne tenue dans le temps et donc aisés à conserver

[76] Fiche rédigée par Luc Verrier et Alain Carou (BnF).

sur un plan purement physique, le MiniDisc ne répond pas aux conditions 1 (nombre de constructeurs très limité, technologie menacée à court terme), 2 (format de stockage propriétaire), 3 (pas d'outil de contrôle existant) et 5 (vitesse d'extraction bridée).

- **Le DAT** permet l'enregistrement en PCM 16 bits/48 Khz. Cependant, l'archivage sur DAT est déconseillé depuis plusieurs années en raison de la fragilité de ce support (condition 4 non remplie). Les conditions 1, 3 et 5 sont également non remplies.
- **Le CD enregistrable** une fois (CD-R) répond aux conditions 1 (technologie universelle), 2 (compatibilité tous formats de fichiers), 3 (existence d'outils d'analyse abordables), 5 et 6. Pour satisfaire à la condition 4, en revanche, des règles strictes sont à observer.

De manière générale, aucun support d'archivage n'offre aujourd'hui de garantie de pérennité sur le long terme, du fait *primo* de leur dégradation naturelle, *secundo* de l'obsolescence plus ou moins rapide des technologies de lecture. L'archivage numérique consiste donc non pas à trouver le support éternel, mais à mettre en œuvre une méthode rationnelle et réaliste de contrôle de support, de veille technologique et de migration (copies ponctuelles et copies en masse) en fonction des nécessités. Alors que dans le monde analogique, chaque génération de copie est source de perte qualitative, le nombre de copies est indifférente dans le monde numérique, du moment qu'elles sont effectuées à temps.

ARCHIVER SUR CD ENREGISTRABLE

Support et format doivent être clairement distingués. Le **support** CD-R permet l'inscription de données dans plusieurs **formats**, notamment le CD audio, lisible sur un lecteur de salon et limité à une résolution audio 16 bits/44.1 kHz ; et le CD-ROM, qui autorise le stockage de tout type de fichier, audio ou autre.

Le CD-R (appelé également CD-WORM, c'est-à-dire enregistrable une seule fois) peut être considéré comme un bon support d'archivage sur étagères (« off-line ») pour une durée de plusieurs années. Cependant, c'est depuis plusieurs années un marché essentiellement grand public, où les industriels cherchent à baisser leurs prix, y compris aux dépens de la qualité. Peu de marques réunissent donc les caractéristiques requises pour satisfaire potentiellement à un objectif d'archivage.

Quelques critères peuvent aider à s'y retrouver :

- **Capacité** : la norme « Orange Book » définit une capacité équivalant à 73 minutes de CD audio. Il est aujourd'hui quasiment impossible de trouver des CD-R de cette durée. Il est fortement recommandé de s'en tenir à ceux qui l'excèdent le moins, à savoir ceux de 80 minutes (Par précaution, on s'abstiendra de remplir le CD jusqu'au bout).
- **Couche métallique** : trois métaux sont employés (aluminium, argent, or). L'or présente la réflectivité la plus élevée, donc les meilleures chances de retrouver correctement l'information à la lecture. Des problèmes de corrosion ont été constatés avec l'argent. La qualité de la métallisation s'est révélée un point critique ces dernières années : un CD présentant une apparence

grêlée, piquée ou cloquée avant ou après gravure ne doit pas être archivé.

- **Couche de pigment** : c'est la couche qui est transformée par le passage du laser graveur. Trois pigments existent : phtalocyanine, cyanine et azo. La phtalocyanine présente une stabilité intéressante. L'identification du pigment enregistrée en en-tête du CD-R et parfois fournie par le logiciel de gravure ne doit pas être considérée comme fiable.
- **Vitesse de gravure optimale** : les CD optimisés pour des vitesses basses (jusqu'à 12x) obtiennent généralement de meilleurs résultats que les autres.
- **Production triée** : le revendeur doit pouvoir garantir que la production a été pré-triée par le fabricant, de manière à éliminer les ratés de fabrication du lot. Cela se traduit en principe par des discontinuités dans les numéros de série des CD achetés.

Cela dit, ces paramètres n'offrent pas des garanties suffisantes. La qualité d'un CD-R gravé est caractérisée par une série de paramètres, délivrés par un analyseur, parmi lesquels on retiendra :

- Le taux d'erreurs corrigibles (BLER, BERL, E22) et incorrigibles (E32).
- Les qualités de pré-traçage de la piste : Push Pull.
- La qualité de la modulation : I3, I11.
- La précision des transitions on/off du laser de gravure : symétrie, jitter.
- La majorité des analyseurs courants n'indiquent que les taux d'erreur. Les préconisations de l'Association internationale des archives sonores et audiovisuelles (IASA) fixent les valeurs à ne pas dépasser lors du contrôle qualité post-gravure :

BERL	< 5
BLER moyen	< 2
BLER max	< 10
E22	= 0
E32	= 0

Les caractéristiques de la gravure varient en fonction de la vitesse de gravure et du graveur utilisé. Aucune marque de CD-R ne peut donc être recommandée pour elle-même indépendamment du graveur avec lequel on la couple. De même, les variations dans la composition, le mode de fabrication et la rigueur du tri obligent à réexaminer très régulièrement ses choix.

Il existe deux grandes familles d'analyseurs sur le marché :

Software (PlexTools, CD Inspector, QA 201…) : ce sont les moins couteux, car ils exploitent les résultats fournis par un lecteur externe. Mais leur fiabilité dépend de celle du lecteur dont on se sert, ce qui représente le plus souvent une inconnue. Il est recommandé au moins, si on sert de ce type d'outils, de comparer les résultats obtenus en se branchant sur deux lecteurs différents.

Hardware (CATS, EC2 …) : ce sont des appareils d'analyse complets, platine de lecture incluse. Des modèles fiables existent à partir de 6 000 euros. L'analyse en vitesse réelle (1x) est vivement recommandée.

Les limites du CD-R deviennent manifestes pour de grandes masses de données : le contrôle sur analyseur et la copie sont des tâches fortement consommatrices de temps, du fait de la non-robotisation de ces tâches ; l'investissement initial est très faible, mais le coût unitaire du CD-R (support vierge + temps-personne) est peu compétitif en comparaison des solutions de stockage de masse.

ARCHIVER EN MASSE SUR BANDES OU DISQUES DURS

L'archivage à base de bandes magnétiques offre aujourd'hui le meilleur rapport entre qualité des données et prix de revient pour la gestion de grandes masses d'informations. Ces supports en cartouches peuvent être déployés dans des robotiques qui en assurent le chargement en lecteurs-enregistreurs. Ceux-ci sont couplés à des serveurs sur disque où les données sont stockées le temps d'une consultation. L'accès et le contrôle du support sont donc largement automatisables.

Plusieurs technologies sont en concurrence (LTO, Storagetek, AIT…), très fiables mais soumises à des cycles d'obsolescence très courts du fait de la densification croissante des capacités (les volumes stockables sur un support doublent tous les dix-huit mois à deux ans). Le renouvellement du parc de lecteurs-enregistreurs et la migration des données sur une nouvelle génération de supports sont à prévoir tous les quatre à six ans, selon les prévisions des fabricants.

Plus couteux, le stockage entièrement en-ligne sur disques durs nécessite une extrême sécurisation de son architecture (bien plus développée que dans les classiques schémas de réplication de données RAID).

DANS TOUS LES CAS, LE STOCKAGE NUMERIQUE EST GENERATEUR DE COUTS RELATIVEMENT IMPORTANTS TOUT AU LONG DE LA VIE DE L'INFORMATION A CONSERVER.

SUPPORTS POUR ENREGISTRER ET ARCHIVER LA VIDEO[77]

Numériser la vidéo pour la sauvegarder

VIE ET MORT DES TECHNOLOGIES VIDEO ANALOGIQUES

Au même titre que les technologies d'enregistrement audio analogiques (et à plus court terme encore peut-être), les technologies vidéo analogiques sont aujourd'hui en voie d'extinction. Devenues sans usage dans le domaine de la production, supplantées qu'elles sont par le numérique et les facilités qu'il offre, évincées par le DVD dans le domaine de l'édition du fait de leur qualité inférieure, les bandes vidéo et les vidéocassettes n'auront bientôt plus de valeur que pour les archives qui les auront collectées patiemment, mais qui n'ont pas d'autre voie que de les numériser pour continuer à en rendre le contenu accessible. Pas davantage que dans le monde de l'audio ou des documents informatiques, les besoins des archives n'ont suffi ni ne suffiront dans l'avenir à prolonger la vie d'une technologie vidéo. Il faut organiser au plus vite, si ce n'est déjà fait, l'entrée dans la sphère numérique des documents existant uniquement sous des formats analogiques. Ce au rythme le plus rapide possible, car les couts de transfert du document augmenteront aussi vite que se raréfieront le matériel de lecture en état de marche et les compétences humaines pour le maintenir.

PLUSIEURS DEGRES D'URGENCE TECHNIQUE

Depuis plusieurs années, il est devenu difficile et couteux de faire transférer des formats totalement obsolètes, tels que 2 pouces et 1 pouce (broadcast), mais aussi EIAJ, VCR (institutionnel), V2000 et Betamax (grand public). L'U-Matic et sa variante le BVU, support très représenté dans les archives institutionnelles et culturelles, est gravement menacé de par sa dégradation intrinsèque et la disparition du matériel de lecture. Apparus dans les années 80 et largement diffusés, les formats plus récents Betacam SP (broadcast) et VHS (grand public) entrent à leur tour dans la zone rouge. L'arrêt de la fabrication du matériel VHS de niveau professionnel en est un signe important. Une fois obsolète, le matériel acquis ne peut être entretenu qu'un temps limité. La disponibilité des pièces détachées est généralement garantie pendant moins de dix ans après la cessation de fabrication.

Ces degrés d'urgence déterminent ainsi des niveaux de priorité technique, qui sont ensuite à croiser avec les priorités documentaires.

OBJECTIF : LE STOCKAGE DE MASSE DANS DES STANDARDS OUVERTS

Numériser, mais dans quel format de données ?

Côté format, le choix d'un standard « ouvert » s'impose, quel que soit le média dont on envisage la numérisation (image fixe, image animée, écrit, son). Autrement dit, les règles d'interprétation informatique des données numériques en signal vidéo doivent être publiques. Un format « propriétaire » (secret industriel) sera difficile, voire impossible à restituer le jour où le matériel qui lui est associé aura disparu.

[77] Fiche rédigée par Alain Carou et Dominique Théron (BnF).

L'HYPOTHESE BETA NUMERIQUE

A ce titre, le transfert sur Betacam numérique (format propriétaire Sony) ne peut être qu'une solution transitoire, une étape de transfert avant le passage à un format numérique pérenne. En fin de vie de cette technologie, il sera nécessaire de relire en vitesse réelle les supports pour passer à un autre format. L'évaluation du cout réel d'une sauvegarde en passant par le Beta numérique (support déjà couteux à la base) doit donc intégrer le cout d'une migration ultérieure. Cette option aura cependant un intérêt dans deux cas :

- Si l'on veut stocker à terme dans un format numérique sans compression (voir § suivant), mais que monter une chaine de numérisation de masse de ce type soit difficile à court terme, le passage par le Beta numérique représente une mesure conservatoire.
- Si l'archive n'est pas du tout prête à un archivage de masse rationnel de fichiers numériques : le Beta numérique permet alors de rester dans une logique traditionnelle de supports sur étagères, le temps de préparer la mutation nécessaire.

COMPRESSION OU NON ?

La compression vidéo repose sur l'élimination de détails pas ou peu perçus par l'œil et l'utilisation des redondances d'une image à l'autre : l'économie faite dans la description des images successives entraine une réduction du volume de données d'un facteur 10, 20 ou 50.

Le choix de la compression s'impose aujourd'hui pour toute opération de numérisation en masse si l'on veut rester dans des échelles de cout raisonnables. Le principe de réalité entre ainsi en conflit avec la règle déontologique, strictement observée dans le monde des archives sonores, qui imposerait une numérisation sans compression. Pour des usages spécifiques de recherche, qui requièrent un maximum de définition et des analyses image par image (exemples : une opération chirurgicale, une interprétation musicale filmée en plan large), un minimum de compression, voire pas de compression du tout, est une option à examiner sérieusement.

Le taux de compression (défini en nombre de bits par seconde) est à choisir en fonction de la qualité du format d'origine : 6 Mbits/sec suffisent amplement pour le VHS, 12 Mbits/sec paraissent nécessaires pour le Betacam SP.

LES NORMES MPEG

Les normes MPEG-2 et 4 répondent à l'exigence de compression et de format ouvert. A l'heure actuelle, le MPEG-2 reste le standard dominant. Le développement de la norme MPEG-4 (qualité analogue, voire supérieure, pour un débit moindre) est cependant à suivre.

La conformité des numériseurs à la norme qu'ils sont supposés produire doit être contrôlée, dans la mesure où elle n'est pas systématiquement assurée (exemple : respect de la résolution de l'écran 720x576). Des outils logiciels d'analyse du flux MPEG existent pour cela.

METADONNEES

Un document numérique, quel qu'il soit, n'est pas pérennisable sans un minimum de métadonnées associées. Les métadonnées minimales sont celles qui permettront l'identification du contenu, la description complète de son mode de production (description de la chaine de numérisation) et les caractéristiques techniques du format qui permettront d'engager des actions de pérennisation (par exemple la migration vers un autre format) en cas de

risque. Pour être exploitables informatiquement, ces métadonnées doivent obéir strictement à une formalisation (par exemple dans le langage de balise XML). Afin de limiter au maximum les saisies manuelles, perte de temps pour les techniciens, les métadonnées devront être générées automatiquement en exploitant les informations déjà connues préalablement (celles issues notamment du travail de préparation documentaire).

D'autres métadonnées pourront par ailleurs être ajoutées à loisir selon l'usage : vignettes périodiquement extraites du document comme aide à la consultation ; image numérisée de jaquettes ou de fiches papier associées au document vidéo ; indexation temporelle du contenu ; ou encore (dans le futur) reconnaissance de la voix permettant une recherche « plein texte », etc.

ORGANISATION DE LA CHAINE DE NUMERISATION

La numérisation se décompose en plusieurs étapes mais a pour règle de base la meilleure relecture possible du document d'origine.

PREPARATION DOCUMENTAIRE ET PHYSIQUE DES ELEMENTS

Le travail commence par une identification du support, du standard couleur (ou NB), de la durée et, si possible, du contenu. Idéalement, un magnétoscope permettant de relire les bandes concernées doit donc être à la disposition de la personne chargée de cette préparation. Mais les archives anciennes (antérieures à 1975) et /ou broadcast (antérieures à 1985), sont des bandes magnétiques sur flasques qui peuvent réclamer l'aide d'un prestataire seul équipé pour cela. Une fois les analyses, tris et classements effectués, une liste dans un tableur devient l'outil de base. Attention à ne pas sous-estimer l'organisation des informations dans cette liste qui servira à alimenter diverses bases de données par la suite.

Il faut alors nettoyer la bande, avec des machines spécialisées quand elles existent, ou grâce à un passage sur un magnétoscope de réforme et un essuyage manuel quand il n'y a pas d'autre solution.

CHAINE DE TRANSFERT

Vient ensuite le magnétoscope : bon état général, têtes de lectures neuves ou récentes, niveaux audio et vidéo correctement réglés sont la base. Le standard couleur, s'il n'est pas en PAL d'origine (mais en Secam ou NTSC), doit impérativement être transcodé dans de bonnes conditions grâce à un appareil spécifique.

Un autre élément incontournable de la chaine de lecture est le TBC (correcteur de base temps), appareil voué à compenser les instabilités et fluctuations temporelles présentes sur le signal vidéo. Le recours à des machines contemporaines des sources à traiter, le plus souvent analogiques, permet de résoudre un certain nombre de problèmes dépassant les normes actuelles qu'un TBC numérique contemporain sera incapable de traiter.

Cette phase de lecture du document ne saurait être complète sans évoquer les différents outils de contrôle nécessaires : oscilloscope, vecteurscope (PAL), moniteur vidéo et audio de bonne qualité.

NUMERISATION, COMPRESSION

Suivent la conversion analogique-numérique proprement dite, et la compression. La performance des cartes d'encodage vidéo en matière de compression varie d'un modèle à l'autre. Il est indispensable de tester leur fiabilité en examinant leur capacité à gérer des images très mouvantes (par

exemple la surface de l'eau, ou une danse à un carnaval) sans générer de défauts (carrés figés, pixellisation).

CONTROLE QUALITE

Un contrôle qualité de la numérisation et des métadonnées s'impose avant la sauvegarde sur support d'archivage. Dans l'intervalle, le fichier reste sur un serveur-tampon (baie de disques sécurisée).

Il porte sur la conformité des noms de fichiers, des métadonnées et sur la qualité du résultat livré. L'attention du vérificateur devra se porter sur les « pertes de synchronisation » (perte de l'image et du son), les problèmes de « tracking » (suivi de piste, réglable sur le scope), la présence des canaux son et le réglage mono/stéréo.

A l'organisation du contrôle puis du versement dans l'archive numérique finale (cf. *infra*) doivent répondre impérativement des capacités serveur et réseau adaptées.

CORRECTION DU SIGNAL

Un traitement du signal peut être souhaitable. Quand on a affaire à des supports de qualité médiocre, un débruitage en amont de la numérisation est indispensable pour permettre une compression MPEG correcte. En effet, le bruit (parasitage aléatoire du signal) représente en numérique une masse considérable d'informations à gérer en plus du signal utile.

D'autres opérations peuvent intervenir en aval de la numérisation, avec des outils très performants. Il convient de dissocier restauration linéaire avec des réglages moyens (colorimétrie par exemple) s'appliquant à tout le document, et restauration plan à plan. Le facteur « temps passé » d'opérateurs spécialisés est discriminant entre la restauration de documents seulement destinés à l'archivage et la restauration de documents ou d'extraits voués à une diffusion commerciale. Dans le cas d'une intervention lourde, il devrait être décidé, pour des raisons déontologiques, d'archiver la copie droite (avant restauration) en plus du résultat final.

En tout cas, la restauration ne doit pas être considérée comme un substitut à une lecture de qualité, c'est une opération complémentaire.

VERSEMENT DANS L'ARCHIVE NUMERIQUE

Il n'existe pas de technologie de stockage numérique pérenne. Le stockage numérique est donc affaire de migrations (copies) en masse périodiques et contrôlées. Rien à voir avec la lourdeur de la copie d'analogique en numérique : la copie de numérique à numérique peut être automatisée et ultra-rapide.

Les technologies de bandes d'archivage offrent les niveaux de sécurité (élevé) et de cout (bas) recherchés. Super DLT, Super AIT, LTO sont des choix correspondant aux besoins de la vidéo, avec des capacités de stockage sur bande allant actuellement de 100 à 400 Go. Des robotiques ou, à moindre échelle, des systèmes auto-loader avec un lecteur permettent de réaliser les opérations de lecture, contrôle d'état des supports et copie sans manipulation humaine. Ces technologies sont soumises à un cycle d'obsolescence rapide, qui contraindra à des migrations de masse tous les cinq ou six ans environ. D'où la nécessité impérieuse de disposer d'une visibilité financière à moyen terme, au-delà de l'opération de passage au numérique, pour garantir la pérennité des investissements engagés et – surtout – l'accès aux fonds qui ne seront bientôt plus accessibles du tout sous leur forme analogique d'origine.

Quelques références complémentaires en ligne (essentiellement en anglais) :

Identifier les formats vidéo à vue d'œil et connaitre le niveau de risque technique :

www.video-id.com

Conserver l'accès aux données numériques :

bibnum.bnf.fr/conservation/infopreservation_fr.pdf

Le format de métadonnées de préservation METS :

www.loc.goc/standards/mets

Numériser sans compression la vidéo scientifique, une démarche pionnière de la Phonogrammarchiv de Vienne (à lire dans un souci prospectif, mais encore difficile à mettre en œuvre dans les limites économiques habituelles) : www.pha.oew.ac.at/phawww/literatur/iasa21_2003.pdf

CODAGES ET FORMATS

Pour les ressources enregistrées, leurs annotations linguistiques et documentaires.

Dans le monde informatique, les données sont codées en suivant des codages explicitement définis et organisés logiquement dans des formats de fichier. Ces derniers sont eux-mêmes stockés sur des supports ayant leur propre organisation physico-logique.

LES GRANDS PRINCIPES GUIDANT LE CHOIX D'UN CODAGE OU D'UN FORMAT

La distinction la plus importante est celle qui est faite entre *propriétaire* et *non-propriétaire*. Un codage propriétaire est un codage qui appartient à une personne ou une société qui en garde secrète la description. Il s'agit en règle générale d'une stratégie commerciale. Un tel codage est à bannir pour la conservation à long terme dans la mesure où les données ainsi codées risquent de disparaitre avec le secret de leur description. Seul un codage non propriétaire et libre permet une conservation dans de bonnes conditions.

Un autre aspect important, étroitement lié à l'aspect non-propriétaire, est la *standardisation* ou la *normalisation*. On peut définir un standard comme un accord entre des fabricants industriels qui défendent leur intérêts (souvent commerciaux), alors qu'une norme est un accord passé au sein d'un État (normes nationales : par exemple l'AFNOR) ou entre des États (normes internationales : par exemple l'ISO). Les organismes de normalisation prévoient aussi des mécanismes d'entretien et de conservation des normes créées, ce qui n'est pas forcément le cas des organismes de standardisation. On privilégiera les normes internationales dans la mesure où elles représentent la meilleure garantie de maintien de la connaissance indispensable à une interprétation correcte des données.

Un autre aspect auquel il faut prêter attention est la possibilité, pour un codage, d'utiliser des techniques protégées par des brevets, ce qui peut en limiter l'usage pendant un certain temps et/ou sur une certaine zone géographique. Par exemple, le groupe de travail « Moving Pictures Experts Group » qui gère, sous les hospices de l'ISO, les standards de compression, de décompression, de codage... pour l'image animée et pour le son, a notamment défini un standard connu sous le nom de « MP3 » ou « MPEG audio Layer 3 ». Ce codage, qui pour l'usager semble libre parce qu'utilisé dans des outils eux-mêmes gratuits, est en fait couvert par un brevet détenu par les sociétés Fraunhofer IIS et Thomson, et n'est ni libre ni gratuit.

LES DONNEES AUDIO

CODAGES

Un codage au sens étroit du terme désigne le type de correspondance que l'on souhaite établir entre chaque valeur du signal analogique et le nombre binaire qui représentera cette valeur. Il existe différents types de codages :

PCM : (Pulse Coded Modulation) c'est la valeur réelle de la mesure qui est représentée ;

Différentiel : c'est la différence entre le niveau du signal à l'instant de l'échantillonnage et le niveau qu'il avait lors de l'échantillonnage précédent qui est représenté.

Prédictif : il prévoit la valeur suivante d'après l'historique des valeurs échantillonnées. Le codage mesure seulement la différence entre la valeur prévue et la valeur réelle.

Adaptatif : il adapte la résolution (nombre de bits) au type de variation sonore détecté.

Le codage le plus simple et le plus répandu est certainement le codage PCM, même si ce n'est pas le plus économique en espace de stockage ou en temps de transfert. En dehors de ces choix de codage, la qualité de l'enregistrement dépendra du matériel de prise de son ou de numérisation, de la situation d'enregistrement, ainsi que des caractéristiques de numérisation : fréquence d'échantillonnage, résolution de l'échantillon et nombre de canaux[78].

Il est aussi d'usage de parler de codages pour les algorithmes de compression que l'on peut appliquer aux données. Ces algorithmes proposés dans des programmes appelés *codec* aboutissent généralement à une perte d'information (MACE, MPEG, u_law, etc.), c'est-à-dire que le résultat de la décompression des données n'est pas identique à l'original. En général, une bonne compression de parole ou de musique propose de supprimer en priorité les informations que la physiologie de l'oreille humaine ne permet pas d'entendre (voir fiche *Codages*). Ces algorithmes ont pour but de diminuer la taille des fichiers ou d'accroître le débit des transferts. Pour de la conservation de document il est bien évident que l'on ne se tournera pas vers de telles solutions. Pour la même raison, on évitera l'utilisation d'outils comme les enregistreurs miniDisc qui appliquent à la source un algorithme de compression.

FORMATS

Un format de fichier définit les règles d'écriture et l'organisation des données encodées. Ces règles sont utilisées par les logiciels pour écrire/enregistrer et pour lire/écouter. Les formats de fichier audio sont assez nombreux (RIFF/wav, AIFF, AU, MP3…). Ils peuvent éventuellement être liés à certains codages (par exemple, le format MP3 est lié au codage MPEG). Comme pour les codages, le choix d'un format reposera sur son aspect propriétaire ou non, normalisé ou non. Une attention particulière sera apportée à l'aspect libre du format. En effet, certains formats sont liés à des techniques soumises à des brevets qui ne pourront pas forcément être acquittés des utilisateurs successifs. De plus l'emploi de ces formats est souvent limité aux seules solutions que le fabricant logiciel qui détient le brevet propose, en général uniquement pour les plates-formes porteuses commercialement (MS-Windows, MacOS, etc.). A plus long terme, si vous ne trouvez pas comment normaliser vos données, vous risquez de ne plus pouvoir les lire (les fabricants de logiciels ne sont pas tenus d'en assurer la maintenance).

[78] Pour ces caractéristiques, nous conseillons d'adopter celles des CD-Audio (bon compromis qualité/quantité), c'est-à-dire : 44 100 Hz, 16 bits, mono ou stéréo (en fonction des conditions de l'enregistrement). L'organisme IASA (International Association of Sound and Audiovisual Archives) préconise actuellement, pour la conservation des données audio, des caractéristiques plus élevées : 96 KHz, 24 bits, format BWF).

LES ANNOTATIONS LINGUISTIQUES

CODAGES

Les annotations linguistiques sont composées de définitions d'objets linguistiques (les mots, les morphèmes, les tours de parole, etc.) et de commentaires sur ces objets. On distingue généralement dans les commentaires, les transcriptions qui donnent une version écrite de l'oral (en utilisant un certain nombre de conventions de notation comme celles de l'API), des autres annotations que sont les traductions, les gloses, les indications de mise en scène, etc. qui utilisent, elles, une métalangue (la plupart du temps, il s'agit de la langue de l'annotateur). Toutes ces annotations requièrent la mise en place d'un système de codage des caractères. Les conventions d'écriture des langues précisent aussi le sens de l'écriture, l'ordre des éléments à utiliser lors d'un tri, les équivalences de casse, l'utilisation de la ponctuation, etc. Depuis 1990 (date de la version 1.0), nous disposons d'un code « universel » qui fédère l'ensemble des codes existants. Ce code (Unicode[79]) est synchronisé sur la norme ISO-10646 qui a le même objectif. Il est déjà largement utilisé et a été adopté notamment par la Toile. Il permet donc de coder des documents multilingues mélangeant des écritures aux caractères et aux propriétés différents, et ceci de manière indépendante de la plate-forme informatique utilisée, ce qui facilite l'échange et le partage des documents. Dans la mesure où il n'y a pas d'autres propositions de codage en concurrence, Unicode est devenu incontournable.

Le reste des annotations concerne les objets de l'analyse. Ces objets doivent à la fois être définis et utilisés de manière identifiable dans les documents. Les codages utilisés en linguistique sont très liés aux théories employées et sont très peu formalisés, de sorte qu'il n'y a pratiquement pas d'implémentation informatique. La plupart du temps, il s'agit tout au plus d'ontologies. A notre connaissance, le travail le plus abouti et accepté comme un standard est certainement la TEI (Text Encoding Initiative). Elle a pour vocation le codage de la structure logique d'un certain nombre de types de documents utilisés dans la littérature, la linguistique, etc., comme par exemple les poèmes, les pièces de théâtre, les dictionnaires, les transcriptions de la parole. Ces propositions de codage ne sont pas forcément adéquates à toutes les analyses possibles, mais il est judicieux, au moment de choisir un codage, de se situer par rapport à celles existantes. Il sera aussi utile de suivre les avancées du groupe de travail de l'ISO/TC 37/SC4 qui porte sur la gestion des ressources linguistiques, qui est aujourd'hui en cours d'élaboration et qui concernera autant le codage des annotations linguistiques que celui des métadonnées documentaires.

FORMATS

Les deux familles principales de format de fichier pour structuration de l'information sont les bases de données relationnelles et les langages de balisage de textes. Nous ne parlerons pas d'une troisième grande famille représentée par l'ensemble des systèmes propriétaires, qu'ils puisent leurs justifications historiquement ou commercialement, ni des outils dont le but n'est pas la structuration de l'annotation mais sa présentation typographique, sa mise en page (logiciels de traitement de texte).

---

[79] Site web du Consortium Unicode (http://www.unicode.org).

Les bases de données sont généralement utilisées pour traiter des données de calcul alors que les systèmes de balisage de texte le sont pour les données textuelles. Ces deux mondes sont beaucoup plus entrelacés que par le passé.

La plus grande révolution a certainement été l'arrivée en 1998 du langage de balisage de texte XML. Ce dernier est un avatar de SGML, lui même normalisé ISO-8879 en 1986. XML est à la fois plus simple et plus moderne que son ancêtre. Il est bien intégré dans le web. Il s'agit en fait de tout un ensemble de technologies (Xpath pour l'identification et la navigation dans une arborescence XML, Xlink et Xpointer pour l'expression des liens, XSL pour la définition de feuilles de styles, Xquery comme langage de requête, DOM comme interface de programmation…). L'ensemble de ces technologies est géré par le consortium W3. Son adoption par les fabricants de logiciel a été très rapide et XML est considéré maintenant comme un standard incontournable pour la structuration, la gestion et l'échange des ressources. Du point de vue des bases de données relationnelles, il est surtout utilisé comme un format d'échange permettant de passer d'un système à un autre. Plus récemment, l'apparition de bases de données natives XML a rendu plus floue la distinction entre ces deux mondes.

Un des grands principes de XML est la séparation de la structure logique de la structure physique (par exemple sa mise en page). Une autre propriété de XML est qu'il permet de définir une syntaxe formelle pour la description de la structure logique des documents que l'on souhaite créer. C'est ce qu'a fait la TEI en définissant une ou des DTD[80].

LES METADONNEES

Les métadonnées servent à décrire des ressources (enregistrements, annotations). Ces descriptions peuvent contenir des informations sur la nature physique des ressources (durée de l'enregistrement, format de fichier, etc.), sur les droits associés, sur la situation d'enquête (lieu, date, participants, etc.). Ces métadonnées correspondent aux renseignements que l'on pourrait trouver dans une notice bibliographique de bibliothèque. Il existe un certain nombre de renseignements communs avec ce type de notice, mais les caractéristique propres des corpus oraux, ainsi que les préoccupations particulières des personnes qui les étudient, ont conduit à la définition de champs tels que l'âge du locuteur ou les conditions d'enregistrement, que l'on aura plus de mal à faire entrer dans une notice classique de bibliothèque. Les métadonnées servent principalement à deux choses : à cataloguer et à échanger. Pour que les échanges soient possibles, il convient de normaliser à la fois la forme des métadonnées mais aussi la procédure d'échange.

CODAGES

Plusieurs codages ont été proposés et sont utilisés pour la description des enregistrements et de leurs annotations. La TEI propose d'écrire toutes ces informations dans un en-tête assez détaillé. Pour les ressources du web, Dublin-Core[81], normalisé ISO-15836 en 2003, propose un jeu de quinze étiquettes qui sont notamment utilisées dans les en-têtes des fichiers HTML. Il existe bien sûr les codages pratiqués par les bibliothèques tels que les standards Marc, US-Marc, etc. qui se sont adaptés pour coder les nouveaux

[80] Document Type Definition.

[81] Site web du Dublin Core Metadata Initiative (http://dublincore.org).

supports informatiques. Il existe aussi des communautés qui ont proposé des recommandations comme par exemple OLAC[82] (basé sur du Dublin-Core enrichi et spécifié pour l'adapter aux ressources linguistiques), ou IMDI[83].

FORMATS

Quelle que soit la manière dont les métadonnées sont encodées (préconisation Dublin-Core, OLAC, TEI ou IMDI), la tendance générale est à l'utilisation de XML comme format d'échange. Le libre choix est laissé aux gestionnaires des métadonnées de les structurer directement en XML, en utilisant une base de données ou toute autre solution. Le choix d'une solution repose sur les critères énoncés précédemment.

LES PROTOCOLES D'ECHANGE

Le protocole Z39.50 est une norme ANSI/NISO, gérée actuellement par la « Library of Congress ». Sa vocation est la recherche automatisée d'informations bibliographiques dans des bases de données réparties. Le but originel était l'interconnexion des systèmes ouverts (OSI). En fait, la plupart des implémentations existantes ont superposé ce protocole sur TCP-IP plutôt que sur les couches définies dans le modèle OSI. De nombreuses bibliothèques universitaires utilisent ce protocole pour échanger leurs notices bibliographiques. Leur nombre est actuellement en forte croissance.

L'OAI (Open Archive Initiative) est une organisation plus récente qui définit entre autres un protocole relativement simple pour la récolte de métadonnées dans les archives ou « réservoirs » de données. A l'origine, il s'agissait de permettre l'interopérabilité entre les différentes archives de pré-prints et e-prints qui avaient chacune leur langage de requête. Ce protocole comprend un petit nombre de requêtes qu'il est possible d'adresser à un détenteur d'archives. Par exemple, on peut obtenir d'un détenteur d'archives son identification, la liste de ses identifiants de ressources, la liste des encodages qu'il utilise pour ses metadonnées, etc. Ce protocole fixe aussi la syntaxe XML des réponses que peut émettre un fournisseur d'archives. Un certain nombre de règles de politesse doivent être implémentées par le fournisseur, comme l'envoi de codes particuliers pour signaler les erreurs de syntaxe des requêtes. Ce protocole a l'avantage d'être simple à mettre en œuvre, et d'utiliser XML pour le formatage. L'objectif de l'OAI est la standardisation de la procédure de collecte des métadonnées afin de permettre à des fournisseurs de services (par exemple des moteurs de recherche) d'effectuer leur travail sur des métadonnées préalablement centralisées. En effet, la recherche directe à travers un ensemble réparti de fournisseurs, comme c'est le cas avec le protocole Z39.50, pose des problèmes de performance lorsque certains nœuds du réseau ralentissent ou bloquent la poursuite.

[82] Open Language Archives Community (http://www.language-archives.org).

[83] EAGLES/ISLE Metadata Initiative (http://www.mpi.nl/IMDI/).

BIBLIOTHEQUE NATIONALE DE FRANCE

LES STATUTS DE LA BIBLIOTHEQUE NATIONALE DE FRANCE :

Dans son article 2, le décret no 94-3 du 3 janvier 1994 « portant création de la Bibliothèque nationale de France » indique que celle-ci :

> *« a pour missions […]*
>
> *de collecter, cataloguer, conserver et enrichir dans tous les champs de la connaissance, le patrimoine national dont elle a la garde, en particulier le patrimoine de langue française*[84] *ou relatif à la civilisation française. A ce titre,*
>
> – elle exerce, en vertu de l'article 5, alinéa 2, de la loi du 20 juin 1992[85] […] les missions relatives au dépôt légal confiées par cette loi et les décrets pris pour son application à la Bibliothèque nationale ; elle gère, pour le compte de l'État, dans les conditions prévues par la loi du 20 juin 1992 susvisée, le dépôt légal dont elle est dépositaire. Elle en constitue et diffuse la bibliographie nationale.
> – elle rassemble, au nom et pour le compte de l'État, et catalogue des collections françaises et étrangères d'imprimés, de manuscrits, de monnaies et médailles, d'estampes, de photographies, de cartes et plans, de musique, de chorégraphies, de documents sonores, audiovisuels et informatiques. […]
>
> *d'assurer l'accès du plus grand nombre aux collections […].A ce titre : - elle conduit des programmes de recherche en relation avec le patrimoine dont elle a la charge […] ; elle coopère avec d'autres bibliothèques et centres de recherche et de documentation français ou étrangers, notamment dans le cadre des réseaux documentaires ; […] elle permet la consultation à distance […] ; elle mène toutes actions pour mettre en valeur ses collections. »*

De cet ensemble d'activités, on retiendra six missions fondamentales de l'activité de la BnF : enrichir le patrimoine national dont elle a la garde, cataloguer ses collections, les conserver, les communiquer au public, les valoriser et enfin coopérer avec d'autres établissements. En ce qui concerne l'enrichissement des collections, nous renvoyons à la partie du chapitre quatre consacrée à la BnF. Rapportées au département de l'Audiovisuel, en charge du patrimoine sonore, vidéographique, multimédia et électronique au sein de la BnF, les cinq autres missions se déclinent comme suit.

[84] Nous soulignons.
[85] Remplacée depuis par les articles L131-1 à L133-1 relatifs au dépôt légal du Code du patrimoine (*Journal officiel* du 24 février 2004).

LA CONSERVATION DES DOCUMENTS SONORES (ET AUDIOVISUELS)

Conservant une collection d'un million d'enregistrements sonores, 120 000 documents vidéographiques et 70 000 documents multimédias et électroniques, le département de l'Audiovisuel a mis en place un plan de sauvegarde de ses collections, visant au transfert progressif des supports fragiles ou obsolètes sur support numérique. En ce qui concerne le son, les cylindres, les disques à gravure directe " Pyral ", les supports magnétiques (bandes, cassettes…) sont reportés en priorité sur support numérique (sur mémoire de masse informatique et sur CD-R). L'ensemble de la vidéo analogique (VHS, U-Matic…) a été intégrée dans le même plan de sauvegarde et a été numérisée.

Le département de l'Audiovisuel est membre de l'IASA (International Association of Sound and Audiovisual Archives, http://www.iasa-web.org dont il suit et relaye en France les préconisations en termes de conservation des supports audiovisuels. Il est également membre actif de l'Association Française des détenteurs de documents Audiovisuels et Sonores (AFAS) qui organise régulièrement des journées d'étude sur les questions de numérisation (voir sur le site de l'association : http://afas.mmsh.univ-aix.fr/).

LE TRAITEMENT DOCUMENTAIRE

L'ensemble des collections du département de l'Audiovisuel fait l'objet d'un traitement documentaire informatisé en format INTERMARC. Le catalogue du département de l'Audiovisuel est intégré au catalogue général de la Bibliothèque, BN-OPALE PLUS, et peut être consulté en ligne à l'adresse : http://www.bnf.fr. A noter toutefois qu'en raison de la complexité de certains fonds, un service de recherche à distance permet de répondre aux questions des usagers : audiovisuel@bnf.fr.

LA CONSULTATION DES DOCUMENTS SONORES ET AUDIOVISUELS

Deux salles audiovisuelles ont été programmées sur le site François-Mitterrand-Tolbiac de la Bibliothèque : l'une au niveau Tout public (salle B), en Haut de Jardin, l'autre au niveau « Recherche » en Rez de Jardin (salle P). Celle-ci est équipée de 54 places audiovisuelles et de 17 cabines d'écoute et de visionnage. Accessible aux chercheurs, sur accréditation, elle offre à la consultation l'ensemble de la collection patrimoniale audiovisuelle du département. Un système audiovisuel constitué de régies manuelle ou robotisée, de serveurs numériques, de postes de consultation permet la communication de l'ensemble de ces documents.

LA VALORISATION DU PATRIMOINE SONORE

A l'heure actuelle le département de l'Audiovisuel offre une trentaine d'heures d'enregistrements sonores issues de ses collections à l'écoute en ligne sur le site de la Bibliothèque : http://www.bnf.fr, notamment dans les programmes « Gallica »,

« Gallica-Voyage en France », http://gallica.bnf.fr/voyagesenfrance/ ;

« Gallica-Voyage en Afrique », http://gallica.bnf.fr/VoyagesEnAfrique/ ;

« Anthologie », http://gallica.bnf.fr/Anthologie/.

Cette offre est évidemment destinée à croitre avec, dans un premier temps, le projet de mise en ligne de l'ensemble des enregistrements produits par les Archives de la Parole entre 1911 et 1914.

LA COOPERATION AU PLAN NATIONAL ET INTERNATIONAL

L'alinéa 4 de l'article 3 du décret du 3 janvier 1994 précise que la Bibliothèque nationale de France peut « coopérer, en particulier par la voie de convention ou de participation à des groupements d'intérêt public, avec toute personne publique ou privée, française ou étrangère, et notamment avec les institutions qui ont des missions complémentaires des siennes ou qui lui apportent leur concours ». Cette coopération prend place au sein du « Département de la Coopération » de la BnF qui travaille en étroite relation avec les départements de collections dont le département de l'Audiovisuel. Les « pôles associés » sont une illustration de cette coopération. A l'heure actuelle, dans le domaine de l'archive sonore, quatre centres (Conservatoire occitan, Dastum, Maison méditerranéenne des sciences de l'homme, Métive) affiliés à la Fédération des Associations de Musiques et de Danses Traditionnelles (FAMDT) sont ainsi pôles associés de la BnF et perçoivent une aide au traitement documentaire de leurs fonds. Aujourd'hui, la coopération s'oriente également vers des actions de numérisation partagée, de mise en place de projets de catalogues collectifs, etc.

LES ARCHIVES : LEGISLATION

LES ARCHIVES DE FRANCE

La direction des Archives de France est une direction du Ministère de la Culture et de la Communication qui assure la mise en œuvre et le contrôle de la loi 79-18 du 3 janvier 1979 sur les archives, aujourd'hui codifiée dans le code du patrimoine (ordonnance du 20 février 2004). Elle coordonne toutes les attributions confiées par la loi à l'administration des archives, à l'exception de celles qui concernent les archives des ministères des affaires étrangères et de la défense, et des services et établissements qui en dépendent ou qui y sont rattachés.

Depuis l'entrée en vigueur de la loi n° 83-663 du 22 juillet 1983, la direction des Archives de France ne gère plus directement les archives départementales, placées désormais sous l'autorité des conseils généraux, mais elle garde sur elles un contrôle scientifique et technique.

La direction des Archives de France comprend (arrêté du 25 mars 2002) :

- l'Inspection générale ;
- la délégation aux célébrations nationales ;
- le département du réseau institutionnel et professionnel ;
- le département de la politique archivistique et de la coordination interministérielle ;
- le département de l'innovation technologique et de la normalisation ;
- le département des publics ;
- le bureau des affaires générales et de la documentation.

LES ARCHIVES AU SENS DU CODE DU PATRIMOINE (LIVRE II)

Les archives sont définies à l'article L 211-1 comme suit :

> *« Les archives sont l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité. La conservation de ces documents est organisée dans l'intérêt public tant pour les besoins de la gestion et de la justification des droits des personnes physiques ou morales, publiques ou privées, que pour la documentation historique de la recherche. »*

La loi distingue deux catégories d'archives : les archives publiques et les archives privées.

LES ARCHIVES PUBLIQUES

« Sont considérées comme archives publiques :

Les documents qui procèdent de l'activité de l'État, des collectivités locales, des établissements et entreprises publics ;

Les documents qui procèdent de l'activité des organismes de droit privé chargés de la gestion des services publics ou d'une mission de service public ;

Les minutes et répertoires des officiers publics ou ministériels (article L 211-4). »

« Les archives publiques, quel qu'en soit le possesseur, sont imprescriptibles » (art. L 212-1).

« Les conditions de leur conservation sont déterminées par décret en Conseil d'État » (art. L 212-2).

« Les archives publiques font l'objet de procédures de sélection et de certaines règles précises d'élimination. »

A l'expiration de leur période d'utilisation courante par les services, établissements et organismes qui les ont produits ou reçus, les documents mentionnés à l'article 211-4 font l'objet d'un tri pour séparer les documents à conserver et les documents dépourvus d'intérêt administratif et historique, destinés à l'élimination :

« La liste des documents destinés à l'élimination ainsi que les conditions de leur élimination sont fixées en accord entre l'autorité qui les a produits ou reçus et l'administration des archives ».

LES ARCHIVES PRIVEES

« Les archives privées sont l'ensemble des documents définis à l'article 1er qui n'entrent pas dans le champ d'application de l'article 211-4 » (art. 211-5)

C'est le mode de production et non pas le type de support ou le sujet qui définit l'appartenance à l'une ou l'autre catégorie.

Exemple : l'enregistrement d'une séance du Conseil général est un document d'archives public tandis que l'enregistrement d'une interview d'un personnage politique à la radio est un document d'archives privé.

MODALITES DE CONSULTATION

Les *modalités de consultation* diffèrent selon la catégorie : pour *les archives publiques,* la communication est encadrée par la loi

« Article 6. – Les documents dont la communication était libre avant leur dépôt aux archives publiques continueront d'être communiqués sans restriction d'aucune sorte à toute personne qui en fera la demande. »

[...]

Tous les autres documents d'archives publiques pourront être librement consultés à l'expiration d'un délai de trente ans ou des délais spéciaux prévus à l'article L 213-2.

Article L 213-2 Le délai au-delà duquel les documents d'archives publiques peuvent être librement consultés est porté à :

> *« a) Cent cinquante ans à compter de la date de naissance pour les documents comportant des renseignements individuels de caractère médical ;*
>
> *b) Cent vingt ans à compter de la date de naissance pour les dossiers de personnel ;*
>
> *c) Cent ans à compter de la date de l'acte ou de la clôture du dossier pour les documents relatifs aux affaires portées devant les juridictions, y compris les décisions de grâce, pour les minutes et répertoires des notaires ainsi que pour les registres de l'état civil et de l'enregistrement ;*
>
> *d) Cent ans à compter de la date du recensement ou de l'enquête, pour les documents contenant des rensei-*

gnements individuels ayant trait à la vie personnelle et familiale et, d'une manière générale, aux faits et comportements d'ordre privé, collectés dans le cadre des enquêtes statistiques des services publics ;

e) Soixante ans à compter de la date de l'acte pour les documents qui contiennent des informations mettant en cause la vie privée ou intéressant la sûreté de l'État ou la défense nationale, et dont la liste est fixée par décret en Conseil d'État. »

POUR LES ARCHIVES PRIVEES

Article 213-6. Lorsque l'État et les collectivités locales reçoivent des archives privées à titre de don, de legs, de cession, de dépôt révocable ou de dation au titre de l'article 1131 et du I de l'article 1716 bis du code général des impôts, les administrations dépositaires sont tenues de *respecter les conditions de conservation et de communication qui peuvent être mises par les propriétaires.*

Leur consultation est donc définie par le propriétaire et spécifiée dans le contrat de don ou dépôt.

Cas particulier : Les Archives privées présentant pour des raisons historiques un intérêt public peuvent être classées comme **archives** *historiques*, sur proposition de l'administration des archives, par arrêté du Ministre chargé de la culture (art. L 212-15). L'article L 212-20 spécifie que « les archives classées comme archives historiques sont imprescriptibles » mais, article 12 que « le classement de documents comme archives historiques n'emporte pas transfert à l'État de la propriété des documents classés » (art. L 212-16).

Enfin, « toute destruction d'archives classées est interdite », article L 212-27/a.

PLACE DES ARCHIVES ORALES AU SEIN DES ARCHIVES NATIONALES ET DES SERVICES D'ARCHIVES DEPARTEMENTAUX ET MUNICIPAUX

L'article premier de la loi sur les archives ne fait pas de distinction par support ou par domaine. Les corpus oraux enregistrés sur support audio ou vidéo ne constituent donc pas une catégorie à part. Ils peuvent selon leur mode de production être des archives publiques ou des archives privées.

Les modes d'intégration dans les collections, peuvent se faire de façon passive ou active :

- L'institution reçoit les versements des administrations dans le cadre de l'exercice de la loi mais c'est le détenteur d'archives orales privées qui, seul, prend l'initiative du versement. Ce dernier a le choix entre le don, le legs, le dépôt révocable, de la cession de droit.
  C'est lui qui décide des conditions de consultation. S'il ne manifeste aucune volonté particulière, les règles des archives publiques seront appliquées aux archives orales privées.
- Le service d'archives peut prendre l'initiative d'un programme de collectes et produire, pour compléter ou se substituer à des archives absentes, des enregistrements de type interviews, témoignages, récits de vie.

MUSEES DE FRANCE : LEGISLATION

Le Code du Patrimoine consacre son Livre IV aux Musées pour lesquels la loi n° 2002-5 du 4 janvier 2002 a créé l'appellation « musées de France » :

> *« Article Premier. – L'appellation « musée de France » peut être accordée aux musées appartenant à l'État, à une autre personne morale de droit public ou à une personne morale de droit privé à but non lucratif ».*
>
> *Est définit « comme musée, au sens du présent livre, toute collection permanente composée de biens dont la conservation et la présentation revêtent un intérêt public et organisé en vue de la connaissance, de l'éducation et du plaisir du public. »*

Les « musées de France » ont pour missions permanentes de :

- Conserver, restaurer, étudier et enrichir leurs collections ;
- Rendre leurs collections accessibles au public le plus large ;
- Concevoir et mettre en œuvre des actions d'éducation et de diffusion visant à assurer l'égal accès de tous à la culture ;
- Contribuer aux progrès de la connaissance et de la recherche ainsi qu'à leur diffusion.

L'application de la loi passe par l'instauration d'un Haut Conseil des musées de France défini à l'article 3. Cette appellation peut être retirée. Si les collections des musées de France sont imprescriptibles, elles doivent, avant leur inscription sur l'inventaire des musées, recevoir l'avis scientifique de commissions spécifiques.

Les textes, la loi et les décrets et arrêtés pris pour l'application de la loi 2002-5 du 4 janvier 2002, favorisent l'organisation de réseau et une politique de dépôts d'œuvres d'un musée à l'autre. Les Directions Régionales des Affaires Culturelles (DRAC) sont chargées de veiller, en région, au contrôle technique de l'application des textes.

Les musées, régis antérieurement par l'Ordonnance de 1949, de par leur contenu et leur mode d'organisation sont d'une infinie variété. Entre l'établissement public du Louvre et un écomusée, pionnier mais de taille modeste, comme celui de la Roudoule dans les Alpes-Maritimes, peu de ressemblance, si ce n'est qu'il s'agit d'un de musée de France dans les deux cas.

INATHEQUE DE FRANCE

Sources de mémoire

L'INSTITUT NATIONAL DE L'AUDIOVISUEL

Créé en 1975, l'Ina est un établissement public à caractère industriel et commercial, chargé de conserver et exploiter le patrimoine audiovisuel français.

L'INATHEQUE DE FRANCE

La loi du 20 juin 1992 instituant un dépôt légal pour la radio et la télévision, représente une date essentielle dans l'histoire de l'audiovisuel français. Pour la première fois, à travers cette loi, l'audiovisuel, tout comme l'écrit, est considéré comme une source majeure d'archives et de mémoire.

Pour mettre en œuvre cette nouvelle mission, l'Ina crée, le 1er janvier 1995, l'Inathèque de France.

SES MISSIONS :

- Assurer la constitution et la conservation du patrimoine audiovisuel national.
- Organiser la consultation des œuvres et documents à des fins de recherche.
- Publier la bibliographie exhaustive des documents conservés au titre du Dépôt Légal.
- Favoriser la production et la diffusion des savoirs sur les images, les sons et les médias afin d'enrichir le débat public.

CONSERVATION ET ENRICHISSEMENT DOCUMENTAIRE

Ce sont 45 chaines de télévision et 17 diffuseurs radio qui sont suivis 365 jours par an.

Ce seront à terme 100 chaines collectées.

Chaque année 380 000 heures de programmes de télévision et 150 000 heures de radio sont identifiées et cataloguées dans les bases de données de l'Ina.

70 000 émissions de télévision et de radio font l'objet d'une description de contenu, sous la forme de mots-clés et de résumés, complétée par tout autre élément d'information nécessaire à l'exploitation de ces documents par les chercheurs.

CONSULTATION

L'Inathèque de France accueille les étudiants et les chercheurs dans son Centre de consultation situé au rez-de-jardin de la Bibliothèque nationale de France.

Le centre dispose de 56 places équipées d'un poste de consultation multimédia (SLAV : Station de Lecture AudioVisuelle) qui permet à la fois la consultation des bases de données de l'Ina, la gestion de corpus de travail, l'écoute ou le visionnage des émissions, leur analyse à l'aide d'outils adaptés.

Plus d'un million d'heures de télévision et de radio sont consultables.

La consultation s'exécute dans le respect du Code de la Propriété intellectuelle et artistique de sorte qu'aucune copie des enregistrements ne peut être effectuée, même à des fins pédagogiques et universitaires.

PROGRAMME « ARCHIVAGE » DU LACITO

Le LACITO (Laboratoire de Langues et Civilisations à Tradition Orale) est un laboratoire du CNRS dont les chercheurs (linguistes, anthropologues et ethno-musicologues) travaillent depuis plus d'une trentaine d'années à la description de langues pour la plupart sans écriture. De leurs enquêtes de terrain, ils ramènent des enregistrements audio, plus rarement vidéo, ainsi que des transcriptions, des traductions, etc. faites sur place avec l'aide de locuteurs. Ces enregistrements et analyses constituent les matériaux de base qui vont servir aux chercheurs pour poursuivre leurs recherches au retour de leur mission.

Le chercheur durant son enquête sera amené à expliquer les buts de sa mission, et tentera d'instaurer une « relation de confiance » entre lui et ses informateurs. Cette confiance est d'autant plus importante que les chercheurs sont parfois amenés à faire d'autres missions sur le même terrain. Elle peut être difficile à obtenir et facile à perdre, y compris par l'intervention ultérieure d'autres catégories d'enquêteurs (missionnaires, etc.) auxquelles les enquêtés risquent d'assimiler le chercheur.

L'*information préalable*, tout comme la *demande d'autorisation*, sont en général, compte tenu de la nature des cultures étudiées, faites sous un mode oral. Dans la pratique des enquêtes, ce n'est que très récemment que les chercheurs se préoccupent de garder une trace de cette autorisation (par exemple sous la forme d'un enregistrement audio). Ce qui prévalait jusqu'à présent, et qui prévaut encore aujourd'hui, c'est la relation de confiance qui lie enquêteurs et enquêtés.

L'information donnée par les chercheurs sur l'utilisation des enregistrements dépend fortement du niveau de culture de leurs interlocuteurs. Il est en effet parfois difficile de faire comprendre les implications de la mise à disposition d'un enregistrement audio sur la Toile à des personnes qui n'ont jamais entendu parler de l'informatique et qui n'ont jamais vu d'ordinateur. De plus, pour les enquêtes pratiquées il y a trente ans ou plus, aucun des chercheurs n'imaginait à l'époque les nouvelles utilisations qu'il pourrait faire de ses données. Dans ces conditions, bien sûr, l'information préalable ne pouvait être complète. Par ailleurs, le suivi des informateurs n'est pas aisé dans tous les pays et retrouver des locuteurs afin de les informer des changements de *finalité* n'est pas toujours possible.

Les enregistrements, jusqu'à récemment, servaient principalement aux chercheurs qui les avaient collectés. Des copies pouvaient en être faites pour des collègues, mais il n'existait ni catalogue, ni organisation pour le stockage, la conservation et la copie. Quand un chercheur disparaissait, toutes ses données accumulées risquaient donc de disparaitre avec lui.

Vers la fin de années 90, un programme s'est mis en place au LACITO pour lutter contre la disparition des données d'enquête. Dans ce programme « Archivage », les enregistrements analogiques sont numérisés afin de les préserver du vieillissement. Ils sont aussi catalogués afin de pouvoir les retrouver. Les notes de terrain (transcription, traductions, etc.) sont elles aussi numérisées et cataloguées. Enfin, des liens sont établis dans le catalogue pour ne pas perdre la relation qui existe entre enregistrements et notes de terrain.

Le programme « Archivage » a deux buts principaux qui sont :

- la préservation et la pérennisation des données d'enquête,
- leur diffusion.

La préservation est assurée par la numérisation des sources. Celle-ci se fait en utilisant des formats et des codages ouverts et libres. Les enregistrements sont numérisés sans compression en qualité CD-Audio dans un format WAV. Les notes de terrain sont structurées avec le langage de balisage de texte XML/Unicode en utilisant une syntaxe inspirée de la TEI. Les transcriptions sont codées la plupart du temps avec l'Alphabet Phonétique International. L'ensemble de ces ressources (fichiers audios et fichiers d'annotations) sont cataloguées au sein d'un document XML. Chacune d'elles est décrite à l'aide de métadonnées codées avec des étiquettes Dublin-Core suivant les spécifications préconisées par OLAC. Les ressources sont stockées sur des CD-ROM (un contrat est actuellement en discussion pour que la BnF soit le conservateur de ces données), elles sont aussi stockées sur un serveur où elles sont régulièrement recopiées sur des supports de sauvegarde.

La diffusion est assurée par un site Internet qui héberge à ce jour quelques 130 documents dans une trentaine de langues (principalement des langues de Nouvelle-Calédonie, du Népal et du Caucase). Une interface de consultation a été définie afin de consulter de manière synchronisée les documents d'enregistrement et leurs annotations. L'accès à l'ensemble de ces données se fait en général par la consultation du catalogue. Cela peut se faire localement en utilisant l'interface du site des archives, soit en consultant des moteurs de recherche spécialisés, puisque l'archive est une « archive ouverte », c'est-à-dire qu'elle utilise le protocole OAI-PMH pour communiquer avec l'extérieur. Tous les outils de consultation comme ceux qui ont été développés pour la création et la diffusion des ressources sont des *logiciels libres*.

CLAPI

Corpus de Langue Parlée en Interaction, laboratoire ICAR

Depuis trente ans, le laboratoire ICAR (ex-GRIC) (UMR 5191 du CNRS) mène à Lyon des recherches sur les interactions. En s'appuyant sur cette longue tradition, le laboratoire a développé une banque de données de Corpus de Langue Parlée en Interaction, CLAPI, dans le but d'assurer la sauvegarde et la gestion des corpus anciennement produits dans le laboratoire et de stimuler la production de nouveaux corpus en accord avec les exigences théoriques et technologiques du laboratoire actuel.

La base CLAPI compte en octobre 2005 :

- 600 h d'enregistrements audio et en partie vidéo, dont 350 h numérisées (2,5 Mo de mots) ;
- 20 h de transcriptions alignées avec le signal sonore et au format XML. (125.000 mots) ;
- 70 corpus (dont 35 décrits par la fiche de métadonnées) ;
- corpus d'interactions dans des situations sociales très variées (de la conversation quotidienne à des activités très spécifiques de travail, ou à des situations institutionnelles diversifiées) ;
- corpus de français et d'autres langues (comme p.ex. les langues régionales et l'arabe) et de situations natifs/non-natifs.

La base CLAPI ne se limite pas à accueillir des corpus ; elle est avant tout fondée sur un savoir-faire développé par une équipe, notamment dans les domaines suivants :

- *Le terrain* : le recueil de données en situation « naturelle » repose sur une approche ethnographique qui prépare les enregistrements et permet d'identifier les lieux et les moments les plus pertinents et propices à la constitution du corpus. Celle-ci ne se limite donc pas à une simple « capture » audio ou vidéo, même si l'enregistrement qui en est issu constitue le fondement du travail ultérieur de traitement et d'analyse.
- *L'enregistrement des données* : les enregistrements - audio et vidéo - visent des situations sociales d'interaction très diverses, recueillies dans un cadre dit « naturaliste », au sens de non orchestré et non provoqué par le chercheur. Des dispositifs d'enregistrement ont été développés qui concilient la capture multi-source et la préservation de la « naturalité » de l'interaction. En outre sont recueillis les documents, artefacts et autres objets manipulés ou produits pendant l'interaction.
- *La transcription* : la convention ICOR a été élaborée, qui assure la représentation, à différents niveaux de granularité, des phénomènes spécifiques à l'oral en interaction tout en garantissant l'homogénéité nécessaire à l'exploitation automatique et à l'interopérabilité ; la convention ICOR se limite pour l'instant aux phénomènes verbaux : la notation du multimodal est à l'étude.
- *L'identification des corpus et les métadonnées*: pour CLAPI a été mis au point un ensemble fonctionnel de descripteurs adaptés aux corpus de LPI (75 rubriques).
- *Les dimensions juridique, déontologique et éthique* : CLAPI a été l'occasion de concevoir des documents juridiques (autorisation de

recueil et de diffusion, convention de dépôt dans la base CLAPI, charte d'hébergement, convention de prêt) en application des dispositions relatives à la protection de la vie privée, à la propriété intellectuelle et au droit des bases de données ; ainsi que d'instaurer des pratiques d'anonymisation sur le signal, les transcriptions et le contenu des descripteurs.

- *L'intégration de corpus dans la base* : la base CLAPI a été conçue pour accueillir aussi bien des corpus anciens que récents, pour sauvegarder des corpus à valeur patrimoniale et historique comme pour inspirer la création de nouveaux corpus répondant aux exigences contemporaines sur le plan technique et scientifique.
- *L'hébergement sécurisé* : un système d'accès informatisés différenciés a été mis au point pour gérer les consultations et les interventions au sein de la base.
- *L'accès aux corpus* : CLAPI a stimulé les expériences sur la négociation de la diffusion des données interactionnelles, et les moyens humains qu'elle requiert, dans le respect des contraintes légales et en accord avec les auteurs des corpus. ICAR a pris l'option de rendre interrogeables en ligne librement, par les outils de la plate-forme, des extraits de corpus choisis par leurs auteurs (en octobre 2005, 15 extraits de corpus soit 3h30, dont 2h15 avec signal).
- *La diffusion de ces savoir-faires* : autour de CLAPI a été mise sur pied l'organisation de journées d'études, de formations internes/externes et aussi des propositions d'assistance technique et d'expertise.
- La conception et les développements d'outils de traitement et d'analyse des corpus :
  - traduction des transcriptions originales, quelle que soit la convention utilisée, au format XML, pour offrir l'homogénéité nécessaire aux analyses automatiques (20h à ce jour),
  - reconnaissance automatique des variantes graphiques d'une même forme générées par l'usage de « l'orthographe adaptée »,
  - phénomènes modélisés à ce jour : productions verbales/tours de parole et leur attribution aux locuteurs, timing, pauses (courtes, longues, quantifiées), chevauchements, tokens/ formes, commentaires/observations),
  - concordancier avec alignement texte/signal,
  - recherche de co-occurrences,
  - interface graphique permettant d'effectuer des requêtes personnalisées multi-critères qui combinent descripteurs et phénomènes,
  - bilan quantitatif des phénomènes par transcription,
  - repérage automatique des répétitions dans une transcription.

La plate-forme CLAPI est consultable à l'adresse suivante : http://clapi.univ-lyon2.fr

PFC

Le projet Phonologie du Français Contemporain (PFC) vise à constituer un vaste corpus de phonologie du français contemporain. Il a démarré sur l'initiative conjointe de Jacques Durand (ERSS, CNRS – Université de Toulouse le Mirail), Bernard Laks (MoDyCo, CNRS – Université Paris X Nanterre), Chantal Lyche (Université d'Oslo). À terme, PFC constituera la plus grosse base de données orales portant sur le français et l'une des plus grosses bases toutes langues confondues.

Le projet part de la constatation qu'il est nécessaire de poursuivre le travail de description entrepris depuis au moins un siècle par tous les spécialistes de la communication parlée pour :

- fournir une meilleure image du français parlé dans son unité et sa diversité, dans la réalité de ses usages attestés et dans sa diversité géographique, sociale et stylistique ;
- mettre à l'épreuve les modèles phonologiques et phonétiques sur le plan synchronique et diachronique ;
- favoriser les échanges entre les connaissances phonologiques et les outils du traitement automatique de la parole ;
- permettre la conservation d'une partie importante du patrimoine linguistique du monde francophone, et ce en contrepoint aux corpus déjà constitués ;
- permettre la constitution de meilleurs matériaux pédagogiques pour la description du français.

À partir d'un protocole d'enquête uniforme et en prenant appui sur des méthodes d'analyse et des outils développés en commun, le projet a pour ambition d'offrir une vision globale et unitaire de la phonologie du français contemporain, dont les principales caractéristiques sont :

- la diversité socio-géographique : une cinquantaine de points d'enquête prévus dans l'espace francophone à partir de groupes issus de réseaux denses. À chaque point d'enquête, 10 locuteurs en moyenne sont interviewés selon un protocole unique et constant ;
- la diversité de registres : 4 situations prises en compte dans le protocole ;
- la diversité des phénomènes phonologiques envisagés (inventaires phonologiques, schwa, liaison…) ;
- l'importance quantitative : environ 500 locuteurs, soit entre 800 et 1 000 h d'enregistrements, dont une centaine de locuteurs prévus pour la Belgique et les autres pays francophones.

En même temps qu'il offre à chaque chercheur la possibilité de suivre ses propres hypothèses et de construire ses propres objets sur la base sonore numérisée, PFC offre un ensemble de préanalyses particulièrement utiles en fournissant des transcriptions stables et vérifiées. Sur les fichiers sonores de ce corpus, est effectué un travail de transcription et d'alignement du texte sur le signal. Les principes de transcription s'inspirent des travaux et des expériences antérieures dans la transcription de gros corpus de français parlé (GARS à Aix-Marseille, VALIBEL à Louvain-la-Neuve), notamment pour la gestion des tours de parole, la transcription des pauses, des reprises, des incises et des erreurs.

S'agissant du schwa (le e muet) et des phénomènes de liaison, PFC offre en plus un ensemble d'analyses finales quantifiées avec un détail et une précision

jamais atteintes : pour chaque locuteur, codage complet et exhaustif, aligné sur le signal, de 3 et 5 minutes de parole en conversation guidée et libre, ainsi que de la lecture de deux éléments codifiées (un texte et une liste de mots).

Le traitement de ces données est basé sur les dernières avancées de la recherche en phonologie, phonétique et sociolinguistique. La structure informatique les stockant et les diffusant se situe, dans la problématique actuelle de diffusion de contenus sur Internet (particulièrement le langage de structuration XML et l'interopérabilité des données).

La difficulté spécifique d'un corpus sonore est, à l'heure actuelle, de pouvoir faire des liens entre une demande documentaire (les locuteurs du même âge ou de la même enquête) et telle partie d'un fichier sonore : trouver les « pointeurs » qui permettent, à partir d'un descripteur, formulé par un ou plusieurs termes, de rentrer à un endroit précis du fichier sonore. C'est un des enjeux du projet PFC : permettre la consultation d'un vaste corpus sonore, plus développé qu'une simple lecture des données.

PFC propose alors une structure de consultation des données recueillies et homogénéisées, via les protocoles Internet. Une base de données fortement structurée et relationnelle est ainsi accessible avec un simple navigateur. L'interface d'interrogation permet des requêtes larges et fines sur ces données avec un croisement inédit entre les données documentaires textuelles et les données sonores numérisées.

Notre objectif majeur est de construire un corpus favorisant différents niveaux d'approche, adapté à différents publics (étudiants, enseignants, chercheurs, ingénieurs). La variété des exploitations possibles est très grande grâce à la mise à disposition d'une ressource à la masse critique importante et aux données standardisées et donc interopérables. L'enseignant ayant besoin d'un tutoriel comparatif de français oral pour des publics, même jeunes, comme l'ingénieur devant construire un système de reconnaissance vocale, pourront se baser utilement sur cette ressource.

http://www.projet-pfc.net

DELIC

L'équipe DELIC (DEscription Linguistique sur Corpus), créée en 1999, a développé un projet qui s'appuie notamment sur l'élaboration et l'exploitation morphosyntaxique de corpus oraux (et aussi écrits). Elle a hérité du corpus du GARS (Groupe Aixois de Recherches en Syntaxe), ce qui l'a confrontée aux difficultés que soulève la récupération de corpus un peu anciens et a développé divers projets de constitution de nouveaux corpus. Cette présentation de l'équipe sera centrée sur les problèmes rencontrés pour témoigner de l'expérience acquise.

LA CONSTITUTION DES CORPUS

La partie ancienne (le corpus du GARS) a été développée pendant une vingtaine d'années[86] et compte 1 700 000 mots restaurés. En l'espace de 20 ans, de nombreux changements sont intervenus (modification des supports d'enregistrement, variations dans les conventions, sensibilisation aux problèmes juridiques, etc.). Tout un travail de restauration a donc été nécessaire.

Les nouveaux corpus tiennent bien évidemment compte de l'expérience acquise :

- les enregistrements sont effectués sur MD (minidisque), puis le son est numérisé ;
- le matériel d'enregistrement (MD et micro) permet de disposer, en général, d'enregistrements d'une bien meilleure qualité ;
- des autorisations sont remplies pour chaque nouvel enregistrement ;
- les conventions ont été revues pour écarter quelques phénomènes qui n'étaient pas utilisés dans les analyses (par exemple, les allongements ne sont plus notés) et restent stables depuis quelques années.

L'équipe a aussi pu bénéficier de l'expérience des membres du GARS qui étaient avertis de certains problèmes liés au recueil des données et elle continue à former des étudiants pour collecter et transcrire les enregistrements.

Pour les transcriptions, deux techniques sont utilisées : soit « à l'ancienne » avec écouteur et papier (ou saisie sur clavier), soit à l'aide du logiciel Transcriber (disponible sur le Net). Les transcriptions sont ensuite vérifiées par des personnes averties. Ce travail d'édition qui demande beaucoup de soin et de temps est indispensable.

LES PROBLEMES DE CONSERVATION

Pour la partie ancienne du corpus, les enregistrements sur cassettes sont quelquefois d'une qualité médiocre. Un certain nombre d'entre eux ont été numérisés, mais ce travail demande un investissement très lourd. D'autre part, dans certains cas, les enregistrements ont été égarés, et pour un grand

[86] Blanche-Benveniste, Cl. (2000) « Corpus de français parlé » dans Bilger ed. *Corpus Méthodologie et applications linguistiques*, rappelle quelques aspects du développement de ce corpus.

nombre de transcriptions il a fallu scanner les textes qui avaient été saisis avec des machines à écrire.

Tout cela a montré l'importance des tâches de gestion et de classement des archives. Une personne de l'équipe s'est spécialisée dans cette activité pour les nouveaux corpus afin de faire correspondre rapidement les divers documents qui doivent être reliés : enregistrement, fiche signalétique, transcription.

Pour le CRFP[87] (Corpus de Référence du Français Parlé) une grande énergie a été engagée dans les problèmes de gestion de ces divers documents et a montré la nécessité d'une organisation conséquente pour éviter (ou du moins limiter) la perte de certains documents (des fiches signalétiques), le travail inutile (corriger la version antérieure d'une transcription), etc.

Les enregistrements existent sous plusieurs formats (MD et fichiers numérisés) ce qui garantit, en partie, leur pérennité.

L'EXPLOITATION

Pour les corpus les plus récents, on procède ensuite à un alignement texte/son (pour lequel on utilise Transcriber). Ce traitement permet d'améliorer la qualité des transcriptions. D'autre part, les corpus peuvent être exploités à l'aide du logiciel *Contextes* réalisé par Jean Véronis. Dans sa dernière version, ce concordancier permet aussi d'écouter le passage sélectionné.

L'équipe, qui regroupe des linguistes et des informaticiens, développe des projets de description morphosysntaxique et d'analyse semi-automatique sur de gros corpus (écrits et oraux).

LA DIFFUSION

L'équipe possède actuellement plusieurs corpus dont on vient de rappeler qu'ils se rattachent à des projets distincts. Ils présentent des caractéristiques différentes (qualité des enregistrements, autorisation, etc.).

- CorpAix (le corpus du GARS) 1 700 000 mots
- CRFP 460 000 mots
- C-ORAL-ROM88 232 000 mots
- Corpus DELIC (en développement depuis 2000) 560 000 mots
- CRFP-2 (projet en cours) : enregistrement du français des médias.

Pour les plus anciens (CorpAix), la diffusion de passages longs est exclue, car ils ont été constitués en dehors d'un cadre juridique (pas d'autorisation). Le CRFP est consultable sous forme d'extraits sur le net (http://www.up.univ-mrs.fr/delic/). Un certain nombre d'autorisations manquent et bloquent sa diffusion. Le corpus C-ORAL-ROM est lui accessible par le biais de l'édition signalée.

Les résultats des analyses conduites sur les divers corpus mentionnés sont diffusés dans les publications des membres de l'équipe.

[87] Ce corpus est présenté dans *Recherche sur le Français Parlé* 18 (2004) Université de Provence.

[88] Ce projet européen qui porte sur 4 langues romanes est présenté dans Cresti, E. & Moneglia, M. *éd.* (2005) *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam, John Benjamins.

ESLO

LES ENQUETES SOCIO-LINGUISTIQUES A ORLEANS, 1968-2008

L'enquête ESLO (Enquête Socio-Linguistique à Orléans), conduite par des universitaires britanniques à des fins didactiques (enseignement du français langue étrangère dans le système public d'éducation anglais) en 1968, comprend environ 200 interviews, toutes référencées, et plus de 300 heures de parole incluant des enregistrements cachés, des conversations téléphoniques, des réunions publiques, des entretiens médico-pédagogiques, etc. Ce corpus constitue, par son ampleur et sa cohérence, le plus important témoignage sur le français parlé avant 1980.

Le premier objectif est de numériser les documents sonores à partir des enregistrements magnétiques et d'en proposer une indexation et un premier balisage afin de mettre les données en ligne sur Internet.

Parallèlement, une exploitation exhaustive d'un sous-ensemble est engagée. Partant de l'expérience acquise, le CORAL (Centre Orléanais de Recherche en Anthropologie et Linguistique) en partenariat avec d'autres laboratoires (CELITH-MODYCO) a mis en chantier une nouvelle enquête dénommée ESLO2. L'objectif est d'évaluer, à une quarantaine d'années de distance, la dynamique sociale du français (des usages de la langue comme des jugements sur son emploi). La prise en compte de la diversité des changements est rapportée aux paramètres sociaux, révélant l'inégalité des résistances ou des propensions à la transformation de la langue, mais aussi la typologie et la dynamique des évolutions.

Cette façon de procéder présente l'avantage de préfigurer la référence attendue dans un domaine qui en est encore à se structurer et dans lequel se manifeste de manière récurrente une demande de définition pour un format standardisé de *collecte*, de *conservation*, de *traitement* et d'*analyse* :

- la *collecte* sur le terrain est première, non seulement dans ses aspects techniques, aujourd'hui bien maîtrisés, mais aussi dans la définition du profil de l'échantillon représentatif et dans la problématisation des interactions entre les témoins et les enquêteurs ;
- la *conservation*, qui inclut la préservation des supports, l'indexation des contenus et l'accessibilité (c'est-à-dire la protection) des données, conditionne le partage des sources à des fins d'étude scientifique et d'expertise politique ;
- le *traitement*, en lien étroit avec le développement des matériels et des langages informatiques, suppose la maîtrise d'une chaîne d'opérations, depuis la conversion numérique des enregistrements jusqu'à une transcription balisée et ouverte à l'ensemble des interrogations pertinentes pour les demandes du linguiste, du sociologue ou des décideurs, des didacticiens voire du grand public ;
- l'*analyse* constitue l'épreuve des théories (et des logiciels) puisqu'elle compare les formalisations et les opérations et qu'elle valide ou infirme les hypothèses en prenant argument de leur compatibilité aux faits.

Avec la constitution et la comparaison de telles enquêtes, les politiques et les acteurs de la transmission linguistique ont à leur disposition un outil d'aide à la décision irremplaçable, qui permet d'appréhender, aussi objectivement que possible, le devenir du français parlé dans toutes ses dimensions (phonologique et prosodique, lexicale et syntaxique, sémantique et pragmatique). La définition d'un standard rigoureux et réaliste devrait orienter les descriptions du français parlé en France au service de la recherche, des applications et de l'expertise.

INVENTAIRE DES CORPUS

INVENTAIRE DE LA DGLFLF

L'aventure du *Guide des bonnes pratiques* a mis en lumière la nécessité de disposer d'une meilleure vision des corpus de langue française qui existent en France et à l'étranger. La DGLFLF a donc piloté un inventaire qui fournit diverses indications dont :

- le nom du corpus ;
- le responsable ;
- la taille, le contenu, l'état de ces données (supports utilisé, etc.) ;
- le type d'accès possible (accès libre, partiel, limité à une équipe, etc.).

La présentation de l'inventaire reprend et développe ces différents paramètres.

Cet inventaire (qui peut être téléchargé sur le site de la DGLFLF) pourrait faciliter les contacts et les échanges entre équipes, permettre d'identifier les manques les plus flagrants dans le domaine des données orales constituées et aider les futurs projets de constitution de grandes banques de données à mieux cerner les forces disponibles et les besoins.

En l'état, le lecteur dispose d'un état des lieux (partiel à cause des oublis) qui peut être complété en fournissant toute information utile à :

Paul.Cappeau@univ-poitiers.fr.

www.dglflf.culture.gouv.fr

BIBLIOGRAPHIE

BIBLIOGRAPHIE

BIBLIOGRAPHIE GENERALE

ABOU-HAIDIR, L., dir. (2002) « Transcription de la parole normale et pathologique », *Revue Parole* 22/23/24.

ACHARD, P. (1991) « Une approche discursive des questionnaires : l'exemple d'une enquête pendant la guerre d'Algérie », *Langage et société* 55 : 5-40.

ADLER, P.A. (1987) *Membership Rules in Field Research*, Sage, Newbury Park.

AIJMER, K. & ALTENBERG, B. eds (1992) *English Corpus Linguistics. Studies in honour of Jan Svartvik,* London/New-York, Longman.

ATKINSON, J.M. & HERITAGE, J. eds (1984) *Structures of Social Action*, Cambridge, CUP.

AUER, P. et al. (1999) *Language in Time. The Rhythm and Tempo of Spoken Interaction*, Oxford, OUP.

BANGE, P. (1983) « Points de vue sur l'analyse conversationnelle », *DRLAV* 29 : 1-28.

BARRAS, C., ADDA, G., ADDA-DECKER, M., HABERT, B., BOULA DE MAREÜIL, P. & PAROUBEK, P. (2004) « Automatic audio and manual transcripts alignment, time-code transfer and selection of exact transcripts », *Proceedings of the fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbonne : 877-880.

BAUDE, O. (2004) « Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques », *Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble) : 7-11.

BEAUD, S. & WEBER, F. (1997) *Guide de l'enquête de terrain : produire et analyser des données ethnographiques*, Paris, La Découverte.

BECKER, H. S. & GEER, B. (1960) « Participant observation : the analysis of qualitative field data », ADAMS & PREISS eds : 267-289.

BERGOUNIOUX, G. dir. (1992) « Enquêtes, Corpus et Témoins », *Langue Française* 93.

BIBER, D. (1985) *Variations across spoken and written language*, Cambridge, CUP.

BIBER, D. (1999) *Longman Grammar of Spoken and Written English*, Londres, Longman.

BILGER, M. dir. (2000) « Linguistique sur corpus, études et réflexions », *Cahiers de l'université de Perpignan*, Perpignan, Presses universitaires.

BILGER, M. ed. (2000) *Corpus, Méthodologie et applications linguistiques*, Paris, Champion.

BLANCHE-BENVENISTE, Cl. & JEANJEAN, C. (1987) *Le français parlé : transcription et édition*, Paris, Didier-Erudition.

BLANCHE-BENVENISTE, Cl. (1997) « Transcription et technologie », *Recherches sur le Français Parlé* 14 : 87-100.

BLANCHE-BENVENISTE, Cl. BILGER, M., ROUGET, C. & VAN DEN EYNDE, K. (1999) *Le Français Parlé : Études grammaticales*, Paris, CNRS-Editions.

BLANCHE-BENVENISTE, Cl., ROUGET, C. & SABIO, F. (2001) *Choix de textes de français parlé : trente-six extraits*, Paris, Champion.

BOURDIEU, P. (1982) *Ce que parler veut dire. L'économie des échanges linguistiques*, Paris, Fayard.

BOURDIEU, P. (1993) *La misère du monde*, Paris, Seuil.

BÜRKI, Y. & DE STEFANI, E *ed.* (à paraître), *Transcryptio*, Berne, Peter Lang.

CAMERON, D., FRAZER, E., HARVEY, P., RAMPTON, M. & RICHARDSON, K. (1991) *Researching Language : Issues of Power and Method*, London, Routledge.

CLIFFORD, J. & MARCUS, G. E. eds (1986) *Writing Culture. The Poetics and Politics of Ethnography,* Berkeley, University of California Press.

CONDAMINE, A. *ed.* (2006) *Sémantique et corpus*, Paris, Hermes.

COUPER-KUHLEN, E. & SELTING, M. *ed.* (1996) *Prosody in Conversation : Interactional Studies*, Cambridge, CUP.

CRESTI, E. & MONEGLIA, M. ed. (2005) *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/Philadelphie, Benjamins.

CRIBIER, F. & FELLER, E. (2003) *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »* rapport présenté à Madame la Ministre déléguée à la Recherche et aux nouvelles technologies,

dactylogr. 2 vol.

et http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf

DEPPERMANN, A. (2000) « Ethnographische Gesprächsanalyse : zur Nutzen und Notwendigkeit von Ethnographie für die Konversationsanalyse », *Gesprächsforschung* 1 : 96-124.

DURANTI, A. (1997) *Linguistic Anthropology,* Cambridge, CUP.

ENCREVE, P., & FORNEL de, M. (1983) « Le sens en pratique », *ARSS* 46, L'usage de la parole.

GADET, F. (2003) *La variation sociale en français*, Paris, Ophrys.

GUILHAUMOU, J., MESINI, B. & PELEN, J.-N. (1997) Récits de vie. « Dynamiques et autonomies des récits de vie dans le champ de l'"exclusion" ». *Cahiers de littérature orale* 41 : 91-126.

GUMPERZ, J. J., & HYMES, D. eds (1972) *Directions in Sociolinguistics : The Ethnography of Communication,* New-York, Hold, Rinehart & Winston.

HABERT, B., NAZARENKO, A. & SALEM, A. (1997) *Les linguistiques de corpus*, Paris, A. Colin.

HAMMERSLEY, M. & ATKINSON, P. (1995) *Ethnography : Principles in Practice*, Londres, Routledge.

HEATH, C. (1997) « Analysing work activities in face to face interaction using video », Silverman *ed.*

HOUTKOOP-STEENSTRA, H. (2000) *Interaction and the Standardized Survey Interview*, Cambridge, CUP.

JACOBSON, M. (2004) « Corpus oraux en linguistique de terrain », *Traitement Automatique des Langues*, 45/2 : 63-88.

JACOBSON, M. (2004) « Les archives sonores au LACITO », *Bulletin de liaison de l'AFAS* 26 (http://afas.mmsh.univ-aix.fr/bulletin/Bulletin AFAS 26.pdf).

JEFFERSON, G. (1973) « A Case of Precision Timing in Ordinary Conversation : Overlapped Tag-Positioned Address Terms in Closing Sequences », *Semiotica* 9 : 47-96.

JEFFERSON, G. *ed.* (1983) « Issues in the transcription of naturally occurring talk : caricature versus capturing pronunciational particulars, Tilburg Papers », *Language and Literature* 34.

JEFFERSON, G. (1985) « An Exercise in the Transcription and Analysis of Laughter », T. van Dijk *ed.* : 25-34.

JEFFERSON, G. (1996) « A case of transcriptional stereotyping », *Journal of Pragmatics* 26/2 : 159-170.

JORDAN, B. & HENDERSON, A. (1995) « Interaction analysis : Foundations and practice », *The Journal of the Learning Sciences* 4/1 : 39-103.

KALLMEYER, W. & SCHÜTZE, F. (1976) « Konversationsanalyse », *Studium Linguistik* 1 : 1-28.

KENNEDY, G. (1998) *An introduction to Corpus Linguistics*, Londres, Longman.

KNOBLAUCH, H., RAAB, J., SOEFFNER, H.-G. & SCHNETTLER, B. *ed.* (2006) *Video analysis*, Berne, Peter Lang.

LABOV, W. (1972) *Sociolinguistic Patterns*, Philadelphie, University of Pennsylvania Press.

LEECH, G. (1992) « The state of the art in corpus linguistics », Aijmer & Altenberg eds : 8-29

MARTIN, Ph. (1987) « Prosodic and Rhythmic Structures in French », *Linguistics* : 925-949.

MAYNARD, D. W., HOUTKOOP-STEENSTRA, H., SCHAEFFER, N. C. & ZOUWEN, J. V. D. *eds.* (2002) *Standardization and Tacit Knowledge. Interaction and Practice in the Survey Interview*, New York, John Wiley.

MITCHELL, R. G. Jr (1991) « Secrecy and disclosure in fieldwork », Shaffir, W.B., Stebbins, R.A. eds : 207-222.

MOERMAN, M. (1988) *Talking Culture : Ethnography and Conversation Analysis*, Philadelphie, University of Pennsylvania Press.

MONDADA, L. (1998) « Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête ». Actes du Colloque de Lausanne 13-14.12.1998, *Cahiers de l'ILSL* 10 : 39-68.

MONDADA, L. (2000) « Les effets théoriques des pratiques de transcription », *Linx* 42 : 131-150.

MONDADA, L. (2001) « Pour une linguistique interactionnelle », *Marges Linguistiques* 1, http://www.marges-linguistiques.com.

MONDADA, L. (2002) « Pratiques de transcription et effets de catégorisation », *Cahiers de Praxématique* 39 : 45-75.

MONDADA, L. (2003) « Observer les activités de la classe dans leur diversité : choix méthodologiques et enjeux théoriques », Perera, Nussbaum, Milian eds : 49-70.

MONDADA, L. (2006) « Video recording as the reflexive preservation-configuration of phenomenal features for analysis », Knoblauch, H., Raab, J., H.-G. Soeffner, Schnettler, B. eds.

MONDADA, L. (2006) « L'analyse de corpus dans la perspective de la linguistique interactionnelle : des analyses de cas singuliers aux analyses de collections », Condamine *ed.*

MONDADA, L. (à paraitre) « La demande d'autorisation comme moment structurant pour l'enregistrement et l'analyse des pratiques bilingues », *Tranel*, Université de Neuchâtel.

MONDADA, L. (à paraitre), « La pertinenza del dettaglio : registrazione e trascrizione di dati video per la linguistica internazionale », Bürki, E. de Stefani (à paraître).

NØLKE, H & ANDERSEN, H.L. *ed.* (2002) « Macro-syntaxe et macro-sémantique », *Actes du Colloque International d'Aarhus, mai 2001*, Berne, Peter Lang.

OCHS, E. (1979) « Transcription as theory », OCHS, E. & SCHIEFELLIN, B.B. (1979) : 43-72.

OCHS, E., & SCHIEFELLIN, B.B. *eds* (1979) *Developmental Pragmatics*, New-York, Academic Press.

OCHS, E., SCHEGLOFF, E. & THOMPSON, S. *eds.* (1996) *Interaction and Grammar*, Cambridge, CUP.

ONG, W. (1988) *Orality and Literacy*, Londres, Routledge.

PERERA, J., NUSSBAUM, L. & MILIAN, M. *ed.* (2003) *L'educacio linguistica en situacions multiculturals i multilingues,* Barcelone, ICE Universitat de Barcelona.

PLATT, J. (1983) « The development of the "participant observation" method in sociology : origin, myth, and history », *Journal of the History of the Behavioral Sciences* 19 : 379-393.

QUERE, L. *et al. ed.* (1984) *Arguments ethnométhodologiques*, Paris, Centre d'Étude des Mouvements Sociaux, EHESS.

Recherches sur le Français Parlé 5 (1984) « Pourquoi le français parlé est-il si peu étudié ? ».

Revue Française de Linguistique Appliquée (1996) 1-2, (1999) IV-1.

SACKS, H. (1972a) « An initial investigation of the usability of conversational materials for doing sociology », Sudnow *ed.* : 31-74.

SACKS, H. (1972b) « On the Analyzability of Stories by Children », Gumperz & Hymes *eds.* : 325-345.

SACKS, H. (1984) « Notes on methodology », J. M. Atkinson & J. Heritage *ed.* : 21-27.

SACKS, H. SCHEGLOFF, E.A., & JEFFERSON, G. (1974) « A simplest systematics for the organization of turn-taking for conversation », *Language* 50 : 696-735.

SACKS, H. (1992) *Lectures on Conversation* [1964-72] (2 Vol.) Oxford, Basil Blackwell.

SANKOFF, D., SANKOFF, G., LABERGE, S. & TOPHAM, M. (1976) « Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale », *Cahiers de Linguistique* 6 : 85-125.

SCARANO, A. ed. (2003) *Macro-syntaxe et pragmatique. L'analyse linguistique de l'oral*, Rome, Bulzoni editore.

SELTING, M. (1995) « Der "mögliche Satz" als interaktiv relevante syntaktische Kategorie », *Linguistische Berichte* 158 : 298-325.

SELTING, M. (1996) « On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation », *Pragmatics* 6 (3) : 371-389.

SELTING, M. (2000) « The construction of units in conversational talk », *Language in Society* 29 : 477-517.

SHAFFIR, W.B. & STEBBINS, R. A. eds. (1991) *Experiencing Fieldwork : An inside View of Qualitative Research*, Londres, Sage.

SILVERMAN, D. ed. (1997) *Qualitative Research. Theory Method and Practice*, Londres, Sage.

SINCLAIR, J. (1991) *Corpus, Concordance, Collocation*, Londres, OUP.

SINCLAIR, J. (1996) *Preliminary recommendations on corpus Typology*, Technical Report, EAGLES.

SINCLAIR, J. & COULTHARD, R. M. (1975) *Towards an Analysis of Discourse*, Londres, OUP.

« SPEECH ANNOTATION AND CORPUS TOOLS », A special issue of *Speech Communication* 33, 1-2 (2001) Steven Bird and Jonathan Harrington.

SPRADLEY, J. P. (1980) *Participant Observation*, New-York, Hold, Rinehart & Winston.

SUDNOW, D. *ed.* (1972) *Studies in Social Interaction,* New York, Free Press.

TEUBERT, W. (1999) « Corpus Linguistics. A Partisan View », *TELRI-Newsletter (Trans-European Language Ressources Infrastructures)* 8 : 4-19.

TIOUKA, A. (2005) « La question du droit autochtone sera-t-elle résolue en France ? » *Ethnies* 31-32.

TRAVERSO, V. (2002) « Transcription et traduction des interactions en langue étrangère », *Cahiers de Praxématique* 39 : 77-99.

VAN DER STRATEN (1998) « Remarques sur la transcription des enregistrements en vidéo », *CALAP* 18 : 161-177.

VAN DIJK, T. *ed.*, *Handbook of discourse Analysis,* Volume 3, New-York, Academic Press.

WELLAND, T. & PUGSLEY, L. *eds.* (2002), *Ethical Dilemmas in Qualitative Research*, Aldershot, Ashgate.

Bibliographie Patrimoine de l'oral et conservation

ARON-SCHNAPPER, D., HANET, D., DEWARTE, S. & PASQUIER, D. (1980) *Histoire orale ou archives orales ? Rapport d'activité sur la constitution d'archives orales pour l'histoire de la sécurité sociale*, Paris, Association pour l'étude de l'histoire de la Sécurité sociale.

CALLU, A. & LEMOINE, H. (2004) *Patrimoine sonore et audiovisuel français : entre archive et témoignage : guide de recherche en sciences sociales*, 7 vol., 1 CD-Rom, 1 DVD-Rom, Paris, Belin.

DESCHAMPS, F. (2001) *L'historien, l'archiviste et le magnétophone*, Paris, Comité pour l'histoire économique et financière de la France.

DOURNON, G. (1996) *Guide pour la collecte des musiques et instruments traditionnels*, Edition augmentée, Paris, UNESCO.

« Musique et son : les enjeux de l'ère numérique. Création musicale, recherche, archivage, transmission », (2002), *Culture et Recherche* 91-92.

DURAND, C. (1999-2000) *Folklore et droit d'auteur*, mémoire de DESS, Propriété intellectuelle et communication, Université Montesquieu-Bordeaux IV.

JOUTARD, P. (1979) « Historiens, à vos micros. Le document oral, une nouvelle source pour l'histoire », *L'Histoire* 12 : 106-113.

JOUTARD, P. (1983) *Ces voix qui nous parlent du passé*, Paris, Hachette.

NORA, P. dir. (1983) *Les lieux de mémoire*, Paris, Gallimard.

PROST, A. (1996) *Douze leçons sur l'histoire*, Paris, Seuil.

RICOEUR, P. (2000) *La Mémoire, l'histoire, l'oubli*, Paris, Seuil.

TOURTIER-BONAZZI (de), C. (1990) *Le témoignage oral aux archives...*, Paris, Archives nationales.

VOLDMAN, D. dir. (1992) « La Bouche de la vérité ? La recherche historique et les sources orales », *Les Cahiers de l'IHTP* 21.

VALLIERE M. (2002) *Ethnographie de la France : histoire et enjeux contemporains des approches du patrimoine ethnologique*, collection Cursus, Paris, Armand Colin.

Revues et periodiques

Bulletin de l'IHTP 1 (juin 1980) « Problèmes de méthode en histoire orale », table ronde de l'Institut d'Histoire du Temps Présent.

Bulletin de l'IHTP 75, (juin 2000), Danièle Voldman, « Le témoignage dans l'histoire du temps présent », Les Cahiers de l'IHTP (Institut d'Histoire du Temps Présent)

Sonorités, bulletin de l'AFAS, Association française des détenteurs de documents audiovisuels et sonores.

International Journal of Oral History

Aspects techniques

BONNEMASON, B., GINOUVES, V. & PERENNOU, V. (2001) *Guide d'analyse documentaire du son inédit pour la mise en place de banques de données*, Parthenay, Modal-AFAS.

CALAS, M.-F. & FONTAINE, J.-M. (1996), *La Conservation des documents sonores*, Paris, CNRS Editions.

GENDRE, C. (1999) *Enregistrement et conservation des documents sonores*, Paris, Eyrolles.

Pour la conservation des données numériques, voir les sites suivants :

Association française des détenteurs de documents audiovisuels et sonores (AFAS) :

http://afas.mmhs.univ-aix.fr/

Le compte rendu et les principales interventions du séminaire commun AFAS / BnF des 7 et 8 octobre 2004 portant sur : « La numérisation des archives sonores au service de la conservation : principes généraux et recommandations pratiques » sont consultables en ligne sur le site de l'Association.

Bibliothèque nationale de France :

http://bibnum.bnf.fr/conservation/infopreservation_fr.pdf

International Association of Sound and Audiovisual Archives :

http://www.iasa-web.org/

Voir notamment :

Bradley, K. dir., Guidelines on the production and preservation of digital objects. International Association of Sound and Audiovisual Archives. ISBN 8799030918 (voir sur le site Internet de l'Association).

Ministère de la Culture et de la Communication :

http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_04.htm

Références techniques sur la conservation :

Pickett et Lemcoe, Preservation and storage of sound recordings, Wahington, 1959.

Gilles Saint-Laurent, Care and handling of sound recordings :

http://palimpsest.stanford.edu/byauth/st-laurent/carefr.html

Cylinder, Disc and Tape Care in a Nutshell :

http://www.loc.gov/preserv/care/record.html

Équipement pour l'enregistrement de terrain :

http://www.vermontfolklifecenter.org/res_audioequip.htm.

Sur les techniques de prise de son et les matériels :

voir collections spécialisées chez Eyrolles et Dunod

Conseils sur le site de l'ASPPAC : www.asppac.com

Recommandations des Archives de France pour la gravure sur CD-R : http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2005.004.recommandations.pdf

D'autres informations pratiques sur le CD (surtout pour qui n'a pas un puissant analyseur) : http://www.mrichter.com/cdr/primer/primer.htm

LANGUES EN DANGER : LIENS UTILES

Les organismes et les institutions finançant la recherche sur les langues en danger mènent des réflexions similaires à celle qui est proposée dans ce guide. Nous donnons ci-dessous à titre informatif quelques adresses de sites Internet :

http://www.unesco.org/culture/heritage/intangible/meetings/paris_march2003.shtml

[Site de l'UNESCO et page du colloque intitulé « *Safeguarding endangered Languages* »]

http://www.mpi.nl/DOBES/INFOpages/applicants/legal-ethics-issues.html

[Site du Max-Planck Institute, et du programme DOBES pour la description des langues en danger – recommandations légales]

http://www.eva.mpg.de/lingua/files/ethics.html

[Recommandations du Département de Linguistique du Max-Planck Institute for Evolutionary Anthropology].

http://sapir.ling.yale.edu/~elf/ethics.html

[Rapport du SALSA Special Colloquium sur *Archiving Language Materials in. Web-Accessible Databases: Ethical Challenges*, 22 avril 2001. By D. H. Whalen, President, Endangered Language Fund].

http://www.hrelp.org/

[Programme de financement de recherches sur les langues en danger de la SOAS (School of Oriental and African Studies), University of London].

http://www.ogmios.org/home.htm

[Site de la Fundation for Endangered Languages, dont la dernière conférence (octobre 2004) a eu pour thème : Endangered Languages and Linguistics Rights]

GLOSSAIRE JURIDIQUE

Sauf mention contraire, les citations sont conformes au Dictionnaire comparé du droit d'auteur et du copyright

Anonymisation :

Opération par laquelle se trouve supprimé d'un ensemble de données recueillies auprès d'un individu ou d'un groupe tout lien permettant l'identification de ces derniers (voir fiche *Données personnelles et anonymisation*)

Auteur :

« Personne physique qui crée l'œuvre. Investie à titre originaire des droits d'auteur quel que soit son statut (indépendant, salarié, etc.) et les circonstances dans lesquelles elle réalise l'œuvre. Seule titulaire du droit moral de son vivant ».

Creative commons :

Le « Creative Commons » est une organisation dévouée à l'expansion des œuvres qui sont libres à la réutilisation et/ou la distribution. C'est dans ce but qu'elle a créé la licence Creative Commons. Cette licence autorise certains usages librement définis par les auteurs, parmi onze possibilités combinées autour de quatre pôles : Attribution (signature de l'auteur initial) ; Commercial (possibilité de tirer profit commercial de l'œuvre) ; No derivative works (possibilité d'intégrer tout ou partie dans une œuvre composite/ samplage) ; Share alike (obligation de rediffuser selon la même licence). Symbole général : cc.

Le mouvement Creative Commons propose des contrats-types d'offre de mise à disposition d'œuvres en ligne. Inspirées par les licences de logiciel libre et le mouvement open source, ces textes facilitent l'utilisation et la réutilisation d'œuvres (textes, photos, musique, sites Internet...). Au lieu de soumettre toute exploitation des œuvres à l'autorisation préalable des titulaires de droits, les licences Creative Commons permettent à l'auteur d'autoriser à l'avance certaines utilisations selon des conditions exprimées par lui, et d'en informer le public.

L'objectif recherché est d'encourager de manière simple et licite la circulation des œuvres, l'échange et la créativité.

Domaine public :

« Sphère d'exploitation libre et gratuite des œuvres de l'esprit qui échappent au monopole de l'auteur lorsque le monopole d'exploitation est expiré. Comprend aussi les éléments de libre parcours qui ne donnent pas prise au droit d'auteur (idées, hypothèses scientifiques…) ».

Données personnelles :

(Loi du 6 aout 2004) Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne.

Droit d'auteur :

« Droit de propriété incorporelle exclusif et opposable à tous, qui comprend l'ensemble des prérogatives morales (*droit de divulgation, droit à la paternité, droit à l'intégrité de l'œuvre, droit de repentir ou de retrait)* et patrimoniales (*droit de reproduction, droit de représentation* et *droit de suite*) dont jouit l'auteur sur son œuvre du seul fait de sa création. Dans la pratique, désigne également la rémunération perçue par l'auteur à l'occasion de l'exploitation de son œuvre ».

Droits de propriété intellectuelle :

(V. vocabulaire Cornu) « Terme générique englobant la propriété industrielle et la propriété littéraire et artistique ».

Droit moral :

« Ensemble des prérogatives extrapatrimoniales qui confèrent à l'auteur sur son œuvre, à l'artiste interprète sur sa prestation, un pouvoir de contrôle, indépendamment de la cession des droits patrimoniaux et de l'extinction du monopole. Comporte plusieurs attributs : pour l'auteur, droit de divulgation, droit à la paternité, droit à l'intégrité, droit de repentir ou de retrait ; pour l'artiste interprète, les seuls droits à l'intégrité, à la paternité. Indisponible, perpétuel, il se transmet à cause de mort aux héritiers du titulaire initial ou aux personnes désignées par lui ».

Droits patrimoniaux :

« Droit d'exploitation qui confère à l'auteur ou ses ayants- droit le pouvoir exclusif d'autoriser ou d'interdire, durant une période limitée, tout mode d'exploitation consistant en la représentation ou la reproduction d'une œuvre de l'esprit. Jouissent également d'un monopole d'exploitation : l'artiste interprète, sur sa prestation, le producteur de phonogrammes ou de vidéogrammes sur son enregistrement, l'entreprise de communication audiovisuelle sur son programme ».

Droit de divulgation :

« Attribut du droit moral de l'auteur d'une œuvre de l'esprit en vertu duquel l'auteur (ou, à sa mort, ses représentants) peut, seul, décider de porter sa création à la connaissance du public, au moment et selon les modalités qu'il détermine librement, ou, au contraire, s'y refuser. L'exercice de ce droit est le préalable nécessaire à l'exploitation patrimoniale de l'œuvre ».

Droit de repentir et de retrait :

« Attribut du droit moral permettant à un auteur, qui regrette sa décision de divulgation d'une œuvre, de remettre en cause l'exécution à venir d'un contrat d'exploitation pourtant régulièrement passé par lui. Il permet à l'auteur : soit de retirer entièrement l'œuvre du commerce (« retrait »), c'est-à-dire faire cesser l'exploitation ; soit de remanier l'œuvre (« repentir »), c'est-à-dire de changer l'objet du contrat, et cela bien que la transformation modifie pour l'exploitant les conditions et l'intérêt du contrat ».

Droit à la paternité :

« Attribut du droit moral qui permet, d'une part, à l'auteur de proclamer le lien qui l'unit à sa création et, d'autre part, à l'artiste interprète d'affirmer le lien qui l'unit à sa prestation. Positivement, droit pour le bénéficiaire d'apposer ses nom et qualités sur l'œuvre ou la prestation, de choisir l'anonymat ou la pseudonymie. Négativement, droit de s'opposer à ce qu'un tiers appose son propre nom sur l'œuvre. Parfois étendu par la jurisprudence à l'usurpation du nom (faux artistique) ».

Droit au respect de l'œuvre :

« Droit à l'intégrité. Attribut du droit moral permettant à un auteur ou un artiste interprète d'imposer à toutes personnes un devoir de respect de son œuvre ou de sa prestation, qu'il s'agisse de tiers (vandales, iconoclastes…) ou de personnes qui ont acquis des droits sur l'œuvre (cocontractant des bénéficiaires, propriétaire du support matériel de l'œuvre). Comporte d'une part le droit au respect de la forme de l'œuvre ou de la prestation qui fait échec à toute suppression, adjonction, destruction ou modification. Inclut d'autre part le droit au respect de l'esprit de l'œuvre ou de la prestation, qui permet de s'opposer à toute altération du sens ou de la destination ».

Droit à la copie privée :

« Reproduction totale ou partielle d'une œuvre de l'esprit strictement réservée à l'usage privé du copiste et non destinée à une utilisation collective. Exception légale au droit de reproduction ».

Droit de citation :

« Exception de citation. Liberté de procéder à de courts emprunts d'une œuvre de l'esprit à des fins critique, polémique, pédagogique, scientifique ou d'information, lorsque l'œuvre est divulguée et à condition d'en respecter l'intégrité, la paternité et la source. »

Droit à l'oubli :

Principe qui limite la conservation des données à caractère personnel à une durée qui n'excède pas celle nécessaire aux finalités pour lesquelles ces données ont été collectées et traitées. Souffre des exceptions quand la conservation a pour finalité des traitements à des fins historiques, statistiques ou scientifiques dans les conditions prévues pour les archives publiques.

Droit pour toute personne physique d'exiger du responsable du traitement des données que celles-ci soient effacées quand la durée de conservation est expirée. (voir fiche *Données personnelles et anonymisation*).

Original :

« Œuvre à partir de laquelle peuvent être réalisées des copies. Dans le domaine des arts graphiques et plastiques, objet matériel dans lequel est incorporée l'œuvre de l'esprit qui, émanant de la main de l'artiste ou réalisée grâce à ses instructions et sous son contrôle donne naissance à un droit de suite. Il peut s'agir d'un objet unique ou d'exemplaires effectués en tirage limité dont le nombre est fixé en fonction de la technique de reproduction et conformément aux usages de la profession ».

Valeur probatoire :

Ce qui mesure la valeur d'un mode de preuve (écrit, témoignage) comme élément de conviction. Détermine la confiance qu'il faut accorder à ce mode de preuve dans la hiérarchie des modes de preuve.

INDEX

E

F

G

I

J

L

T

U

V

X

TABLE DES MATIERES

WWW.BNF.FR WWW.CECOJI.CNRS.FR WWW.ILF.CNRS.FR WWW.TYPOLOGIE.CNRS.FR

Mise en page Pascale Racaud, Presses Universitaires d'Orléans.

Titre	***Interoperability of audio corpora : the case of the french corpora***
Type	Actes
Editeur	LREC
Année	2006
Référence	Baude, O., Jacobson, M., Tchobanov, A., Walter, R. (2006), « interoperability of audio corpora : the case of the french corpora, LREC 2006, Genova, Italy.

INTEROPERABILITY OF AUDIO CORPORA :

THE CASE OF THE FRENCH CORPORA

Olivier Baude
Laboratoire Coral**,** UFR Lettres, Langues et Sciences Humaines de l'Université d'Orléans, 10, rue de Tours, 45072 Orléans cedex 02, olivier.baude@univ-orleans.fr
Michel Jacobson
Laboratoire Lacito, UMR 7107, 7 rue Guy Môquet, Bât. D, 94801 Villejuif, jacobson@idf.ext.jussieu.fr
Atanas Tchobanov
Laboratoire MoDyCo, UMR 7114, Université Paris 10, 200, avenue de la République, 92000 Nanterre cedex, atanas.tchobanov@u-paris10.fr
Richard Walter
Laboratoire MoDyCo, UMR 7114, Université Paris 10, 200, avenue de la République, 92000 Nanterre cedex, richard.walter@u-paris10.fr

Abstract

We present here the choices which were made within the framework of three oral corpora projects : Socio-linguistics studies on Orleans (ESLO), Phonology of the Contemporary French (PFC), The Archivage corpus of the LACITO lab. This comparative presentation of three corpora of audio linguistic resources comes from a analysis about the options the project have to operate to describe them for discovery purposes and to compare the contents. The aim is to illustrate the interest to think the interoperability and the methodology of codings and the metadata. Through this step, we want to simplify the technical creation of audio corpora and thus the constitution of linguistic resources, usable by enlarged academic and industrial communities.

Today, labs, institutions, and firms working on digitizing and diffusing audio corpora are much more numerous than in the past. This can be explained by, amongst other things, the maturity and low cost of digitization techniques. The difficulty of conserving analog audio has also led many institutions to digitize their data in order to preserve it. But preserving, sharing and exchanging data requires more know-how than to simply digitize or computerize it. One must create a corpus carefully, with considerable attention to the question of how to diffuse the data.

The multiplicity of audio corpora in linguistics has led to a confusing heterogeneity of codings, formats, and methods of cataloguing, referencing and diffusion. This diversity slows down the appropriatation of the data by communities of users. Software projects relying on critical volumes of data cannot be carried out without joining different resources together. Clearly, heterogeneous data is leading to exorbitant costs for such projects. Ultimately, this will endanger the plurality of knowledge, and the transfer towards other academic as well as industrial fields.

Interoperability is the key concept which makes it possible to create a convergence of techniques and formats, resulting in a diversity of the practices. It is important to establish a dialogue and to promote exchanges and transfers of experience between the various initiatives. The specific processing of each corpus will be thus improved. With a minimum of interworking, average users of the corpus will be able to work with it independently of the tools developed by the authors. Data could be added to already existing grid of codings.

We need some standardization in order to be able to satisfy the greatest number of users and in the perennial possible way. However, interoperability requirement should not be an additional hassle for corpora designers. One should not force the producers of resources brutally to change working method; it is rather a question ofsetting up import/export facilities in common formats. The aim is to provide the user with the necessary tools to have a diachronic and synchronic glance on big corpora.

We present here the choices which were made within the framework of three oral corpora projects. Both projects rely on linguistic resources, but historically, scientifically and technologically, they differ on many aspects.

Each project will be presented by the person in charge for the software development of the corpus. Convergence and divergence points will be stressed. It is indeed important to be able to establish comparative analyses between these various audio corpora, as well from the point of view of the contents as of the techniques and methodologies.

Socio-linguistics studies on Orleans (ESLO)

The Socio-linguistics studies on Orleans (ESLO) is an investigation carried out to the end of the Sixties by British academics to didactic aims (teaching of French in the english public system of education). The investigation was into a sample representative of the urban "orléanaise" community, includes approximately 200 interviews, all referred, and more than 300 hours of sound recordings recordings (which hidden recordings, telephone conversations, public meetings, medico-teaching talks...). This corpus constitutes, by its width and its coherence, most important testimony on French spoken before 1980.

The first objective otf this project is to digitize the sound documents starting from the tape recordings and to propose of it an indexing and a first transcription in order to put the data in an organization of storage and consultation.

In parallel, an exhaustive exploitation of a subset is committed. With this experience and the analys of the first results, a recent socio-linguistics investigation, called ESLO2, is in hand on the same geographical area. The objective is to evaluate, forty away years, the social dynamics of French (the uses of the language and the judgements on its employment). This project will provide diachronic linguistic resources for the same area. The diversity of the changes is reported to the social parameters, revealing the inequality of resistances or transformation of the language, but also the typology and the dynamics of the evolutions.

Phonology of the Contemporary French (PFC)

The Phonology of the Contemporary French project (PFC) aims at constituting a vast corpus of phonology of contemporary French, through the whole francophonie and according to precise geographical, social and linguistic criteria. The corpus is composed of recordings, annotations, socio-linguistic information and codings of certainphonological phenomena (schwa, prosody, etc).

With the help of some forty researchers and fifteen PhD students, PFC tries to document and describe the pronunciation of French in its diversity and on the basis of attested usage. The main general objectives of PFC are the following:

1. test phonological and phonetic models from a synchronic and diachronic point of view, giving pride of place to intra speaker and inter-speaker variation;
2. develop a close collaboration between phonologists, experimental phoneticians and specialists in NLP;
3. allow for the conservation of a representative part of French usage across the world;
4. allow the development of better pedagogical material on the basis of usage-based data.

Launched in 1999, the first phase of the project involved the large-scale gathering of data (surveys, digitalisation and transcription of the recordings) on the basis of a uniform methodology which permits a strict comparability of the results. Simultaneously, we devised several systems of annotation and coding for various phenomena (phonological inventories, schwa, liaison) and developed a first family of tools for the partial exploration of the corpus.

The ambition of PFC project is first of all to constitute our data as one of the reference corpora of spoken French. This requires completing the network of survey points, introducing a prosodic level in our coding system and making the corpus inter-operable. To this end, we must systematise our encoding norms as well as adapt and develop various tools for the treatment and the manipulation of the data. Secondly, we are ready to exploit our data on a large scale in order not only to provide better descriptions but also to engage in the current theoretical debates between various approaches

such as stochastic models, principles and parameters, optimality theory or laboratory phonology.

The Archivage corpus of the LACITO

The Archivage corpus of the LACITO lab has been a programme of data safeguard and diffusion of languages with strong oral tradition and on reduced geographical areas. Data collection has been carried out for more than 40 years by researchers in linguistics, anthropology and ethnomusicology all over the world. The corpus is composed of recordings, annotations and transcriptions aligned to the audio.

The main and first goal of the « Archivage » project from the LACITO was to preserve the data harvested by the researchers from this laboratory. This project concerned only the primary data, those which was harvested on the field, i.e. the speech recordings, together with annotations of differents kinds like: transcriptions, translations, morpho-phonemic analysis etc. done during the fieldwork with the help of the speakers.

A digitalization policy has been defined in the laboratory to save some of the recordings (old analog tapes) which was degrading, to catalogue and to document them. The choices concerning the digitalization process where quite comparable of those chosen for the two other projects (ESLO and PFC), i.e. wav/pcm files, 44.1 Khz, 16 bits. On the other hand, for the encoding of the linguistic annotations, we have chosen to create a specific formal syntax especially build for this project but inspired by another one (the Text Encoding Initiative DTD). This syntax defines the objects and concepts the researchers used, i.e. the texts, the word lists or sentence lists, the corpora, the sentences ou breath group, the words, the morphemes, the transcriptions, the translations, etc. This syntax is implemented in an XML DTD and all the annotation's file we have done until now conforms to this syntax.

To disseminate these data, we have created a web architecture, build on the concepts defined by the Open Archive Initiative (OAI). The result of this work is one open archive, i.e. an archive harvestable with the use of the protocole OAI-MHP. This archive provide a free access to all the metadata which describes the resources of the archive in a Dublin-Core coding and in a OLAC (Open Langage Archives Community) coding. This archive at this day disseminate freely not only the metadata of some 150 documents (about 30 languages, most of them unwritten) but also the documents themselves (recordings and linguistic annotations).

This program growth so that a number of others laboratories sharing the same preoccupations came tu us adding some other users of our tools and infrastructure.

Convergence and divergence points

This work allows us to identify the actors and functions indispensables to define conservative organisation for sharing this kind of data. What this work has teach us was from different orders:

1. In technical terms this work allows us to understand that it is not possible to separate the preoccupations of safety from those of choising the formats and the encoding of the informations. The concepts of encoding without any loosing of information and the free access to the description of the encoding and to the format are primordiaux. That is for this reason that we do nont accept any proprietary formats or any formats with legal restrictions. Are ignored to all the formats and codec for audio compressions with loosing. In revanche all free, open and standardised formats are welcome. The main problem today comes from the absence of any standard for linguistic pratices.
2. In organisational terms this work allows us to identify the two main missings wich are the porduction side (The digitisation of the old tapes represent today the only alternative for the preservation of the date. The struggle against the loosing of the colected data is a struggle against the time becose the amount of data to treat is so big and the 'moyens' we have usualy in our kind of laboratories so small that we progress with ant's foot step). The second side we miss is the long time preservation. The disitization is not an end and do not 'garantie' the preservation of the data. What it does is just facilitate the preservation in the mesure of thta duplicate the data can be done without any loosing. We can't think about preservation outside of an institution in which preservation is is its main goal. Only the organisation, the technology 'veille' on new supports, encoding, formats, etc. can avoid the obsolescence of the data.

Interoperability with a architecture of cataloguing

Sound recording standards being defined better, interworking between the three projects is already practically established. On the other hand, the descriptors of resources are not the same ones according to projects. The coding of the metadata varies because the names of the fields and their possible values were selected for the needs for the investigation. Needs will be indeed different for sociolinguists, ground linguists, dialectologists, phonologists, specialists in TAL or linguistic engineering. The linguistic analyses and thus their codings can also vary. Arranged orthographical transcriptions are the minimum, but linguists will certainly be happy with phonetic transcriptions, morpho-syntactic cuttings, lematisations, etc.

Because of this diversity, interworking between these corpora is currently quasi impossible. Facing this failure, we decided not to change methods and working tools. This is currently impossible, because part of the investigations are already finished or are under development. Instead we work on the definition of common export format and to adopt a common architecture of cataloguing. Exporting data in a standardized common format allows the exploitation with generic tools. The choice of XML in this context is quite natural. On the otherhand there is not standard for the structure of XML documents (schemas or DTD) for the annotation of speech. The chapter devoted to the transcription in the TEI is at the same time too poor and unsuited to the existing practices. And existing metadata standards (DCMI, OLAC, IMDI, MARC, etc.) do not cover completely the needs of oral corpora management.

Interoperability with a tank of data

The results of these projects (PFC, ESLO and Archivage) direct us towards the definition of an organization of storage and cataloguing of the data centered, around the concept of tank of data, in the definition given by the OAI, i.e. moissonable, with a strict separation of the data and metadata. This tank, in the course of construction, will accept only data and metadata in formats and codings, open and free of right, whose definition will also have to be stored in the tank.

This objective is dictated by the need to maintain the data in the medium and long term. Another key concept for the conservation and the mutualisation of the data is the separation of the logic structure and the typographical structure. The first represents the abstract or scientific of the physical structure; the second represents the usual forms of consultation for this type of data. This separation makes it possible to carry out calculations on the data with terms which the linguists can handle; at the same time, it makes it possible to make evolve easily the various representations of information (on paper, screen, multi-media, adaptation to a handicap, etc).

The CatCod Initiative

This comparative presentation of three corpora of audio linguistic resources aims to illustrate the interest to think up stream the interoperability and the methodology of codings and the metadata. Through this step, we want to simplify the technical creation of audio corpora and thus the constitution of linguistic resources, usable byenlarged academic and industrial communities.

The vocation of the CatCod initiative is to organize in France the community of the oral around a common practice of coding and cataloguing for the oral corpora. The purpose of the CatCod group which gathers participants of various laboratories and French universities is to describe the current practices. The result of this work should be to define a standard and its formalization in order to propose them to the Text Encoding Initiative (TEI) consortium.

References

Baude O., Jacobson M., Tchobanov A., Walter R. (2005), Interopérabilité des corpus sonores. In *Phonological Variation : The Case of French, Bulletin PFC 5* (http://www.projet-pfc.net).

Bray, T., Paoli, J. et Sperberg-McQueen, C. M. (Eds) (1998). Extensible Markup Language (XML) Version 1.0, In *Word Wide Web Consortium*.

Jacobson, M. (2004). Corpus oraux en linguistique de terrain. *Traitement automatique des langues*, 45/2, pp. 63-88.

Jacobson, M., Lowe, J. B. & Michailovsky, B. (2001). Linguistic documents synchronizing sound and text, *Speech Communication*, vol. 33, n° 1-2, pp. 79-96.

Sperberg-McQueen, C. M., & Burnard, L. (1994), *TEI Guidelines for Electronic Text Encoding and Interchange (P3)*, Chicago and Oxford : ACH/ACL/ALLC Text Encoding Initiative.

Délégation générale à la langue française et aux langues de France, Ministère de la culture et de la communication (2005), *Guide des bonnes pratiques pour la constitution, l'exploitation, la diffusion et la conservation des corpus oraux*, Paris : CNRS éditions (2006).

The Open Archives InitiativeOAI: http://www.openarchives.org

The Dublin Core Metadata Initiative DCMI: http://dublincore.org

The Open Language Archives Community OLAC: http://www.language-archives.org

The CatCod initiative: http://icar.ens-lsh.fr/wiki/index.php

The Phonology of the Contemporary French (PFC) : http://www.projet-pfc.net

The Archivage corpus of the LACITO lab: http://lacito.vjf.cnrs.fr/archivage

Titre	***« Corpus oraux les bonnes pratiques d'une communauté scientifique »***
Type	Actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification, juin 2006
Editeur	Presses universitaires de Toulouse
Année	2007
Référence	Baude, O. (2007) « Corpus oraux les bonnes pratiques d'une communauté scientifique »**,** in actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification, juin 2006, Presses universitaires de Toulouse, 61-66.

CORPUS ORAUX : LES *BONNES PRATIQUES* D'UNE COMMUNAUTE SCIENTIFIQUE

Olivier Baude (*pour le groupe de travail*[1])
CORAL – Université d'Orléans EA 3850 / Délégation Générale à la Langue Française et aux Langues de France

SOMMAIRE

0. Introduction

Les problèmes juridiques liés à la diffusion des corpus oraux ont été l'occasion d'une démarche originale adoptée par une communauté scientifique ouverte à un travail pluridisciplinaire. Cette démarche a comporté plusieurs étapes. Une lecture croisée des textes juridiques par les linguistes et les juristes a permis de repérer les problèmes. Les chercheurs ont ensuite accepté d'expliciter leurs pratiques au regard de la législation. Cette étape fondée sur la réflexivité a permis d'élaborer des propositions pour de bonnes pratiques partagées par la communauté scientifique et de repérer des aspects juridiques qui posent des difficultés dans l'état actuel du droit.

Ce travail s'est concrétisé par la rédaction de l'ouvrage *Corpus oraux, guide des bonnes pratiques 2006*[2]. Rédigé par un groupe de travail constitué de linguistes, juristes, informaticiens et conservateurs, cet ouvrage a pour vocation explicite, d'éclairer la démarche des chercheurs, de repérer les problèmes et les solutions juridiques et de favoriser l'émergence de pratiques communes pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux.

Le résultat de ce travail interdisciplinaire ouvre les portes d'une réflexion sur les pratiques des chercheurs en sciences sociales et leurs relations aux données, à l'heure de l'exploitation et de la diffusion en masse de celles-ci.

1. Contextes pour une diffusion de la recherche

[1] L'ouvrage *Corpus oraux, guide des bonnes pratiques 2006* a été rédigé par O. Baude (-coordinateur- Coral et DGLFLF), C. Blanche-Benveniste (EPHE et université de Provence), M-F Calas (DMF), P. Capeau (Université de Poitiers), P. Cordereix (BnF), L. Goury (Cnrs-Celia), M. Jacobson (Cnrs-lacito), I. de Lamberterie (Cnrs-Cecoji) C. Marchello-Nizia (Cnrs-Ilf et ENS-LSH-Lyon) et Lorenza Mondada Cnrs-Icar et Université Lyon 2). Cet article est une présentation des travaux de ce groupe, la paternité de la grande majorité du contenu en revient donc à l'ensemble des auteurs à qui le rédacteur exprime toute sa gratitude.

[2] *Corpus oraux, guide des bonnes pratiques 2006*, CNRS éditions, Paris.

1.1 La linguistique de corpus et l'oral

Depuis plus de 30 ans le domaine de la linguistique de corpus s'est considérablement développé autour des corpus écrits, aussi bien en ce qui concerne la masse des données disponibles que l'élaboration d'outils de traitement automatique de celles-ci. La situation est totalement différente pour les corpus oraux[3]. Pourtant, les toutes nouvelles technologies en matière de stockage, de diffusion mais aussi d'exploitation des enregistrements sonores, couplées aux outils (transcriptions synchronisées sur le signal, annotations, etc.) ouvrent des perspectives prometteuses pour les études sur les corpus de langues parlées. De nombreux corpus ont été constitués ou sont en cours de constitution et leur diffusion posent des problèmes juridiques et éthiques que la communauté scientifique doit prendre en charge. Pourquoi et comment?

1.2 Une politique de diffusion

Depuis 1982 et la loi pour la recherche et le développement technologique en France[4], la diffusion des résultats fait partie des missions des chercheurs. Plus récemment, la déclaration de Berlin signée par la plupart des Directeurs Généraux des Établissements Publics à caractère Scientifique et Technologique (EPST) le 22 octobre 2003 plaide pour la constitution de bases de connaissances en libre accès[5]. Enfin, les programmes de numérisation patrimoniale comprennent un volet de valorisation des ressources numérisées (cf. texte de Lund de 2001 prônant la mise en place des standards d'interopérabilité).

1.3 Les initiatives de mutualisation

Cette dernière notion de standards d'interopérabilité se retrouve dans différentes initiatives internationales (TEI, groupe de travail ISO TC37 SC4 pour la gestion des ressources linguistiques, protocole d'échange OAI, norme ANSI/NISO Z39.50, projet Open Language Archive Community, etc.) ainsi que dans des choix techniques (utilisation du langage de balisage XML par exemple). Dans le même cadre de valorisation de la recherche et de mutualisation des ressources, le Cnrs s'est doté, en 2005, d'une direction de l'information scientifique, et développait un an plus tard des centres de ressources numériques.

Dans le même temps des laboratoires de recherche lançaient différentes initiatives pour la diffusion et l'accessibilité des corpus oraux (Base Clapi du laboratoire Icar[6], projet Corpus Oraux de l'EPML 50[7], programme Archivage du Lacito[8], constitution de grands corpus disponibles comme le projet Phonologie du Français contemporain[9], C-oral-Rom[10], etc.).

1.4 Le Guide des bonnes pratiques

C'est dans ce contexte que la Délégation générale à la langue française (direction du ministère de la culture) et le CNRS ont constitué un groupe de travail pluridisciplinaire qui a pour mission de favoriser la collecte et l'exploitation de corpus oraux.

Ce groupe de travail comporte des linguistes experts et des chercheurs de "terrain" porteurs de projets actuels, des représentants des fédérations de laboratoire du CNRS, des juristes, des représentants des grands organismes de conservation sous la tutelle du Ministère de la Culture et des juristes de ces institutions. L'objectif premier était de permettre un travail en commun sur un objet scientifique, de favoriser sa conservation et surtout sa diffusion (diffusion auprès de différentes équipes de recherche mais aussi auprès d'un public

---

[3] Pour plus de commodités et selon l'usage nous utiliserons les termes *corpus oraux* comme termes génériques définissant des collections ordonnées d'enregistrements de productions linguistiques orales et multimodales.

[4] Art 5 de la Loi n°82-610 du 15 juillet 1982 modifiée d'orientation et de programmation pour la recherche et le développement technologique de la France, aujourd'hui art. L 111-1 du code de la recherche. *JO* du 16-07-1982, p. 2273 et ss.

5 Corpus oraux, Guide des bonnes pratiques op. cité, p 36.

6 Clapi-Icar http://clapi.univ-lyon2.fr

7 EPML50 (ex Asila)

8 Archivage du Lacito : http://lacito.vjf.cnrs.fr/archivage/index_fr.html

9 PFC http://www.projet-pfc.net

10 C-Oral-Rom 2005.

plus large). Or, il est très vite apparut que les aspects juridiques étaient les premiers obstacles à la diffusion de l'oral transcrit (qui est propriétaire de quoi? Qui est responsable de la diffusion? Quel sont les autorisations à recueillir? Qu'en est-il du droit d'auteur?, etc.). Enfin, ce travail sur les aspects juridiques a très vite été lié à une réflexion sur l'éthique du chercheur et l'occasion d'une démarche réflexive sur ses méthodes.

Dans un premier temps, le groupe de travail s'est orienté vers l'élaboration par la communauté scientifique "de bonnes pratiques" avec les contraintes suivantes: premièrement il n'existe pas de réponses juridiques simples à l'exploitation de l'oral et à la transcription des données et deuxièmement les solutions passent systématiquement par un travail réflexif sur la démarche du chercheur, seul moyen pour qualifier le statut des enregistrements et les objets exploités. Les "bonnes pratiques " consistent donc à clarifier les questions juridiques, mais aussi - et c'est là un point fondamental - à porter une réflexion sur le travail scientifique des linguistes dans le respect d'une éthique validée par la communauté scientifique.

2. Aspects juridiques

D'une façon très schématique la réponse aux questions juridiques consiste à définir le statut juridique de l'objet "corpus" par ses conditions d'élaboration et sa composition, afin de procéder à la gestion contractuelle des droits des personnes concernées et de définir les responsabilités de ceux qui vont intervenir dans la vie du corpus (créateurs, hébergeurs, diffuseurs,…).

2.1 Définition de l'objet

Pour des raisons épistémologiques et techniques, la forme des corpus oraux est relativement complexe. Dans la majorité des cas les corpus oraux sont constitués :

- d'enregistrements (analogiques ou numériques) qui en cas de supports analogiques ont une durée de vie très courte avec une perte de qualité lors des migrations,
- de données contextuelles sur les locuteurs et la situation d'enquête qui peuvent être en partie des données personnelles (nom propre, profession, adresse, lieu,…),
- de transcriptions (sous la forme de fichiers indépendants ou permettant une synchronisation sur le signal ; transcription phonétique, orthographique, multilinéaire, etc.),
- d'annotations "secondaires" (informations sur les conditions de production des énoncés, précisions sur les phénomènes sonores tels que les rires et les bruits),
- d'annotations enrichies (étiquetage morphologique, syntaxique, annotations prosodiques pragmatiques, …),
- d'une documentation.

2.2 Domaines juridiques concernés

Pour définir le statut juridique de l'objet scientifique "corpus oral" et les droits des personnes concernées il faut tout d'abord connaitre les conditions d'élaboration du corpus et de ses différentes composantes. Il s'agit ensuite de définir si le corpus est constitué d'informations du domaine public et/ou s'il est le produit d'une ou plusieurs créations intellectuelles susceptibles d'être protégées par le droit d'auteur. Il convient enfin de vérifier si le corpus contient des données personnelles qu'il faudra alors traiter. Ces statuts juridiques déterminés et les droits qui en découlent connus, il convient de s'enquérir des modalités de la gestion contractuelle de ces droits et de savoir si les titulaires de ceux-ci se sont prononcés sur les conditions de mise à disposition et de réutilisation des corpus - en apportant par exemple, leur consentement d'une manière formelle.

2.3 Diffusion scientifique et droit d'auteur

Seule une explicitation rigoureuse de la démarche du chercheur permet de savoir si un corpus est protégé par le droit d'auteur. Si tel est le cas, quels sont ces droits?

Il convient de distinguer les droits patrimoniaux des prérogatives du droit moral. Les droits patrimoniaux se résument en un droit exclusif au profit de l'auteur (ou des titulaires) ou des ayants droits (bénéficiaires d'une cession, héritiers...) d'autoriser ou interdire la reproduction ou la communication au public de l'oeuvre protégée. Quant aux prérogatives du droit moral, toujours attachées à la personne physique créatrice de l'œuvre protégée, elles sont au nombre de quatre : le droit de divulgation, le droit de repentir et de retrait, le droit à la

paternité et le droit au respect de l'œuvre. En réalité, il existe une possibilité intermédiaire où les corpus protégés par le droit d'auteur peuvent être mis en libre accès dans le cadre d'une licence accordée par les titulaires de droits autorisant l'utilisation et l'exploitation des résultats (c'est le cas des Creatives commons). Sans être dans le domaine public, ces corpus sont – de par la volonté de leurs créateurs – libres d'accès et d'utilisation. Néanmoins, si les créateurs peuvent renoncer à exercer leurs droits patrimoniaux, il ne leur est pas possible de renoncer à leur droit moral qui reste imprescriptible.

3. Eléments pour de *bonnes pratiques*

3.1 Expliciter la démarche du chercheur

Les objectifs scientifiques, liés à la constitution, à l'exploitation, à la conservation et à la diffusion des corpus oraux sont très diversifiés, et le respect de ceux-ci, ainsi que leur hétérogénéité, impliquent que soit reconnu la diversité des démarches qui peuvent être adoptées par les chercheurs et par les utilisateurs ultérieurs de ces corpus.

Le Guide des bonnes pratiques n'a pas vocation à contraindre cette démarche en prescrivant une méthodologie type, mais souhaite fournir toutes les informations nécessaires au repérage des points juridiques et éthiques « sensibles ». Seule l'identification précise et détaillée des éléments de la situation en jeu et notamment de la forme des données et de leurs supports, des pratiques de terrain, mais aussi des différentes étapes du traitement, permet d'apporter à la fois des éléments de réponses juridiques correspondant à la situation, et une évaluation des « risques » éventuels. Enfin, une analyse réflexive sur la démarche liée à la constitution et aux traitements des corpus oraux est le premier élément de l'élaboration d'une éthique reconnue par l'ensemble d'une communauté scientifique.

3.2 Le recueil de consentement

Le geste éthique le plus classique de la démarche du chercheur-enquêteur est le recueil de consentement du témoin. En réalité cette pratique est peu maîtrisée et souvent réduite à un formulaire de demande d'autorisation qui évoque en une phrase " le cadre d'un programme de recherche". Or sans informations préalables précises la demande d'autorisation n'a pas d'objet ni de sens. Pour que cette autorisation soit pertinente il conviendrait de concevoir le recueil d'un consentement "éclairé" qui démontre que le signataire est informé des finalités de la recherche et des conséquences à son égard d'une participation au projet.

Dans le cadre du recueil de données et notamment d'enregistrement pour des corpus oraux, le consentement devrait tenir compte de l'adéquation au destinataire (les informations fournies, pour être comprises doivent être adaptées aux compétences de compréhension du destinataire), et de l'explicitation des finalités de l'enquête (qui toutefois ne doivent pas renforcer le paradoxe de l'observateur en pointant l'objet de l'observation).

De plus, les explications sur le projet scientifique, doivent être complétées par des informations précises comme par exemple : les responsables de l'enquête et leur affiliation institutionnelle, ainsi que les financeurs ; une adresse de contact, les personnes qui auront accès aux données et qui travailleront sur elles, la façon dont les données seront anonymisées, le fait que les données seront transcrites selon des conventions particulières, la façon dont les données seront archivées une fois l'enquête terminée, les modalités d'accès aux informations relatives au projet et concernant tout particulièrement les données/analyses faisant référence à la personne (possibilité d'accès aux fichiers et informations concernant tout particulièrement la personne), les droits de la personne, notamment le droit de rétractation, les risques éventuels ainsi que les retombées positives, morales ou matérielles, de l'étude.

Enfin, le consentement devra préciser l'objet de la demande : les actions effectuées par les chercheurs dans le cadre du projet, les formats et les conditions de l'enregistrement, les conditions de diffusion des données et des résultats, les contextes de diffusion des données et des résultats. Il est à noter que les formes de l'autorisation ne sont pas imposées par le législateur et qu'une demande orale enregistrée peut être valide et même parfois indispensable.

Sur le plan juridique, la collecte de données sensibles sans recueil de consentement est possible à la condition particulière que les données soient anonymisées dans un très bref délais. La procédure d'anonymisation est également très importante pour obtenir l'accord des témoins a fortiori dans le cas d'une diffusion des données primaires.

3.3 L'anonymisation

Les pratiques actuelles des chercheurs en terme d'anonymisation se réduisent la plupart du temps à une opération de masquage d'un nom propre, d'une adresse ou d'un numéro de téléphone. Afin de vérifier la validité de ces pratiques et d'en définir les modalités, il convient de reposer avec précision la question légale qui est celle de l'impossibilité d'identifier des personnes. En effet, l'objectif est de protéger la vie privée des personnes enregistrées en dépersonnalisant les données, ce qui a amené le législateur à ne pas réduire cette identification à la simple présence de données nominatives.

Ainsi, si techniquement l'anonymisation consiste au remplacement ou au codage des données sensibles par des éléments neutres selon les supports concernés (remplacement par un blanc ou un pseudo à l'écrit, par un bip dans les fichiers sons et par floutage des visages sur les enregistrements vidéos), il serait erroné de penser que cette solution ne demande pas une expertise plus approfondie des risques d'exploitation d'éléments "dénommant".

3.4 Structure du corpus

Il existe d'autres possibilités que l'anonymisation par cryptage. Celles-ci reposent sur des limitations techniques prévues par la structure du corpus. La loi québécoise « concernant le cadre juridique des technologies de l'information » propose de protéger l'anonymat non pas en modifiant les données, mais en limitant les possibilités de recherche, voire en les adaptant à la personne qui consulte la base selon des critères bien précis (sa profession, une autorisation, sa présence dans le fichier, etc.)

Cette dernière perspective offre pour la constitution et l'exploitation de corpus oraux la possibilité de faire coïncider les obligations légales avec les nécessités du travail de recherche. Toute donnée étant potentiellement sensible, une anonymisation systématique s'avère de plus en plus complexe ; elle peut même mettre en danger l'intérêt de certaines recherches. En effet, des détails concernant les personnes comme par exemple le nom, ou le lieu d'habitation peuvent constituer un élément important du corpus, ainsi que des résultats que l'on peut en tirer. C'est pourquoi la possibilité de ménager des niveaux d'accès selon des critères stricts (ex : chercheur ou non, présence d'autorisation, but de la consultation etc.) semble une alternative efficace. Il existe d'autres procédés à inventer. En effet, l'article 11-2 de la nouvelle loi ouvre la possibilité de faire certifier des techniques nouvelles par la CNIL.

4. Conclusion

La démarche originale présentée ici a plusieurs intérêts. Outre le fait qu'elle offre les garanties d'une diffusion des corpus pour la recherche et pour d'autres finalités, elle impose une posture éthique aux collecteurs, utilisateurs et diffuseurs de corpus. C'est aussi l'occasion de porter un regard réflexif sur des pratiques et sur une démarche scientifique peu souvent explicitée. Enfin, il s'agit de permettre la constitution de corpus dont la mutualisation est la première étape d'une démarche scientifique rigoureuse qui ouvre les portes de l'analyse et de l'interprétation.

BIBLIOGRAPHIE

Baude, O. (2006), *Corpus oraux, Guides des bonnes pratiques, 2006*, CNRS-Editions et Presses Universitaires d'Orléans.

Baude, O. (2004) « Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques », *Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble) : 7-11.

Biber, D. (1985) *Variations across spoken and written language*, Cambridge, CUP.

Biber, D. (1999) *Longman Grammar of Spoken and Written English*, Londres, Longman.

Bilger, M. dir. (2000) « Linguistique sur corpus, études et réflexions », *Cahiers de l'université de Perpignan*, Perpignan, Presses universitaires.

Bilger, M. ed. (2000) *Corpus, Méthodologie et applications linguistiques*, Paris, Champion.

Blanche-Benveniste, Cl. & Jeanjean, C. (1987) *Le français parlé : transcription et édition*, Paris, Didier-Erudition.

Callu, A. & Lemoine, H. (2004) *Patrimoine sonore et audiovisuel français : entre archive et témoignage : guide de recherche en sciences sociales*, 7 vol., 1 CD-Rom, 1 DVD-Rom, Paris, Belin.

Cameron, D., Frazer, E., Harvey, P., Rampton, M. & Richardson, K. (1991) *Researching Language : Issues of Power and Method*, London, Routledge.

Condamine, A. *ed.* (2006) *Sémantique et corpus*, Paris, Hermes.

Cresti, E. & Moneglia, M. ed. (2005) *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/Philadelphie, Benjamins.

Cribier, F. & Feller, E. (2003) *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »* rapport présenté à Madame la Ministre déléguée à la Recherche et aux nouvelles technologies,

dactylogr. 2 vol.

et http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf

Encreve, P., & Fornel de, M. (1983) « Le sens en pratique », *ARSS* 46, L'usage de la parole.

Habert, B., Nazarenko, A. & Salem, A. (1997) *Les linguistiques de corpus*, Paris, A. Colin.

Jacobson, M. (2004) « Corpus oraux en linguistique de terrain », *Traitement Automatique des Langues*, 45/2 : 63-88.

Jacobson, M. (2004) « Les archives sonores au LACITO », *Bulletin de liaison de l'AFAS* 26 (http://afas.mmsh.univ-aix.fr/bulletin/Bulletin AFAS 26.pdf).

Joutard, P. (1979) « Historiens, à vos micros. Le document oral, une nouvelle source pour l'histoire », *L'Histoire* 12 : 106-113.

Kennedy, G. (1998) *An introduction to Corpus Linguistics*, Londres, Longman.

Labov, W. (1972) *Sociolinguistic Patterns*, Philadelphie, University of Pennsylvania Press.

Leech, G. (1992) « The state of the art in corpus linguistics », Aijmer & Altenberg eds : 8-29

Mondada, L. (1998) « Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête ». Actes du Colloque de Lausanne 13-14.12.1998, *Cahiers de l'ILSL* 10 : 39-68.

Mondada, L. (2006) « Video recording as the reflexive preservation-configuration of phenomenal features for analysis », Knoblauch, H., Raab, J., H.-G. Soeffner, Schnettler, B. eds.

Mondada, L. (à paraitre) « La demande d'autorisation comme moment structurant pour l'enregistrement et l'analyse des pratiques bilingues », *Tranel*, Université de Neuchâtel.

Quéré, L. *et al. ed*. (1984) *Arguments ethnométhodologiques*, Paris, Centre d'Étude des Mouvements Sociaux, EHESS.

Recherches sur le Français Parlé 5 (1984) « Pourquoi le français parlé est-il si peu étudié ? ».

Revue Française de Linguistique Appliquée (1996) 1-2, (1999) IV-1.

Sacks, H. (1984) « Notes on methodology », J. M. Atkinson & J. Heritage *ed.* : 21-27.

Shaffir, W.B. & Stebbins, R. A. eds. (1991) *Experiencing Fieldwork : An inside View of Qualitative Research*, Londres, Sage.

Silverman, D. ed. (1997) *Qualitative Research. Theory Method and Practice*, Londres, Sage.
Sinclair, J. (1991) *Corpus, Concordance, Collocation*, Londres, OUP.
Sinclair, J. (1996) *Preliminary recommendations on corpus Typology*, Technical Report, Eagles.
Sinclair, J. & Coulthard, R. M. (1975) *Towards an Analysis of Discourse*, Londres, OUP.
« Speech Annotation and Corpus Tools », A special issue of *Speech Communication* 33, 1-2 (2001) Steven Bird and Jonathan Harrington.
Welland, T. & Pugsley, L. *eds.* (2002), *Ethical Dilemmas in Qualitative Research*, Aldershot, Ashgate.

Titre	***« Constituer et exploiter un grand corpus oral, choix et enjeux théoriques le cas des Eslos »***
Type	Actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification juin 2006
Editeur	Presses universitaires de Toulouse
Année	2007
Référence	Abouda, L., **Baude, O.** (2007) « Constituer et exploiter un grand corpus oral, choix et enjeux théoriques le cas des Eslos »**,** in actes du colloque *Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification juin 2006, Presses universitaires de Toulouse: 161-168.

CONSTITUER ET EXPLOITER UN GRAND CORPUS ORAL : CHOIX ET ENJEUX THEORIQUES. LE CAS DES ESLO[1]

Lotfi ABOUDA – Olivier BAUDE

CORAL – Université d'Orléans

SOMMAIRE

0. Introduction

Contrairement aux corpus de français écrit il n'existe pas de grand corpus de français oral disponible pour l'ensemble de la communauté scientifique. La présentation du projet de diffusion des corpus ESLO (Enquêtes Socio-Linguistique à Orléans) à l'ensemble des acteurs de la recherche, qu'ils viennent des sciences cognitives ou de l'anthropologie, de la physique (traitement du signal) ou des études de genre (gender studies), de la dictionnairique ou du TAL, est l'occasion d'interroger les raisons complexes d'une telle situation.

Un regard épistémologique, notamment sur la place des données en linguistique, apporte des éléments d'explication qu'il convient de prendre en compte avant de proposer des pistes pour une méthodologie favorable à l'exploitation de grand corpus oraux. Les principaux choix théoriques et techniques opérés lors de l'exploitation scientifique (numérisation, transcription, annotation, diffusion, analyses) du corpus ESLO1, vu comme étape liminaire à ESLO2, répondent à un objectif précis : participer à la réflexion sur l'évolution des modèles et des méthodes de constitution et d'exploitation des corpus oraux destinés à des finalités linguistiques.

1. Quelques considérations sur le linguiste et ses corpus

Sans oser une présentation épistémologique de la place des corpus en linguistique, nous souhaitons présenter quelques considérations sur l'usage des corpus en linguistique qui sont à l'origine du projet de constitution, d'exploitation et de diffusion des corpus ESLO.

[1] Enquête Socio-Linguistique à Orléans.

1.1 Données attestées et situées VS masse de données

La linguistique connaît actuellement un bouleversement méthodologique amorcé il y a plus de 30 ans. Les possibilités offertes par le traitement automatique du langage et notamment les techniques d'exploitation des documents numériques ont permis des développements théoriques fondées sur l'exploitation de corpus, mettant ceux-ci aux centres de la description et de l'analyse linguistique.
Une ambiguïté demeure cependant. En effet, les corpus étaient utilisés bien avant le développement du domaine du TAL. Travailler sur corpus consistait alors à considérer l'objet d'étude comme une collection ordonnée de productions attestées et situées. Cette définition de l'objet impliquait une démarche empirique de description des faits qui s'opposait à une démarche hypothético-déductive fondée sur l'intuition du chercheur. La méthodologie de travail sur corpus était donc un acte scientifique fort et fondateur de certains domaines (sociolinguistique, analyse de la conversation, ethnolinguistique, etc.) centrés sur la conception de "corpus de langue parlée".
Depuis les années 1980, la linguistique de corpus s'est définie autour de grands corpus de langue écrite traités informatiquement comme l'ont décrit Kennedy (Kennedy, 1998) pour l'anglais et Habert pour le français (Habert et al., 1997).

Ainsi les possibilités offertes par le traitement informatique de masse de données sont devenues l'atout principal de la linguistique de corpus. Toutefois trois questions, selon nous centrales, se trouvent biaisées dans ce contexte :

- le corpus est souvent constitué de productions très normés (romans, articles de presses, textes officiels) dont le traitement requiert une standardisation (orthographe, étiquetage, etc.) ;
- le corpus est souvent considéré comme représentatif d'une hétérogénéité des pratiques de part le simple fait qu'il constitue une masse de données ;
- la disponibilité de vaste corpus (FRANTEXT) permet dans de nombreux travaux d'éluder la question pourtant centrale de la constitution du corpus comme première étape d'une théorie linguistique.

1.2 Place des corpus oraux

Si la linguiste du corpus s'est massivement développée, force est de constater que la linguistique dispose de peu de corpus oraux. C'est un paramètre qu'il est facile d'expliquer : la tradition littérale est continue depuis l'Antiquité quand les modes de conservation du son ont moins d'un siècle et demi d'existence. Mais ce n'est pas l'unique raison. L'oral s'accommode beaucoup moins d'un traitement excluant les variations. L'écrit est normalisé par sa présentation même en chaîne de caractères. Il est le produit d'une transcription déjà effectuée, que la source en soit assignée au mental ou au signal. Avant tout retravail par les instruments et les outils du TAL, une homogénéisation de la présentation et des formes a été accomplie à divers niveaux : orthographe, découpe des mots et des phrases, ponctuation…
La recherche en intelligence artificielle a été facilitée, quand elle se donnait les langues pour objet, par la saisie d'énoncés écrits, avec pour conséquence l'élaboration de techniques et d'approches dont l'extension à des corpus oraux (enregistrements, transcriptions phonétiques) était malaisée. On est en présence d'un cas d'école concernant l'ajustement réciproque des données et des outils qui pâtit de l'extension des processus à de nouvelles catégories d'objets.

Les problèmes de l'extension des méthodes éprouvées sur des corpus scripturaux à des corpus oraux se situent sur différents plans :

- insuffisance des corpus oraux, que ce soit en termes quantitatifs de disponibilité globale, ou qualitatifs de fiabilité scientifique ou de prétraitement ;

- dissymétrie des champs d'application de l'enquête (opposition des études linguistiques de terrain - field linguistics), orientées vers les langues sans tradition écrite, et des linguistiques de bureau (armchair linguistics), centrées sur les textes de référence et l'écrit -, les départements informatiques étant plus souvent confrontés à celles-ci pour lesquelles existe de surcroît une forte demande des industries de la langue ;
- parcellisation des enquêtes et des standards retenus pour la collecte, la conservation et la codification : ainsi, le très important travail d'archive orale entrepris dans les deux dernières décennies par les historiens et les sociologues a souvent été entrepris sans finalité externe qui aurait pu assurer une exploitation ouverte des fonds ;
- faible exigence de prescription : les corpus sont constitués sur des objectifs *ad hoc*, ciblant leur finalité en fonction d'objectifs circonscrits, par exemple la reconnaissance vocale ou la fouille de textes ;
- pratiques lacunaires de catalogage et de description des ressources : la bibliothéconomie des archives sonores reste aujourd'hui encore balbutiante et c'est un chantier international où il importe que soient formulées des propositions pour tout ce qui a trait à la standardisation des produits, à l'indexation et à la consultation (représentativité des éléments de catalogage par rapport aux contenus en fonction de pertinences multiples).

1.3 Corpus disponibles *VS* corpus fantômes

Nous l'avons précisé, si la linguistique de corpus s'est considérablement développée, ce n'est pas pour autant, et le fait mérite d'être grassement souligné, que les corpus eux même soient disponibles. En effet, à l'exception notamment de FRANTEXT et du *Monde*, les corpus sont toujours évoqués dans les travaux, mais ne sont que très rarement diffusés. Ils jalonnent les articles et les thèses comme les fantômes hantent les couloirs et les tours : toujours évoqués comme preuve mais n'apparaissant à nul autre qu'à celui qui en parle.
Cette situation ne mérite pas d'être caricaturée et on peut esquisser une typologie des corpus en fonction d'un critère de disponibilité :

- Certains corpus ont été constitués dans le cadre d'une recherche précise et n'ont de pertinence que pour celle-ci. Les conditions de collecte ou le travail très spécifique d'annotation ne permet pas la diffusion de ces données.
- D'autres corpus ne sont pas disponibles par volonté des chercheurs qui souhaitent garder une priorité scientifique sur un travail de collecte couteux et laborieux.
- Enfin il existe des corpus conçus comme des bases de données qui prennent le statut de corpus de référence par le simple fait que ce sont les seuls disponibles. Ces corpus sont alors utilisés simplement … parce qu'ils sont là.

Nous ne pouvons terminer cette courte typologie sans évoquer les corpus totalement fantômes qui fondent certains travaux sans qu'aucune information ne précise les raisons de l'absence de l'accès aux données, pourtant seule garantie d'un travail scientifique en principe ouvert à la falsification.
Le programme ESLO se situe résolument dans une démarche scientifique pour laquelle un corpus non disponible n'existe pas.

En bref, la linguistique de corpus a dans un premier temps peu pris en charge le domaine de la langue parlée et des données situées. Cependant les technologies récentes permettant de numériser le son et d'avoir une synchronisation temporelle entre le signal et une ou des transcriptions ainsi que les initiatives de normalisation de structuration des corpus (*TEI*), des métadonnées (*Dublin core* et *Olac* par exemple) et des données liées ouvrent de nouvelles perspectives pour la linguistique de corpus. Toutefois il n'existe pas

actuellement de grand corpus français de langue parlée disponible pour la communauté scientifique. Le projet ESLO (cf. infra) souhaite répondre à cette demande.

2. Les corpus ESLO : de la collecte à l'exploitation d'un corpus oral

2.1 ESLO1, un corpus à reconstruire

L'Enquête Socio-Linguistique à Orléans (désormais : ESLO1) a été conduite en 1968 par des universitaires britanniques avec une visée didactique : l'enseignement du français langue étrangère dans le système public d'éducation anglais. Il s'agit d'un vaste corpus estimé à plus de 300 heures (environ 4 500 000 mots).

Elle comprend environ 200 interviews, toutes référencées (caractérisation sociologique des témoins, identification de l'enquêteur, date et lieu de passation de l'entretien), mais aussi une gamme d'enregistrements variés (conversations téléphoniques, réunions publiques, transactions commerciales, repas de famille, entretiens médico-pédagogiques, etc.). Certains des enquêtés ont ainsi été enregistré dans des situations très différentes. ESLO 1 couvre l'ensemble des catégories socioprofessionnelles, hommes et femmes, avec plusieurs locuteurs originaires de différentes régions. C'est un échantillon des formats de la communication, des tâches linguistiques, des types de discours selon une approche essentiellement dialogique. Ce corpus représente, par son ampleur, sa rigueur et sa cohérence, le plus important témoignage disponible sur le français parlé avant 1980. Si les fins de sa constitution étaient linguistiques, ESLO 1 est un témoignage unique sur les jugements concernant mai 68 vu de la province ou sur les représentations collectives de la cité à cette époque.

Le Coral (Centre orléanais de recherche en anthropologie et linguistique) a réussi à récupérer en 1993 l'ensemble de documents originaux composés des bandes magnétiques, un catalogue dactylographié, quelques centaines de feuillets de transcription manuscrites (d'une qualité inégale) et les fiches d'identification des locuteurs.

L'opportunité offerte par la numérisation des originaux arrivés en fin de vie a permis au Coral de consacrer un projet à la conservation et à la valorisation du corpus. L'opération de numérisation n'était pas en l'occurrence anodine ; c'est une véritable reconstruction du corpus et sa transformation en un nouvel objet scientifique qui a été opérée. Les documents sonores ont été récolligés et complétés (la conservation avait été défectueuse), numérisés à partir des enregistrements et une indexation et un premier catalogage informatisé a pu être réalisé. Parallèlement, l'exploitation exhaustive d'un sous-ensemble a été entreprise au point de rencontre de données linguistiques variationnistes et cognitives (description d'une tâche). L'étape suivante consiste à transcrire et baliser l'intégralité du corpus.

L'enjeu de cette reconstruction n'est pas neutre. Il s'agit d'établir des principes ayant valeur de normalisation afin de mettre l'ensemble des données à la disposition de la communauté scientifique dans un format qui en permette une exploitation fiable, optimale et intensive, y compris pour des applications industrielles après sélection des contenus.

2.2 ESLO 2, un corpus à anticiper

En partant des acquis d'ESLO 1, une nouvelle enquête, dénommée ESLO 2, a été mise en chantier par le CORAL. Il s'agit, à quarante années de distance, de constituer un corpus comparable dans le produit attendu et dans les modalités de la collecte : l'objectif a été fixé à 400 heures environ de documents sonores qui totaliseraient approximativement 6 000 000 de mots. Réunis, ESLO 1 et ESLO 2 formeront une collection de 700 heures d'enregistrement, soit plus de 10 000 000 de mots, ce qui est considéré aujourd'hui comme une valeur repère pour les investigations projetées.

ESLO 2 a été conçu pour préfigurer la référence attendue dans un domaine qui en est encore à se structurer et dans lequel se manifeste de manière récurrente une demande de définition pour un format standardisé de *collecte*, de *conservation*, de *traitement* et d'*analyse* :

- la *collecte* sur le terrain est première, non seulement dans ses aspects techniques, aujourd'hui bien maîtrisés, mais dans la définition du profil de l'échantillon représentatif et dans la problématisation des interactions entre les témoins et les enquêteurs ;
- la *conservation*, qui inclut la préservation des supports, l'indexation des contenus et l'accessibilité (c'est-à-dire la protection) des données, conditionne le partage des sources à des fins d'étude scientifique ou didactique ;
- le *traitement*, en lien étroit avec le développement des matériels et des langages informatiques, suppose la maîtrise d'une chaîne d'opérations, depuis la conversion numérique des enregistrements jusqu'à une transcription balisée et ouverte à l'ensemble des interrogations pertinentes pour les demandes du linguiste, du sociologue ou des décideurs, des didacticiens voire du grand public ;
- l'*analyse* constitue l'épreuve des théories (et des logiciels) puisqu'elle compare les formalisations et les opérations et qu'elle valide ou infirme les hypothèses en prenant argument de leur compatibilité aux faits.

Les acquis en matière de conservation, de traitement et d'analyse seront reportés sur ESLO1 comme le requiert la comparabilité attendue.

3. Choix pour la mutualisation et l'interopérabilité d'un grand corpus oral

Quels sont les choix et les enjeux contenues dans l'objectif de mutualisation et d'interopérabilité d'un grand corpus oral de type ESLO ? Nous nous bornerons ici à présenter une démarche suffisamment générale qui interroge l'exploitation des corpus en sciences humaines.

3.1 Rôle des métadonnées pour l'interopérabilité

Les corpus constituent des ressources numériques sur lesquelles se fondent la majorité des travaux en linguistique actuellement sans que la question de la forme de ceux-ci et des limites qui bornent cette collection ordonnée soit toujours clairement résolue.
Un corpus est constitué de données brutes et/ou annotés (Véronis 2000, Habert et Fuchs 2004). Dans le cas des corpus oraux, les enregistrements de la parole constituent les données primaires, la transcription et les autres annotations éventuelles représentent des données secondaires. L'ensemble de ces données sont décrites par des métadonnées chargées de documenter le corpus.
Ce sont ces dernières informations, particulièrement importantes pour rendre une ressource disponible mais aussi pour expliciter les critères de sélection et d'organisation des données et donc des bornes du corpus, qui manquent souvent dans les corpus disponibles. Or il s'agit ni plus ni moins de poser ainsi la question de la représentativité du corpus. C'est en effet l'explicitation des bornes du corpus (conditions de productions, de réception, contexte des usages, informations sociologiques sur les producteurs, genre, etc.) qui permet de juger de la représentativité de corpus qui du statut d'échantillon de la langue passe très souvent à celui de corpus de référence (même si ce statut référentiel est implicite) sans aucun regard réflexif sur la forme de celui-ci.

Dans le cas du corpus ESLO1, l'équipe a souhaité conserver le travail de catalogage et de documentation déjà anticipé par les auteurs du corpus. La démarche actuelle et validée par la communauté consiste à utiliser des métadonnées *Dublin Core* et les extensions préconisées par le programme OLAC. Ce jeu d'étiquettes permet des opérations de

catalogage tout à fait satisfaisantes en termes de description d'une ressource qu'on souhaite répertorier pour la rendre accessible.

Cependant cette procédure n'est pas suffisante pour documenter un corpus conçu comme un réservoir qui doit permettre à un chercheur de construire son propre corpus répondant aux exigences de sa recherche. Dans le cadre de cette extraction/construction, les informations permettant de borner le corpus doivent répondre à une granularité très fine. Dans le cas du corpus des ESLO, nous avons déjà pointé l'importance accordée aux informations sur les locuteurs, la situation de collecte et l'échantillonnage dont le but était de constituer un corpus représentatif.
Ce travail méthodologique permet de considérer que la collecte a correspondu à un travail méthodologique rigoureux, source de données représentatives et qu'ainsi le chercheur est sûr d'être confronté à des données pertinentes. Il convient d'aller plus loin et de considérer qu'un corpus doit contenir une riche documentation sur les données mais aussi sur les contextes de production de ces données. Ces contextes concernent aussi bien les données sur les locuteurs et la situation de collecte que l'explicitation de la démarche du chercheur.

Quelles sont les possibilités pour mettre à disposition un corpus qui contient en lui-même les informations sur ses bornes constitutives ? Le standard XML offre des éléments de réponse.
Techniquement ce standard sépare la représentation physique et logique des documents (les données et les métadonnées). Tout document XML comporte donc l'identification des éléments possibles et leurs relations possibles (Définition de Type de Document) et les données identifiées selon cette DTD. C'est alors la notion même de données brutes qui est redéfinie. Ainsi la TEI rend obligatoire la constitution d'un header (en-tête) en début de corpus qui recense les informations sur le contexte de production des données. Cependant le chapitre de la TEI consacré à l'oral est actuellement beaucoup trop succinct pour permettre une véritable normalisation de cette démarche.

Il y a donc un enjeu à considérer les métadonnées comme des éléments de description des données au sens linguistique de celle-ci et non simplement en termes de documentation de ressources. Les métadonnées doivent permettre d'expliciter la démarche du chercheur en proposant une description fine des ses choix théoriques "encapsulés" dans des choix techniques. Les opérations de transcription sont en ce sens un exemple particulièrement éclairant.

3.2 Des données exploitables : le cas de la transcription

La difficulté la plus importante rencontrée par les initiateurs d'ESLO 1 a été l'ampleur de la tâche de transcription. Sur ce point aussi, et même principalement, l'avancée technologique bouleverse l'objet scientifique.

Depuis quelques années, alors que la manipulation du son numérique devenait très aisée (capacité de stockage, rapidité d'accès, débit suffisant pour une transmission en réseau…), des logiciels permettent la synchronisation du son et de la transcription (*Praat*, *Transcriber*, *Winpitch*, *soundedit*, etc.).
Ces innovations ont des répercussions méthodologiques importantes sur le travail du linguiste. En effet, avec des transcriptions alignées sur le signal sonore, l'oral devient physiquement l'objet d'étude et est systématiquement disponible en même temps que la transcription. Le retour aux données peut alors être systématique, ce qui est de nature à faciliter les procédures de vérification, étape essentielle du travail scientifique, malheureusement souvent rendue impraticable de par l'inaccessibilité des corpus.

Parallèlement, la synchronisation, qui permet l'annotation de segments temporels, offre une base de référence pour de la multi annotation et donc de la multi transcription. On peut

concevoir, pour un même segment, une multitude de transcriptions, opérées dans des cadres théoriques distincts et/ou avec des granularités différentes, dont chacune répond à un besoin scientifique spécifique. Ici, la transcription n'est plus la vérité d'un chercheur (au mieux) ou d'un transcripteur, elle devient cumulative.

Face à l'ampleur de la tâche, les choix pour la transcription d'ESLO1 ont été fondés sur la volonté de mettre à disposition une transcription de l'intégralité du corpus le plus rapidement possible sans que celle-ci n'implique une théorie linguistique très déterminée (même si toute transcription est une formalisation impliquant une théorie).
Cette première transcription est conçue comme une transcription de base avec un simple statut d'outil de navigation au sein du corpus sonore et de repérage de phénomènes selon une granularité grossière. L'outil sélectionné a été *Transcriber* pour sa simplicité d'utilisation, sa robustesse face à des fichiers longs, et sa sortie en un format de fichier XML qui nous a semblé être une garantie d'interopérabilité.

Les conventions de transcriptions ont donc été réduites au minimum : La segmentation se fait sur une unité intuitive de type "groupe de souffle et/ou unité syntaxique pertinente. Le tour de parole a été défini par les changements de locuteurs uniquement, les pauses indiquées automatiquement par leur durée (précision du centième de seconde).

3.3 Corpus mutualisé pour des analyses multi-domaines : le test d'eslomelette

Le groupe du CORAL qui travaille sur ESLO est composé de chercheurs dont les sensibilités théoriques sont diverses, et les domaines de compétence en linguistique assez variés, allant de l'épistémologie à la syntaxe, en passant par le TAL, la phonologie, la pragmatique, la sociolinguistique, etc.
Cette diversité théorique est vue comme un atout majeur pour ce projet, dont l'objectif premier est la mise à disposition d'un corpus, laquelle ne peut pas être conçue sans multi-opérabilité.
Or, quelle meilleure garantie pour un corpus qui se veut disponible pour la communauté linguistique dans son ensemble que d'être conçu par des chercheurs dont les centres d'intérêts sont assez divers ?
Pour mettre à l'épreuve cette première piste, l'équipe a choisi un échantillon de ce corpus sur lequel elle a décidé d'opérer toutes les étapes du travail linguistique, de l'identification du corpus jusqu'à l'analyse, opérée dans des domaines divers (syntaxe, pragmatique, lexique, phonologie…), en passant par l'annotation, qui comporte elle-même différentes phases (transcription, annotation, métadonnées). L'équipe a donc constitué un sous-corpus composé des 90 réponses à la question "comment-faites-vous une omelette ?". Les couples questions réponses ont été transcrits selon les conventions testées. Ces fichiers de transcription ainsi que l'ensemble des méta données constituent une collection de documents intégrés à une base de donnée XML native. Une interface (xquery) a été réalisée dans le cadre du projet GRICO[2] et du CRDO[3] après un travail conjoint d'informaticiens spécialisés dans la gestion de corpus oraux et les chercheurs en linguistique de l'équipe.

Cette première expérience est intéressante à plus d'un égard. D'abord, elle permet de voir sur un petit échantillon toutes les erreurs (d'annotation, de structuration), qu'il est encore temps d'éviter pour la totalité du corpus. Ensuite, elle précise l'utilité d'un corpus situé.

Pour ne donner ici qu'un exemple, on peut citer une recherche en pragmatique opérée par des membres de l'équipe. L'analyse pragmatique de la question de l'omelette montre qu'à partir de la question zéro, telle qu'elle figure dans le questionnaire – i.e. «

[2] Groupe de Recherche sur l'Interopérabilité des Corpus Oraux. Michel Jacobson (Lacito-CRDO) et Richard Walter (Modyco).
[3] Centre de Ressources pour la Description de l'Oral. http://crdo.vjf.cnrs.fr:8080/exist/crdo/

« Comment est-ce qu'on fait une omelette ? Pourriez-vous m'expliquer comment on fait ? » - les enquêteurs, visiblement gênés par la question, développent toutes sortes de modalisation. Après le relevé systématique des différentes marques de distanciation vis-à-vis de la question, qui se distinguent en fait en deux groupes, à savoir d'une part les « stratégies de justification » (évocation des écarts culturels entre la France et l'Angleterre, contrôle de la qualité du son, etc.) et, d'autre part, les « stratégies d'atténuation » (emploi du conditionnel, l'enchâssement de la question, l'emploi de l'atténuation autonymique, etc.), on peut se poser une série de questions que la nature et la structuration du corpus permettent, et qui auraient été tout simplement impossibles ailleurs. Par exemple, y a-t-il dans ce dégradé de modalisation une variable sociologique ? Autrement dit, l'enquêteur utilise-t-il plus ou moins de modalisation selon le profil de l'enquêté (son âge, son sexe, son niveau sur l'échelle AM) ? Ce type de questions, combien intéressante d'un point de vue linguistique, est tout simplement impossible dans d'autres corpus. Autre interrogation : y a-t-il une variable individuelle ? Autrement dit, les enquêteurs se distinguent-ils les uns des autres vis-à-vis de leur relation avec la question ? Et, d'ailleurs, un enquêteur quelconque utilise-t-il au fil du temps que dure l'enquête (en l'occurrence presque un an) les mêmes stratégies de modalisation ? Toutes ces interrogations, et bien d'autres, auraient été fastidieuses ailleurs : ici, elles sont non seulement possibles, grâce à la fois aux outils du TAL et à la disponibilité des métadonnées, mais en plus utiles : par exemple, les interrogations naïves qui viennent d'être évoquées permettent de poser des questions cruciales, concernant la réflexivité de l'enquête, son statut, son degré de figement et d'interaction, etc. Derrière ces questions, il s'agit ni plus ni moins que de poser la question de la pertinence et de la validité de données non situées.

Cet enjeu n'est pas restreint à la pragmatique et à la sociolinguistique, le travail entrepris par des chercheurs aux objectifs très différents permet de tester les possibilités de réappropriation de contraintes méthodologiques (par exemple, la normalisation recherchée par le chercheur en TAL est-elle compatible avec le linguiste variationniste ?).

Conclusion

Les enjeux inhérents à l'exploitation d'un grand corpus oral ne se résument pas à des choix techniques imposés par les outils du traitement automatique du langage et de la linguistique de corpus. L'exemple des corpus d'ESLO ne mettent en évidence que ce qu'on savait déjà : "on ne peut dissocier l'accumulation des données et la critique de leur constitution".

Cette évidence interroge la linguistique sur la constitution même de son objet mais aussi l'ensemble des sciences sociales sur l'exploitation de la masse de données. La réponse passe nécessairement par la maîtrise de la totalité de la chaîne : de la collecte des données à leurs organisation à des fins d'analyses variées.

Bibliographie

Abouda L., 2004, « Deux types d'imparfait atténuatif », *Langue française*, 142, p. 58-74,

Baude O., Jacobson M., Tchobanov A., Walter R., à paraître, « Interopérabilité des corpus sonores : le cas des corpus en français », *Colloque international Phonological variation : the case of French*, 25-27 août 2005, Tromso.

Baude O, 2004 : « Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques *», Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble) : 7-11.

Baude O. (ed), 2006, *Corpus oraux. Guide des bonnes pratiques 2006*, Paris, Cnrs éditions – Orléans, PUO.

Bergounioux G., 1992, « Les enquêtes de terrain en France », *Langue française*, 93, p. 3-21.

Bergounioux G., Baraduc J., Dumont C., 1992, « L'Etude socio-linguistique sur Orléans (1966-1991), 25 ans d'histoire d'un corpus », *Langue française*, 93, p. 74-93.

Blanche-Benveniste C., Jeanjean C., 1987, *Le français parlé, transcription et édition*, Paris, Didier érudition.

Blanc M., Biggs P., 1971, « L'enquête sociolinguistique sur le français parlé à Orléans », *Le français dans le monde*, 85, p. 16-25.

Delais-Roussarie E. et Durand J. (ed), 2003, *Corpus et variation en phonologie du français, méthodes et analyses*, Toulouse, PUM.

EAGLES, 1996, Preliminary Recommendations on Spoken Texts, EAG-TCWG-SPT/P, Pise, Consiglio Nazionale delle Ricerche, Instituto di Linguistica Computazionale.

Habert, B., et al., 1997, *Les linguistiques de corpus*, Paris, Armand Colin.

Habert, B., Fuchs, C., (2004) "Introduction le traitement automatique des langues : des modèles aux ressources", *le français moderne traitement automatique et ressources numérisées pour le français*, pp 1-13.

Mertens P., 2002 « Les corpus de français parlé ELICOP : consultation et exploitation », in Binon, J., Piet; Elen, J., Mertens, P., Sercu, Lies (eds) (2002) *Tableaux Vivants*, Opstellen over taal-en-onderwijs aangeboden aan Mark Debrock. Leuven, Universitare Pers.

Pierrel, J-M, ed, (2000) *Ingénierie des langues*, Paris, Hermès sciences.

Rastier, F. (2004). « Enjeux épistémologiques de la linguistique de corpus ». *Texto !* [en ligne], juin 2004. Rubrique Dits et inédits. Disponible sur : <http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html>.

Sinclair J., 1996, « Preliminary recommendations on corpus Typology », Technical Report, Eagles.

« Speech Annotation And Corpus Tools », A special issue of Speech Communication Volume 33, numbers 1-2, 2001, Edited by Steven Bird and Jonathan Harrington.

Véronis, J., (2000), "Annotation automatique de corpus : panorama et état de la technique". In Pierrel J-M (ed), pp 111-130.

Wynne M., 2005, *Developing Linguistic Corpora : a Guide to Good Practice*, AHDS, http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm. Visité le 01 juillet 2006.

Titre	***Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux***
Type	Article
Editeur	De Werelt
Année	2007
Référence	Baude, O., (2007) « Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux », in *Corpus : état des lieux et perspectives.*RFLA XII-1, éditions De Werelt, Amsterdam, 71-84.

Aspects juridiques et éthiques
de la conservation et de la diffusion des corpus oraux

Olivier Baude.
Université d'Orléans / DGLFLF

Résumé : *La numérisation des corpus de données sonores et multimodales ouvre de larges perspectives pour les sciences du langage. Toutefois, la conservation et l'exploitation de ces corpus oraux posent de nouveaux problèmes éthiques et juridiques que la communauté scientifique doit prendre en compte. Cet article présente les résultats d'un groupe de travail interdisciplinaire qui a rédigé un* Guide des bonnes pratiques *pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux.*

Abstract : *The digitalization of spoken language corpora opens large perspectives for linguistics. However, the archiving and the exploitation of these spoken corpora raise new ethical and legal problems that the scientific community must take into account. This article presents the results of an interdisciplinary working group which wrote a* Guide of good practices *for the constitution, the exploitation, the archiving and the diffusion of spoken language corpora.*

0. Introduction

Si depuis plus de 30 ans le domaine de la linguistique de corpus s'est considérablement développé autour des corpus écrits (masse de données disponibles, élaboration d'outils de traitement automatique de celles-ci), la situation est totalement différente pour les corpus oraux[1] et ce n'est que très récemment que les questions de conservation et de diffusion de ceux-ci se posent. Les toutes nouvelles technologies en matière de stockage, de diffusion mais aussi d'exploitation des enregistrements sonores, couplées aux outils de traitement automatique du langage (transcriptions synchronisées sur le signal, annotations, etc.) ouvrent des perspectives prometteuses pour les études sur les corpus de langues parlées et devraient permettre de renouveler les sciences du langage en modifiant le champ de la linguistique de corpus. Toutefois cette situation ne va pas sans poser de nombreuses questions techniques, méthodologiques et théoriques mais aussi juridiques et éthiques. Ce sont principalement ces deux derniers aspects qui seront

[1] Pour plus de commodités et selon l'usage nous utiliserons les termes *corpus oraux* comme termes génériques définissant des collections ordonnées d'enregistrements de productions linguistiques orales et multimodales.

abordés dans cet article consacré au témoignage d'un travail interdisciplinaire qui s'est concrétisé par la rédaction d'un *Guide des bonnes pratiques* pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux[2]. Il s'agira donc, après avoir présenté le contexte - scientifique, épistémologique et politique - d'une telle initiative, de repérer concrètement les problèmes juridiques et éthiques afin d'éclairer la démarche des chercheurs et de proposer des solutions sous la forme d'un travail réflexif anticipant l'élaboration de bonnes pratiques.

1. Contextes

1.1. Objets émergents et émergence des problèmes

Il y a bien entendu des raisons épistémologiques propres au champ de la linguistique pour expliquer le retard pris en France pour la constitution et l'exploitation de corpus de données orales (Blanche-Benveniste 1987 ; Bergounioux 1992 ; Baude 2004). Cependant, et parallèlement à une modification du champ qui rend ses lettres de noblesse à la description de données attestées, le développement des nouvelles technologies facilitant la manipulation de masses de données sonores va transformer radicalement l'objet « données orales ».

En effet, si les outils informatiques permettent depuis peu de numériser et conserver en masse des enregistrements de la voix, de les diffuser facilement par les réseaux internet et autres, et de manipuler aisément des fichiers sons volumineux, les outils de reconnaissance et de synthèse de la parole mais aussi la transcription synchronisée sur le signal offrent aux corpus oraux l'accès aux outils de traitement automatique des corpus écrits créant ainsi un nouvel objet scientifique du fait même de cette synchronisation son/texte. Ce sont ces nouveaux objets, qui mêlent l'écrit et l'oral, qui ont fait émerger avec insistance des problèmes de conservation et de diffusion des corpus pour deux raisons principales.

- La présence de données primaires telle que la « voix » dans un corpus rappelle qu'il n'y a pas d'oralité sans « locuteur ». Il devient alors beaucoup plus difficile d'oublier que ce locuteur est une personne dont il convient pour des raisons tant juridiques qu'éthiques de respecter les droits.

- La linguistique de l'oral est constituée de différents domaines ; or si celui de la parole est méthodologiquement moins éloigné des corpus écrits, ce n'est pas le cas du français parlé, de la sociolinguistique, de la langue en interaction, de l'ethnolinguistique, etc., où le locuteur-témoin est particulièrement présent, par les données sociologiques souvent intégrées aux corpus, mais aussi par la réflexion sur les conditions de collecte des enregistrements (et sur l'impact de la diffusion des analyses pour la communauté observée).

---

[2] L'ouvrage *Corpus oraux, guide des bonnes pratiques 2006* est le résultat des travaux d'un groupe de réflexion réuni autour d'Isabelle de Lamberterie. Il a été rédigé par O. Baude coordinateur (Université d'Orléans/DGLFLF), C. Blanche-Benveniste (EPHE/Université de Provence), M-F. Calas (DMF), P. Cappeau (Université de Poitiers), P. Cordereix (BnF), L. Goury (CNRS), M. Jacobson (CNRS), I. de Lamberterie (CNRS) C. Marchello-Nizia (ILF/ENS-LSH-Lyon) et Lorenza Mondada, (Université Lyon 2).
Cet article est une présentation des travaux de ce groupe, la paternité de la grande majorité du contenu en revient donc à l'ensemble des auteurs à qui le rédacteur exprime toute sa gratitude.

1.2. Définition de l'objet : composition d'un corpus oral

Pour ces raisons épistémologiques et techniques, la forme des corpus oraux est relativement complexe. Dans la majorité des cas les corpus oraux sont constitués :

- d'enregistrements (analogiques ou numériques) et qui en cas de supports analogiques ont une durée de vie très courte avec une perte de qualité lors des migrations, de productions de locuteurs ;
- de données contextuelles sur les locuteurs et la situation d'enquête qui peuvent être en partie des données personnelles (nom propre, profession, adresse, lieu,…) ;
- de transcriptions primaires (sous la forme de fichiers indépendants ou permettant une synchronisation sur le signal ; transcription phonétique, orthographique, multilinéaire, etc.) ;
- d'annotations secondaires (informations sur les conditions de production des énoncés, précisions sur les phénomènes sonores tels que les rires et les bruits) ;
- d'annotations enrichies (étiquetage morphologique, syntaxique, annotations prosodiques pragmatiques)

1.3. Cadres politiques de diffusion de la recherche

La complexité du contenu des corpus oraux et notamment la présence de données personnelles mais aussi le lourd investissement que représente la constitution comme le traitement (ne seraient-ce que les fastidieuses opérations de transcription) et l'enrichissement du corpus ont longtemps été invoqués pour expliquer l'absence de disponibilité et de diffusion de ceux-ci.

Récemment ces contraintes se sont confrontées aux cadres politiques de la diffusion de la recherche. Ainsi, depuis 1982 et la loi pour la recherche et le développement technologique en France[3], la diffusion des résultats fait partie des missions des chercheurs. De plus nous sommes dans une dynamique d'échange de l'information scientifique comme le confirme la déclaration de Berlin signée par la plupart des Directeurs Généraux des Établissements Publics à caractère Scientifique et Technologique (EPST) le 22 octobre 2003 sur *le Libre Accès à la Connaissance en Sciences exactes, Sciences de la vie, Sciences humaines et sociales*, dont l'objectif est de promouvoir Internet « comme instrument fonctionnel au service d'une base de connaissance globale de la pensée humaine » (*Corpus oraux*, 36). Enfin, les programmes de numérisation patrimoniale comprennent un volet de valorisation des ressources numérisées (cf. texte de Lund de 2001 prônant la mise en place des standards d'interopérabilité).

Cette dernière notion de standards d'interopérabilité se retrouvent dans différentes initiatives internationales (TEI, groupe de travail ISO TC37 SC4 pour la gestion des ressources linguistiques, protocole d'échange norme ANSI/NISO Z39.50, Open Language Archive Community, etc.) ainsi que dans des choix techniques (utilisation du langage de balisage XML par exemple). Dans le même cadre de valorisation de la recherche et de mutualisation des ressources, le CNRS se dotait en 2005 d'une direction de l'information scientifique et développait un an plus tard des centres de ressources numériques.

[3] Art 5 de la Loi n°82-610 du 15 juillet 1982 modifiée d'orientation et de programmation pour la recherche et le développement technologique de la France, aujourd'hui art. L 111-1 du code de la recherche. JO du 16-07-1982, pp. 2273 et ss.

Conjointement à ce cadre politique, des laboratoires de recherche lançaient également différentes initiatives pour la diffusion et l'accessibilité des corpus oraux (Base CLAPI du laboratoire ICAR[4], projet Corpus Oraux de l'EPML 50[5], programme Archivage du LACITO[6], constitution de grands corpus disponibles comme par exemple, le projet Phonologie du Français contemporain[7]).

Face à cette volonté de conserver et diffuser des corpus oraux par des communautés scientifiques aux pratiques très différentes, de nombreuses questions notamment juridiques constituent un frein : Quelles autorisations doit-on faire signer à des locuteurs enregistrés dans le but de constituer un corpus de transcriptions exploitées et diffusées ? Est-ce qu'il faut que le corpus respecte l'anonymat ? A qui appartient le corpus ? Qui peut le vendre ? Qui est responsable du traitement et de la diffusion du corpus ? Quelles données (et métadonnées) peut-on conserver et diffuser sous la forme de fichiers informatiques ? Qui peut décider de la diffusion ou de la non diffusion des corpus ? Que doit on faire pour assurer la conservation des enregistrements (et qui peut assurer cette conservation sur du long terme) ? etc.
Les réponses à ces questions existent maintenant sous la forme d'un *Guide des bonnes pratiques* dont l'originalité de la démarche mérite d'être soulignée.

1.4. Une initiative d'élaboration de « bonnes pratiques »

Dans le cadre de son programme « corpus de la parole », la DGLFLF (Délégation Générale à la Langue Française et aux Langues de France, Ministère de la Culture) a créé un groupe de travail pluridisciplinaire regroupant des juristes, linguistes, conservateurs, informaticiens et représentants des fédérations de recherche en linguistique du CNRS. Ce groupe de travail s'est donné comme objectifs de recenser les pratiques actuelles et de définir en priorité les contraintes méthodologiques et théoriques liées à la recherche, de diffuser une synthèse sur la législation existante, d'établir des recommandations, et, le cas échéant, en cas de vide ou de flou, de formuler des propositions pour l'élaboration de normes et règles juridiques (notamment européennes) (*Corpus oraux*, 20).

Au fil du travail les propositions ont pris la forme d'un *Guide des bonnes pratiques pour la constitution, l'exploitation, la diffusion et la conservation des corpus oraux*.

2. Les aspects juridiques des corpus oraux en cinq questions[8]

D'une façon très schématique la réponse aux questions juridiques consiste à définir le statut juridique de l'objet « corpus » par ses conditions d'élaboration et sa composition, afin de procéder à la gestion contractuelle des droits des personnes concernées et définir les responsabilités de ceux qui vont intervenir dans la vie du corpus (créateurs, hébergeurs, diffuseurs…). Ces aspects juridiques peuvent se réduire à cinq questions essentielles.

Q1 : *Quel est le statut juridique de l'objet « corpus oral » ?*

---

[4] CLAPI-ICAR http://clapi.univ-lyon2.fr.
[5] EPML50 (ex Asila).
[6] Archivage du LACITO : http://lacito.vjf.cnrs.fr/archivage/index_fr.html
[7] PFC http://www.projet-pfc.net
[8] Cette partie de l'article reprend les éléments principaux rédigés par I. de Lamberterie pour le *Guide*.

Pour définir le statut juridique de l'objet scientifique « corpus oral » et les droits des personnes concernées, il faut tout d'abord connaitre les conditions d'élaboration du corpus et de ses différentes composantes. Il s'agit ensuite de définir si le corpus est constitué d'informations du domaine public et/ou s'il est le produit d'une ou plusieurs créations intellectuelles susceptibles d'être protégées par le droit d'auteur. Il convient enfin de vérifier si le corpus contient des données personnelles qu'il faudra alors traiter. Ces statuts juridiques déterminés et les droits qui en découlent connus, il convient de s'enquérir des modalités de la gestion contractuelle de ces droits et de savoir si les titulaires de ceux-ci se sont prononcés sur les conditions de mise à disposition et de réutilisation des corpus en apportant par exemple leur consentement d'une manière formelle.

Q2 : *Qui est le propriétaire d'un corpus ?*

Le statut juridique d'un corpus dépend des données qu'il contient. Soit il dépend du domaine public, et est alors libre d'exploitation pour tout le monde, soit il appartient à un *auteur*, et est alors protégé.

Le domaine public recouvre non seulement les idées de liberté d'accès et de gratuité d'utilisation des données, mais aussi la possibilité pour chacun de les exploiter. Il se caractérise, en outre, par l'absence de monopole puisque les informations qui tombent dans le domaine public deviennent de facto des « choses communes » (*Corpus oraux*, 38).

Bien entendu la distinction n'est pas définitive, ainsi les œuvres protégées par le droit de la propriété intellectuelle, notamment par le droit d'auteur ou les brevets, finissent par entrer dans le domaine public. En France, c'est 70 ans après la mort de son auteur qu'une œuvre tombe dans le domaine public, même si à l'expiration de ce délai, d'autres types de protection peuvent subsister sur les œuvres de l'esprit : les droits patrimoniaux d'une part, les attributs imprescriptibles du droit moral d'autre part. En conséquence certaines œuvres (ou éléments de ces œuvres) tombées dans le domaine public, peuvent encore être protégées par les dispositions qui concernent le droit moral.

Les corpus oraux, ou tout au moins les enregistrements et les transcriptions relèvent-ils des considérations évoquées ? Comme nous allons le voir, la réponse n'est pas simple et les enregistrements linguistiques suscitent ainsi de nombreuses hésitations. Il est en effet complexe de savoir si le contenu d'une langue, son expression phonique, font ou non partie du domaine public. En outre, ce fonds commun est-il universel ou bien seulement commun à une petite communauté ? Aujourd'hui, il fait de plus en plus l'objet de revendications identitaires qui soulèvent de nouvelles interrogations (*Corpus Oraux*, 38).

La protection du droit d'auteur demande également une définition précise des objets et des opérations de traitement de ces objets. En effet pour qu'un corpus soit protégé par le droit d'auteur il faut que trois conditions soient remplies:

Il faut en premier lieu qu'il corresponde à l'exigence d'une activité créatrice : un travail de compilation d'informations n'est pas protégé en soi. Pour être protégé, il est par ailleurs indispensable que le corpus ait une forme définie. Ce qui est protégé ce n'est pas le contenu du corpus mais son enveloppe, son architecture. Enfin, la forme du corpus doit répondre à la condition d'être originale. L'auteur est en principe la (ou les) personne(s) physique(s) sous le nom de laquelle (ou desquelles)

l'œuvre est divulguée. Le travail scientifique suppose l'intervention de nombreux acteurs dont bon nombre sont susceptibles de revendiquer la qualité d'auteur sur les résultats de la recherche. Certains corpus oraux, comme les autres produits de la recherche, peuvent rester l'œuvre d'un auteur unique, alors que d'autres peuvent être l'œuvre de plusieurs auteurs. Dans le cas de pluralité d'auteurs, le droit distingue les œuvres de collaboration des œuvres collectives. Pour les premières, chaque co-auteur dispose des mêmes prérogatives. D'autres œuvres - telles que les bases de données ou les dictionnaires - peuvent être qualifiées d'œuvre collective lorsqu'elles sont créées « sur l'initiative d'une personne physique ou morale qui l'édite, la publie et la divulgue sous sa direction et sous son nom et dans laquelle la contribution personnelle des divers auteurs ... se fond dans l'ensemble » (Art. L 113-2 du CPI). Dans ce dernier cas, c'est la personne physique ou morale qui a pris l'initiative de l'œuvre qui dispose des droits d'auteur.

Par ailleurs, le contexte de la création ou le statut de l'auteur peuvent avoir des incidences sur la détermination du titulaire des droits d'auteur. L'œuvre a-t-elle été créée dans le cadre d'une mission de service par un employé ou un fonctionnaire ? Quels sont les droits respectifs de l'auteur et de son employeur ? Si la question est résolue le plus souvent par le contrat de travail, elle reste plus délicate quand le créateur est un fonctionnaire.

Q3 : *Quels droits pour l'auteur d'un corpus ?*

Il convient de distinguer les droits patrimoniaux des prérogatives du droit moral. Les droits patrimoniaux se résument en un droit exclusif au profit de l'auteur (ou des titulaires) ou des ayants droits (bénéficiaires d'une cession, héritiers...) d'autoriser ou interdire la reproduction ou la communication au public de l'œuvre protégée. Quant aux prérogatives du droit moral toujours attachées à la personne physique créatrice de l'œuvre protégée, elles sont au nombre de quatre : le droit de divulgation, le droit de repentir et de retrait, le droit à la paternité et le droit au respect de l'œuvre. En réalité, il existe une possibilité intermédiaire où les corpus protégés par le droit d'auteur peuvent être mis en libre accès dans le cadre d'une licence accordée par les titulaires de droits autorisant l'utilisation et l'exploitation des résultats. Sans être dans le domaine public, ces corpus sont – de par la volonté de leurs créateurs – libres d'accès et d'utilisation. Néanmoins, si les créateurs peuvent renoncer à exercer leurs droits patrimoniaux, il ne leur est pas possible de renoncer à leur droit moral qui reste imprescriptible.

Q4 : *Qui est responsable (et de quoi) ?*

Tout traitement de données doit avoir un responsable, sa mission est d'éviter ou de circonvenir les risques inhérents à la gestion et l'utilisation des données recueillies. La loi lui fixe donc des obligations. La directive européenne 95/46/CE (« Flux de données transfrontières ») dans son article 2, repris pour la refonte de la loi « Informatique et libertés » donne la définition suivante : « *Le responsable d'un traitement de données à caractère personnel est, sauf désignation expresse par les dispositions législatives ou réglementaires relatives à ce traitement, la personne, l'autorité publique, le service ou l'organisme qui détermine ses finalités et ses moyens* ». Le responsable du traitement se doit de veiller à la qualité des données (adéquate, pertinente, non excessive et exacte, ce qui implique un droit d'accès, d'opposition et de rectification ouvert aux personnes concernées), au respect de leur(s) finalité(s) annoncées au préalable et au respect du principe de licéité (les

données doivent avoir été recueillies loyalement, ce qui implique le recueil d'un consentement éclairé). Il est, de plus, responsable de la déclaration à la CNIL, du recueil du consentement et du respect des règles de conservation et notamment de la confidentialité des données.

Q5 : *Comment se conformer au respect de la vie privée ?*

Si un corpus contient des données personnelles il dépend de la loi « Informatique et liberté ». La création d'un corpus passe le plus souvent par la collecte de données. Celles-ci pouvant être des données personnelles, cette collecte doit être faite dans le respect de la loi « Informatique et libertés » : licéité et loyauté, information préalable, obtention du consentement des personnes concernées, respect des finalités annoncées[9]… Si les données sont « anonymisées » de manière irréversible, elles sortent du champ de la loi et peuvent être conservées (voir annexe anonymisation). Toutefois dans la recherche, le besoin de « traçabilité » nécessite souvent de sauvegarder les données personnelles. L'anonymisation ne consiste pas simplement à faire disparaître les noms propres mais plutôt à gérer les données personnelles afin de ne pas permettre l'identification des locuteurs. Ainsi, La CNIL préfère parler de données personnelles : « *…pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne* ».

De fait l'identification d'une personne peut se faire de manière directe ou indirecte. Le caractère personnel d'une donnée dépend des moyens de tri, de rapprochement qui pourraient être mis en œuvre. Cela conduit donc à une évolution constante du champ des données personnelles, la technique mettant à la disposition du plus grand nombre des outils de requêtes plus en plus performants.

Certains auteurs avancent qu'il suffit qu'il y ait une probabilité suffisante de rapprochement à une personne pour qu'une donnée acquière un caractère personnel indirect. Les analyses ne sont pas toujours explicites mais il n'est possible de négliger cet argument (Lamy, *Droit de l'informatique et des Réseaux*, 508). Dans le domaine statistique, la CNIL a imposé des seuils au-delà desquels des rapprochements d'agrégats de données – pourtant individuellement anonymes – sont interdits.

Le caractère personnel d'une information dépend de l'objet qu'elle décrit, du contexte dont elle provient, mais aussi de la personne qui la reçoit. Pour pouvoir identifier un individu ou un groupe, nous avons besoin des informations, mais nous n'y arriverons pas sans un élément de connaissance propre qui déclenchera le mécanisme d'association. Lorsqu'on réfléchit à l'anonymisation, il convient donc de connaitre les éléments à traiter, mais aussi les opérations que vont subir les données (pour les corpus oraux : les données primaires (signal), les métadonnées, la transcription, l'annotation, le balisage, la diffusion, la conservation, etc.).

3. Eléments de réponse

[9] http://www.cnil.fr/fileadmin/documents/approfondir/textes/CNIL-78-17_definitive-annotee.pdf

Les réponses à ces cinq questions sont loin d'être évidentes et la gestion des points juridiques et éthiques sensibles ne peut se restreindre à l'application de contraintes mais doit respecter les pratiques des chercheurs. Les objectifs, notamment scientifiques, liés à la constitution, à l'exploitation, à la conservation et à la diffusion des corpus oraux, sont très divers, et le respect de ceux-ci, ainsi que leur hétérogénéité, impliquent que soit reconnue la diversité des *pratiques* qui peuvent être adoptées par les chercheurs et par les responsables de la diffusion et de la conservation de ces corpus. Seule l'identification précise et détaillée des éléments de la situation en jeu (forme des données et de leurs supports, pratiques de terrain, étapes de leurs traitements), permettent d'apporter à la fois des éléments de réponses juridiques correspondant à la situation, et une évaluation des « risques » éventuels. Enfin, une analyse réflexive sur la démarche liée à la constitution et aux traitements des corpus oraux est le premier élément de l'élaboration d'une éthique reconnue par l'ensemble d'une communauté scientifique.

3.1. Expliciter la démarche du chercheur

Comme nous l'avons évoqué plus haut, la définition du statut juridique d'un corpus ne peut se faire qu'après avoir explicité les éléments qu'il contient ainsi que les conditions de son élaboration et de son exploitation. Seul ce travail d'explicitation permet d'anticiper les problèmes juridiques et éthiques de conservation et de diffusion. La grille suivante permet de conduire ce travail en sept étapes descriptives : type de données, techniques d'enquêtes, rôles des participants et modes d'approche, lieu de l'enquête, dispositif d'enregistrement, annotations, exploitation-diffusion.

De quels types sont les données composant un corpus oral ?

On peut grossièrement opérer une dichotomie entre données primaires (enregistrements audio et vidéo mais aussi documents lus ou écrits durant l'action enregistrée) et données secondaires (annotations (et en particulier descriptions) des données primaires (métadonnées)). Cette distinction a deux utilités : caractériser la source et donc un état original (de base) du corpus et définir les niveaux du travail scientifique du chercheur (collecte, transcription, annotation, analyses,..).

Les données sont enregistrées et conservées sur des supports (physiques, magnétiques, optiques) selon un mode analogique ou numérique. La description du support et du mode est importante pour évaluer les possibilités de conservation et leurs conséquences sur l'objet « original ». Enfin l'explicitation de la structure des données et donc des codages, des formats et des formalismes utilisés permet de définir si un corpus répond à une technologie explicite, acceptée par une communauté et/ou ayant fait l'objet d'une normalisation et qui peut être ouverte à la communauté ou posséder un caractère propriétaire (brevets ou formats propriétaires).

Comment ces données ont-elles été collectées, quelles ont été les conditions d'élaboration du corpus ?

Pour répondre à cette question il convient d'abord de définir les techniques d'enquêtes utilisées : l'enregistrement en laboratoire selon un protocole expérimental produit des données et une situation aux caractéristiques juridiques bien différentes de l'enregistrement de récits de vie, de contes, de conversations professionnelles,

d'émissions radio ou télédiffusées, d'entretiens etc. Il faut ensuite caractériser les relations entre les enquêteurs et les enquêtés qui dépendent généralement du mode d'approche du chercheur (le témoin est-il rémunéré ? a-t-il exprimé son consentement ? est-il captif ?) et du rôle des participants ainsi que du lieu de l'enregistrement (est-ce un lieu public ou privé?). Enfin, il s'agit de décrire le dispositif d'enregistrement (audio ou vidéo, voyant ou caché, géré par le chercheur ou par les témoins, etc.) qui matérialise très souvent les aspects juridiques et éthiques de la relation observateur-observé.

Quels sont les traitements subis par les données primaires collectées ?

L'exploitation par les chercheurs des enregistrements oraux et multimodaux modifient considérablement l'objet. Ainsi les opérations d'annotations et notamment la transcription intègrent de nombreux aspects théoriques et interprétatifs mais aussi juridiques et éthiques. De fait,

> *dans le passage de l'oral à l'écrit graphico-visuel, de nombreuses opérations de catégorisation sont effectuées, soit quant aux formes linguistiques, segmentées visuellement en unités (Blanche-Benveniste & Jeanjean, 1987 ; Mondada, 2000), soit quant à l'identité des locuteurs eux-mêmes (Mondada, 2003). Du point de vue de la protection de l'image et de l'identité des personnes enquêtées et enregistrées, il convient d'apprécier ces effets pour éviter la surinterprétation, la stéréotypisation (Jefferson 1996) et la stigmatisation des locuteurs et de leurs façons de parler* (*Corpus oraux*, 73).

Les opérations d'annotations ne sont donc jamais neutres en termes d'effet de la recherche sur des données produites par les locuteurs. La description rigoureuse des traitements permet d'évaluer les effets de ceux-ci sur des productions dont on doit respecter les droits de leurs auteurs. De plus cette description fournit la possibilité d'attribuer la paternité et la responsabilité de niveaux de traitement à un ou plusieurs auteurs.

Quelles sont les pratiques de conservation et de diffusion ?

L'explicitation de la démarche de conservation et de diffusion revient à déterminer ce qui sera accessible et les modalités de cet accès (qui, quand, comment?).

L'explicitation de la démarche du chercheur permet donc de définir les données qui composent un corpus ainsi que les conditions d'élaboration de celui-ci. Ce travail fait, il est beaucoup plus aisé de repérer les questions juridiques qui se posent dans les deux grands domaines juridiques concernés : les droits d'auteur (moraux et patrimoniaux) et le respect de la vie privée. Cette description de la démarche est une première étape dans l'élaboration de bonnes pratiques qui doit logiquement être complétée par deux aspects qui semblent indispensables pour le respect de l'éthique du chercheur et qui permettront d'anticiper les problèmes de conservation et de diffusion. Ces deux aspects sont le recueil du consentement de l'enquêté et les procédures d'anonymisation des données.

3.2. Les bonnes pratiques pour un consentement éclairé

Pour des raisons éthiques, les chercheurs collectant des enregistrements produits par des informateurs souhaitent obtenir leur consentement. Dans le cas où le corpus

contient des données protégées par le droit d'auteur ou étant considérées comme des *données personnelles*[10] cette collecte doit obligatoirement être faite dans le respect de la loi « Informatique et libertés » et donc respecter les notions de licéité et loyauté, d'information préalable, du respect des finalités annoncées et a fortiori d'obtention du consentement des personnes concernées [11].

Le recueil de consentement n'est pas une opération banale dans les pratiques des chercheurs, parfois inexistante et souvent réduite à un formulaire de demande d'autorisation qui évoque en une phrase " le cadre d'un programme de recherche". Or sans informations préalables précises la demande d'autorisation n'a pas d'objet ni de sens. Pour que cette autorisation soit pertinente il conviendrait de concevoir le recueil d'un consentement "éclairé" qui démontre que le signataire est informé des finalités de la recherche et des conséquences à son égard d'une participation au projet.

Dans le cadre du recueil de données et notamment d'enregistrement pour des corpus oraux, le consentement devrait tenir compte de *l'adéquation au destinataire* (les informations fournies, pour être comprises doivent être adaptées aux compétences de compréhension du destinataire), et *de l'explicitation des finalités de l'enquête* (qui toutefois ne doivent pas renforcer le paradoxe de l'observateur en pointant l'objet de l'observation).

De plus, les explications sur le *projet scientifique***,** doivent être complétées par *des informations précises* comme par exemple : les *responsables* de l'enquête et leur affiliation institutionnelle, ainsi que les financeurs ; une *adresse* de contact, les *personnes qui auront accès aux données* et qui travailleront sur elles, la façon dont les données seront *anonymisées***,** le fait que les données seront transcrites selon des *conventions particulières*, la façon dont les données seront *archivées* une fois l'enquête terminée, les *modalités d'accès* aux informations relatives au projet et concernant tout particulièrement les données/analyses faisant référence à la personne (possibilité d'accès aux fichiers et informations concernant tout particulièrement la personne), les droits de la personne, notamment le droit de *rétractation*, les *risques* éventuels ainsi que les retombées positives, morales ou matérielles, de l'étude.

Enfin, le consentement devra préciser l'objet de la demande : *les actions* effectuées par les chercheurs dans le cadre du projet, *les formats* et les conditions de l'enregistrement, *les conditions de diffusion* des données et des résultats, *les contextes* de diffusion des données et des résultats. Il est à noter que les formes de l'autorisation ne sont pas imposées par le législateur et qu'une demande orale enregistrée peut être valide et même parfois indispensable.

Sur le plan juridique, la collecte de données sensibles sans recueil de consentement est possible à la condition particulière que les données soient anonymisées dans un

[10] Selon la Loi du 6 aout 2004, « Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne ».

[11] http://www.cnil.fr/fileadmin/documents/approfondir/textes/CNIL-78-17_definitive-annotee.pdf

très bref délais. La procédure d'anonymisation est également très importante pour obtenir l'accord des témoins a fortiori dans le cas d'une diffusion des données primaires.

3.3. Les bonnes pratiques de l'anonymisation

Les pratiques actuelles des chercheurs en termes d'anonymisation se réduisent la plupart du temps à une opération de masquage d'un nom propre, d'une adresse ou d'un numéro de téléphone. Afin de vérifier la validité de ces pratiques et d'en définir les modalités, il convient de reposer avec précision la question légale qui est celle de *l'impossibilité d'identifier des personnes*. En effet, l'objectif est de protéger la vie privée des personnes enregistrées en dépersonnalisant les données , ce qui a amené le législateur à ne pas réduire cette identification à la présence de données nominatives. Ainsi, les textes législatifs abandonnent les termes de « données nominatives » au profit de « données personnelles » :

> *...pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne (directive 95/46/CE).*

Une attention toute particulière doit donc être apportée aux données personnelles contenues dans un corpus qui permettraient d'identifier directement un témoin (formes nominatives, données personnelles, profession, statut, titres, activités sociales, parenté, réseaux, référence à des lieux, référence à des caractéristiques de la personne, caractéristiques physiques, etc.) mais aussi à tout ce qui peut permettre indirectement une identification (notamment les possibilités de recoupement d'informations).

Les procédures d'anonymisation concernent les données premières audio et/ou vidéo, les données premières textuelles, les données secondaires et les données secondaires visuelles en évitant une anonymisation sur les données premières originales et en gardant éventuellement la possibilité de travailler sur des données partiellement anonymisées dans un cadre de recherche précis même si en revanche, la diffusion implique une anonymisation rigoureuse.

Techniquement l'anonymisation consiste principalement au remplacement ou au codage des données sensibles par des éléments neutres selon les supports concernés (remplacement par un blanc ou un pseudo à l'écrit, par un bip dans les fichiers sons et par floutage des visages sur les enregistrements vidéos).

Cependant le codage des données personnelles n'est pas la seule procédure d'anonymisation. La possibilité d'élaborer des bases de données séparées est utilisée en France. Plus intéressant encore « *les limitations techniques* » qui permettent de protéger l'anonymat sans modifier les données mais en limitant les possibilités de recherche et de croisement des données (ne pas permettre de croiser le lieu d'habitation et la profession par exemple). Cette dernière possibilité est particulièrement intéressante dans le cadre de recherche sur la langue parlée où les informations sociologiques pertinentes sont indispensables.

3.4. Des corpus oraux au patrimoine sonore, l'enjeu d'une normalisation

Les enregistrements de la voix, à la base des corpus oraux, ont été constitués en fonds sonores depuis presque un siècle (Calas 1996). La numérisation de ces fonds

sonores et l'informatisation des catalogues n'ont fait qu'accroître les relations entre les corpus oraux - objets scientifiques - et les archives orales - objets patrimoniaux. Ainsi plusieurs projets actuels de grands corpus oraux (Phonologie du Français Contemporain, Archivage du LACITO, Eslo) ont, en vue du dépôt de leurs données, sollicité les institutions de conservation qui elles mêmes développent la diffusion en ligne de leurs archives (Gallica pour la BnF, archives pour tous de l'INA). Le transfert de la gestion de la conservation et de la diffusion des corpus à une institution fait apparaître les mêmes problèmes que rencontrent les chercheurs : principe de cohérence des fonds, gestion des métadonnées, des formats et des normes, questions de propriétés intellectuelles et de protection de la vie privée objectifs de. Or si ces problèmes, qui sont nouveaux pour les chercheurs, n'ont pas été résolus par les institutions de conservation, c'est aussi parce que la question du statut de l'oral comme objet patrimonial est indissociable du statut de l'oral dans le champ scientifique.

En ce sens, les documents numériques et les technologies de conservation et de diffusion en masse posent d'une façon cruciale le problème de la normalisation des données (supports, formats, codages etc.) pour des usages diversifiés.

Les bonnes pratiques et la normalisation des données (*Corpus oraux*, 79-95)

La réponse technique à un problème méthodologique et éthique démontre l'importance des relations entre l'éthique, le théorique et le technique dans les questions d'exploitation de corpus. Les objectifs de partage de données, d'échange et d'interopérabilité sont ainsi indissociables des aspects de diffusion et de conservation des corpus et posent de fait la question des pratiques de normalisation même s'il ne peut y avoir de travail sur la normalisation des corpus sans prise en compte de l'impact de celle-ci sur les analyses. Il suffit pour s'en persuader d'évaluer l'enjeu de la normalisation pour un locuteur ou pour une communauté lié à la normalisation provoquée par les opérations de transcription.

4. Conclusion

Les objectifs de conservation et de diffusion des corpus ne reposent pas simplement sur des questions techniques et stratégiques de normalisation (pour permettre l'échange, la multifinalité et l'interopérabilité), il s'agit également d'élaborer des bonnes pratiques dans le respect de l'éthique et du juridique qui obligent le chercheur à *savoir ce qu'il fait* et la communauté scientifique à *prendre en charge les objets scientifiques qu'elle construit*. Ce n'est alors pas étonnant de découvrir que les questions juridiques se concentrent sur deux domaines : la propriété des données et le respect de la vie privée. Les réponses ne sont pas simples et ne sont pas systématiques, les contraintes de la recherche font que certains corpus ne peuvent à la fois suivre des bonnes pratiques respectant l'observé et des bonnes pratiques de diffusion des données. Il est fondamental de respecter l'hétérogénéité des pratiques de la recherche fondée sur les corpus oraux et de considérer que certains corpus ne sont pas fait pour être diffusés ni même accessibles en dehors de la relation enquêteur-enquêté. Cependant pour la grande majorité des cas, l'élaboration de bonnes pratiques par la communauté scientifique doit permettre de donner un véritable statut aux corpus oraux en imposant une démarche éthique qui reconnaît les droits des observés, en facilitant l'exploitation, la diffusion et la conservation par

une réflexion sur la normalisation des données et métadonnées et en anticipant les changements de finalités. Pour atteindre cet objectif, il revient au chercheur d'expliciter sa démarche, de la collecte à la diffusion, de gérer les droits repérés (de propriété intellectuelle et de respect de la vie privée) et de structurer les corpus pour faciliter l'interopérabilité et le cumulatif tout en permettant par une limitation technique de protéger l'accès aux différentes données.

Ainsi, l'élaboration de *bonnes pratiques*, loin d'être restreintes à l'application de contraintes juridiques est aussi l'opportunité pour une communauté scientifique de s'acquitter de la dette que le chercheur contracte envers son terrain, en offrant un véritable statut à son objet d'étude.

Olivier Baude
Université d'Orléans et DGLFLF
UFR LLSH, 10 rue de Tours, 45072 Orléans.
<olivier.baude@univ-orleans.fr>

Références

Baude, O. (2004). Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques. In *Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble), 7-11.

Becker, H.S. & Geer, B. (1960). Participant observation : the analysis of qualitative field data. In Adams & Preiss (eds.), 267-289.

Bergounioux, G. (ed.) (1992). Enquêtes, Corpus et Témoins. *Langue Française* 93.

Biber, D. (1999) *Longman Grammar of Spoken and Written English*. Londres, Longman.

Bilger, M. (ed.) (2000). *Corpus, Méthodologie et applications linguistiques*. Paris, Champion.

Blanche-Benveniste, C. (1997). Transcription et technologie. *Recherches sur le Français Parlé* 14, 87-100.

Bourdieu, P. (1993). *La misère du monde*. Paris, Le Seuil.

Calas, M-F. & Fontaine, J-M (1996). *La conservation des documents sonores*. Paris, CNRS Editions.

Callu, A. & Lemoine, H. (2004). *Patrimoine sonore et audiovisuel français : entre archive et témoignage : guide de recherche en sciences sociales*. Paris, Belin, 7 vol., 1 CD-Rom, 1 DVD-Rom.

Condamines, A. (ed.) (2006). *Sémantique et corpus*. Paris, Hermes.

Cresti, E. & Moneglia, M. (eds.) (2005). *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam, Benjamins.

Cribier, F. & Feller, E. (2003). *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »*. Rapport au Ministre délégué à la Recherche et aux nouvelles technologies, dactylogr., 2 vol. http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf

Encreve, P., & Fornel (de) M. (1983). Le sens en pratique. *ARSS* 46, L'usage de la parole.

Gumperz, J.J., & Hymes, D. (eds.) (1972). *Directions in Sociolinguistics : The Ethnography of Communication*. New-York, Hold, Rinehart & Winston.

Habert, B., Nazarenko, A. & Salem, A. (1997). Les linguistiques de corpus. Paris, A. Colin.

Habert, B., (2005*), Instruments et ressources électroniques pour le français*, Ophrys, Paris.

Jacobson, M. (2004). Corpus oraux en linguistique de terrain. *Traitement Automatique des Langues*, 45/2, 63-88.

Jacobson, M. (2004). Les archives sonores au LACITO. *Bulletin de liaison de l'AFAS* 26 (http://afas.mmsh.univ-aix.fr/bulletin/Bulletin AFAS 26.pdf).

Kennedy, G. (1998). *An introduction to Corpus Linguistics*. Londres, Longman.

LAMY, Droit de l'informatique et des réseaux (S. Marcellin, L. Costes & *al.* eds., Paris, 2004).

Leech, G. (1992). The state of the art in corpus linguistics. In Aijmer & Altenberg (eds.), 8-29

Mondada, L. (1998). Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête. Actes du Colloque de Lausanne (13-14.12.1998), *Cahiers de l'ILSL* 10, 39-68.

Mondada, L. (2006). Video recording as the reflexive preservation-configuration of phenomenal features for analysis. In Knoblauch, H., Raab, J., Soeffner, H.G., Schnettler, B. (eds.)

Quéré, L. & *al.* ed. (1984) *Arguments ethnométhodologiques*, Paris, Centre d'Étude des Mouvements Sociaux, EHESS.

Recherches sur le Français Parlé 5 (1984). Pourquoi le français parlé est-il si peu étudié ?.

Revue Française de Linguistique Appliquée (1996) I-2, (1999) IV-1.

Sacks, H. (1984). Notes on methodology. In J.M. Atkinson & J. Heritage (eds.), 21-27.

Sankoff, D., Sankoff, G., Laberge, S. & Topham, M. (1976). Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale. *Cahiers de Linguistique* 6, 85-125.

Silverman, D. (ed.) (1997). *Qualitative Research. Theory Method and Practice*. Londres, Sage.

Sinclair, J. (1996). *Preliminary recommendations on corpus Typology*. Technical Report, Eagles.

Speech Communication (2001) Speech Annotation and Corpus Tools. Vol. 33, 1-2, S. Bird & J. Harrington (eds.).

Welland, T. & Pugsley, L. (eds.) (2002). *Ethical Dilemmas in Qualitative Research*. Aldershot, Ashgate.

Titre	***Les corpus de la parole, patrimoine immatériel et langues de France***
Type	Article de vulgarisation
Editeur	MCC
Année	2008
Référence	Baude, O., Alessio, M., (2008) «Les corpus de la parole, patrimoine immatériel et langues de France», in *Culture & Recherche*, n° 116 et 117, Ministère de la Culture et de la communication, Paris. Pp 42-43.

Les corpus de la parole :
patrimoine immatériel et langues de France

Olivier Baude et Michel Alessio
MCC / Délégation générale à la langue française et aux langues de France

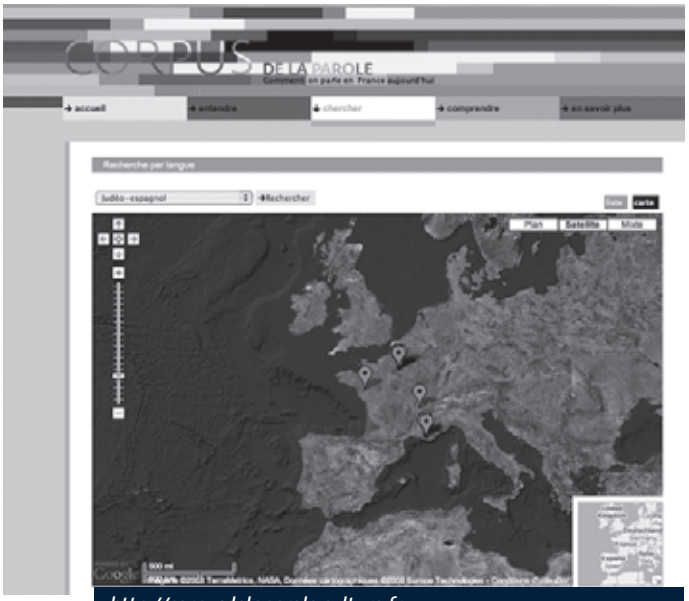
Verba volant, scripta manent. Dans un pays si fortement marqué par la tradition écrite et les usages littéraires, longtemps la langue parlée n'a pas été perçue dans toute son importance. C'est l'auteur de la grande *Histoire de la langue française,* Ferdinand Brunot, qui, le premier, s'est préoccupé d'enregistrer et de conserver les traces sonores de faits de langue, en créant les fameuses *Archives de la parole* en 1911. Ces enregistrements sur rouleaux de cire, pieusement conservés, forment le fonds premier du département de l'audiovisuel à la Bibliothèque nationale de France. Jusque-là, la parole vivante apparaissait curieusement – et paradoxalement – comme une forme subalterne, dérivée, et pour tout dire dégradée de la langue écrite. Effet de culture : telle était la représentation dominante qu'on se faisait du langage en France. Ç'a été le travail de la linguistique du XXe siècle que de rétablir l'ordre des choses en se fondant notamment sur la description de données orales constituées en collections et ordonnées par des critères scientifiques : les corpus oraux.

Ce travail ne se fait pas d'un trait. À la suite des *Archives de la parole,* sont créées en 1932 la phonothèque du musée de l'Homme et en 1938 la Phonothèque nationale. C'est toutefois le musée national des Arts et Traditions populaires qui possédait le plus de témoignages oraux. Mais ceux-ci étaient exclusivement réalisés par des ethnologues, et destinés à leurs recherches. L'oral n'était pas encore collecté pour lui-même ; il n'était que le support d'études en sciences humaines et sociales. D'ailleurs, lorsqu'André Malraux lance l'Inventaire général des monuments et richesses artistiques de la France en 1964, l'oral n'y figure nulle part.

À l'ère du numérique, la sauvegarde de l'oral prend véritablement son essor. L'informatique permet de faciliter la classification des enregistrements et leur accès. Mais cela ne résout pas tous les problèmes. Un enregistrement isolé ne présente guère d'intérêt en soi. Il n'y a patrimoine de l'oral que lorsque plusieurs documents sont regroupés autour d'un thème. Enfin, la validation scientifique et le traitement des données enregistrées marquent la porte d'entrée du domaine du patrimoine. Celui-ci en effet ne s'étend pas aux données sonores brutes qu'un particulier a pu collecter à des fins personnelles, lors d'une conversation ou d'un entretien. Le « sceau de la science » doit garantir que les corpus oraux ont été correctement composés, c'est-à-dire indexés, transcrits, éventuellement traduits, balisés, annotés, catalogués.

En France, les premières grandes enquêtes sur le français ont été effectuées dans les années 1950 à des fins didactiques. Or, plus de cinquante ans après, nous ne disposons toujours pas d'un véritable corpus de référence qui permette toutes sortes de recherches (descriptions, analyses, applications) et on mesure le retard pris par notre pays, y compris vis-à-vis d'autres zones francophones, en comparant ces résultats avec les données engrangées ailleurs.

Le ministère de la Culture et de la Communication / Délégation générale à la langue française et aux langues de France (DGLFLF), en partenariat avec les chercheurs des universités et du CNRS, a entrepris de combler ce retard, dans une perspective de valorisation de la



http://corpusdelaparole.culture.fr

diversité. La France dispose en effet d'une grande richesse de langues. À côté du français, langue nationale, langue commune, présente sur les cinq continents, les langues de France constituent un patrimoine culturel unique : il y a sur le territoire de la République des langues romanes, des langues germaniques, le breton, langue celtique, le basque, qui n'est pas une langue indo-européenne, des créoles, des langues amérindiennes, des langues polynésiennes, des langues austronésiennes, etc. Plus de 75 langues sont reconnues comme « langues de France », c'est-à-dire parlées par des citoyens français en France depuis assez longtemps pour faire partie du patrimoine culturel national, et qui par ailleurs ne sont langue officielle d'aucun État.

Ce patrimoine est trop souvent méconnu, et si des archives sonores existent désormais pour la quasi-totalité de ces langues, la richesse qu'elles représentent n'était jusqu'ici accessible ni à l'ensemble de la communauté scientifique ni au grand public. Plus grave encore, de nombreux documents sonores conservés sur des supports physiques à bout d'usage (comme les enregistrements sur bandes magnétiques) sont voués à disparaître dans des délais très courts. Or, il s'agit souvent des derniers ou des seuls documents dont nous disposions sur des langues de France – comme pour certaines langues de Guyane ou de Nouvelle-Calédonie –, mais aussi sur le français. Ainsi, au ministère de la Culture, la DGLFLF a numérisé les seuls enregistrements de français constitués par des linguistes dans les années 1970.

Aujourd'hui, avec le progrès des nouvelles technologies, la numérisation offre non seulement la possibilité de sauver ce patrimoine mais aussi l'occasion de le valoriser en transformant les documents originaux en de véritables ressources linguistiques numé-

riques. En créant, en partenariat avec le CNRS, le programme *Corpus de la parole,* le ministère de la Culture et de la Communication/DGLFLF s'est engagée depuis 2004 dans une triple démarche.
La première étape de ce programme était d'ordre méthodologique ; il s'agissait de définir les conditions dans lesquelles les productions verbales devaient être recueillies à des fins d'études et de recherches, et c'est ainsi qu'a été entreprise l'édition de l'ouvrage *Corpus oraux, Guide des bonnes pratiques* (CNRS-Editions et PUO, 2006), consacré à la constitution, la conservation et l'exploitation des corpus oraux. Ce guide s'inscrit à l'exact croisement d'une démarche scientifique et d'une politique culturelle ; il constitue aujourd'hui une proposition de charte pour tous les chercheurs, auxquels il fournit les instruments, y compris les instruments de prescription d'ordre juridique, qui permettent de constituer ces données brutes en objets de savoir.
La seconde étape a consisté à lancer un vaste chantier de numérisation dans le cadre du plan national de numérisation du ministère de la Culture (DDAI/MRT). Ce plan a déjà permis de sauvegarder des centaines d'heures d'enregistrement.
La troisième étape a consisté à rendre les données accessibles à tous, d'abord à la collectivité des chercheurs, mais aussi, au-delà des chercheurs, au grand public, et c'est désormais possible avec le site *Corpus de la parole* (**http://corpusdelaparole.culture.fr**), dont la première version est en ligne depuis le début de l'année 2008. Ce site donne accès à un catalogue collectif de corpus oraux en français et en langues de France, sous la forme de fonds sonores transcrits et numérisés.
La sauvegarde et l'exploitation de ces enregistrements en français et en langues de France sont un enjeu de première importance. C'est un enjeu pour la recherche, pour le développement de l'ingénierie linguistique et pour l'enseignement, mais aussi pour le développement d'une politique culturelle qui reconnaisse les faits de langue comme éléments du patrimoine immatériel dans toute sa variété.

Titre	***Le droit de la parole***
Type	chapitre
Editeur	Presses universitaires de Perpignan
Année	2008
Référence	Baude, O. (2008) **«**Le droit de la parole», in M. Bilger (ed), *Données orales, les enjeux de la transcription*, Presses universitaires de Perpignan, p 23-34

Olivier Baude
CORAL-Université d'Orléans et DGLFLF-MCC
(baude@wanadoo.fr)

1.3 Le droit de la parole[1] [2]

Les aspects juridiques de la constitution et de l'exploitation des corpus oraux ont très longtemps été négligés même si une préoccupation éthique a toujours été au cœur des travaux des chercheurs de terrain - notamment chez les ethnolinguistes et les sociolinguistes. La situation change avec les nouveaux usages et l'engouement actuel pour la diffusion des corpus oraux. Ainsi, la numérisation des données ouvre des possibilités de stockage, de diffusion et d'utilisation seconde – ressources pour des outils didactiques et pour l'ingénierie – tout en suscitant de nombreuses questions juridiques : Comment recueillir le consentement des personnes enregistrées ? Faut-il anonymiser les données ? A qui appartiennent les enregistrements, les transcriptions, les annotations ? La transcription est-elle couverte par le droit d'auteur? Etc.

Les questions sont complexes et les réponses ne vont pas de soi. Trois grands domaines juridiques sont concernés : le droit d'auteur et la propriété intellectuelle, les données personnelles et le respect de la vie privée et les responsabilités des "exploitants" et diffuseurs. Il convient donc dans un premier temps de définir le statut juridique d'un corpus oral afin de procéder à la gestion contractuelle des droits des personnes repérés et enfin de définir les responsabilités de tous ceux qui interviendront lors de l'exploitation du corpus (de la collecte à la diffusion).

1.3.1 Quel est le statut juridique d'un corpus et de ses transcriptions?

Définir le statut juridique d'un corpus nécessite de décrire avec précision son contenu puis les conditions d'élaboration et d'exploitation de celui-ci. La forme d'un corpus oral est très variable et son contenu diffère selon les méthodologies, les théories et les usages de ses concepteurs. Dans la majorité des cas les corpus oraux sont constitués d'enregistrements analogiques ou numériques – et qui dans le cas de supports analogiques ont une durée de vie très courte avec une perte de qualité, lors des migrations – de productions de locuteurs ; de données contextuelles sur les locuteurs et la situation d'enquête qui peuvent être en partie des données personnelles (nom propre, profession, adresse, lieu,…) ; de transcriptions "annotations-primaires" phonétiques, orthographiques, multilinéaires, etc. (sous la forme de fichiers indépendants ou offrant une synchronisation avec le signal) ; d'annotations secondaires (informations sur les conditions de production des énoncés, précisions sur les phénomènes sonores tels que les rires et les bruits) ; d'annotations enrichies (étiquetage morphologique, syntaxique, annotations prosodiques pragmatiques).

[1] Les problèmes juridiques liés à la diffusion des corpus oraux ont été l'occasion d'une démarche originale adoptée par une communauté scientifique ouverte à un travail pluridisciplinaire mené dans le cadre du programme *Corpus de la Parole* de la Délégation générale à la langue française et aux langues de France – Ministère de la Culture et de la Communication. Cette démarche a comporté plusieurs étapes. Une lecture croisée des textes juridiques par les linguistes et les juristes a permis de repérer les problèmes. Les chercheurs ont ensuite accepté d'expliciter leurs pratiques au regard de la législation. Cette étape fondée sur la réflexivité a permis d'élaborer des propositions pour de bonnes pratiques partagées par la communauté scientifique et de repérer des aspects juridiques qui posent des difficultés dans l'état actuel du droit.

[2] Ce texte présente principalement les résultats du groupe de travail publiés sous la forme d'un guide : *Corpus oraux, guide des bonnes pratiques 2006* rédigé par O. Baude coordinateur (Université d'Orléans/DGLFLF), C. Blanche-Benveniste (EPHE/Université de Provence), M-F. Calas (DMF), P. Cappeau (Université de Poitiers), P. Cordereix (BnF), L. Goury (CNRS), M. Jacobson (CNRS), I. de Lamberterie (CNRS) C. Marchello-Nizia (ILF/ENS-LSH-Lyon) et Lorenza Mondada, (Université Lyon 2). La paternité de la grande majorité du contenu de ce texte en revient donc à l'ensemble des auteurs à qui le rédacteur exprime toute sa gratitude.

Le statut juridique d'un corpus se définit également par les conditions d'élaboration et d'exploitation de celui-ci. En effet, l'exploitation par les chercheurs des enregistrements oraux et multimodaux modifient considérablement l'objet : les opérations d'annotations et notamment la transcription intègrent de nombreux aspects théoriques et interprétatifs mais aussi juridiques et éthiques. De fait, "*dans le passage de l'oral à l'écrit graphico-visuel, de nombreuses opérations de catégorisation sont effectuées, soit quant aux formes linguistiques, segmentées visuellement en unités (Blanche-Benveniste & Jeanjean, 1987 ; Mondada, 2000), soit quant à l'identité des locuteurs eux-mêmes (Mondada, 2003). Du point de vue de la protection de l'image et de l'identité des personnes enquêtées et enregistrées, il convient d'apprécier ces effets pour éviter la surinterprétation, la stéréotypisation (Jefferson 1996) et la stigmatisation des locuteurs et de leurs façons de parler*"[3].
Les opérations d'annotations ne sont donc jamais neutres en termes d'effet de la recherche sur des données produites par les locuteurs. La description rigoureuse des traitements permet d'évaluer les effets de ceux-ci sur des productions dont on doit respecter les droits de leurs auteurs. De plus, seule une description rigoureuse fournit la possibilité d'attribuer la paternité et la responsabilité de niveaux de traitement à un ou plusieurs auteurs.

Cette étape de description du corpus est donc fondamentale pour différencier les composantes susceptibles d'être protégées par le droit d'auteur ainsi que les composantes contenant éventuellement des données personnelles. C'est à l'issue de cette étape qu'il conviendra de procéder à la gestion contractuelle des droits repérés et aux traitements nécessaires à ceux-ci comme le recueil de consentement ou l'anonymisation.

1.3.2 Droit d'auteur et propriété intellectuelle

Les transcriptions d'un corpus, tout comme les autres éléments de celui-ci, sont elles protégées par le droit d'auteur ? D'un point de vue juridique il convient de classer les composantes d'un corpus comme relevant du domaine public – dans ce cas son exploitation est totalement libre –, ou comme étant protégées par le droit d'auteur .

Pour qu'un corpus ou qu'une composante d'un corpus soit considéré comme protégé par le droit d'auteur il faut qu'il remplisse trois conditions : qu'il corresponde à l'exigence d'une *activité créatrice*, qu'il ait une *forme définie* et que cette forme soit *originale*[4]. Si tel est le cas l'auteur, qui est en principe la (ou les) personne(s) physique(s) sous le nom de laquelle (ou desquelles) l'œuvre est divulguée[5], accède à des droits patrimoniaux et aux prérogatives du droit moral. "*Les droits patrimoniaux se résument en un droit exclusif au profit de l'auteur (ou des titulaires) ou des ayants droits (bénéficiaires d'une cession, héritiers...) d'autoriser ou interdire la reproduction ou la communication au public de l'œuvre protégée. Quant aux prérogatives du droit moral toujours attachées à la personne physique créatrice de l'œuvre protégée, elles sont au nombre de quatre : le droit de divulgation, le droit de repentir et de retrait, le droit à la paternité et le droit au respect de l'œuvre*"[6].

Ces précisions juridiques ne permettent pas une réponse tranchée mais précisent les questions juridiques : le contenu d'une langue, son expression phonique de même que son expression graphique font-ils ou non partie du domaine public ? Le travail scientifique de collecte comme celui de transcription et d'annotation doivent-il être considérés comme des activités

[3] *Corpus oraux, Guide des bonnes pratiques, 2006*, p73.
[4] *Corpus oraux, Guide des bonnes pratiques, 2006*, p39.
[5] Le travail scientifique suppose l'intervention de nombreux acteurs dont bon nombre sont susceptibles de revendiquer la qualité d'auteur sur les résultats de la recherche. Certains corpus oraux, comme les autres produits de la recherche, peuvent rester l'œuvre d'un auteur unique, alors que d'autres peuvent être l'œuvre de plusieurs auteurs.
[6] *Corpus oraux, Guide des bonnes pratiques, 2006*, p 40.

créatrices d'œuvres à la forme définie et originale ? Cette dernière question est d'autant plus complexe que le statut d'auteur-fonctionnaire n'est pas clairement définit actuellement.

Seule une réflexion sur les pratiques des chercheurs, menée par l'ensemble de la communauté scientifique permettra de construire des réponses. Il est alors d'autant plus important de signaler une solution intermédiaire où les corpus protégés par le droit d'auteur peuvent être mis en libre accès dans le cadre d'une licence accordée par les titulaires de droits autorisant l'utilisation et l'exploitation des résultats[7].

1.3.3 Le respect de la vie privée

Le second grand volet des aspects juridiques concerne la gestion d'éventuelles données personnelles relevant du respect de la vie privée. La présence de données personnelles dans un corpus implique obligatoirement de se conformer à la *loi informatique et liberté (licéité et loyauté,information préalable, obtention du consentement),* ou de procéder à l'anonymisation *irréversible* de celles-ci. Avant de présenter les modalités du recueil de consentement et les techniques d'anonymisation il est nécessaire de définir plus précisément ce que sont "*les données personnelles*".

La loi informatique et liberté ne restreint pas le respect de la vie privée aux précautions à prendre pour gérer des données "nominatives". La question légale est celle, plus vaste, de l'impossibilité d'identifier des personnes. En effet, l'objectif est de protéger la vie privée des personnes enregistrées en dépersonnalisant les données, ainsi, les textes législatifs abandonnent les termes de « données nominatives » au profit de « données personnelles » :

> *…pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne (directive 95/46/CE).*

Une attention toute particulière doit donc être apportée aux données personnelles contenues dans un corpus qui permettraient d'identifier directement un témoin (formes nominatives, données personnelles, profession, statut, titres, activités sociales, parenté, réseaux, référence à des lieux, référence à des caractéristiques de la personne, caractéristiques physiques, etc.) mais aussi à tout ce qui peut permettre indirectement une identification (notamment les possibilités de recoupement d'informations).

1.3.4 Techniques d'anonymisation des données primaires et des transcriptions

L'anonymisation n'est pas une opération obligatoire. Elle est toutefois indispensable dans les cas où le corpus contient des données personnelles sans que le consentement des personnes concernées ait été recueilli. Ainsi, le nom propre n'étant pas le seul élément devant donner lieu à un traitement spécifique, il serait plus juste de parler de "*dépersonnalisation*" des données.

La première phase consiste donc à repérer les données permettant l'identification directe et indirecte mais aussi celles qui pourraient porter préjudice.

Dans une seconde phase les données primaires sont traitées au moyen d'opérations techniques (bippage, effacement, déformation, etc.).

Enfin la troisième phase concerne la "dépersonnalisation" des transcriptions. Le principe général est celui de la substitution. L'information est remplacée par un segment vide, un hyperonyme ou une abréviation (NN, NPersonne), par des caractères spéciaux (****), par un pseudonyme (Pierre à la place de Paul).

Une éventuelle quatrième phase concerne l'anonymisation des métadonnées par cryptage ou structuration de bases de données séparées.

[7] *Corpus oraux, Guide des bonnes pratiques, 2006*, p 39.

L'objectif de l'impossibilité d'identification est totalement irréaliste dans le cadre des corpus contenant des enregistrements (il faudrait systématiquement déformer la voix des locuteurs), il est donc fondamental de prendre en compte la notion de *"moyens susceptibles d'être raisonnablement mis en œuvre*" (Cf. supra) ; et d'avoir une démarche éthique et scientifique suffisamment rigoureuse pour permettre une exploitation sereine de corpus oraux. Enfin la communauté scientifique se doit d'inventer des procédés prenant en compte ses usages d'exploitation et de diffusion. Ainsi *la loi québécoise « concernant le cadre juridique des technologies de l'information » propose de protéger l'anonymat non pas en modifiant les données, mais en limitant les possibilités de recherche, voire en les adaptant à la personne qui consulte la base selon des critères bien précis (sa profession, une autorisation, sa présence dans le fichier, etc.). Cette dernière perspective offre pour la constitution et l'exploitation de corpus oraux la possibilité de faire coïncider les obligations légales avec les nécessités du travail de recherche. Toute donnée étant potentiellement sensible, une anonymisation systématique s'avère de plus en plus complexe ; elle peut même mettre en danger l'intérêt de certaines recherches. En effet, des détails concernant les personnes comme par exemple le nom, ou le lieu d'habitation peuvent constituer un élément important du corpus, et des résultats qui peuvent en être obtenus. C'est pourquoi la possibilité de ménager des niveaux d'accès selon des critères stricts (ex : chercheur ou non, présence d'autorisation, but de la consultation, etc.) semble une alternative efficace.*
Il existe d'autres procédés à inventer[8]*".*

1.3.5 Consentement

Le recueil du consentement des personnes enregistrées reste la meilleure solution éthique et juridique, pour autant que celle-ci ne soit pas réduite à une vague demande d'autorisation. Or sans informations préalables précises la demande d'autorisation n'a pas d'objet ni de sens. Pour que cette autorisation soit pertinente il faut établir un consentement "éclairé" qui démontre que le signataire est informé des finalités de la recherche et des conséquences à son égard d'une participation au projet.

Dans le cadre du recueil de données et d'exploitation de corpus oraux, le consentement devrait tenir compte de l'adéquation au destinataire - les informations fournies, pour être comprises doivent être adaptées aux compétences de compréhension du destinataire -, et de l'explicitation des finalités de l'enquête - qui toutefois ne doivent pas renforcer le paradoxe de l'observateur en pointant l'objet de l'observation -. Le consentement devra préciser l'objet de la demande : les actions effectuées par les chercheurs dans le cadre du projet, les formats et les conditions de l'enregistrement, les conditions de diffusion des données et des résultats, les contextes de diffusion des données et des résultats[9].
De plus, les explications sur le projet scientifique, doivent être complétées par des informations précises notamment sur la façon dont les données seront anonymisées (le cas échéant) et sur la forme que prendront les énoncés transcrits. En effet, la transcription est empreinte de lourds enjeux éthiques. *"La représentation écrite de la langue surprend souvent les locuteurs, et peut même leur déplaire considérablement. Il arrive qu'ils refusent l'image de leur langue transmise par la transcription, qu'ils désavouent le chercheur et qu'ils refusent son travail. Ainsi le recours à l'API et à l'orthographe adaptée peut produire des effets de stigmatisation et d'asymétrie à l'encontre de locuteurs « non-natifs »* "[10].

[8] *Corpus oraux, Guide des bonnes pratiques, 2006*, p 109
[9] Il est à noter que les formes de l'autorisation ne sont pas imposées par le législateur et qu'une demande orale enregistrée peut être valide et même parfois indispensable.
[10] *Corpus oraux, Guide des bonnes pratiques, 2006*, p 76

Il n'existe pas de réponses juridiques toutes faites aux questions posées par le droit de la parole. Les solutions ne consistent pas à appliquer simplement une législation existante, mais passent par l'élaboration de bonnes pratiques dans le respect d'une démarche éthique. Cela implique que le chercheur *sache ce qu'il fait* et qu'il soit donc capable d'expliciter sa démarche dans un véritable travail réflexif. Cette explicitation est indispensable à la définition même d'un corpus constitué d'enregistrements, de transcriptions et d'annotations et à la connaissance de ses conditions d'élaboration et d'exploitation mais aussi à l'évaluation des enjeux empirico-théoriques. Ainsi, l'élaboration de bonnes pratiques, loin d'être restreintes à l'application de contraintes juridiques est aussi l'opportunité pour une communauté scientifique de s'acquitter de la dette que le chercheur contracte envers ceux qui produisent les données, en prenant en charge, dans toutes leurs dimensions – y compris juridiques et éthiques - les objets scientifiques qu'elle construit.

Bibliographie

Baude, O., coord. (2006). *Corpus oraux, Guide des bonnes pratiques 2006*, CNRS éditions et PUO, Paris et Orléans.

Baude, O. (2004). Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques. In *Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble), 7-11.

Beaud, S. & Weber, F. (1997). *Guide de l'enquête de terrain : produire et analyser des données ethnographiques.* Paris, La Découverte.

Becker, H.S. & Geer, B. (1960). Participant observation : the analysis of qualitative field data. In Adams & Preiss (eds.), 267-289.

Bergounioux, G. (ed.) (1992). Enquêtes, Corpus et Témoins. *Langue Française* 93.

Biber, D. (1985). *Variations across spoken and written language*. Cambridge, CUP.

Biber, D. (1999) *Longman Grammar of Spoken and Written English*. Londres, Longman.

Bilger, M. (ed.) (2000). Linguistique sur corpus, études et réflexions. *Cahiers de l'université de Perpignan*, Perpignan, Presses Universitaires de Perpignan.

Bilger, M. (ed.) (2000). *Corpus, Méthodologie et applications linguistiques*. Paris, Champion.

Blanche-Benveniste, C. & Jeanjean, C. (1987). *Le français parlé : transcription et édition*. Paris, Didier-Erudition.

Blanche-Benveniste, C. (1997). Transcription et technologie. *Recherches sur le Français Parlé* 14, 87-100.

Blanche-Benveniste, C., Bilger, M., Rouget, C. & van den Eynde, K. (1999). *Le Français Parlé : Études grammaticales*. Paris, CNRS-Editions.

Blanche-Benveniste, C., Rouget, C. & Sabio, F. (2001). *Choix de textes de français parlé : trente-six extraits*. Paris, Champion.

Bourdieu, P. (1982)**.** *Ce que parler veut dire. L'économie des échanges linguistiques*. Paris, Fayard.

Bourdieu, P. (1993). *La misère du monde*. Paris, Le Seuil.

Calas, M-F. & Fontaine, J-M (1996). *La conservation des documents sonores*. Paris, CNRS Editions.

Callu, A. & Lemoine, H. (2004). *Patrimoine sonore et audiovisuel français : entre archive et témoignage : guide de recherche en sciences sociales*. Paris, Belin, 7 vol., 1 CD-Rom, 1 DVD-Rom.

Cameron, D., Frazer, E., Harvey, P., Rampton, M. & Richardson, K. (1991). *Researching Language : Issues of Power and Method*. London, Routledge.

Condamines, A. (ed.) (2006). *Sémantique et corpus*. Paris, Hermes.

Cresti, E. & Moneglia, M. (eds.) (2005). *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages.* Amsterdam, Benjamins.

Cribier, F. & Feller, E. (2003). *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »*. Rapport au Ministre délégué à la Recherche et aux nouvelles technologies, dactylogr., 2 vol.
http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf

Encreve, P., & Fornel (de) M. (1983). Le sens en pratique. *ARSS* 46, L'usage de la parole.

Gadet, F. (2003). *La variation sociale en français*. Paris, Ophrys.

Guilhaumou, J., Mesini, B. & Pelen, J.N. (1997) Récits de vie. Dynamiques et autonomies des récits de vie dans le champ de l'exclusion. *Cahiers de littérature orale*, 41, 91-126.

Gumperz, J.J., & Hymes, D. (eds.) (1972). *Directions in Sociolinguistics : The Ethnography of Communication*. New-York, Hold, Rinehart & Winston.

Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*. Paris, A. Colin.

Jacobson, M. (2004). Corpus oraux en linguistique de terrain. *Traitement Automatique des Langues*, 45/2, 63-88.

Jacobson, M. (2004). Les archives sonores au LACITO. *Bulletin de liaison de l'AFAS* 26 (http://afas.mmsh.univ-aix.fr/bulletin/Bulletin AFAS 26.pdf).

Joutard, P. (1979). Historiens, à vos micros. Le document oral, une nouvelle source pour l'histoire. *L'Histoire*, 12, 106-113.

Kennedy, G. (1998). *An introduction to Corpus Linguistics*. Londres, Longman.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphie, University of Pennsylvania Press.

LAMY, Droit de l'informatique et des réseaux (S. Marcellin, L. Costes & *al.* eds., Paris, 2004).

Leech, G. (1992). The state of the art in corpus linguistics. In Aijmer & Altenberg (eds.), 8-29

Mondada, L. (1998). Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête. Actes du Colloque de Lausanne (13-14.12.1998), *Cahiers de l'ILSL* 10, 39-68.

Mondada, L. (2006). Video recording as the reflexive preservation-configuration of phenomenal features for analysis. In Knoblauch, H., Raab, J., Soeffner, H.G., Schnettler, B. (eds.)

Mondada, L. (à paraitre) « La demande d'autorisation comme moment structurant pour l'enregistrement et l'analyse des pratiques bilingues », *Tranel*, Université de Neuchâtel.

Quéré, L. & *al.* ed. (1984) *Arguments ethnométhodologiques*, Paris, Centre d'Étude des Mouvements Sociaux, EHESS.

Recherches sur le Français Parlé 5 (1984). Pourquoi le français parlé est-il si peu étudié ?.

Revue Française de Linguistique Appliquée (1996) I-2, (1999) IV-1.

Sacks, H. (1984). Notes on methodology. In J.M. Atkinson & J. Heritage (eds.), 21-27.

Sankoff, D., Sankoff, G., Laberge, S. & Topham, M. (1976). Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale. *Cahiers de Linguistique* 6, 85-125.

Shaffir, W.B. & Stebbins, R.A. (eds.) (1991). *Experiencing Fieldwork : An inside View of Qualitative Research*. Londres, Sage.

Silverman, D. (ed.) (1997). *Qualitative Research. Theory Method and Practice*. Londres, Sage.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Londres, OUP.

Sinclair, J. (1996). *Preliminary recommendations on corpus Typology*. Technical Report, Eagles.

Sinclair, J. & Coulthard, R.M. (1975). *Towards an Analysis of Discourse*. Londres, OUP.

Speech Communication (2001) Speech Annotation and Corpus Tools. Vol. 33, 1-2, S. Bird & J. Harrington (eds.).

Welland, T. & Pugsley, L. (eds.) (2002). *Ethical Dilemmas in Qualitative Research*. Aldershot, Ashgate.

Titre	***Du Français Fondamental aux ESLO***
Type	Actes
Editeur	EME-Louvain
Année	2009
Référence	Abouda, L., **Baude, O.** (2009) «Du Français Fondamental aux ESLO», in Bruxelles, Mondada, Simon, Traverso «Grand corpus de français parlé, Bilan historique et perspectives de recherche, *Cahiers de Linguistique Revue de sociolinguistique et de sociologie de la langue française 33/2*, EME, Louvain, 131-146.

Du Français Fondamental aux ESLO

ABOUDA, Lotfi & BAUDE, Olivier
Université d'Orléans

L'Enquête Socio-Linguistique à Orléans (ESLO), réalisée en 1968-71, marque une rupture méthodologique et théorique avec le Français Fondamental sur différents points qui sont représentatifs des différents usages des corpus oraux en linguistique. Ainsi, la définition des enjeux de l'exploitation scientifique - numérisation, transcription, annotation, diffusion, analyses - du corpus ESLO conçue comme étape liminaire à la réalisation d'une nouvelle enquête variationniste en cours d'élaboration (ESLO2), permet d'interroger l'évolution des modèles et des méthodes de constitution et d'exploitation des corpus oraux destinés à des finalités linguistiques et didactiques.

Mots-clés : corpus oraux, linguistique variationniste, numérisation, transcription

The "*Enquête Socio-Linguistique à Orléans*" (ESLO), conducted from 1968 to 1971, marks a methodological and theoretical break with the "*Français Fondamental*" on the different levels corresponding to the different uses made of oral corpora. Hence, re-defining the scientific stakes - digitalization, transcription, annotation, distribution, analyses - involved in the analysis of the ESLO corpus (in fact the first stage of a new project of variationist data collecting, ESLO 2) will allow us to question the models and methods employed in constituting and exploiting oral corpora intended for linguistic and didactic purposes.

Keywords : Spoken corpora, variationist linguistics, digitalization, transcription

0. Introduction

Une dizaine d'années après la réalisation du Français Fondamental (FF), une équipe d'universitaires britanniques a entrepris la constitution d'un corpus de français oral, à visée à la fois linguistique et didactique : l'Enquête Socio-Linguistique à Orléans (désormais ESLO).

L'ESLO marquera une rupture avec plusieurs décisions ayant guidé la réalisation du Français Fondamental : souci d'une identification sociologique raisonnée en termes de CSP, conservation des enregistrements, adaptation du corpus à des interrogations multiples, préservation de la cohérence discursive, réflexivité de l'enquête, observation de l'interaction et des conduites linguistiques…

L'objet de cet article sera de présenter les principales étapes du travail de reconstruction d'ESLO1, actuellement en cours au sein du Centre Orléanais de Recherche en Anthropologie et Linguistique (CORAL), reconstruction vue comme une étape liminaire à la réalisation d'une nouvelle enquête variationniste à Orléans, ESLO2.

En comparant ESLO1 reconstruit au FF, nous pouvons mesurer tout le chemin parcouru en cinquante ans, ce qui ne se résume pas à une simple évolution technique, même si celle-ci y a joué un rôle décisif.

1. Du français fondamental à l'ESLO

1.1. Situations et contextes historiques

Souvent comparé, voire vu comme une réponse, au *Basic English*, le « Français Fondamental » (FF) a été créé au début des années 1950, pour promouvoir le français dans des colonies qui allaient bientôt devenir indépendantes.

La décision de recueillir du français oral, située dans son contexte épistémologique – très largement dominé par l'écrit –, peut s'expliquer par l'évolution de l'enseignement du français langue étrangère (FLE). En effet, l'efficacité de l'apprentissage du FLE, devenu une véritable technique évaluable en termes de coût, « suppose, ainsi que l'écrivent Bergounioux et al. (1992, p. 75-76), une méthodologie particulière, fondée sur des relevés quantifiés moins dépendants de la littérature classique ».

En 1966, un groupe de linguistes britanniques se réunit pour établir un bilan de l'enseignement du FLE en Angleterre. Si, dans ce pays, le rôle de l'oral dans l'enseignement des langues étrangères était reconnu depuis longtemps, il manquait pour le français un ensemble cohérent, systématique et actualisé de matériaux pédagogiques.

L'inexistence, y compris en France, de matériaux exploitables poussa ces chercheurs à entreprendre la constitution d'un nouveau corpus. L'entreprise, qui dura plus de cinq ans entre la conception et la réalisation, donna naissance à l'un des corpus les plus vastes de français oral : 350 bandes magnétiques représentant quelques 317 heures d'enregistrements, ce qui correspond à quelques ± 4 500 000 mots.

Mais, ainsi qu'on le verra, la taille de ce corpus ne constituera pas son seul intérêt.

1.2. A la recherche de données spontanées et situées

Si les objectifs déclarés de l'ESLO sont multiples, sa principale raison d'être reste tout de même didactique – l'enseignement du FLE –, objectif qu'elle partage avec le FF. Mais au-delà de toutes les différences didactiques, nombreuses, qui opposent les deux projets, la principale opposition concerne la nature elle-même des matériaux recueillis.

En effet le FF, s'il a bien intégré la langue orale, le fait dans le but de configurer un état moyen du français. L'oral a été ici homogénéisé, figé, fragmenté, ramené à des moyennes, avant de totalement disparaître.

Sur ce point, les initiateurs d'ESLO ont pris des décisions diamétralement opposées, en choisissant de représenter la variété des usages.

Cette variété des usages, revendiquée – dans le champ alors balbutiant de la sociolinguistique – et concrètement intégrée dans ESLO comme autant de variables, concernait aussi bien les différences générationnelles et communautaires, que les différences sociales, sans négliger les différences liées aux conditions de production du discours (Blanc & Biggs 1971, p. 16).

1.2.1. Quand le « Portrait sonore d'une ville » cache une « communauté d'auditeurs »

Pour faire accepter l'enquête auprès de la population locale, les enquêteurs ont dû mettre en avant le concept de « portrait sonore d'une ville ».

En réalité, le choix d'Orléans fut, sur le plan géographique, un non choix. Pour écarter des variables incontrôlables, il fallait une ville, de la taille d'Orléans, dont on pouvait reconstruire la dynamique des formes linguistiques simultanément présentes, une cité « assez vaste pour que la variation y soit accusée et perpétuée à travers des réseaux linguistiques d'échanges autonomes, et assez restreinte pour que n'importe quel membre de cette communauté linguistique ait dû interférer dans les circuits de communication des autres groupes. » (Bergounioux et al. 1992, p. 79). Cela écarte l'hypothèse qui ferait du choix d'Orléans celui d'une forme standardisée du français, représenté par une variété régionale non marquée. Et de fait, on rencontre parmi les locuteurs enregistrés un nombre non-négligeable de témoins qu'une enquête dialectologique aurait écartés[1]. L'accent de certains de ces témoins était vu par les enquêteurs non pas comme une imperfection, mais comme un intérêt supplémentaire de l'enregistrement.

Ces faits appuient fortement l'hypothèse formulée par Bergounioux et al. (1992) quand ils écrivent (op. cit.) : « C'est une communauté d'auditeurs qui est construite, autant qu'une communauté de locuteurs, à notre connaissance pour la première fois en France. » En bref, il s'agissait pour les enquêteurs de saisir les variétés du français qu'on pouvait entendre à un moment donné dans un lieu donné, non marqué géographiquement.

A l'intérieur de ce cadre, l'objectif des enquêteurs n'était pas de rechercher « cet individu mythique, l'Orléanais moyen » (Blanc & Biggs 1971, p. 23), mais d'appréhender dans toute son hétérogénéité une communauté de locuteurs-auditeurs.

De telles exigences nécessitaient le recours à une véritable enquête linguistique dont la méthodologie était à l'époque encore à construire.

[1] Trois Pied-Noirs d'Algérie, deux Aquitains, deux Bretons, un Lorrain et dix Parisiens.

1.2.2. Panel et échantillonnage

Une fois la ville choisie, il fallait affronter le délicat problème de la détermination de l'échantillon de population sur lequel porterait l'enquête.

Les initiateurs d'ESLO, retenant la méthode de l'échantillonnage au hasard, ont fait appel à l'INSEE pour réaliser une sélection aléatoire de 600 témoins, répartis, en plus du sexe, en 3 tranches d'âges (18/30, 31/50, 51 et +) et 6 catégories socioprofessionnelles. Cet échantillon ne se voulait pas représentatif mais diversifié et offrant un nombre égal suffisant de témoins pour une *étude linguistique*.

Si, sur le principe, la démarche se distinguait nettement de celle du FF, les résultats concrets n'étaient pas à la hauteur des espérances, puisque le taux de refus, prévu autour de 50%, a été très largement dépassé (au final, ils n'ont obtenu que 147 entretiens, soit moins de 25%). Ce rendement, faible, ne pouvait que nuire à l'équilibre de l'échantillonnage (par exemple, il y a eu peu d'ouvriers, dont la plupart étaient de surcroît rompus à la prise de parole publique).

Mais la technique d'échantillonnage, tout en ouvrant la porte à une enquête rigoureuse, ne pouvait qu'endommager les méthodes d'enquêtes.

Ainsi, dans un même souci de comparabilité, l'équipe privilégiera la situation de l'entretien en face à face, situation certes très formelle, mais qui avait l'avantage d'être pour eux contrôlable. « *Les mêmes questions sont posées par les mêmes personnes dans les mêmes conditions* », écriront Blanc & Biggs (op.cit., p. 17).

1.2.3. A la recherche de formes spontanées

Cette volonté de rigueur méthodologique dans le contrôle de la situation et dans la restriction des variables s'est confrontée à l'objectif de recueillir du français spontané dans un bel exemple de ce que la sociolinguistique nommera par la suite "le paradoxe de l'observateur"[2].

Les initiateurs d'ESLO chercheront alors à compléter le corpus d'entretiens par le recours à des situations plus naturelles, qu'ils paraissaient toutefois ne pas totalement maîtriser, à la fois sur le plan de la méthodologie de l'enquête et sur celui de la qualité technique de l'enregistrement.

Ces emplois plus spontanés représentent in fine 2/3 des enregistrements regroupés en cinq catégories:

1. des reprises de contacts informelles (15 reprises de contacts dans des situations variées : discussion entre amis,…)
2. enregistrements au hasard en micro caché (rue, magasins, etc.)
3. interviews de personnalités du monde syndical, politique, universitaire et de l'administration d'Orléans (maire, évêque, …)
4. tables rondes, conférences-débats (sur des thèmes variés : condition de la femme, promotion sociale,…)

[2] Labov 1973.

5. entretiens au Centre Médico Psychopédagogique d'Orléans (entretien entre une assistance sociale et des parents).

1.2.4. De Bernstein à Bourdieu

A ce tâtonnement (technique et théorique) dans la recherche du recueil de français spontané en interaction, s'est ajouté celui d'une sociolinguistique applicable à l'enquête.

L'entretien en face-à-face a été élaboré sur la base de 3 questionnaires : le premier (ouvert) devait permettre de recueillir les positions du témoin sur son expérience personnelle et divers types de discours déterminés. Le second, semi-fermé, intitulé "questionnaire socio-linguistique" et confié à un élève de Pierre Bourdieu – Bernard Vernier –, s'il porte encore les traces des théories de Bernstein sur la langue[3], enferme un recueil des représentations du témoin sur la norme linguistique et culturelle. La sociologie naissante de Bourdieu se rencontre également dans le troisième questionnaire, fermé, qui porte, parallèlement à l'état civil, sur les pratiques déclarées des habitudes culturelles.

Cette importance donnée au capital culturel s'est également concrétisée dans l'élaboration d'une nouvelle grille, l'échelle AM (de son concepteur Alix Mullinaux, qui comprend cinq agrégats (notés de A à E). Complémentaire à celle de l'INSEE, cette grille tente de rendre compte, parallèlement aux critères habituels, des pratiques et références culturelles ainsi que de la mobilité géographique potentielle des témoins.

Ce n'est pas le moindre des intérêts du projet ESLO que de porter les traces de la fin d'une sociolinguistique militante (et naïve) qui se bornait à corréler variations sociales et hiérarchisation de la compétence linguistique, et les prémices d'une nouvelle sociologie qui donnera toute sa place à la distinction apportée par le capital culturel.

1.3. Disponibilité des données

1.3.1. Français fondamental et données volatiles.

Le français fondamental était fondé dès son origine sur le constat de la pauvreté des enregistrements du français parlé disponibles, que ce soit dans les fonds des institutions de conservation du patrimoine que dans ceux des archives scientifiques[4].

Mais, lorsqu'ils ont entrepris de recueillir leurs propres données, ils n'ont pas hésité à effacer les enregistrements audio au fur et à mesure qu'ils les ont transcrits.

Comment expliquer ce choix, étonnant, qui revenait de fait à dénier à ces enregistrements toute reconnaissance comme objet scientifique et patrimonial ?

[3] « L'origine sociale et le niveau d'éducation du témoin concourent avec ses activités socioprofessionnelles à déterminer sa compétence linguistique. Selon cette hypothèse plusieurs types d'attitudes seraient en corrélation avec les niveaux socio-culturels ». Blanc & Biggs (op. cit., p. 18)

[4] « Entre le peu d'intérêt du Musée de la Parole pour des enregistrements de français, la dominante folklorique des documents conservés au Musée des arts et traditions populaires (2 documents exploitables) et le caractère gourmé des interventions en dépôt aux Archives de la radio (6 documents exploitables), la moisson fut maigre ». Bergounioux et al. (1992, p. 76).

La première raison est purement technologique. Les enregistrements du Français Fondamental ont été réalisés sur des disques de papier magnétique particulièrement fragiles[5], et considérés comme coûteux.

Mais la vraie raison est sans doute ailleurs : l'équipe du FF pensait que la transcription, qu'elle voulait rapide, était un moyen de conserver ce corpus, sans que ne soit évoquée la question de la réutilisation des données primaires, par d'autres ou par l'équipe elle-même, ne serait-ce que par soucis de contrôle.

1.3.2. L'ESLO : des données résolument disponibles

Sur ce point, aussi, le projet ESLO exposera des choix diamétralement opposés. Il s'agissait bel et bien de constituer un corpus *disponible* pour de multiples travaux, aussi bien en linguistique qu'en didactique du FLE, dans une démarche inhabituelle en France.

La conservation et la possibilité de réutilisation des matériaux recueillis étaient dès le départ considérées par les initiateurs d'ESLO comme deux objectifs fondamentaux. Ce choix s'est concrétisé de différentes manières.

1. D'abord par le **catalogage** et l'**indexation** : l'équipe de d'ESLO a publié en 1974 un catalogue descriptif et analytique[6] qui répertoriait les enregistrements avec : résumé du contenu, indexation des questions, organisation du questionnaire, catégorisation sociologique précise des locuteurs et description de la situation d'enquête.
2. **La conservation des données primaires (enregistrements et documents d'enquête).**
3. **Les transcriptions.** Bien qu'une transcription intégrale fût difficilement envisageable pour un corpus estimé à plus de 4 millions de mots, l'équipe a entrepris immédiatement la transcription d'extraits qui se voulaient représentatifs et qui recouvraient toutes les catégories des témoins (INSEE et AM).
4. **La diffusion du corpus**

   Outre l'annonce systématique dans les articles de la disponibilité du corpus, le catalogue précise dès la page 4:

> "*Les transcriptions et enregistrements sont disponibles à tout chercheur intéressé, contre remboursement des frais de matériaux et de copiage; (…) Des listes de transcriptions et enregistrements sont disponibles à ceux qui s'adressent à nous.*" (Lonergan,et al.,1974, p. 4).

[5] Cependant cette limitation de la technologie est presque présentée comme un avantage : « L'appareil Recordon que nous avons choisi présentait pour notre travail des avantages certains. Son poids était peu élevé : un peu plus de 6 kg. […]. L'interruption dans les conversations que nous imposait le changement des disques toutes les six minutes, au terme de leur durée maxima d'enregistrement, ne présentait pas non plus d'inconvénient. Nous ne nous souciions pas non plus de la conservation des disques. Nous profitions largement des possibilités qu'offrent les disques en papier d'être effacés et de servir ainsi à plusieurs enregistrements successifs. Il aurait été beaucoup trop coûteux de conserver tous les enregistrements comme beaucoup de bons esprits nous le suggéraient » Gougenheim et al. (1964, p. 63).

[6] Il s'agit d'un important volume de 218 pages, reproduit en 1993 par l'université d'Orléans.

1.4. Un corpus peu exploité

Les initiateurs d'ESLO auront donc tout fait pour rendre leur corpus disponible, et exploitable. Cela n'a pas empêché que son exploitation didactique fût faible[7], et son exploitation scientifique pendant longtemps quasi inexistante, en tout cas en France.

Le diagnostic de cet échec – eu égard à l'ambition initiale du projet – fait apparaître plusieurs causes.

D'abord, ESLO constitue un corpus particulièrement encombrant et lourd à manipuler : à part le catalogue qui a été dactylographié, la plupart des documents étaient manuscrits, et les transcriptions ne concernaient qu'une petite partie d'un corpus dont les enregistrements représentent plus de 300 bandes magnétiques .

Ensuite, ESLO n'a été que très partiellement transcrit. La transcription d'un corpus d'une telle taille constitue un défi colossal. En plus de l'absence de cadre théorique sur la transcription du français parlé – le travail théorique sur cette question débutera véritablement avec les travaux de Claire Blanche-Benveniste dans les années 1980 –, le temps de transcription peut être estimé à 20 000 heures de travail pour un coût qui dépassait de loin les moyens dont pouvaient disposer une équipe. Il n'y aura donc pas de transcription de l'intégralité du corpus.

Enfin, et surtout, l'absence d'exploitation de ce corpus peut s'expliquer par une raison épistémologique profonde : le français parlé est loin d'être un domaine légitime dans le champ de la linguistique. Cette situation constatée à plusieurs reprises notamment par Claire Blanche-Benveniste (Blanche-Benveniste et Jeanjean (1987))[8] est due à trois raisons principales. Premièrement, la linguistique en tant que discipline universitaire a *incorporé* au sein même de son organisation, la domination symbolique de la forme écrite de la langue en faisant la part belle à la grammaire normative (Bergounioux 1992). Deuxièmement, la linguistique s'est construite sur une lecture des dichotomies proposées par Ferdinand de Saussure qui a située, hors de la discipline, la description des variations en général et des formes orales en particulier. L'étude du français parlé s'est alors trouvée exclue du champ d'une science vouée à la recherche d'invariants structurés en système. Enfin, l'histoire même du français en France confirme la place attribuée à l'écrit, longtemps considéré comme seul véritable capital symbolique légitime.

2. Reconstruire ESLO1, penser ESLO2

L'apparition des nouvelles technologies et du traitement informatique des corpus va relancer des possibilités d'exploitation du corpus ESLO. Pour les raisons épistémologiques évoquées plus haut, il n'est pas étonnant que celles-ci se fassent en dehors de la France.

[7] On relève seulement deux utilisations du corpus dans le domaine du FLE : (i) la méthode anglaise diffusée sous le titre « Les Orléanais ont la parole » (livre du maître, livre de l'élève et support de cours sur bandes), et (ii) la méthode du BELC (12 cassttes).

[8] « Mais qui s'intéresse au français parlé ? […] peu de gens y voient un objet légitime d'étude (même chez les linguistes) pour bon nombre de ceux-ci la langue parlée c'est bon pour l'exotisme ; la description de la langue parlée vaut pour les dialectes et les patois du français ; elle vaut aussi pour les langues sans écritures dites "exotiques" ; mais pas pour une langue de culture comme le français. » Blanche-Benveniste et Jeanjean (1987).

On doit en effet aux universités de Louvain et d'Amsterdam, dans le cadre des deux projets ELILAP puis ELICOP, la diffusion d'un corpus informatisé[9] comprenant une partie du corpus d'Orléans transcrit, avec un étiquetage morphosyntaxique et un concordancier. La partie disponible représente près de 80 heures (900 000 mots) de transcription orthographique et une dizaine d'heures de transcription phonétique[10].

Depuis, ce corpus a donné lieu à de nombreuses exploitations et continue d'être utilisé et développé à l'université de Louvain[11].

Le CORAL, qui détenait l'intégralité des enregistrements originaux, a entrepris récemment la numérisation des bandes magnétiques à des fins de conservation et de diffusion.

Le traitement en 2005 d'un corpus vieux de 35 ans n'est pas une chose aisée et implique une véritable reconstruction, reconstruction d'autant plus nécessaire qu'elle devait répondre à un objectif de comparabilité avec une nouvelle enquête sociolinguistique à Orléans, ESLO2.

Cette nouvelle enquête, en cours d'élaboration au sein du CORAL[12], consiste à constituer un corpus suffisamment analogue à ESLO1, y compris sur le plan quantitatif, et en même temps adapté à la situation contemporaine. L'enquête portera sur 200 témoins, les enregistrements débuteront fin 2006.

Reconstruire ESLO1, c'est donc aussi penser ESLO2 avec l'évaluation de 35 ans d'évolution technologique, méthodologique et théorique. Une évaluation qui permet d'anticiper les questions de conservation, de transcription, d'annotation, de structuration des métadonnées, de multi-exploitation et surtout l'impact de ces différents choix sur l'analyse linguistique.

2.1. Quand la numérisation transforme l'objet scientifique

La conservation d'archive consiste à trouver les garanties d'une non altération du support d'origine et, à défaut – systématique à ce jour, à dupliquer l'original sur un nouveau support avec le moins de perte possible. En ce qui concerne les bandes magnétiques d'ESLO, la copie devenait indispensable – 35 ans est une durée critique pour ce type de support. Or, la numérisation ne consiste pas simplement en un changement de support. La digitalisation transforme l'objet en en facilitant la manipulation et la diffusion, mais surtout en offrant des possibilités de traitement informatique des données primaires et des métadonnées. Cependant ces traitements demandent de repenser la structure du corpus et des objets qui le composent.

[9] Dans le cadre d'un projet de recherche mené de 1980 à 1983 (Le français parlé. Banque de données automatisée ; analyse linguistique fondamentale et applications, sous la direction de Josse De Kock, Mark Debrock, Nicole Delbecque et Ellen Bas), le Département de Linguistique de la K.U.Leuven a reçu la gestion de l'ensemble de ces enregistrements, soit près de 500 heures. Ce premier projet est connu sous le sigle ELILAP (*Etude LInguistique de la LAngue Parlée*). Les responsables du projet ont voulu rendre accessibles les données sous forme informatisée. Plusieurs parties des trois corpus ont été transcrites et sont actuellement disponibles sous forme de transcriptions graphiques (± 100 heures) et phonétiques (± 12 heures) automatisées. L'ensemble des corpus compte actuellement plus d'un million de mots et peut être considéré comme constituant un échantillon représentatif de la langue parlée

[10] http://bach.arts.kuleuven.be/elicop/

[11] autour du Professeur Piet Mertens.

[12] en partenariat avec le CELITH et MODYCO

Après la première phase consacrée à la numérisation des enregistrements selon les normes en cours[13], il a fallu déterminer les opérations de traitements applicables à ces données informatiques. De fait, la chaîne de traitement élaborée doit permettre de transcrire les données primaires pour pouvoir disposer de tous les outils développés dans le domaine du TAL et de la linguistique de corpus. En ce sens la transcription doit suivre des conventions suffisamment proches des corpus écrits afin de permettre notamment un étiquetage morphologique, prosodique, syntaxique ou tout autre annotation. Nous reviendrons par la suite sur les effets de cette normalisation de l'oral pour les outils de la linguistique de corpus pour nous intéresser maintenant à un effet de la numérisation qui est souvent ignoré : le rôle des métadonnées[14].

2.2. Numérisation des métadonnées : la réapparition du locuteur

2.2.1. Traitement informatique du témoin

Numériser un corpus implique également de traiter différemment la documentation et les descriptions qui peuvent éventuellement accompagner celui-ci et qui sont, nous l'avons souligné, volontairement très riches dans le cas d'ESLO. Dans le cas d'un corpus numérique, il est aisé d'établir des relations entre les données primaires et les principes d'élaboration du corpus, la normalisation et les formalismes choisis, les techniques utilisées et de nombreuses autres informations (ou méta-informations). Or, cette documentation est notamment le lieu pour fournir de précieux renseignements sur la situation de collecte et le profil des témoins. Cette opération a été repérée comme étant fondamentale depuis le développement de la linguistique de corpus : " *La documentation doit couvrir deux volets distincts : les sources utilisés et la responsabilité éditoriale de constitution du corpus d'une part, les conventions d'annotation d'autre part*" (Habert et al.1997[15]). Récemment, le langage XML apporte une solution convaincante en séparant les données et les informations sur la structure des données, alors décrites dans l'en-tête du document (recommandations de la TEI[16]).

La gestion de ces informations souvent répertoriées sous le terme de métadonnées rend nécessaire une uniformisation du traitement comme le proposent différentes initiatives centrées sur la gestion, la diffusion et la réutilisation des corpus (EAGLES[17], OLAC[18]).

Dans le cas du corpus d'ESLO, nous disposons d'un exemple particulièrement intéressant car les métadonnées avaient déjà été répertoriées pour la publication du catalogue en 1974[19]. Or, la transformation du catalogue en une base de données offre des perspectives infinies de requêtes dans d'excellentes conditions. Ainsi, les données sociologiques ont été intégrées à des bases de données relationnelles et deviennent facilement disponibles comme champs que

[13] Les choix qui ont prévalu à la numérisation des enregistrements d'ESLO correspondent aux recommandations de l'IASA (association internationale d'archives sonores et audiovisuelles) et de l'AFAS[13] qui diffusent en France ces recommandations. Une copie droite de chaque bande a été réalisée en WAV, 44100 htz, 16 bits. Une attention particulière a été apportée aux choix de codage ouvert, de format non propriétaire, de normes pérennes et de l'évolution de la technologie (fréquence largement supérieure à la fréquence d'enregistrement initiale).

[14] Pour une présentation des méthodes de constitution de corpus oral numérique : Delais-Roussarie 2000.
[15] Habert et al. 1997, p.156.
[16] Text Encoding Initiative
[17] EAGLES, 1996, Preliminary Recommendations on Spoken texts
[18] Open Language, Archive Community
[19] Catalogue ESLO, 1974

l'on peut croiser avec des requêtes sur la transcription et l'annotation des données linguistiques.

Outre les possibilités techniques de requêtes croisées, nous avons tenu à réintroduire le locuteur dans les données primaires, continuant ici notre démarche qui a consisté à rendre indissociable la transcription à la voix du locuteur.
Le corpus reprend ainsi son statut de données situées par la réapparition du locuteur dans les corpus oraux, et par la reconstruction du profil sociologique de ce locuteur comme témoin aux caractéristiques bornées par l'échantillonnage et l'enquête sociologique.

Pour ESLO2, le travail théorique sur la sociologie applicable à une enquête linguistique débute. Cependant, la gestion des métadonnées d'ESLO1 permet d'ors et déjà de penser que non seulement une intégration en nombre de ce type de données est réalisable, mais qu'il y a ici la possibilité de rendre à la linguistique la méthodologie d'une véritable science des données attestées et situées (depuis ESLO1 l'apport de la sociolinguistique et de l'ethnométhodologie modifie la donne).

2.2.2. Reconnaissance juridique du locuteur (protection des données personnelles, propriété intellectuelle)

La reconnaissance du locuteur passe également par celle de son statut. La reconstruction du corpus d'Orléans 35 ans après a donné aussi lieu à un travail sur les aspects juridiques de l'exploitation de corpus. Ce travail a été mené parallèlement aux initiatives actuelles et notamment le groupe de travail sur le *Guide des bonnes pratiques* pour la constitution et l'exploitation des corpus oraux[20].

Le cas du corpus d'Orléans est un cas d'école : aucune autorisation n'avait été demandée à l'époque, ni sur les enregistrements, ni sur leur exploitation. Le corpus contient des données privées (nom, profession), des données sensibles (récit de vie nominatif, préférence politique, religieuse, enregistrements confidentiels au CMPP). De plus, les différentes phases d'exploitation multiplient les questions de propriétés (Essex, Orléans, Louvain,…).

A ces questions juridiques, le CORAL a apporté un certains nombres de réponses :

- les enregistrements seront anonymisés (bipage en temps réel des données personnelles, lors de la consultation)
- la structure de la base de données permet différents niveaux d'accès (toutes données pour un certain type de recherches selon une charte de confidentialité, et pour la conservation, données partielles pour une diffusion du corpus à la communauté scientifique, et données publiques pour une diffusion large).
- Le CORAL, qui a obtenu des financements de l'Etat pour ce programme, s'est engagé à suivre les règles de mise à disposition des corpus pour la communauté scientifique et pour les institutions patrimoniales. Il sera ainsi librement disponible et déposé à la BnF.

Là encore l'expérience d'ESLO1 nous as permis d'entreprendre un travail en profondeur sur l'élaboration des autorisations que rempliront les enquêtés afin de permettre le recueil d'un consentement le plus éclairé possible. Les opérations d'anonymisation et la gestion des droits de propriétés intellectuelles seront traités uniformément sur les deux corpus.

[20] Baude O (coord.) 2006,

2.3. Transcription synchronisée : la réapparition du locuteur - suite

2.3.1. La transcription synchronisée et le retour à la source

Nous l'avions déjà précisé, la difficulté la plus importante rencontrée par les initiateurs d'ESLO 1 a été l'ampleur de la tâche de transcription. Sur ce point aussi, et même principalement, l'avancée technologique bouleverse l'objet scientifique.

Depuis quelques années, alors que la manipulation du son numérique devenait très aisée (capacité de stockage, rapidité d'accès, débit suffisant pour une transmission en réseau…), des logiciels permettent la synchronisation du son et de la transcription (*Praat*, *Transcriber*, *Winpitch*, *soundedit*, etc.).
Ces innovations ont des répercussions méthodologiques importantes sur le travail du linguiste.

D'abord, les outils du traitement automatique des corpus écrits deviennent utilisables sur des données orales, qui d'un coup rattrapent 25 ans de recherche. Ensuite, avec des transcriptions alignées sur le signal sonore, l'oral devient physiquement l'objet d'étude et est systématiquement disponible en même temps que la transcription. Le retour aux données peut alors être systématique, ce qui est de nature à faciliter les procédures de vérification, étape essentielle du travail scientifique, malheureusement souvent rendue impraticable de par l'inaccessibilité des corpus.

Parallèlement, la synchronisation, qui permet l'annotation de segments temporels, offre une base de référence pour de la multi annotation et donc de la multi transcription. On peut concevoir, pour un même segment, une multitude de transcriptions, opérées dans des cadres théoriques distincts et/ou avec des granularités différentes, dont chacune répond à un besoin scientifique spécifique. Ici, la transcription n'est plus la vérité d'un chercheur (au mieux) ou d'un transcripteur, elle devient cumulative.

2.3.2. La transcription de degré 0 comme alternative à la transcription figée

Face à l'ampleur de la tâche, les choix actuels du CORAL ont été fondés sur la volonté de mettre à disposition une transcription de l'intégralité du corpus le plus rapidement possible sans que celle-ci n'implique une théorie linguistique très déterminée (même si toute transcription est une formalisation impliquant une théorie).
Nous avons conçu cette première transcription à un degré le plus proche du zéro, en lui donnant uniquement le statut d'outil de navigation au sein du corpus sonore. L'outil sélectionné a été *Transcriber* pour sa simplicité d'utilisation, sa robustesse face à des fichiers long, et sa sortie en un format de fichier XML qui nous a semblé être une garantie d'interopérabilité.

Les conventions de transcriptions ont donc été réduites au minimum. Cependant, même à ce niveau "zéro", de nombreuses questions restent présentes comme la structuration des segments et leur granularité – qu'est-ce qu'un mot ? une phrase ? un tour de parole ? –, le choix des évènements à transcrire, la gestion des chevauchements et des pauses.

Ce choix de transcription est actuellement testé sur des extraits de corpus[21]. Seule une évaluation rigoureuse des contraintes de ce choix sur les autres niveaux d'annotation (morphologique, syntaxique, prosodique, etc.) et sur les analyses linguistiques qui en découleront permet de le valider ou non.

3. Conclusion

Reconstruire ESLO1 consiste, avant toute possibilité d'analyse linguistique, à rendre le corpus disponible pour la communauté scientifique et implique l'anticipation de finalités qui n'ont pas été prévues lors de l'élaboration du projet ni même lors de l'étape de numérisation réalisée actuellement. Penser ESLO2 répond au même objectif même si la compatibilité entre deux corpus ne nous est pas apparue suffisante pour élaborer une standardisation qui n'existe pas actuellement.

Derrière ces choix méthodologiques et techniques, l'enjeu du domaine même de la linguistique pointe. Il s'agit surtout pour nous de concevoir la linguistique comme nécessairement une linguistique de données tout en ne perdant pas l'occasion de rendre à la linguistique la prise en compte de la nature sociale de la langue.

Constituer des grands corpus n'est pas une fin en soi, d'autant que ceux-ci peuvent contenir des données normées ou normalisées en masse. L'engouement actuel pour les corpus oraux, s'il constitue une chance pour la linguistique, comporte un risque majeur, celui de normaliser les données orales plutôt que saisir consciemment de la variation. De même, le recours à d'autres méthodes que la prospection des données peut paraître nécessaire. On ne doit pas exclure a priori d'autres méthodologies complémentaires : on peut aussi s'autoriser à manipuler des données et les transformer en fonction d'hypothèses falsifiables.

Le corpus du Français Fondamental représentait déjà une prise de position sur l'objet de la linguistique. L'évolution des méthodologie du traitement des données a fait apparaître de nombreuses questions qui méritent d'être abordées sans concession au moment où le champ de la linguistique s'ouvre enfin aux corpus. L'objectif de l'analyse des données d'ESLO 1&2 est de participer à ce débat.

[21] Des membres du CORAL entreprennent actuellement, à partir de différents points de vue disciplinaires (sociolinguistique, syntaxe, TAL, phonologie, pragmatique, etc.), une série de recherches linguistiques qui visent un même objet, i.e. le corpus de l'omelette. Il s'agit d'une petite sous-partie du corpus ESLO1, composé des réponses de 90 témoins à la question « Comment faites-vous une omelette ? » soit au total 120 minutes environ.
Ce mini-corpus qui offre l'avantage de l'unité thématique joue pour nous le rôle de test : il s'agit de tester sur une petite échelle l'ensemble du travail que nous nous proposons d'entreprendre sur la totalité du corpus ESLO1 mais aussi sur ESLO2, qu'il s'agisse de la transcription, ou de la faisabilité de certains types de recherches linguistiques, en passant par la structuration de la base de données.

Bibliographie

Abouda L., 2004, « Deux types d'imparfait atténuatif », *Langue française*, 142, p. 58-74,

Association française des détenteurs de documents audiovisuels et sonores (AFAS) : [http://afas.mmsh.univ-aix.fr/] voir notamment Bradley, K. (dir) *Guidelines on the production and preservation of digital objects*. International Association of Sound and Audiovisual Archives.

Baude O., Jacobson M., Tchobanov A., Walter R., à paraître, « Interopérabilité des corpus sonores : le cas des corpus en français », *Colloque international Phonological variation : the case of French*, 25-27 août 2005, Tromso.

2004 : « Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques », *Actes du Colloque international du GRESEC « La publicisation de la science »* (Grenoble) : 7-11.

Baude O. (ed), 2006, *Corpus oraux. Guide des bonnes pratiques 2006*, Paris, Cnrs éditions – Orléans, PUO.

Bergounioux G., 1992, « Les enquêtes de terrain en France », *Langue française*, 93, p. 3-21.

Bergounioux G., Baraduc J., Dumont C., 1992, « L'Etude socio-linguistique sur Orléans (1966-1991), 25 ans d'histoire d'un corpus », *Langue française*, 93, p. 74-93.

Blanche-Benveniste C., Jeanjean C., 1987, *Le français parlé, transcription et édition*, Paris, Didier érudition.

Biggs P, Dalawood M., 1976, *Les orléanais ont la parole : Teaching Guide and Tapescript*, (livre du maître), Londres, Longman.

Biggs P, Dalawood M., 1976, *Les orléanais ont la parole : Teaching Guide and Tapescript*, (livre de l'élève), Londres, Longman.

Blanc M., Biggs P., 1971, « L'enquête sociolinguistique sur le français parlé à Orléans », *Le français dans le monde*, 85, p. 16-25.

Calas M.-F., Fontaine J-M, 1996, *La Conservation des documents sonores*, Paris, CNRS Editions.

, *Orléans Archive*, Language center, University of Esssex, Coldchester.

A. Condamine (éd.), 2005, *Sémantique et corpus*, Paris, Hermès.

Coste D. (dir), 1984, *Aspects d'une politique de diffusion du français langue étrangère depuis 1945*, Paris, Hatier.

Delais-Roussarie E. et Durand J. (ed), 2003, *Corpus et variation en phonologie du français, méthodes et analyses*, Toulouse, PUM.

EAGLES, 1996, Preliminary Recommendations on Spoken Texts, EAG-TCWG-SPT/P, Pise, Consiglio Nazionale delle Ricerche, Instituto di Linguistica Computazionale.

Encrevé P., 1976, « Labov, linguistique, sociolinguistique », in *Labov 1976*, Paris, éditions de Minuit.

Gougenheim G., Michéa R., Rivenc P., Sauvageot A., 1964, *L'élaboration du français fondamental (1e degré)*, Paris, Didier.

Habert, B., et al., 1997, *Les linguistiques de corpus*, Paris, Armand Colin.

Mertens P., 2002 « Les corpus de français parlé ELICOP : consultation et exploitation », in Binon, J., Piet; Elen, J., Mertens, P., Sercu, Lies (eds) (2002) *Tableaux Vivants*, Opstellen over taal-en-onderwijs aangeboden aan Mark Debrock. Leuven, Universitare Pers.

Mondada L., 2005, « L'analyse de corpus dans la perspective de la linguistique interactionnelle : des analyses de cas singuliers aux analyses de collections », In. A. Condamine (éd.), *Sémantique et corpus*.

Mondada L., 1998, « Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête », Actes du Colloque de Lausanne 13-14 décembre1998, *Cahiers de l'ILSL*, 10, p. 39-68.

Sinclair J., 1996, « Preliminary recommendations on corpus Typology », Technical Report, Eagles.

« Speech Annotation And Corpus Tools », A special issue of Speech Communication Volume 33, numbers 1-2, 2001, Edited by Steven Bird and Jonathan Harrington.

Wynne M., 2005, *Developing Linguistic Corpora : a Guide to Good Practice*, AHDS, http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm. Visité le 10 mai 2006.

Titre	***Corpus oraux, Guide des bonnes pratiques 2006* (version anglaise)**
Type	Livre (traduction)
Editeur	En ligne (HAL)
Année	2009
Référence	

VERSION ALLEMANDE

CORPUS ORAUX

Guide des bonnes pratiques 2006

coordonné par **Olivier BAUDE**

PRESSES UNIVERSITAIRES ORLÉANS

CNRS EDITIONS

Dieses Werk ist das Ergebnis der Überlegungen einer um Isabelle **de LAMBERTERIE** gesammelten Arbeitsgruppe.
Es wurde von Olivier **BAUDE** koordiniert.

Olivier **BAUDE** (*DGLFLF und CORAL – Universität Orléans*)
Claire **BLANCHE-BENVENISTE** (*EPHE und Universität der Provence*)
Marie-France **CALAS** (*DMF*)
Paul **CAPPEAU** (*Universität Poitiers*)
Pascal **CORDEREIX** (*BnF)*
Laurence **GOURY** (*CNRS – CELIA*)
Michel **JACOBSON** (*CNRS – LACITO*)
Isabelle **de LAMBERTERIE** (*CNRS-CECOJI*)
Christiane **MARCHELLO-NIZIA** (*CNRS-ILF und ENS-LSH-Lyon*)
Lorenza **MONDADA** (*ICAR, CNRS, Universität Lyon2*)

In Zusammenarbeit mit :

Gilles **ADDA** (*für das Kollectiv COPTE LIMSI-CNRS*), Michel **ALESSIO** (*DGLFLF*), Alain **CAROU** (*BnF*), Ibrahim **COULIBALY** (*CDF – Universität Grenoble*), Valérie **GAME** (*BnF*), Fabrice **MOLLO** (*CNRS-CECOJI*), Michel **RAYNAL** (*INA*), Jean **SIBILLE** (*DGLFLF*), Dominique **THERON** (*BnF*), Luc **VERRIER** (*BnF*).

Übersetzung : Gisèle **DESOTEUX**

VORSTELLUNG DER VERFASSER

OLIVIER BAUDE
Dozent in Sprachwissenschaften an der Universität Orléans, Mitglied des *CORAL* (EA-3850), Forschungszentrum von Orléans in Anthropologie und Linguistik. Sekretär des wissenschaftlichen Rates der Überwachungsstelle der sprachlichen Praktiken, *DGLFLF*, oberster Rat für die Überwachung des Französisches und der Sprachen Frankreichs.

CLAIRE BLANCHE-BENVENISTE
Verdiente Professorin, Praktische Hochschule von Paris und Universität der Provence. Forscherin im Gebiet der französischen Linguistik : geschriebene und gesprochene Sprache, Syntax, Morphologie, Aufbau von Korpora der gesprochenen Sprache.

MARIE-FRANCE CALAS
Generalkonservatorin des Erbes. Generalinspektorin der Museen, Direktion der Museen Frankreichs. Spezialist für das Tongebiet als umfangreiches interdisziplinäres Gebiet verstanden, das Geschichte, Verwaltung, Bewahrung und Aufwertung der gesprochenen und musikalischen Aufnahmen, der Töne der Umgebung, die heute ein integrierender Bestandteil des immateriellen Erbes sind, einbezieht.

PASCAL CORDEREIX
Oberkonservator der Bibliotheken. Verwalter der Tondokumente in der audiovisuellen Abteilung der *BnF*, Nationalbibliothek Frankreichs ; außerdem Vizepräsident des französischen Vereins der Besitzer von audiovisuellen und Tondokumenten (*AFAS*). Seine Grundtätigkeit richtet sich nach der Problematik der Archivistik des Tones.

LAURENCE GOURY
Forschungsbeauftragte im *IRD*, Forschungsinstitut für die Entwicklung, Mitglied des *CELIA*, Forschungszentrum der einheimischen Sprachen Amerikas), Terrainlinguistik und Typologie (insbesondere Kreolsprachen).

MICHEL JACOBSON
Informatikingenieur im Labor für « Sprachen und Kulturen mit verbaler Tradition » des nationalen Forschungszentrums. Mitverantwortlicher des Programms « Archivierung ». Spezialist für die Verwaltung der Korpora der gesprochenen Sprache.

ISABELLE DE LAMBERTERIE
Forschungsdirektorin im *CNRS*, verantwortlich für den Stab « Normativität und Informationsgesellschaft » des Forschungszentrums über die juristische internationale Zusammenarbeit (*CECOJI* – UMR 6224), Mitglied des Ethikkomitees des *CNRS*.

CHRISTIANE MARCHELLO-NIZIA
Professorin in Sprachwissenschaften an der *ENS-LSH* (Lyon), Leiterin des Instituts der Französischen Sprache (*CNRS*) : Linguistik, Geschichte, Geschichte des Französisches, Theorien der Sprachentwicklung.

LORENZA MONDADA

Professorin in Sprachwissenschaften an der Universität Lyon 2 und Mitglied des Labors *ICAR* (*UMR CNRS* 5191). Arbeitet im Gebiet der interaktionnellen Linguistik über die Korpora der gesprochenen Sprache in Interaktion so wie über die multimodale Analyse von Videokorpora.

VORWORT VON XAVIER NORTH,
GENERALDELEGIERTER FÜR DIE FRANZÖSISCHE SPRACHE UND DIE SPRACHEN FRANKREICHS

In der Wissenschaftsgeschichte oder in den kulturellen Politiken kommt es selten vor, dass gesamte rohe Angaben und ungenaue Stoffe sich in ein Wissensobjekt verwandeln. Die Veröffentlichung dieses Handbuches zählt zu solchen Momenten, denn es bietet jedem Forscher die Hilfsmittel, die « guten Praktiken » an, die ihm ermöglichen werden, diese Metamorphose durchzuführen : die Verwandlung mündlicher Produktionen in ein Korpus der gesprochenen Sprache, das studiert und aufbewahrt werden und infolgedessen seinen Platz in das kulturelle Erbe der Nation nehmen kann.

Die sprachlichen Produktionen von literarischen Werken oder geschichtlichen Dokumenten in ihrer *geschriebenen*, festen und endgültigen Form waren zweifellos immer im Kern der vom Kultusministerium eingesetzten Politiken, ob es sich um das Buch oder um die Archivalien handele. Erst ganz neulich kam man aber auf den Gedanken, dem lebendigen Aspekt der Sprache in ihr spontanes Sprudeln, in ihre tägliche, gewohnte Formulierung und in die außerordentliche Vielfältigkeit ihrer Mundarten Interesse zu widmen... So zeichnet sich für das erste Mal die Möglichkeit ab, ein echtes Archiv des Wortes auf eine gesicherte Grundlage anzulegen. Der Fortschritt der Technologien sollte dazu beitragen.

Ein Korpus der gesprochenen Sprache ist eigentlich keine einfache Sammlung von Aufnahmen der mündlichen Äußerungen des Menschen, es ist ein « entworfenes » Objekt : die Ausarbeitung der Daten (Digitalisierung, Transkribierung, Indexierung) erlaubt nicht nur ihre Bewahrung sondern gibt ihnen den neuen Status von Forschungs- und Aufwertungsobjekt. Man sollte sich wenigstens auf zusammenhängende und leicht anwendbare Methodevorschriften stützen können.

Das « Handbuch der guten Praktiken » öffnet der Neugierde der Forscher von nun an ein neues und umfangreiches Gebiet. Durch seine *Überwachungsstelle der sprachlichen Praktiken* hat der oberste Rat der Überwachung des Französischen und der Sprachen Frankreichs den Impuls gegeben und er hat dann die Energiequellen und die verschiedenen Ressourcen zusammengebracht und koordiniert, ob sie aus der Welt der Forschung kommen oder aus den durch diese Initiative betroffenen Horizonten des Kultusministeriums.

Die Entwicklung der Korpora der gesprochenen Sprache, ihre Verbreitung und ihre Bewahrung zu sichern, heißt auch, sie zugänglich zu machen, das französische linguistische Erbe in seiner Vielfältigkeit, in seinem Reichtum und in seiner Wahrheit hören zu lassen. Das heißt dazu, sich ein kostbares Instrument der Kenntnis der sprachlichen Praktiken zu geben, die für die Bestimmung der Sprachpolitiken aber auch der Erziehungs- und Sozialpolitiken nützlich sind.

Dieses Unternehmen hat mehrere Monate lang Juristen, Sprachwissenschaftler, Konservatoren und Informatiker zusammengebracht, die die neuen Wege der Kultur und der Forschung mit dem Respekt des Rechtes in Einklang bringen wollten. Das Ergebnis gemeinsamer Denkbemühungen stellen wir heute vor, in der Hoffnung, dass es seinerseits manche Arbeiten befruchtet.

VORWORT VON BERNARD MEUNIER, PRÄSIDENT DES CNRS

Das Mündliche und das Schriftliche. Diese zwei Wörter haben eine starke Aussagekraft. Wir denken an die Art, wie die Kulturen sich durch die gesprochene Praxis und dann durch das Schaffen von Schriften gebildet haben, die es erlaubten, die Worten von den einen oder den anderen durch Raum und Zeit besser zu übermitteln.

Mein Forscherblick über die respektive Rolle des Mündlichen und des Schriftlichen in die Verbreitung der wissenschaftlichen Kenntnisse lässt mich nicht vergessen, dass die mündliche Darstellung vor Berufskollegen oder vor ein breites Publikum weit über die erstgangige Rolle des Geschriebenen immer wesentlich ist, um zu verbreiten, überzeugen, Ideen teilen zu lassen. Die gesprochene Sprache behält eine Überzeugungskraft, die es erlaubt, die Mehrzahl zu erreichen, sobald sie aufgenommen und durch die aktuellen audiovisuellen Mittel übertragen werden kann.

Das Einsammeln und die Verwendung der Korpora der gesprochenen Sprache sollen im Respekt der « guten Praktiken » gemacht werden, wie für die Korpora der gechriebenen Sprache gemacht wird. Wir wissen alle, wie ein Satz, der aus seinem Kontext entnommen und hemmungslos verbreitet wird, für seinen Autor, eine Gruppe Leute oder eine Gemeinschaft gefährlich werden kann.

Die Verfasser dieser merkwürdigen Arbeit haben alle juristische Aspekte des Einsammelns und der Verwendung der Korpora der geschriebenen Sprache angesprochen. Ich wünsche, dass dieses Werk bei den Verfassern und Benutzern der Korpora der gesprochenen Sprache, die wir alle eines Tages sind, die beste Verbreitung genießt.

VORWORT VON JEAN-NOËL JEANNENEY, PRÄSIDENT DER NATIONALBIBLIOTHEK FRANKREICHS

Die Nationalbiliothek Frankreichs freut sich darauf, bei der Ausarbeitung dieses *Handbuches* mitgemacht zu haben. Sie steht zwar in altem und engem Kontakt mit den gesprochenen Sprachen, ihrer Erhaltung und ihrer Verbreitung. Ihre audiovisuelle Abteilung ist der Erbe der *Archives de la Parole* (Archiv des Wortes) von Ferdinand Brunot, die schon 1911 geschaffen wurden. Seit diesem Datum hat sich unsere Einrichtung ständig darum gekümmert, die besten Erhalts- und Bewahrungsbedingungen mündlicher Ausdrücke jeder Art zu sichern, wie ihre Verbreitung in die weiteste Öffentlichkeit.

Die Digitaltechnologien verstärken heute dieses historische und wissenschaftliche Band. Was die Bewahrung betrifft, wurde ein hochfliegender Plan der Digitalisierung unserer Sammlungen engagiert, von dem die Ton- und audiovisuellen Dokumente besonders profitieren. Außerdem wird der Verbreitung dieser Reichtümer zwischen unseren Wänden und aus der Ferne der phantastische Aufschwung unserer digitalisierten online-Bibliothek « Gallica » zustatte kommen, der jedem Internauten erlaubt, Zugang zu diesen Grundquellen des Wissens zu erlangen, egal wo er ist und was der Gegenstand seiner Forschung oder seiner Neugierde ist.

Dieses *Handbuch* ist das Ergebnis einer vertrauensvollen Zusammenarbeit und bezeugt die Komplementarität der Wissen der Sprachwissenschaftler, Juristen, Konservatoren und Informatiker, Ton- und Bildtechniker : ich freue mich darauf, dass die Nationalbibliothek Frankreichs zu diesem innovativen und fruchtbaren Unternehmen beigetragen hat.

1 VORSTELLUNG

1.1 DIE ZIELE

Viele Grundlagen- oder angewandte Forschungen beruhen zur Zeit auf der Auswertung von „Korpora der gesprochenen Sprache“ (geordneten Sammelwerken der Aufnahmen von mündlichen und multimodalen sprachlichen Produktionen). Dieses *Handbuch der guten Praktiken* entsteht aus der Erkenntnis von Sprachwissenschaftlern, die darauf bedacht sind, den Fortbestand der Quellen und einen verschiedenartigen Zugang zu den mündlichen von ihnen produzierten Produktionen zu sichern ; es schneidet zuerst die „Korpora der gesprochenen Sprache“ an, die von Sprachwissenschaftlern und für sie geschaffen und verwendet wurden. Die durch die Erschaffung und die dokumentarische Auswertung dieser Korpora hervorgerufenen Fragen trifft man aber in vielen Fächern : die Völkerkunde, die Anthropologie, die Soziologie, die Psychologie, die Demographie, die mündlich überlieferte Geschichte gebrauchen vor allem die verbale Befragung, die Aussage, das Interview, die Lebensgeschichte. Dieses *Handbuch* beruft sich auf das Verfahren der Sprachwissenschaftler, es stimmt aber mit den Beschäftigungen anderer Forscher überein, die Korpora der gesprochenen Sprache (z. B. in Sprachsynthese und -entzifferung) gebrauchen, auch wenn ihre spezifischen Bedürfnisse im vorliegenden Dokument nicht systematisch angeschnitten werden.

Dieses *Handbuch* hat sich als erstes Ziel gesetzt, die für den Korporaaufbau von mündlichen oder multimodalen Daten erforderlichen *Informationen* zu liefern und *Vorschläge* anzubieten, die sich nicht nur auf die juristischen Punkte beziehen sondern auch auf die materiellen Aspekte, was ebenso das Einsammeln betrifft wie die Strukturierung und die Formgebung der Daten, die Auswertung, Mitteilung und Verwaltung dieser Daten.

Das zweite Ziel dieses *Handbuch* besteht darin, den Forschern, die Korpora der gesprochenen Sprache anlegen oder erweitern, dabei zu helfen, bestimmte „verzögerte Schwierigkeiten“ *vorauszusehen*, die die Auswertung und dann die zukünftige Entwicklung ihres Korpus schwer zu belasten drohen. Bestimmte Anfangsentscheidungen, bestimmte Mängel können sich in späteren Etappen des Prozesses als wichtig erweisen, wenn es zu spät ist, um irgendwas zu ändern.

Das dritte Ziel heißt, das Auftauchen *gemeinsamer Praktiken* zu begünstigen, um sowohl beim Aufbau wie bei der Verwendung der Daten den aktuellen Vorschriften der Bewahrung und der Interaktionen zwischen den Korpora, der Bewertung und der Ethik zu entsprechen.

1.2 DIE BEDINGUNGEN DER BEARBEITUNG

In Frankreich z. B. hat sich der wissenschaftliche Rat der Überwachungsstelle der sprachlichen Praktiken (*DGLF*) gewünscht, die Aktionen für die Bewahrung, den Aufbau und die Aufwertung der mündlichen und multimodalen Korpora stark zu fördern, und dies für folgende Gründe :

- die Bewahrung eines reichhaltigen Erbes der Sprachpraktiken zu ermöglichen ;
- zum Aufbau umfangreicher Referenzkorpora beizutragen, die für die Forschung, den Unterricht, die Sprachindustrien aber auch für das Erbe angelegt werden ;
- zur Entwicklung der Informatikmittel beizutragen, die die Bearbeitung, die Erweiterung und die Aufwertung der Korpora ermöglichen ;
- die Bereitstellung dieser Korpora begünstigen.

1.3 DIE JURISTISCHEN ASPEKTE

Die juristischen Aspekte, die mit dem Aufbau und der Verwendung der Korpora der gesprochenen Sprache gebundenen sind, erwiesen sich sehr schnell als ein häufiges und wesentliches Hindernis.

Diese juristischen Aspekte betreffen hauptsächlich die Fragen der moralischen und erblichen Rechte und des Eigentums der Daten, die bei jeder der vier wichtigen Etappen der Korpusarbeit zu begegnen sind :

- das Einsammeln der Daten und die Aufnahme (Recht auf das Bild, auf die Stimme, Befragungssituation, Erlaubnisse...) ;
- die Verwendung und die elektronische Auswertung der Daten (Archivierung, Datenbank für die Forschung, den Unterricht, die Projektplanung...) ;
- die Verbreitung und das In-Umlauf-Bringen der Daten (Rechte, Rechte auf Zitat, Online-Verbreitung...) ;

– die Verwaltung der Daten.

In Anbetracht der vielen betroffenen Bereiche hat die DGLF die Schaffung eines Komitees mit Fachleuten von verschiedenen Fächern hervorgerufen. Dieses Komitee hat eine Arbeitsgruppe errichtet. Sie soll den Forschungsmannschaften dabei helfen, die Praktiken der Korporaeinsammeln und –Auswertung zu normalisieren, vom Standpunkt der Gesetzgebung und unter Berücksichtigung der gesamten mit der Forschung gebundenen Zwänge. Das hierbei vorgestellte Handbuch ist das Ergebnis der fünfzehnmonatigen Arbeit dieser Gruppe.

Diese Arbeitsgruppe sollte natürlich Rechtswissenschaftler einschließen, die Spezialisten für das Forschungsrecht sind, aber nicht nur : der Bedarf an Kompetenzen bezüglich des Aufbaus der Korpora, der Verwendung und der Verwaltung haben dazu geführt, den Rechtswissenschaftlern Linguisten zuzuteilen, die „Korpuslinguistik" ausüben und mit mündlichen Angaben arbeiten, Vertreter der hohen Institutionen für die Verwaltung des Erbes (IDS, DWDS) und Informatiker, die in Korpusverwaltung spezialisiert sind.

Um ihre Aufgabe zu erfüllen hat sich diese Arbeitsgruppe unter anderem folgende Ziele gesetzt :

– den Bestand der aktuellen Praktiken aufzunehmen und die methodologischen und theoretischen Zwänge zuerst zu bestimmen, die mit der Forschung gebunden sind ;
– eine Synthese über die aktuelle Gesetzgebung zu verbreiten ;
– Empfehlungen zusammenzustellen ;
– wenn nötig, im Falle einer Lücke oder einer Unklarheit, Vorschläge für die Ausarbeitung von juristischen (unter anderem europäischen) Normen und Regelungen zu formulieren.

Deswegen sollte man zuerst :

– den Bestand der betroffenen juristischen Gebiete aufnehmen ;
– die Risiken bestimmen und quantitativ erfassen ;
– die bestehenden Antworten erkennen ;
– und dann diese Antworten in Form von Empfehlungen der guten (juristischen und ethischen) Praktiken aufbauen.

Infolgedessen hat die Gruppe entschlossen, mit mehreren Zeugeteams, die mündliche oder audio-visuelle Daten sammeln oder gesammelt haben, eng zu arbeiten.

Gezielt wurde dadurch das Erlangen einer „Typologie der Situationen" und das Aufzählen aller schon gebrauchten Praktiken und Lösungen, in Deutschland oder anderswo.

1.4 DIE ANDEREN ASPEKTE

Der Arbeitsgruppe wurde unterwegs klar, daß es keine befriedigende Antwort auf die vorhandenen Schwierigkeiten wäre, lediglich eine Liste Empfehlungen oder Lösungen juristischer Natur vorzuschlagen.

Zwar hat es sich gezeigt, daß die Schwierigkeit oder die Lösung ja oft mit der Art der Einsammeln- oder Verwendungpraktik gebunden waren ; daß einige Lösungen durch technische Wege zu erreichen waren, die die Angaben selbst beeinflussten (Anonymisierung oder Verschwommung) ; daß es überhaupt nicht gleichgültig war, dieses oder jenes juristische Problem in diesem oder im anderen Moment zu lösen. Kurz gesagt, Lösungen zu juristischen Fragen vorzuschlagen, hieß soviel wie den Prozess selbst des Einsammelns oder der Formgebung, der Weiterleitung oder der Verwendung dieser Art Daten anzuschneiden.

Schließlich stellte sich über den Respekt hinaus, den man den Rechten der aufgenommenen Personen schuldet, die Frage der „Urheberschaft" dieser Art von Daten : welches sind die Rechte der Sammler dieser Daten ? Wer ist juristisch verantwortlich dafür, wer darf sie weiterleiten ? Die juristischen Aspekte, die mit dem wissenschaftlichen Urheberschaft oder der strafrechtlichen Verantwortung gebunden sind, waren auch, wie man sieht, von der Einsammeln- und Verwendungspraktik der Daten nicht zu trennen.

War es infolgedessen nicht besser, die Sachkenntnis des geplanten „Handbuches" zu erweitern und nicht nur die juristischen Praktiken zu erwähnen, sondern auch die gesamten Praktiken, die diese Art Korpus einsetzt ? Diese Entscheidung wurde getroffen, denn so war es möglich, alle Aspekte gebunden zu halten, so wie sie in der Wirklichkeit sind.

1.5 DIE METHODE

Die von der Arbeitsgruppe gewählte Methode zeichnet sich durch folgende Züge aus :

– die Überzeugung, dass man nicht glauben lassen darf, dass es für jede Situation vorgefertigte Antworten gibt ;
– den Wille, die Forscher nicht zu „zügeln“ (z. B. wenn man bestimmte Praktiken verbietet) ;
– die Einhaltung der Methodologie des Forschers und der mit der Beobachtung gebundenen Zwänge (die Forscher wünschen, Situationen aufzunehmen, ohne dass die vor allem technischen und juristischen Zwänge sie ändern ;
– die Notwendigkeit des Zusammenbringens der Kenntnisse, die bei den verschiedenen Etappen erforderlich sind (Sprachwissenschaftler, Juristen, Verwalter) für die Ausarbeitung und die Verfassung dieses Handbuches ;
– die Kundgebung eines auf dem Respekt des Gesetzes und der Ethik beruhenden Verfahrens ;
– die Notwendigkeit, durch dieses *Handbuch* ein Gutachteninstrument der Risiken (um sie ausfindig zu machen aber auch zu schätzen) zu beschaffen.

1.6 DER FRANZÖSISCHE JURISTISCHE RAHMEN

Etliche Fragen und Lösungen betreffen die *Einwilligung* der Befragten aber auch die Verantwortung der *Copyrightbesitzenden* Instanzen. Da ist sicher ein Hauptpunkt. Er ist aber nicht der einzige auf dem Spiel und die Antworten auf diese Frage haben sich komplex erwiesen.

Die Art und Weise, die Einwilligung und die Erlaubnis heute aufzunehmen, ist verschiedenartig. Es gibt keine anerkannten Normen und viele Schwierigkeiten.

Zuerst soll die Einwilligung *aufgeklärt* sein (Rahmen, Zwecke, „Risiken“ für den Befragten).

Die Aufnahme der Einwilligung a priori kann manchmal die Befragung stören (Paradox des Beobachters), indem sie eine Situation formalisiert, während „natürliche“, dem umgänglichen Gespräch nahe Daten gewünscht sind.

So hat sich zum Beispiel die Praktik interessant und wirksam erwiesen, die darin besteht, den Befragten, (zu der Aufnahme der Einwilligung) ein Dokument zu lassen, das den Rahmen, die Zwecke, die Risiken, den Zugang und die Informationen erläutert, die es erlauben, das zugrunde liegende Bezugssystem der Veröffentlichungen und der Ergebnisse im nachhinein wiederzufinden.

Die Schwierigkeit besteht auch in einem *Widerspruch* zwischen dem Zwang, die Zwecke der Befragung zu erklären, um die Einwilligung aufzuklären, und der Unmöglichkeit, alle Zwecke und *die zukünftigen Gebrauchsmöglichkeiten der Daten* im voraus vorherzusehen, *in Anbetracht der aktuellen Bemühung*, eine *maximale* Interaktionsmöglichkeit zu erlangen.

Dazu ist auch daran zu denken, dass einige verbale Kulturen (und nicht nur am Ende der Welt) keine Möglichkeit geben, eine schriftliche Spur der Einwilligung vorzuschlagen und zu bewahren.

Alle andere Fragen juristischer Natur bieten dieselbe Komplexität an : Anonymität, Verschlüsselung, Verschwommung, Bestimmung der Haftungen, Depotzwang, Mitteilungen, usw., alle Praktiken, die zwangsläufig mit dem Aufbau und dem Bestehen eines Korpus der gesprochenen Sprache gebunden sind. Keiner dieser Aspekte beruht auf einer einzigen, klar definierten und überall anerkannten Praktik.

Jede dieser Etappen wird allmählich mit technischen Wahlen, mit sozialen oder wissenschaftlichen Praktiken eng gebunden, denn das alles ist schwer voneinander zu trennen.

Daher die Entscheidung der Arbeitsgruppe, ein Handbuch anzubieten, das nicht nur ein “juristischer Leitfaden“ ist, sondern auch eine praktische, zuverlässige Hilfe, die alle Aspekte des Prozesses in Betracht zieht.

1.7 EIN „HANDBUCH DER GUTEN PRAKTIKEN“ ?

Dieses Handbuch berücksichtigt den juristischen Rahmen in Frankreich (und in einigen Ländern in Europa) und stützt sich auf die Befragungen der Forscher, die zu seinem Aufbau beigetragen haben. Diese haben versucht, die Grundlagen der geltenden juristischen Regeln und das, was mit ihrem Respekt und ihrer Umsetzung aufs Spiel gesetzt wird, zu verstehen. Eine *dynamische Vision der juristischen Regulation* bildet also durch das von den Forschern gefolgte Verfahren die Grundlage für dieses Handbuch. Die Autoren dieses Handbuchs sind auch selbst in die betroffenen Forschungsgebiete verwickelt und haben sich darum bemüht, Praktiken und Sitten vorzuschlagen, die das bestehende Recht einhalten. Infolgedessen soll das Verfahren des Forschers darin bestehen, diese Rechte und die Zwänge, die sich davon ableiten lassen, zu kennen. Dann wird es darum gehen, die Konsequenzen dieser Zwänge in die Phase des Einsammelns der Daten so wie in die der Aufwertung zu ziehen.

Damit ein solches Verfahren auf eine strikte und glaubwürdige Art und Weise vorgestellt werden kann, soll es zuerst in seinem wissenschaftlichen, politischen, juristischen oder institutionnellen Kontext dargestellt werden. Die vorgeschlagenen Gebräuche und Praktiken werden im Laufe des Handbuchs durch diesen Kontext „aufgeklärt" werden, so dass man besser verstehen kann, was durch den Respekt oder durch die Respektlosigkeit dieser Gebräuche und Praktiken aufs Spiel gesetzt wird.

1.8 EINIGE HÄUFIGE FRAGEN

Das erste Ziel dieses Handbuchs heißt, Informationen und Elemente der Antworten auf die Fragen zu geben, die sich all die Forscher oder Verantwortlichen für den Aufbau, die Auswertung, die Verwaltung und die Verbreitung dieser Korpora stellen.

Um dieses Ziel zu erreichen, wurde das Handbuch mit vielen Verweisen, die ebenso viele mögliche Lesenswege bilden, konzipiert. Die folgenden Fragen bilden die traditionnellen Befragungen am Anfang eines Forschungsprojekts und schlagen also ein erstes Wegbeispiel vor.

Titre	***Corpus oraux, Guide des bonnes pratiques 2006* (version allemande)**
Type	Livre (traduction)
Editeur	En ligne (HAL)
Année	2010
Référence	

Spoken Corpora, Good Practice Guide, 2006

Olivier Baude

CORPUS ORAUX

Guide des bonnes pratiques 2006

coordonné par **Olivier BAUDE**

PRESSES UNIVERSITAIRES ORLÉANS

CNRS EDITIONS

Délégation générale à la langue française et aux langues de France
6, rue des Pyramides75001 PARIS
http://www.dglflf.culture.gouv.fr

ISBN 2-271-06425-2 (CNRS ÉDITIONS)
ISBN 2-913454-30-5 (PUO)
EAN 9782271064 257 (CNRS ÉDITIONS)
EAN 9782913454 309 (PUO)

This book /work is the result of a working group brought together by Isabelle **de LAMBERTERIE**.
It was coordinated by Olivier **BAUDE**.

Olivier **BAUDE** (*DGLFLF et CORAL – Orléans University*)
Claire **BLANCHE-BENVENISTE** (*EPHE and the University of Provence*)
Marie-France **CALAS** (*DMF*)
Paul **CAPPEAU** (*Poitiers University*)
Pascal **CORDEREIX** (*BnF)*
Laurence **GOURY** (*CNRS – CELIA*)
Michel **JACOBSON** (*CNRS – LACITO*)
Isabelle **de LAMBERTERIE** (*CNRS-CECOJI*)
Christiane **MARCHELLO-NIZIA** (*CNRS-ILF and ENS-LSH-Lyon*)
Lorenza **MONDADA** (*ICAR, CNRS,Lyon2 University*)

With the collaboration of:

Gilles **ADDA** (*for the COPTE LIMSI-CNRS*), Michel **ALESSIO** (*DGLFLF*), Alain **CAROU** (*BnF*), Ibrahim **COULIBALY** (*CDF – Grenoble University*), Valérie **GAME** (*BnF*), Fabrice **MOLLO** (*CNRS-CECOJI*), Michel **RAYNAL** (*INA*), Jean **SIBILLE** (*DGLFLF*), Dominique **THERON** (*BnF*), Luc **VERRIER** (*BnF*).

Traduction : Caroline **SARRE**

LIST OF CONTRIBUTORS

OLIVIER BAUDE
Senior Lecturer in Language Sciences at Orléans University, member of the *Centre Orléanais de Recherche en Anthropologie et Linguistique (EA-3850).* Secretary of the scientific committee of the *Observatoire des pratiques linguistiques, Délégation générale à la langue française et aux langues de France.*

CLAIRE BLANCHE-BENVENISTE
Emeritus Professor, *École Pratique des Hautes Études* in Paris and Université de Provence. Researcher in the field of French linguistics: written and spoken language, syntax, morphology, collection of oral language corpora.

MARIE-FRANCE CALAS
Chief Curator of the national heritage. General Inspector of French Museums, *Direction des Musées de France.* Specialist in the field of oral documents, which is a wide pluridisciplinary field comprising the history, management, conservation and promotion of spoken language, music and environmental sound recordings, all considered today as parts of the immaterial heritage.

PASCAL CORDEREIX
Chief librarian, in charge of the oral document section at the audiovisual department of the *Bibliothèque nationale de France.* He is also the vice president of the *Association française des détenteurs de documents audiovisuels et sonores* (French association of owners of audiovisual and sound documents). Most of his work deals with the problems of sound archiving.

LAURENCE GOURY
Research fellow at the IRD *(Institut de Recherche pour le Développement),* member of the CELIA *(Centre d'Etude des Langues Indigènes d'Amérique),* field linguistics and typology (Creole languages in particular).

MICHEL JACOBSON
Computer engineer in the « *Langues et Civilisations à Tradition orale* » research team at the *Centre National de la Recherche Scientifique.* specialising in the management of oral corpora.

ISABELLE DE LAMBERTERIE
Research supervisor (*Directrice de recherche)* at the *CNRS*, in charge of the team *« Normativité et société de l'information »* at the *Centre d'études sur la coopération juridique internationale* (CECOJI – UMR 6224), member of the CNRS Committee on Ethics.

CHRISTIANE MARCHELLO-NIZIA
Professor in Language Sciences at the ENS-LSH (Lyon), Director of the *Institut de Linguistique Française (CNRS)*: historical linguistics, history of the French language, theories on the evolution of languages.

LORENZA MONDADA
Professor in Language Sciences at Lyon 2 University, member of the research team ICAR (UMR CNRS 5191). Specialist of interactional linguistics, she works on corpora of oral languages in interaction, as well as on the multimodal analysis of video corpora.

PREFACE BY XAVIER NORTH,
GENERAL DELEGATE FOR THE FRENCH LANGUAGE AND FOR FRENCH LANGUAGES

Rare are the moments, in the history of science or cultural politics, where an ensemble of raw data and uncertain material becomes an object of knowledge. The publication of this guide does just that, since it offers to every researcher the tools, a guide for "good practice" which will allow him to proceed in this metamorphosis: The transformation of verbal productions into oral corpora, likely to be studied and kept and resultantly to take its place in the Nation's cultural heritage.

Language productions in their written form, being both fixed and definitive, literary works or historical documents, have always been at the heart of the politics put into place by the Minister for Culture, whether in the form of books or archives. But it is only recently that we have started to become interested in the living aspect of language, in its spontaneous springing, in its ordinary daily enunciation, and in the extraordinary variety of its parlances… For the first time, it has become possible to make real archives of the spoken word based on solid ground. Technological advances should contribute to this.

Indeed, an oral corpus is not just a simple collection of recordings of human speech, but it is a tangible object that has been "constructed": processing data (digitisation, transcription, and indexation) allows us not only to conserve it but also to give it a new status, i.e. that of research material and promotion. But this implies using the prescriptions of methodology that are coherent and easy to put into place.

Thanks to "A Guide to good practice", a new and vast domain has now become available to researchers. Through its *Observatoire des pratiques linguistiques,* the *Délégation générale à la langue française et aux langues de France* initiated this work, and then strived to gather and coordinate the various resources, both human and material, which produced this book, whether they originated from the world of research or the different horizons of the Ministry for Culture involved in this initiative.

Ensuring the development of oral corpora, their distribution and their preservation is also making the French linguistic heritage available to listen to in its diversity, richness and truth. It is also creating a precious tool of knowledge of language use which is necessary for the definition of linguistic politics as well as the politics of education and sociology.

For several months, this research brought together lawyers, linguists, librarians and computer experts all working conjointly with the common goal of making it possible to explore new areas of culture and research while respecting the law. It is the result of a common effort that we present in this book today, in the hope that, in its turn, it will generate numerous works.

PREFACE BY BERNARD MEUNIER, PRESIDENT OF THE CNRS

The spoken word and the written word. These two elements possess a powerful evocative force. We think about the way in which civilizations became structured by their oral practices and then by the creation of writings which led to a better transmition in space and time of the words spoken by one or another of us.

As a researcher looking at the respective roles of the spoken and written word in the dissemination of scientific knowledge, I don't fail to remember that, far beyond the essential role of the written word, delivering an oral presentation in front of our peers or a wide audience is always essential for circulating, convincing or sharing ideas. The spoken word maintains a power of conviction, allowing the largest number of people to be reached provided that it can be recorded and transmitted with the help of current audio-visual means.

The collection and use of oral corpora should be done in compliance with a code of "good practices", in the same way as it is done for the collection of written corpora. We all know how a sentence which has been taken out of context and broadcast without reserve can become dangerous for the person who produced it, for a group of people or a community.

The authors of this outstanding book have examined in depth all the legal aspects involved in the collection and use of written corpora. I hope that this book will get the best possible circulation among the actors and users of oral corpora, something that we all are at one time or another.

PREFACE BY JEAN-NOËL JEANNENEY, PRESIDENT OF THE *BIBLIOTHÈQUE NATIONALE DE FRANCE*

The Bibliothèque nationale de France is happy to have contributed to the elaboration of this *Guide*. Indeed, it has had a long-lasting and close link with spoken languages their preservation and their distribution. Its audiovisual department ensues from Ferdinand Brunot's Spoken Archives (*Archives de la Parole)*, created as early as 1911. From then on, our institution's constant concern has been to ensure the best conditions possible for recording and preserving oral expressions of every type, as well as their distribution to as large an audience as possible.

Today digital technology is reinforcing this historical and scientific link. In terms of conservation, an ambitious plan to digitise our collections has been initiated, from which sound and audiovisual documents benefit in particular. Furthermore, the distribution of these precious resources within our walls and from a distance is further enhanced by the rapid expansion of our on-line digital library, "Gallica", which allows every internet user, wherever they may be based or whatever the purpose of their research or interest may be, to have access to fundamental sources of knowledge.

The fruit of a faithful collaboration, this *Guide* shows how complementary knowledge is between linguists, lawyers, librarians, computer scientists, sound and picture technicians: I am delighted that the *Bibliothèque nationale de France* has been able to contribute to this innovative and fruitful undertaking.

1 PRESENTATION

1.1 OBJECTIVES

There is currently a vast amount of fundamental or applied research, which is based on the exploitation of oral corpora (organized recorded collections of oral and multimodal language productions). Created as a result of linguists becoming aware of the importance to ensure the durability of sources and a diversified access to the oral documents they produce, this *Guide to good practice* mainly deals with "oral corpora", created for and used by linguists. But the questions raised by the creation and documentary exploitation of these corpora can be found in numerous disciplines: ethnology, anthropology, sociology, psychology, demography, oral history notably use oral surveys, testimonies, interviews, life stories. Based on a linguistic approach, this *Guide* also touches on the preoccupations of other researchers who use oral corpora (for example in the field of speech synthesis and recognition), even if their specific needs aren't consistently dealt with in the present document.

The *Guide* that we have put together for you primarily aims at providing the necessary *information* for making corpora of oral or multi-modal data, and at offering useful *suggestions* concerning the judicial as well as the material aspects involved just as much with the collection, structuring and transcription of data, as with the exploitation, communication and preservation of the data.

The second objective of this guide is to help researchers who are making or contributing to oral corpora to *anticipate* certain "delayed difficulties" which could seriously jeopardize the exploitation and the future of their corpus. Certain choices made at the beginning, certain missing elements can turn out to be important at later stages of the process once it is too late to make any changes.

The third objective is to encourage the definition of *common practices* in order to fulfill the current requirements of conservation and interoperability of corpora, of evaluation and ethics as much in the constitution as in the use of the data.

1.2 CONTEXT OF ELABORATION

The scientific committee of the *Observatoire des pratiques linguistiques (Délégation générale à la langue française et aux langues de France)* has wished to strongly encourage all measures of preservation, constitution and promotion of oral and multimodal corpora for the following reasons:

- To allow for the maintenance of a rich national heritage about language uses in France;
- To help develop large reference corpora, for research, teaching, the language industries and also for the linguistic heritage;
- To help in the development of computer tools for processing, enriching and promoting corpora;
- To encourage the availability and accessibility of corpora;

1.3 LEGAL ASPECTS

It quickly became apparent that the judicial aspects linked to the consitution and use of oral corpora represented a recurrent and major obstacle.

These legal aspects mainly concern questions of moral and property rights, and data ownership, which arise in each of the four main stages of corpus work:

- Collecting data and recording it (the right to one's image and voice, interview situations, authorisations…);
- The use and computerized exploitation of data (archiving, use of database for research, teaching, engineering…);
- The distribution and publication of data (rights, the right of quotation, online publication of data…);
- The conservation of data.

In view of the fact so many domains were involved, the DGLFLF initiated the creation of a committee made up of experts for a diversity of fields. This committee set up a working group whose objective was to help research groups to standardise their ways of collecting and exploiting data in compliance with the

law whilst taking into account the many constraints inherent to the research. This guide is the result of fifteen months working in this group.

Of course, this working group had to include legal experts in research law, but that wasn't enough. We needed collaborators with specialized skills in collecting, using and preserving corpora, hence why the working group took on linguists working in the field of corpus linguistics and oral data, representatives of the most important organizations for the preservation of the national heritage *(INA, INSI, BnF)*, and computer scientists specialized in corpus management, alongside the legal experts.

To achieve its goal this working group gave itself the following objectives:

- To look at current practices and as a matter of priority to define the methodological constraints and theories bound to research;
- To circulate a synthetic document about existing legislation;
- To make recommendations;
- And if necessary, where there is a gap or something unclear in the law, to formulate suggestions for the creation of legal norms and rules (notably those in Europe).

In order to do this, it was first necessary to:

- Review the judicial domains concerned;
- Identify and quantify the risks;
- Work out the existing responses;
- And then to formulate the responses in the form of a series of recommendations for good practices (both legal and ethical).

For this purpose, the group decided to work closely together with many control teams who were collecting or had previously collected oral or audiovisual data. In this way, the goal was to come up with a "typology of situations", and to examine all the practises and solutions already being used in France as well as elsewhere.

1.4 OTHER ASPECTS

Whilst working towards these goals, the group realised that making a simple list of recommendations or solutions of a legal nature wouldn't be enough to effectively overcome the difficulties that were encountered.

It actually became clear that the difficulties or the solutions were linked to the practices used when collecting or using the data and that certain solutions had to be found through examining technical measures which had an impact on the data itself (anonymisation or blurring). It also appeared that solving a legal problem at one particular stage rather than another did matter. In short, offering solutions for legal questions meant examining the very process of collecting, transcribing, circulating or using this type of data.

Finally, over and above respecting the legal rights of the people who had been recorded, the question of "the right of ownership" of this type of data arose: What rights do the people who collect this data have? Who is legally responsible, who has the right to disseminate it and in what form? As we can see, the legal aspects linked to scientific ownership or penal responsibility were also inseperable from the collection process and use of the data.

With this in mind, would it not be better to enlarge the field of the proposed "Guide" and deal not only with the legal practices but also with all the practices involved in this type of corpus? This is the choice that we made because it allowed all aspects to be intertwined as they are in reality.

1.5 METHODOLOGY

The methodology on which this group has agreed has the following characteristics:

- The conviction that you cannot let people believe that there are ready made answers to every type of situation;
- The eagerness not to hold back researchers (by prohibiting certain practices for example);
- The respect of the researcher's methodology and of the constraints linked with observation (researchers want to record situations that should not be altered by technical or legal constraints).

– The need to elaborate and compile this guide by bringing together the skills required at the different stages (linguists, lawyers, librarians);
– The display of a procedure founded on the respect of the law and ethics;
– The need to provide through this *Guide* a tool for risk assessment (pinpointing and also evaluating risks).

1.6 THE FRENCH JUDICIAL FRAMEWORK

A large number of questions and solutions revolve around the notion of *consent* of the interviewees but also around the responsibility of the *owning* institutions. It is certainly a nodal point, but it is far from being the only thing in question and besides the answers to such a question proved to be complex.

Current practices for gaining consent and authorisation are very varied. No specific norms exist and there are multiple difficulties.

In the first instance, consent should be *informed* (framework, objectives, "risks" for the interviewee).

But it would appear that gaining consent can sometimes hinder the study (the observers's paradox) in formalizing a situation when what is desirable is to obtain » natural » data that is as close as possible to ordinary conversation.

In this way, for example, one practice which proved interesting and efficient (in addition to collecting authorisation) consists in handing out to the interviewees a document explaining the framework, the objectives, the risks, the accessibilty, and the details allowing the references of publications and results to be subsequently found.

The difficulty also comes from a *contradiction* between the need to specify the objectives of the study in order to 'inform' the consent, and the impossibility to anticipate all the objectives and *the future possiblities for the use of the data, considering the current concern with coming up with maximimum* interoperability.

Finally, it should be noted that certain spoken cultures (and not just on the other side of the world) don't offer the possibility to propose and keep a traceable written consent.

All other questions of a legal nature also have the same complexity: anonymisation, encryption, blurring, defining responsibilites, depositing, papers, etc., all the necessary practices linked to the constitution and existence of an oral corpus. None of these aspects rests on one specific practice which is clearly defined and accepted everywhere.

Each of these steps is closely linked to technical choices, to social or scientific practices, all these elements being very difficult to dissociate.

This is why the choice of the working group was to offer a Guide which would not only be a "judicial memento", but also a practical and reliable tool covering all aspects of the process.

1.7 A "GUIDE TO GOOD PRACTICE"?

Taking into account the existing legal framework in France (and more generally in different parts of Europe), this guide relies on the questions asked by researchers who participated in its elaboration. They tried to comprehend the foundations of the judicial rules to abide by and the stakes linked to the respect of these rules and to their implementation. *A dynamic vision of legal regulations* has therefore served as a framework for this guide, through the procedure used by researchers. The authors of the guide, involved themselves in the fields of research dealt with, were concerned with proposing practices and uses which respected the existing laws. For this, the research process should consist in knowing the existence of the laws and of the constraints which surround them. Then, consequences of these constraints need to be identified as much in the stage of data collection as in that of data promotion.

To present such a procedure in a credible and rigourous way, it first has to be put in its context whether it be scientific, political, judicial or institutional. Throughout, the suggested uses and practices will be "clarified" by the context, with a view to better understanding what the implications of respecting or not these uses and practices are.

1.8 SOME FREQUENTLY ASKED QUESTIONS

The first objective of this guide is to provide information and elements to answer the questions asked by all researchers or people in charge of collecting, exploiting, conserving and circulating corpora.

To reach this objective, the guide includes numerous cross-references which make up many possible reading paths. The following questions are representative of the queries which traditionally arise at the beginning of a research project and in this way suggest a first example of reading paths.

Titre	***L'observatoire des pratiques linguistiques» et « entretien avec Pierre Encrevé »***
Type	Article de vulgarisation
Editeur	MCC
Année	2010
Référence	Baude, O., M., Sibille, J. (2010) «L'observatoire des pratiques linguistiques» et « entretien avec Pierre Encrevé », in *Culture & Recherche*, n° 122 et 123, Ministère de la Culture et de la communication, Paris. Pp 82-83

CULTURE ET RECHERCHE

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Ministère Culture Communication

N° 122-123 PRINTEMPS-ÉTÉ 2010

50 + UN

1959-2010
La recherche au ministère de la Culture

L'Observatoire des pratiques linguistiques

Susciter et soutenir des recherches en sociolinguistique sur les pratiques langagières contemporaines, conserver et faire connaître le patrimoine linguistique de la France, notamment par la numérisation et la valorisation de corpus oraux, telles sont les grandes missions de cet observatoire créé au sein de la Délégation générale à la langue française et aux langues de France.

OLIVIER BAUDE et JEAN SIBILLE
MCC / Délégation générale à la langue française et aux langues de France / Observatoire des pratiques linguistiques

www.dglf.culture.gouv.fr/observatoire/observatoire_accueil.htm

1. www.corpusdelaparole.culture.fr

2. Olivier Baude coord., *Corpus oraux. Guide des bonnes pratiques*, Paris : CNRS Éditions - Presses universitaires d'Orléans, 2006.

Créé en 1999 au sein de la Délégation générale à la langue française, l'Observatoire des pratiques linguistiques a pour objectif de recenser, de développer et de rendre disponibles les savoirs relatifs à la situation linguistique en France, afin notamment de fournir des éléments d'information utiles à l'élaboration des politiques culturelles, éducatives ou sociales. Il a également pour but de faire mieux connaître un patrimoine linguistique commun constitué par l'ensemble des langues et des variétés linguistiques parlées en France, qui concourent à la diversité culturelle de notre pays. Doté d'un comité scientifique composé de linguistes, il soutient des projets ou des programmes de recherches dans le cadre d'appels à propositions thématiques ou de partenariats avec le CNRS ou les universités.

Le champ de l'observation est celui de la sociolinguistique et concerne les pratiques actuelles, qu'il s'agisse du français ou des autres langues parlées sur le territoire national : langues « régionales » ou langues issues des différentes vagues de migration.

L'activité de l'observatoire s'organise autour de trois axes :
– le soutien à des travaux d'étude et de recherche, la coordination et l'organisation en réseau de ces travaux ;
– la diffusion des informations recueillies auprès des spécialistes, des responsables de politiques publiques et d'un large public ;
– la conservation, la constitution, la mise à disposition et la valorisation de corpus oraux enregistrés. Ces corpus constituent un outil de travail pour la recherche, mais acquièrent également, avec le temps, un caractère patrimonial.

Depuis 1999, l'observatoire a procédé à cinq appels à propositions thématiques. En dehors du cadre des appels à propositions, il a soutenu notamment, en partenariat avec l'INSEE et l'INED, la conception et l'exploitation du volet linguistique de « l'enquête famille » annexée au recensement de 1999 ; ou encore un programme de recherche de l'Institut de recherche et développement (IRD) et du CNRS sur les langues de Guyane.

Depuis 2004, un des axes majeurs de l'activité de l'observatoire est le développement, dans le cadre d'un partenariat Culture-CNRS, du programme « Corpus de la parole » qui a pour objectif la numérisation et la valorisation des corpus oraux (collections ordonnées d'enregistrements de productions linguistiques orales et multimodales réalisées par des chercheurs) afin de permettre leur conservation et leur transformation en de véritables ressources linguistiques numériques, pour la recherche en sciences humaines, l'enseignement et l'ingénierie des langues. Ce programme a permis, de 2006 à 2009, avec le soutien du plan de numérisation du ministère de la Culture, de constituer et de numériser une collection de corpus oraux en français et en langues de France, qui est mise à la disposition du public sur le site Internet *Corpus de la parole*[1], ouvert en février 2008. Un guide des bonnes pratiques[2] destiné aux chercheurs a également été réalisé. Ce programme doit permettre, non seulement le développement d'une base de données patrimoniales sur l'oral, mais aussi le développement d'outils de traitement automatique des langues et d'ingénierie linguistique rendant possible l'interopérabilité des bases de données de grands corpus.

Les publications de l'Observatoire des pratiques linguistiques

La première phase d'activité de l'Observatoire a consisté à mobiliser les chercheurs et à favoriser l'émergence de réseaux. La seconde phase consiste à créer des espaces nouveaux de diffusion de l'information et d'échange.
Pour cela, la DGLFLF publie un bulletin, *Langues et Cité*, et une collection : les « Cahiers de l'Observatoire des pratiques linguistiques ».
Créé en 2002, *Langues et cité*, est un bulletin d'information de 12 à 16 pages. Quinze numéros ont actuellement été publiés, douze numéros thématiques : n° 1 *Observer les pratiques linguistiques. Pour quelles politiques ?*, n° 2 *Les pratiques linguistiques des jeunes*, n° 3 *Les langues en Guyane*, n° 4 *La langue des signes française*, n° 5 *Les créoles à base française*, n° 6 *Corpus de la parole*, n° 7 *Les rectifications orthographiques de 1990*, n° 9 *La langue romani*, n° 10 *L'occitan* ; 2008, n° 11 *L'arménien en France*, n° 13 *Plurilinguisme et migrations*, n° 15 *L'arabe en France*, et trois numéros non thématiques. Le numéro 16, paru en mars 2010, a pour thème les *Langues en contact*.

Plus techniques, les « Cahiers de l'Observatoire des pratiques linguistiques », présentent des synthèses de 100 à 150 pages sous forme de recueils d'articles rédigés par des chercheurs. Deux volumes sont parus : *Les rectifications orthographiques de 1990 : analyse des pratiques réelles*, et *Migrations et plurilinguisme en France* ; deux sont en préparation et paraîtront en 2010 : *Langues de France, langues en danger : aménagement et rôle de linguistes*, et *Recherches récentes sur la langue des signes française*.

Publications téléchargeables sur le site de la DGLFLF : **www.dglf.culture.gouv.fr > Publications** ; on peut aussi s'y abonner gratuitement.

Le développement de la partie éditoriale du site Internet *Corpus de la parole* permettra, en outre, de constituer un nouvel espace d'information et d'échanges.

Recherches sociolinguistiques et politique des langues

Entretien avec Pierre Encrevé

Linguiste et historien de l'art, Pierre Encrevé est directeur d'études à l'EHESS, directeur du CELITH (Centre de linguistique théorique). Il a été membre des cabinets de Michel Rocard, Premier ministre, puis de Catherine Trautmann, ministre de la culture et de la communication. Il a participé à l'action gouvernementale en tant que linguiste ; il a notamment présenté les principes de rectification de l'orthographe, il est à l'origine de l'extension aux langues de France des compétences de la Délégation générale à la langue française, et a présidé le Comité pour la simplification du langage administratif (COSLA). Auteur du catalogue de l'œuvre de Pierre Soulages en plusieurs volumes, il a été commissaire de l'exposition « Soulages » présentée au Centre Pompidou en 2009. Il est président du conseil scientifique de l'Observatoire des pratiques linguistiques de la Délégation générale à la langue française et aux langues de France (DGLFLF).

Depuis quand le ministère de la Culture s'intéresse-t-il aux recherches sur les langues dans la culture et la société ?

Il y a eu un intérêt pour la diversité linguistique en France au moins depuis le ministère Lang, qui a très vite fait réaliser par un chercheur un rapport sur les langues régionales, et créé un Conseil national des langues régionales réunissant des spécialistes ; initiatives heureuses mais peu suivies d'effets. À l'époque, ce qui touchait à la langue française n'était pas rattaché au ministère de la Culture mais au Premier ministre, et c'est du Premier ministre qu'est partie en 1986 l'initiative, hélas sans succès, en faveur de la féminisation, puis en 1989 l'initiative pour une rectification de l'orthographe, qui a mis en contact les chercheurs les plus pointus de France et de Belgique avec la politique de la langue ; mais on aura attendu jusqu'à 2009 pour voir ses résultats officiellement pris en compte par l'Éducation nationale. Ce n'est qu'à partir de 1993 qu'on a rattaché la Délégation générale à la langue française (DGLF) – créée en 1989 – au ministère de la Culture, et ce n'est que sous le gouvernement Jospin que la prise en compte de diverses questions sociolinguistiques au sens large, comme celle de la féminisation des titres et noms de métiers et celle de la Charte européenne des langues régionales ou minoritaires, a vraiment lancé la pratique d'une relation systématique des chercheurs de l'université et du CNRS avec le ministère de la Culture. C'est dans ce cadre que le ministère de M^me^ Trautmann a pris les initiatives qui ont conduit à la création de l'Observatoire des pratiques linguistiques, mais aussi à l'élargissement du domaine de la DGLF, qui sera transformée en Délégation générale à la langue française et aux langues de France (DGLFLF) en 2001.

Pourquoi cette nécessité d'une réflexion sur le lien entre savoirs et politique linguistique ?

Dans l'intérêt général, à tous égards. Il serait paradoxal qu'un État qui finance des recherches en matière de langues et de linguistique ne cherche pas à établir un lien entre ces savoirs et la politique linguistique qu'il met en œuvre ! C'était évident pour l'orthographe, qui comptait des spécialistes éminents, mais il était tout aussi normal et nécessaire que le gouvernement confie à l'Institut national de la langue française (INALF) le soin d'étudier la féminisation, ou l'établissement de la liste des langues de France. Qui, sinon les tenants des savoirs indispensables pour la mener à bien pouvait être mieux désigné pour cette tâche ? On souffre beaucoup en France des frontières administratives, notamment entre administrations relevant de ministères différents. Il est impérieux que des chercheurs relevant des ministères de l'Éducation nationale ou de l'Enseignement supérieur et de la Recherche, mais aussi d'autres ministères comme c'est le cas de l'Institut national d'études démographiques (INED), travaillent directement avec les services du ministère de la Culture pour tous les domaines les concernant. Il ne faut pas oublier qu'il y a aussi beaucoup de recherches et de savoirs au sein du ministère de la Culture, et qu'il faut tout faire pour resserrer la relation avec les chercheurs des autres ministères. C'est ce qu'on a fait aussi pour l'histoire de l'art en créant, dans les locaux du ministère, l'Institut national d'histoire de l'art (INHA).

La création de l'Observatoire des pratiques linguistiques, l'évolution de la Délégation générale à la langue française (DGLF) en Délégation générale à la langue française et aux langues de France (DGLFLF) ont-elles un lien avec les recherches en sociolinguistique ?

Oui, par définition, dans la logique de ce que je viens de rappeler. L'observation des pratiques linguistiques est indispensable à l'État s'il veut mener une politique linguistique informée et cohérente, et c'est la tâche propre des sociolinguistes, qui ne limitent évidemment pas leurs recherches aux langues officielles mais à toutes les pratiques linguistiques observables en France, en métropole et outre-mer.

Quels types de recherche en linguistique vous semblent aujourd'hui nécessaires pour éclairer les politiques publiques ?

Toutes les recherches théoriques et empiriques sur les pratiques des locuteurs. Ce qui conduit à se pencher sur certains types d'usages (par exemple, l'extension de l'anglophonie dans la recherche scientifique ou dans les entreprises multinationales, mais tout autant les usages linguistiques des adolescents des cités ou des locuteurs amérindiens de Guyane), aussi bien que sur la constitution des corpus langagiers et leur mise à disposition, tant du point de vue juridique qu'empirique et théorique. Sans jamais perdre de vue que ces recherches intéressent prioritairement le ministère auquel est rattachée la DGLFLF dans la mesure où elles peuvent contribuer à éclairer les pouvoirs publics dans la définition et la mise en œuvre de leurs politiques linguistiques.

Titre	***Diversité des langues et plurilinguisme***
Type	Article de vulgarisation
Editeur	MCC
Année	2010
Référence	Baude, O., Alessio, M., (2010) «Diversité des langues et plurilinguisme», in *Culture & Recherche*, n° 124, Ministère de la Culture et de la communication, Paris. Pp 4-5

La pluralité des langues est de mieux en mieux perçue comme une donnée essentielle à la compréhension du fait humain. La recherche sur le plurilinguisme et ses modalités connait en conséquence un fort développement : les évolutions de la recherche sont toujours l'écho des transformations de la société.

MICHEL ALESSIO
et OLIVIER BAUDE
MCC / DGLFLF

Jusqu'à ces dernières décennies, en France, l'intérêt pour les études linguistiques portait principalement sur les modèles théoriques et négligeait l'articulation avec les usages sociaux concrets, lesquels se manifestent toujours sous l'angle de la diversité : problèmes de contacts de langues, de traduction, de bilinguisme, de variation et d'hétérogénéité des pratiques, de leur transmission… Vaste domaine de ce qu'on appelle la sociolinguistique, où une large place est faite en outre aux *représentations*, tel le mythe sacralisé du pays à langue unique, contre l'évidence du plurilinguisme immémorial des Français. Les langues sont pour une large part des objets intellectuellement construits, tributaires de conditions historiques déterminées, qui ne préexistent pas aux pratiques sociales et aux productions culturelles qui leur donnent corps.

Toute politique repose sur des savoirs : ce principe est la règle d'or de l'action linguistique du ministère de la Culture. Il fonde la place qui y est faite à la recherche.

À travers des enquêtes comme celle menée par l'INED au niveau national à l'occasion du recensement de 1999 ou, à l'échelle d'une ville, le programme « Enquêtes sociolinguistiques à Orléans », l'observation des pratiques linguistiques révèle une France où l'on parle des centaines de langues, du basque, depuis 20 000 ans, au pulaar, depuis deux ou trois lustres. De même, évoquer « les » vingt-trois langues de l'Union européenne ne rend pas justice à la diversité intrinsèque de l'ensemble considéré : il s'agit là des langues officielles, mais il y en a beaucoup d'autres…

On considère qu'il y a entre trois et six mille langues dans le monde. Ces chiffres, le classement, le dénombrement et même la dénomination des langues sont déjà en eux-mêmes un objet de recherche et de débat (de représentations), mais nous prenons conscience qu'un grand nombre d'entre elles sont en danger, fragilisées notamment par les formes que prend le mouvement de globalisation du monde.

Le langage est un élément primordial de notre humanité. Comme l'indiquait Henri Meschonnic, c'est dans et par le langage que chaque être humain se constitue dans son histoire, et nous vivons en permanence dans l'implication réciproque des problèmes du langage et des problèmes de la société.

Or, on constate dans la société un appétit de connaissance pour ces questions. Il peut s'exprimer d'une façon aberrante ou naïve, qu'il faut recadrer : il existe un faux savoir sur les langues. En relèvent les idées de hiérarchie ou de beauté des langues, de génie des langues, de clarté de la langue française, d'identification de la langue à la Nation, la confusion langue-orthographe, les formules ossifiées comme « il n'y a pas de voyelles en arabe », « un mot qui n'est pas dans le dictionnaire n'existe pas », etc. La recherche permet de donner des idées justes et claires, et de répondre au désir de connaissance du public, nouveau en France, à cette volonté d'être éclairé et de combler un manque.

Dans le besoin de problématiser la coexistence des langues, le sentiment de perte d'influence du français dans le monde joue un rôle. Même si ce recul est relatif

(en valeur absolue, le français n'a jamais eu autant de locuteurs), il peut avoir un effet positif et servir de tremplin à des entreprises visant à « outiller » la langue pour la création, les communications numériques, le dialogue interculturel, la traduction et l'ensemble des enjeux et défis auxquels nous nous trouvons aujourd'hui confrontés, à commencer par la place grandissante de l'anglais dans le monde. Dans cette partie, la question des ressources, des corpus, des bases de données est décisive. Plusieurs contributions témoignent de l'importance des nouvelles techniques dans les transformations en cours et montrent les voies de l'avenir, des dictionnaires à la traduction automatique. Autant d'éléments qui font l'objet de recherches et informent les politiques linguistiques.

La langue des signes française (LSF) est à l'heure actuelle un domaine de recherches particulièrement actif où, de nouveau, il faut voir un lien direct avec les avancées sociales, et où les nouveaux outils trouvent immédiatement à s'employer. La loi de février 2005 *sur l'égalité des droits et des chances, la participation et la citoyenneté des handicapés* a reconnu la LSF comme langue à part entière, notamment à l'école. Le ministère de la Culture participe du mouvement en soutenant des travaux sur le bilinguisme français-LSF, pour une forme écrite de la langue, pour l'élaboration de dictionnaires collaboratifs en ligne, etc. Dans une large mesure, c'est grâce à la LSF que se transforment les manières traditionnelles de penser la pluralité en France…

Cela consiste souvent à porter un regard nouveau sur des réalités anciennes. À mettre en question(s) ce qui semblait aller de soi, à penser ce qui n'était pas pensé, comme les artistes donnent à voir ce qu'on n'avait jamais vu (et en cela la recherche rejoint la démarche artistique). Ainsi, les langues ont toujours été en contact. De tout temps, dans leur quête de moyens pour parler à l'autre, les hommes ont eu recours à la traduction, à l'interprétariat et, sans le théoriser, à ce que nous appelons l'intercompréhension. Et de tout temps, les contacts de langues ont donné lieu à des phénomènes d'emprunt, d'alternance de codes, d'hybridation, d'échanges, mais aussi de conflit, de domination. L'histoire des langues récapitule l'histoire des hommes.

Ce qui ressort des travaux contemporains, c'est que le cerveau est fait pour connaitre plusieurs langues et fonctionne au mieux de ses capacités dans la mise en œuvre de plusieurs langues. L'aptitude humaine au langage ne se réalise qu'au pluriel : aussi haut qu'on remonte, nous sommes toujours en présence d'une humanité à plusieurs langues, et chaque langue est faite elle-même de pluralité interne : la variation et l'instabilité sont le mode normal de fonctionnement des langues naturelles. De toutes et de chacune. C'est ce que disent les textes rassemblés dans ce dossier.

Sur tous ces points, le ministère de la Culture finance et encourage nombre de recherches à travers l'Observatoire des pratiques linguistiques.

Depuis toujours des langues apparaissent, évoluent, se transforment et disparaissent. Ce n'est pas une catastrophe tant que la pluralité est préservée. Ce n'est pas mourir qui est grave, c'est d'être tué. Pas d'attachement nationaliste à une langue : c'est bien à la pluralité qu'il faut s'attacher.

Qu'est-ce qui se joue dans la pluralité des langues ? Si les langues n'étaient que des moyens de communication, de simples canaux de transmission pour véhiculer un contenu unique sous différents habillages de sons, elles pourraient aisément se substituer les unes aux autres, sans dommage, et leur réduction à une seule serait une bonne affaire pour l'humanité. Pour assurer une bonne communication entre les hommes, en effet, une seule langue suffirait.

« Il faut maintenir la diversité des langues, qui est une diversité de points de vue

Mais une langue n'est pas un instrument neutre de communication des idées, indifférent à leur formation ; la langue intervient *dans la production* même de la pensée[1]. Dans une certaine mesure, c'est une manière chaque fois différente de percevoir et de penser le monde, et donc chaque fois une possibilité de le transformer. En russe, le bleu du ciel n'est pas la même couleur que le bleu de la mer. L'anglais *river* implique une vision des cours d'eau différente de celle qu'exprime le français avec les mots *rivière* et *fleuve*.

On a du mal à imaginer de nouvelles avancées de l'esprit humain dans une situation de langue unique. Car les besoins de l'intelligence et de la sensibilité humaines excèdent toujours les possibilités d'expression d'une langue particulière. C'est pourquoi il faut maintenir la diversité des langues, qui est une diversité de points de vue : chaque langue ne peut penser qu'une petite partie de ce qui est pensable, ne peut dire qu'une petite partie de ce qui est dicible. C'est dans la création artistique et les inventions de pensée que cela apparait : la valeur d'une langue, c'est d'abord ce qui s'invente en elle, les œuvres originales dont elle est l'occasion et le matériau. Il n'y a pas la langue d'un côté et la culture de l'autre, mais des langues-cultures.

1. Voir Marc Crépon, Les géographies de l'esprit, chap. IX : « La diversité des langues selon Wilhelm von Humboldt », Paris, Payot, 1996.

États généraux du multilinguisme dans les Outre-Mer

Cayenne (Guyane), décembre 2011

Rencontres coordonnées par la DGLFLF.

Ces États généraux rassembleront, pendant deux jours et demi, quelque 250 participants venus de la Guyane et de l'ensemble des territoires d'Outre-Mer, de métropole et des pays voisins, avec pour objectif de formuler des préconisations pour la définition d'une politique des langues spécifique pour les Outre-Mer.
www.2011-annee-des-outre-mer.gouv.fr

Titre	***(Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ?***
Type	Article
Editeur	
Année	2011
Référence	Baude O., Duga C. (2011) « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? » *Corpus 10*, *Varia*, 99-118.

(Re)faire le corpus d'Orléans quarante ans après : *quoi de neuf, linguiste ?*

Olivier BAUDE
Céline DUGUA
LLL[1]

Résumé : La comparaison de deux corpus d'enquêtes sociolinguistiques réalisés à quarante ans d'intervalle permet de mettre en perspective certains aspects centraux de la constitution des données et d'interroger, par delà une description des différents choix méthodologiques et théoriques opérés, la place des données dans la linguistique de corpus.

Abstract : Comparing two corpora based on sociolinguistic studies and carried out forty years apart enabled to underline certain central questions of corpus constitution, and question the description of méthodological and theorical choices, and more generally the status of the data in linguistic corpora.

Mots clés : corpus oral, variation, transcription, liaison
Key words : oral corpora, variation, transcription, liaison

Introduction[2]

Que font les linguistes quand ils constituent et exploitent des corpus oraux ? La comparaison de deux projets d'enquêtes sociolinguistiques à des fins de constitution de corpus de référence, réalisés à quarante années d'intervalle offre l'opportunité d'aborder concrètement cette problématique.

L'enquête sociolinguistique à Orléans réalisée en 1968-71 (désormais ESLO1) avait pour objectif de fournir un corpus

[1] Laboratoire Ligérien de Linguistique, Université d'Orléans, EA3850.

[2] Les travaux présentés ont été soutenus par l'ANR *Corpus en SHS* et par la DGLFLF.

représentatif du français tel qu'il est parlé. Ce corpus relevait d'un certain nombre de choix théoriques, méthodologiques et technologiques qui étaient pour certains explicites, pour d'autres révélés lors des différentes opérations de mise à disposition du corpus et pour les derniers totalement implicites. Quarante ans plus tard, le laboratoire ligérien de linguistique de l'université d'Orléans a entrepris un double projet : diffuser largement le corpus ESLO1 dans un format correspondant aux outils de traitement des données actualisés et réaliser un nouveau corpus représentatif du français parlé à Orléans dans les années 2010 (désormais ESLO2), en prenant en compte l'expérience d'ESLO1 et l'évolution des cadres théoriques et méthodologiques de la constitution et de l'exploitation de grands corpus oraux à visée variationniste. C'est bien en effet la variation qui est au cœur de la problématique de la linguistique de corpus fondée sur l'enquête. Ainsi, dans sa préface de Labov (1976), Pierre Encrevé soulignait que « *le premier geste de la reconstruction labovienne, c'est de reposer les questions des* données *de la sociolinguistique* »[3].

Or il n'est pas évident que la constitution de grands corpus réponde à l'objectif « d'adéquation observationnelle » souhaitée par la sociolinguistique et ce pour deux raisons que nous aborderons dans cet article. Premièrement, la méthodologie de corpus implique un degré de « figement » des données d'autant plus contraignant que la masse d'informations est importante et que des outils de traitement sont requis. Deuxièmement, la reconnaissance de l'hétérogénéité des données se heurte à la nécessité de catégoriser les constituants (locuteurs, situations) de cette pratique sociale, pour borner la représentativité du corpus. Dans cet article, nous nous intéressons aux opérations de transcription et de classification sociale des locuteurs. Nous défendons l'idée que le *figement* des données inhérent au travail de constitution et d'exploitation de

[3] « Les « données » de la langue dans son usage quotidien, telle que veut l'étudier Labov, ne seront « produites » qu'au terme d'un long chemin d'aveuglette où se construit pas à pas une science de l'enquête linguistique qui est la première conquête de la sociolinguistique. » (Labov, 1976: 13).

corpus est paradoxalement une opération génératrice de variations.

1. D'ESLO1 à ESLO2

ESLO1 a été conduite à partir de 1968 par des universitaires britanniques avec une visée didactique : l'enseignement du français langue étrangère. Les buts du projet sont clairement exprimés dans le texte de présentation du catalogue des enregistrements :

> (…) dès le début il s'agissait d'autre chose que d'une simple chasse aux images sonores ; bien sûr, il fallait fixer des propos vivants, mais d'une façon systématique, afin de permettre des études fondamentales dans le domaine de la linguistique descriptive, sans lesquelles le renouveau de la pédagogie ne serait, au mieux que superficiel (Lonergan, *et al.*, 1974 : 1).

Il n'est pas anodin que le catalogue des enregistrements soit le document de référence de présentation de cette enquête. Il s'agit ici du cœur du projet : rendre disponible l'ensemble du corpus « à tout chercheur intéressé ».

Le second objectif était de constituer un corpus « sociolinguistique » :

> Selon nous une recherche sociolinguistique impliquait une étude de la langue dans sa diversité plutôt que comme un tout homogène et figé. En effet, même si on étudie un état de langue à un moment précis de l'histoire, il n'empêche qu'il offre une variété à plusieurs niveaux : différences entre les générations ; différences dialectales entre communautés ; différences entre les milieux sociaux ; différences liées aux conditions de production du discours. (Blanc & Biggs, 1971: 16)

Ces deux objectifs ont permis de construire le corpus autour du concept de « portrait sonore d'une ville » afin de croiser représentativité et variations au sein d'une communauté d'auditeurs dans un espace géographique et socioéconomique clairement défini (Bergounioux, *et al.*, 1992 : 79).

De quoi est composé ce corpus ? Selon le catalogue (Lonergan, *et al.*, 1974), il y a 487 enregistrements divisés en huit catégories générales, allant de l'entretien à la conférence en passant par des enregistrements spontanés sur les marchés ou dans la rue. L'ensemble des situations représente 315 heures d'enregistrement évaluées à 4 500 000 mots[4].

Dans sa version d'origine, le corpus comprend, outre des bandes magnétiques, un catalogue reprographié des enregistrements de 265 pages, des fiches d'identification des locuteurs, des fiches relevant les réponses au questionnaire sur les pratiques culturelles et 3365 feuillets présentant les extraits de transcriptions manuscrites ou tapuscrites. Dans les années 1980-90, une partie du corpus a été transcrit et étiqueté puis mis à disposition sur la toile dans le cadre du projet ELILAP / LANCOM[5].

En 2003, l'équipe de l'université d'Orléans entreprend la numérisation du corpus ESLO1 afin de le rendre disponible dans son intégralité. Loin d'un simple transfert de support, il s'agit véritablement de reconstruire le corpus, c'est d'ailleurs sur ce constat que le projet de réalisation d'ESLO2 s'est concrétisé autour de la volonté de maitriser l'ensemble de la chaine, de la constitution à l'exploitation d'un corpus. A terme ESLO2 comprendra plus de 350 heures d'enregistrement afin de former avec ESLO1 un corpus de plus de 700 heures et atteignant les dix millions de mots.

Au-delà d'une visée cumulative qui consisterait simplement à accroître la quantité de données pour fournir des éléments d'analyse et assurer des comparaisons avec d'autres corpus, l'enjeu des enquêtes conduites dans ESLO2 est aussi réflexif (accompagner la campagne de collecte, traiter et exploiter les données pour contribuer à la définition des normes). La mise en œuvre de cette conception implique :

- une prospective sur l'exhaustivité des usages,
- un inventaire des techniques de collecte (formats d'enregistrement et numérisation),

[4] Environ 70 % du corpus présente une qualité acoustique suffisante pour une transcription.

[5] ELILAP 1980-83 puis LANCOM 1993-2001, voir Mertens (2002).

- une politique de formation des enquêteurs et d'information des témoins afin d'intégrer dans les critères de variation celle liée à l'enquêteur,
- un recueil des données en conjonction avec le recueil des métadonnées,
- un codage et un catalogage anticipant les principales requêtes,
- une transcription avec alignement sur le signal,
- un étiquetage, avec catégorisation et lemmatisation,
- une analyse syntaxique (*parsing*), en particulier pour la co-référence anaphorique,
- une procédure d'anonymisation,
- un stockage, avec archivage et indexation,
- une procédure de mise à disposition sur la toile,
- des données partagées par interopérabilité.

Derrière ces objectifs se dresse la volonté d'élaborer ESLO2 en écho à ESLO1. Nous allons maintenant décrire deux aspects caractéristiques de cette évolution des corpus.

2. Procédures de transcription

La plus grande difficulté à laquelle ont été confrontés les auteurs d'ESLO1 a été indéniablement la phase de transcription. Face à la taille du corpus, l'équipe était démunie à la fois sur le plan technologique (le corpus n'était pas informatisé) et sur le plan théorique (les grands travaux sur la transcription n'étaient pas encore publiés). Cependant, force est de constater que, sur ce point comme sur de nombreux autres, les choix et intuitions de l'équipe ont été novateurs.

Le corpus ESLO1 a été transcrit en plusieurs étapes et à différentes époques avec des objectifs eux aussi différents. Les premières transcriptions datent du moment du recueil et des quelques années qui ont suivi. Trente-six bandes, puis cinquante-six, tout ou en partie ont été transcrites à cette époque. Les transcripteurs travaillaient alors sur papier et sur machine à écrire. Dans un deuxième temps (années 1993-2001), une partie du corpus a été repris par des chercheurs de l'Université de Louvain (Debrock*, et al.*, 2000) dans le cadre du

projet Elicop[6]. Enfin, depuis 2003, le LLL (Orléans) s'est donné pour objectif de transcrire et rendre disponible l'intégralité du corpus en y associant le son, des annotations – dont la transcription – et des métadonnées. Certes l'évolution des transcriptions sur 40 ans dépend fortement des technologies, mais c'est aussi la définition même de l'écriture de l'oral qui va être bouleversée. Dès l'origine du projet, les chercheurs ont conscience de la difficulté de la tâche de transcription qui consiste à « essayer de rendre par écrit un discours sonore [ce qui] suppose l'adoption d'un code » (Blanc & Biggs, 1971 : 19). Dans les premières transcriptions d'ESLO1, le code adopté est dit semi-orthographique :

> les écarts par rapport à l'orthographe traditionnelle sont : (1) l'absence de ponctuation, qui implique la non-délimitation des phrases, la suppression des lettres majuscules : à sa place, des signes marquant les pauses, évaluées sur une échelle de durée à trois degrés : minime, moyen, long ; (2) la notation des liaisons non évidentes, des élisions, des allongements de syllabes, des *e* normalement élidés mais prononcés dans ce cas (Blanc & Biggs, 1971 : 20).

Ces choix relèvent d'un souci fort louable : formaliser le signal acoustique (tour de parole, énoncés, mots) afin d'offrir le meilleur matériau de référence pour l'analyse. A ce stade, le son peut disparaitre et les transcriptions devenir les « données primaires » ; c'est d'ailleurs ce qui arrivera par la suite dans le projet Elicop où, sur un total de 118h30, 80 heures proviennent du corpus ESLO1 (64 entretiens, environ 900.000 mots) (Mertens, 2002). Dans ce projet, la mutualisation des corpus et la volonté de pouvoir les interroger sous forme de concordancier et de corpus étiqueté ont déterminé les types de codage utilisés, à savoir un codage orthographique et/ou phonétique, le balisage de certaines particularités de l'oral comme les pauses, les liaisons, les élisions et l'annotation de certains éléments au format SGML (Mertens, 2002).

Dans le projet du LLL, la transcription n'est plus conçue uniquement comme le préalable à une étude sur corpus

[6] http://bach.arts.kuleuven.be/elicop/

oraux, elle est une façon de mettre en perspective les conditions de productions des données. En ce sens, elle constitue une étape qui reflète le champ de la linguistique, les théories, l'inscription du chercheur dans son domaine, et également les « attitudes » et les représentations des transcripteurs[7]. A partir des archives sonores et des transcriptions issues du projet Elicop[8], l'équipe du LLL a transcrit l'ensemble des entretiens (151 enregistrements), les entretiens de personnalités (34 enregistrements), ainsi que d'autres situations (84 enregistrements)[9] afin de disposer d'un panorama varié des différents types d'enregistrements, dans la mesure où les qualités acoustiques permettent une transcription.

La contrainte forte qui apparait et qui détermine nos choix d'annotation est la volonté de transcrire et rendre disponible une grande quantité de paroles (environ 700h). Nos conventions de transcription ont été élaborées à partir de la comparaison des conventions des différents grands projets de corpus oraux francophones dégageant ainsi les principes communs et affinant nos particularités au regard des objectifs propres du projet. Nous avons adopté des principes de base généralement partagés à savoir une transcription orthographique qui conserve les spécificités de l'oral (amorces, disfluences, répétitions, etc.), sans usage de la ponctuation, et avec la segmentation des tours de paroles. Un élément fondamental réside dans la synchronisation entre la transcription et le signal sonore par l'ajout de jalons temporels à l'aide du logiciel Transcriber[10]. Les conventions propres au projet ont été réduites au minimum, en suivant Ochs (1979 : 44) « a more useful transcript is a more selective one ».

---

[7] Les questions liées aux attitudes des transcripteurs ne seront pas abordées ici par manque de place, voir Hriba (en cours) pour plus de précisions.

[8] Les premières transcriptions tapuscrites que nous avons récupérées sous format papier n'ont finalement pas été utilisées, en revanche, celles issues de Elicop ont été intégrées à notre procédure de transcription.

[9] Ces chiffres renvoient au nombre de transcriptions en version B (voir infra) disponibles en mai 2011, et correspondent à 230 heures d'enregistrements.

[10] http://trans.sourceforge.net/en/presentation.php

Transcrire est une tâche fastidieuse et complexe qui implique de mettre en œuvre simultanément un nombre important de compétences, notamment : écouter, segmenter en tours de parole, en questions, attribuer la parole au bon locuteur, utiliser le clavier, respecter l'orthographe, respecter les conventions, anonymiser, etc. Par ailleurs, tous les transcripteurs n'ont pas la même connaissance des normes orthographiques et de nos propres conventions, ils ont aussi un rapport à l'écrit et à la norme différent. Conscients de ces contraintes inhérentes à toute tâche de transcription de l'oral, nous mettons en œuvre une procédure dont l'objectif est de rendre explicite notre démarche afin de fournir au chercheur les outils qui lui permettront de « reconstruire » le travail de transcription.

Une manière de garder la trace de la procèdure d'élaboration de nos transcriptions consiste à conserver les trois versions produites ainsi que les versions du *Guide du transcripteur* (document qui s'étoffe au fur et à mesure pour atteindre 36 pages dans la quatrième version), mais aussi les questions posées par les transcripteurs lors de leur formation et de leur travail et les réponses apportées. Une fois la maitrise du logiciel satisfaisante, l'essentiel des questions porte sur la façon de graphier certaines particularités orthographiques ou typiques de l'oral comme les néologismes, les régionalismes, les mots issus du verlan, l'usage ou non des majuscules, etc. Les réponses à ces questions sont recensées dans un lexique qui répertorie l'ensemble des décisions prises.

Avec cette méthode des trois versions, nous évaluons le temps de transcription à 10 fois pour une première version brute, 5 fois pour une deuxième et autant pour une troisième.

La définition de ce que sont les trois versions de transcription s'est progressivement affinée. Au départ, la première était une version dite brute qui fournissait une transcription alignée au son ; cette version était ensuite relue par un deuxième transcripteur qui corrigeait les segmentations et les variations orthographiques et de conventions ; deuxième version validée enfin par un dernier transcripteur dont la tâche consistait essentiellement à vérifier l'orthographe et les conventions dans le but de rendre la forme écrite acceptable. De

plus, à chaque niveau, le transcripteur pouvait être amené à modifier des formes en raison de différences d'écoute ou de perception. Plutôt que des rajouts ou des suppressions, il s'agit essentiellement de modifications, telles que celle observée dans l'entretien ESLO1_079 (4'15), où un même segment est transcrit sous trois formes différentes dans les versions A, B, C, respectivement : « enfin reprendre à travailler après », « enfin elle recommence à travailler après », « enfin elle reprend le travail après ». Plus généralement, Hriba (en cours), à partir de neuf entretiens ESLO1 (13h d'enregistrements, environ 105 000 mots), transcrits sous les 3 versions, a ont comptabilisé le nombre de modifications apportées à une transcription dans ses passages entre les trois versions. Il en ressort que pour un entretien, en moyenne, 790 modifications sont apportées : 290 entre les deux premières versions et 500 entre les deux suivantes.

Une deuxième procédure de transcription, en phase de test, consiste à définir plus précisément en quoi consiste chacune des versions en termes de tâches. Nous avons répertorié quatre tâches principales : l'écoute, la segmentation, la transcription et l'anonymisation, dont l'importance se répartit différemment selon les versions de transcription. Clarifier les tâches et les répartir dans les trois versions de transcription devrait permettre d'être plus efficace tant au niveau du temps de travail que de la qualité des transcriptions puisque nous limitons le nombre de compétences à mettre en œuvre pour une version donnée. Ainsi, la première version s'attachera essentiellement à la segmentation (en tours de paroles, codage des questions, attribution des codes locuteurs) et à la transcription brute des passages les plus facilement audibles ; la deuxième à affiner les incertitudes d'écoute et à vérifier les conventions et règles d'orthographe ; enfin la dernière à relire précisément l'orthographe et les conventions et à anonymiser les noms de famille ainsi que les passages dits délicats. Les transcriptions rendues disponibles au grand public seront les troisièmes versions, que nous savons être provisoires ; les utilisateurs auront d'ailleurs la possibilité de proposer des corrections tant orthographiques que d'écoute.

Ce travail réflexif sur la procédure de transcription est instructif. La reconnaissance de la légitimité de l'oral a orienté les recherches vers la volonté d'avoir une transcription fidèle reliée à un cadre théorique fort (notamment depuis Blanche-Benveniste). L'analyse des différentes transcriptions des ESLO[11] et des différences dans les versions de transcription montrent que cette opération est génératrice de variations. Face à ce problème nous proposons de concevoir la transcription comme une simple annotation de bas niveau qui permet de faciliter l'accès à la source sonore. Pour contrer les biais importés par cette annotation, il convient de lui redonner son rôle (celui de simple outil) et de la documenter avec beaucoup de précisions.

La transcription n'est pas la seule opération qui « instrumentalise » véritablement un corpus. La construction de la représentativité des données est également au cœur du travail du linguiste. Nous allons voir que là aussi il est nécessaire de porter une attention particulière aux effets des pratiques scientifiques qui sont loin d'être toujours explicites.

3. Représentativité des données

Le travail le plus novateur d'ESLO1 a sans conteste été l'approche sociologique de la représentativité des données. Cette question a été abordée selon deux axes : l'échantillonnage des locuteurs et la diversité des situations de communication enregistrées. Il est étonnant de constater que le deuxième axe a été traité dans un flou théorique relativement important voire avec une certaine naïveté alors que le premier révèle ce que la sociologie et l'enquête statistique portaient de plus rigoureux. Sur ces deux axes les choix d'ESLO2 sont sensiblement différents.

Le premier choix de l'équipe d'ESLO1 a été de limiter « pour des raisons d'ordre théorique et pratique, pour écarter aussi des variables incontrôlables » (Blanc & Biggs, 1971 : 16-17) le nombre de variables recherchées dans le cadre d'un

[11] Les choix de transcription dans ESLO2 sont également ceux du LLL dans sa nouvelle transcription d'ESLO1.

corpus sociolinguistique. Le choix s'est alors porté sur « le portrait sonore d'une ville ».

> C'est une communauté d'auditeurs qui est construite, autant qu'une communauté de locuteurs, à notre connaissance pour la première fois en France (…) On ne cherche pas cet individu mythique, l'Orléanais moyen (Blanc & Biggs, 1971 : 23).

Comment ne pas voir dans ces choix une similitude d'approche avec la linguistique variationniste se développant en France à la même époque (Labov, 1976), qui situe la langue du côté de la réception et qui lui confère de par sa nature sociale des variations instituées socialement ? Si le choix d'une délimitation sociogéographique d'une ville paraissait pertinent à la fin des années soixante, il est apparu aux auteurs d'ESLO2 qu'il convenait d'avoir, 40 ans plus tard, une définition des contours de la ville qui accepte des locuteurs vivant ou travaillant dans l'ensemble des communes limitrophes.

La collectivité choisie, se pose alors la question de l'élaboration de l'échantillon. Pour ESLO1 le cadre théorique sociologique est clairement celui des enquêtes statistiques de l'INSEE :

> Ce sont les services de l'INSEE, qui sur des instructions des membres de l'équipe de l'enquête, ont procédé au tirage au sort de six cents témoins répartis également entre six catégories socioprofessionnelles (Blanc & Biggs, 1971 : 17).

Les autres critères sont le sexe et l'âge. Cette méthodologie offrait clairement la possibilité d'analyses de la co-variation en croisant stratification sociale et variations linguistiques.

ESLO1 témoigne néanmoins des prémices d'un changement théorique en sociologie. Ainsi Alix Mullineaux (Mullineaux & Blanc, 1982) proposa une échelle (dorénavant échelle AM) qui complète les critères de l'INSEE par le diplôme et l'âge de fin d'étude des témoins. Cette nouvelle grille comprend cinq agrégats (de A à E). Or cette première étape, qui annonçait un travail conjoint avec le centre de Sociologie Européenne développant alors les travaux de Pierre

Bourdieu autour du capital culturel et de la distinction, est restée inachevée. Les entretiens furent toutefois complétés par un questionnaire sociolinguistique destiné à capter les pratiques et les représentations des locuteurs en ce qui concerne la langue, puis par un questionnaire sur l'ensemble des pratiques culturelles.

Cette approche était présentée comme particulièrement prometteuse dans la préface du catalogue de 1974.

> L'échelle de catégories socio-culturelles construite par Alix Mullineaux constitue une tentative de classement de la population française en fonction des paramètres de mobilité sociale et de niveaux de culture, et par là, marque un pas important vers l'élaboration des échelles proprement sociolinguistiques indispensables à la nouvelle discipline. (Lonergan, *et al.*, 1974 : 1)

A notre connaissance ce travail, pourtant central dans ESLO1 n'a pas été poursuivi et l'échelle AM qui visait l'amélioration de la catégorisation des témoins par le croisement des critères socioéconomiques avec les critères de niveau scolaire et les critères de capital culturel (dont linguistique) n'a pas été reprise.

En revanche, cette perspective apparait clairement dans l'élaboration d'ESLO2. En premier lieu le bilan de l'échec de la représentativité des critères de l'INSEE dans ESLO1 a été pris en compte. En effet, sur les 600 personnes sélectionnées au hasard, seules 144 ont accepté de participer à l'enquête et bien évidemment selon une répartition déséquilibrée.

L'échantillon d'ESLO2 a été conçu sur la base des critères de l'INSEE couplés à ceux de l'âge et du sexe, simplement comme point d'entrée afin d'orienter la sélection des 150 personnes qui participeront à des entretiens. La mise en place des entretiens a, quant à elle, bénéficié de nouvelles théories développées depuis les années 1970 :

- les développements de l'analyse de la conversation en linguistique et de l'ethnométhodologie qui se construisent sur une opposition frontale entre données provoquées et non provoquées par le chercheur ;

- les recherches en sociologie et en anthropologie sur les techniques d'enquête (Beaud & Weber, 1997 ; Bourdieu, *et al.*, 1968) ;

- les travaux sur la linguistique des genres (Biber, *et al.*, 1998) et sur une typologie des productions liée à une typologie des situations de communications (Koch & Oesterreicher, 2001) qui restreignent les entretiens à un contexte de production linguistique particulier ;

- les travaux sur les effets du dispositif technologique (qualité et discrétion du dispositif d'enregistrement, traitement des données numériques et même développement de la vidéo favorisant les travaux en linguistique interactionnelle (Goodwin, 1994 ; Mondada, 2006).

La méthodologie de l'enquête que nous ne développerons pas davantage ici s'appuie sur la nécessité de produire des entretiens qui permettront, après analyses, de procéder à une classification sociologique des locuteurs. Pour cela, l'équipe s'est tout d'abord appuyée sur un riche « portrait socio-économique de territoire[12] » rédigé par l'INSEE et comprenant 75 cartes illustrant des informations obtenues lors du recensement de 1999. Ensuite le mode d'approche des locuteurs et la « menée » de l'entretien ont été particulièrement étudiés afin de faciliter le recueil d'informations sur les itinéraires de vie et les pratiques culturelles des locuteurs tout en favorisant une parole dans un style le moins formel possible. Ainsi, les entretiens favorisent des discussions libres autour de la vie quotidienne (le quartier, les commerces, les loisirs, etc.).

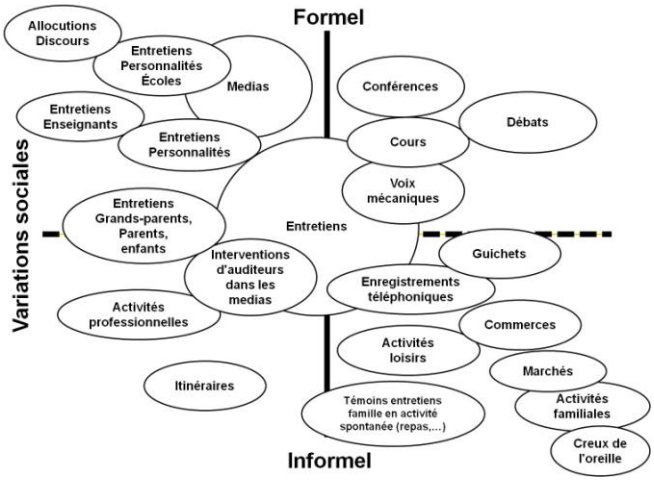
Il est alors clair que les entretiens de l'ESLO2 ne formeront pas un échantillon sociologiquement équilibré mais plutôt une archive – ou un « réservoir » – qui permettra de sélectionner à l'aide d'outils statistiques grossiers (INSEE) mais aussi à l'aide de descripteurs fins après une analyse de contenus réalisée ultérieurement, un véritable corpus à des fins d'études linguistiques.

La représentativité des données a été également abordée dans ESLO1 selon l'axe de la variation diaphasique. La variation diaphasique a cependant bénéficié d'un traitement

[12] Portrait de Territoire INSEE : 75 cartes d'analyses issues du recensement.

bien moins rigoureux que la variation diastratique mais néanmoins avec une forte intuition au regard de l'évolution des théories linguistiques (typologie des genres et linguistique interactionnelle) qui s'est développée par la suite. Dans ESLO1 cette variation a été appréhendée en enregistrant une quinzaine de témoins dans diverses situations (au téléphone, sur leur lieu de travail, en réunion, lors d'un repas de famille, etc.) et par des enregistrements en micro-cachés « au hasard des rencontres, dans la rue dans les magasins, etc. » (Blanc & Biggs, 1971 : 19). Outre les entretiens, deux autres situations avaient été clairement déterminées : l'enregistrement de tables rondes ou/et débats / conférences et des entretiens au CMPP. Pour des raisons techniques liées à la qualité acoustique mais aussi pour des raisons déontologiques et juridiques (micro-cachés) beaucoup de ces enregistrements posent des problèmes de transcription et d'exploitation. Cette partie du corpus, passionnante dans les perspectives recherchées, reste donc quasiment inexploitée.

Pour ESLO2, nous avons réparti un ensemble d'une vingtaine de modules selon deux axes : l'un prend en compte le degré de formalité de la situation et l'autre celui de la classification sociale des locuteurs :



Dans cette architecture du corpus la place des entretiens est relativisée dans une nouvelle perspective de typologie des genres de discours affinée.

La définition des « genres » repose également sur un ensemble de descripteurs contextualisant les productions linguistiques enregistrées. Afin de préserver l'accessibilité de ces descripteurs, il convient de développer des formats de codage et de catalogage. Il y a donc un enjeu à considérer ces métadonnées comme des éléments de description des données linguistiques et non simplement en termes de documentation des sources. Elles doivent permettre d'expliciter la démarche du chercheur en proposant une description fine de ses choix théoriques « encapsulés » dans des choix techniques. Si dans ESLO1 les « métadonnées » se résumaient en un catalogue constitué de fiches tapuscrites (Lonergan, *et al.*, 1974), les ESLOs bénéficient des expériences actuelles de communautés scientifiques (OLAC et DUBLIN-CORE pour le catalogage, TEI pour le codage, pour ne citer qu'elles) qui ne permettent cependant pas de structurer toutes les « données situantes » utilisées par les équipes de recherche à l'origine des deux enquêtes. Néanmoins, ces métadonnées ont été intégrées à une base de données qui permet de faire des requêtes sur le corpus en croisant les recherches sur les transcriptions et les descripteurs des locuteurs et des situations.

Ces choix de constitution et d'exploitation des corpus ESLO ont-ils un réel impact sur le potentiel d'analyse du corpus ? Afin de répondre à cette question nous nous sommes livrés à une analyse test sur le phénomène de la liaison.

4. L'exemple de la liaison

Depuis les années 1970 et tout au long des 40 ans qui se sont écoulés, la liaison a fait l'objet d'études essentielles, dans des domaines variés de la linguistique : phonologie, sociolinguistique, étude de corpus, acquisition, pour n'en citer que quelques-uns. Une des plus fameuses recherches menées sur ESLO1 porte sur la réalisation des liaisons selon une approche à la fois phonologique et sociolinguistique (De Jong, 1988, 1994). Il s'agit d'un travail descriptif d'étude de corpus qui vise à rendre compte des fréquences objectives de réalisation des liaisons, en fonction des catégories grammaticales, du lexique et des caractéristiques sociologiques

des locuteurs. En se limitant à certaines formes verbales, par exemple la forme « est », De Jong (1994) note un effet de la catégorie socio-économique, un effet du sexe (les femmes réalisent davantage la liaison) et un effet de l'âge (les témoins plus avancés en âge font davantage la liaison). Pour d'autres formes verbales, comme les auxiliaires de mode, les résultats diffèrent. Sur cet exemple en particulier, seul l'effet de l'échelle AM est significatif.

Nous avons sélectionné un sous-corpus afin de le confronter aux résultats généraux sur l'usage de la liaison analysés par De Jong. Ce sous-corpus est représentatif de la variété des situations composant ESLO1 (7 des 8 catégories sont représentées, de la prise de contact au repas de famille en passant par des visites professionnelles et des entretiens) et de la variété des locuteurs (de A à D sur l'échelle AM), en accordant une priorité aux enregistrements recoupant ces deux axes de variations – ainsi un même locuteur est enregistré dans 5 situations différentes. Les enregistrements ont une durée qui varie de 1 à 89 minutes pour un total de 11h08, soit environ 100 000 mots. A partir du travail de De Jong (1994), nous avons sélectionné 4 contextes lexicaux pouvant entrer en contexte de liaison (*est*+X, *sont*+X, *ont*+X, *quand*+X), des formes repérées comme fréquentes dans son corpus, qui proviennent de catégories grammaticales variées et qui présentent des taux de réalisation de la liaison relativement contrastés : une forme du verbe *avoir* après laquelle la liaison est peu réalisée (*ont*+X : 8.7% de liaisons réalisées), un « complémenteur » (terminologie utilisée par De Jong) après lequel la liaison est très fortement réalisée (*quand*+X : 96.3%) et deux formes du verbe *être* présentant un taux de réalisation moyen (*est*+X et *sont*+X, respectivement 69% et 46%). Dans notre sous-corpus, nous avons relevé 985 occurrences de liaisons possibles dans ces contextes. Une première analyse confirme les grandes tendances décrites par De Jong : les liaisons sont fortement réalisées dans le contexte *quand*+X (95,7%), moyennement dans le contexte *est / sont*+X (64%) et faiblement dans le contexte *ont*+X (20%).

Toutefois les variations entre les différents locuteurs sont fortes : *quand*+X de 90% à 100% ; *est / sont*+X de 36% à

100% et *ont*+X de 17% à 33%. Nous confirmons également qu'en groupant l'ensemble des situations nous constatons une co-variation entre l'échelle AM et le taux de réalisation des liaisons (de A : taux de liaison plus fort à D : taux de liaison plus faible). De même les variations pour un même locuteur selon les situations de communication sont extrêmement contrastées.

Ainsi, le locuteur 1134 (entretien 024) réalise 100% des liaisons après *est / sont* pendant l'entretien mais ce taux descend à 50% et même 19% en situation informelle lors d'enregistrements avant et après l'entretien[13]. Le locuteur BA725 (entretien 001) réalise quant à lui 75% de liaisons après *est / sont* en entretien et un taux relativement proche de 67% hors entretien. Enfin, un troisième locuteur, 1268 (entretien 029), réalise 78% de liaisons en entretien dans le même contexte et 75% hors entretien. Ces chiffres, éclairés par les informations sur le profil sociologique des locuteurs ramené à l'échelle AM, trouvent une explication compatible avec les théories développées en linguistique variationniste. La locutrice 1268, une jeune femme étudiante issue de la grande bourgeoisie et classée A par l'échelle AM, présente le taux de liaison le plus stable. Quant au locuteur 1134, vendeur de 48 ans ayant arrêté ses études à 13 ans (D sur l'échelle AM), il produit la plus forte variation avec un taux particulièrement faible dans la situation d'après entretien.

Le cas du second locuteur (BA725) est particulièrement intéressant car bien que d'un profil proche de 1134 – boucher de 57 ans ayant arrêté ses études à 14 ans (également D sur l'échelle AM) – son taux de liaison réalisée est relativement stable selon les situations et donc proche des productions de 1268 (A sur l'échelle AM). Toutefois, une analyse du contenu des entretiens permet de pondérer la classification AM. Ainsi 1134 était « charron », « rêvait d'être boulanger », a un « fils militaire et une fille mariée », « ne prend pas de vacances sauf

[13] Il est vraisemblable que les enquêteurs n'attiraient pas l'attention du locuteur sur le fait que l'enregistreur fonctionnait avant et après l'entretien. Outre le fait que cette technique était utilisée à l'époque par W. Labov, la qualité acoustique confirme cette hypothèse.

la pêche », a été « une fois au cinéma en 17 ans » et ne connait pas « le dictionnaire utilisé par [son] enfant ». BA725 souhaite devenir « gérant de plusieurs boucheries », il aime « la lecture et la musique », compte « visiter des musées quand [il sera] à la retraite »… Évidemment, ces quelques informations ne peuvent résumer le contenu de l'entretien ; cependant même si ces deux locuteurs sont classés avec le même code de CSP, et le même code de l'échelle AM (même niveau d'études et même âge de fin d'études), on repère aisément deux trajectoires différentes. Ces trajectoires sont cohérentes avec l'habitus linguistique de ces locuteurs, habitus qui entraine 1134 à adapter son taux de liaison au marché linguistique (entretien *vs* hors entretien) alors que BA725 possède la même stabilité que la locutrice 1268. Il ne s'agit ici que d'hypothèses qui ne peuvent être totalement étayées. Les faiblesses dans l'architecture d'ESLO1, le manque d'informations sur les cadres théoriques et surtout sur la méthodologie mise en œuvre par l'équipe de chercheurs tout comme la difficulté à adopter une démarche réflexive sur les effets de figement des données ne permettent pas d'explorer totalement ces pistes d'analyses.

Conclusion

Est-ce que les corpus ESLO1 et ESLO2 répondent aux objectifs de reconstruction de la linguistique à partir des données, tel que proposé dans le programme de la sociolinguistique ? D'une manière générale est-ce une bonne idée que de faire des corpus ? Ces questions méritent d'être abordées à l'aide d'analyses linguistiques. Ainsi, ce que nous apprend l'étude comparée de deux corpus à quarante années d'intervalle c'est surtout que le dialogue entre sociolinguistique et linguistique de corpus est des plus fructueux sous réserve que toutes les étapes de la collecte à l'analyse soient abordées avec la même rigueur et le même souci de réflexivité.

L'exemple d'analyse des variations du taux de liaisons réalisées selon le profil sociologique du locuteur et la situation (marché linguistique) illustre la difficulté pour le linguiste à manipuler les données d'un corpus. Si un « tamisage » (transcription et codage du signal, classification socio-

économique et échelle AM, catégorisation des situations) permet d'organiser les données afin de dégager des pistes d'études sous la forme de grandes tendances, l'analyse doit pouvoir s'appuyer sur une observation minutieuse des productions linguistiques en situation. Pour cela, il faut que l'élaboration du corpus favorise et anticipe ce retour aux données primaires non dégradées. C'est bien ce que visent les principaux choix de constitution d'ESLO2.

Références bibliographiques

Site ESLO : http://eslo.in2p3.fr

Beaud S. & Weber F. (1997). *Guide de l'enquête de terrain: produire et analyser des données ethnographiques*. Paris : La Découverte.

Bergounioux G., Baraduc J. & Dumont C. (1992). « L'étude socio-linguistique sur Orléans (1966-1991) : 25 ans d'histoire d'un corpus », *Langue française* 93 : 74-93.

Biber D., Conrad S. & Reppen R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge : Cambridge University Press.

Blanc M. & Biggs P. (1971). « L'enquête socio-linguistique sur le français parlé à Orléans », *Le français dans le monde* 85 : 16-25.

Bourdieu P., Chambordon J.-C. & Passeron J.-C. (1968). *Le métier de sociologue*. Paris : Mouton de Gruyter/Bordas.

De Jong D. (1988). *Sociolinguistic aspects of French liaison*, Academisch proefschrift. Amsterdam : Vrije Universiteit Amsterdam.

De Jong D. (1994). « La sociophonologie de la liaison orléanaise », in C. Lyche (éd.) *French Generative Phonology: Retrospective and Perspectives*. Salford : ESRI, 95-129.

Debrock M., Mertens P., Truyen F. & Brosens V. (2000). *ELICOP, Etude LInguistique de la COmmunication Parlée : Constitution et exploitation d'un corpus de*

français parlé automatisé. K.U.Leuven : Departement Linguïstiek.

Goodwin C. (1994). « Recording human interaction in natural settings », *Pragmatics* 3 : 181-209.

Hriba L. (en cours). *Identification automatique des locus de variation dans un corpus de français parlé*, Thèse de doctorat. Université d'Orléans, Orléans.

Koch P. & Oesterreicher W. (2001). « Langage parlé et langage écrit », in G. Holtus *et al.* (eds) *Lexikon der romanistischen Linguistik*. Tübingen : Max Niemeyer Verlag, 584-627.

Labov W. (1976). *Sociolinguistique*. Paris : Editions de Minuit.

Lonergan J., Kay J. & Ross J. (1974). *Etude sociolinguistique sur Orléans, catalogue des enregistrements*. Colchester : Multigraphié.

Mertens P. (2002). « Les corpus de français parlé ELICOP : consultation et exploitation », in J. Binon *et al.* (eds) *Tableaux Vivants. Opstellen over taal-en-onderwijs aangeboden aan Mark Debrock*. Leuven : Universitaire Pers.

Mondada L. (2006). « Video recording as the preservation of fundamental features for analysis », in H. Knoblauch *et al.* (éd.) *Video Analysis*. Bern : Lang, 51-68.

Mullineaux A. & Blanc M. (1982). « The problems of classifying the population sample in the socio-linguistic survey of Orléans (1969) in terms of socio-economic, social and educational categories », *Review of Applied Linguistics* 55 : 3-37.

Ochs E. (1979). « Transcription as theory », in E. Ochs & B. Schieffelin (eds) *Developmental pragmatics*. New York : Academic Press, 43-72.

Titre	***Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012***
Type	Article
Editeur	
Année	2011
Référence	Eshkol-Taravella I., **Baude O**., Maurel D., Hriba L., Dugua C., Tellier I., (2011) Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012. in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 17-46.

Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales[1] sur le français et les langues de France

Michel Jacobson * — Oliver Baude,*****

** Service interministériel des archives de France*
56, rue des Francs-Bourgeois
75003 Paris
michel.jacobson@culture.gouv.fr

*** Laboratoire Ligérien de Linguistique*
Université d'Orléans
olivier.baude@univ-orleans.fr

**** Délégation générale à la langue française et aux langues de France*

RÉSUMÉ. *Le programme «* Corpus de la parole *» est un projet en collaboration entre le ministère de la Culture et de la Communication et le CNRS qui vise à constituer une collection de ressources orales sur le français et les langues de France. Un portail Web offre un accès éditorialisé à cette collection. Cet article présentera les points principaux de l'organisation de ce programme, de la collecte des corpus aux aspects de pérennisation en passant par l'accès et la diffusion des données numériques.*

ABSTRACT. *"Corpus de la parole" is a collaborative project between the Ministry of Culture of France and the CNRS, which aims to build a collection of resources on French and other languages of France. A Web site provides an editorialised access to this collection. This article presents the main points of the organization of this program: the data collection, the access, dissemination and sustainability aspects of the digital data.*

MOTS-CLÉS : Corpus de la parole*, oral, archives ouvertes, OAIS.*

KEYWORDS: *Oral corpora, open archives, OAIS.*

1. Par commodité nous utiliserons le terme « orales » pour désigner des ressources orales mais aussi multimodales.

TAL. Volume 52 – n° 3/2011, pages 47 à 69

1. Contexte

L'origine du programme *Corpus de la parole* provient d'une volonté de l'Observatoire des pratiques linguistiques de la DGLFLF[2] de renouveler les initiatives de conservation et de diffusion des archives orales des linguistes commencées en 1911 avec la naissance des Archives de la parole de Ferdinand Brunot. Une centaine d'années plus tard, les questions de conservation et de diffusion relèvent également des aspects scientifiques de la constitution d'objets et de savoirs par une communauté de chercheurs.

Créé en 1999 au sein de la Délégation générale à la langue française, l'Observatoire des pratiques linguistiques a pour objectif de recenser, de développer et de rendre disponibles les savoirs relatifs à la situation linguistique en France, afin notamment de fournir des éléments d'information utiles à l'élaboration des politiques culturelles, éducatives ou sociales. Il a également pour but de faire mieux connaître un patrimoine linguistique commun constitué par l'ensemble des langues et des variétés linguistiques parlées en France, qui concourent à la diversité culturelle de notre pays. Doté d'un comité scientifique composé de linguistes, il soutient des projets ou des programmes de recherche dans le cadre d'appels à propositions thématiques ou de partenariats avec le CNRS ou les universités.

Le champ de l'observation est celui de la sociolinguistique et concerne les pratiques actuelles, qu'il s'agisse du français ou des autres langues parlées sur le territoire national : langues régionales ou langues issues des différentes vagues de migration. Depuis 2004, un des axes majeurs de l'activité de l'Observatoire est le développement, dans le cadre d'un partenariat Culture-CNRS, du programme *Corpus de la parole* qui a pour objectif la numérisation et la valorisation des corpus oraux (collections ordonnées d'enregistrements de productions linguistiques orales et multi-modales réalisées par des chercheurs) afin de permettre leur conservation et leur transformation en de véritables ressources linguistiques numériques, pour la recherche en sciences humaines, l'enseignement et l'ingénierie des langues. Il a permis, de 2006 à 2009, dans le cadre notamment du plan de numérisation du ministère de la Culture et de la Communication piloté par l'ex-MRT[3], de constituer et de numériser une collection de corpus oraux en français et en langues de France, mise à la disposition du public sur le site Internet *Corpus de la parole*[4], ouvert en février 2008.

Un *Guide des bonnes pratiques* (Baude, 2006) destiné aux chercheurs, a également été réalisé.

La démarche qui a présidé à la rédaction de guide était expressément centrée sur les pratiques des chercheurs afin de permettre une prise en compte des contraintes

2. Délégation générale à la langue française et aux langues de France.
3. Mission recherche et technologie du ministère de la Culture et de la Communication.
4. http://corpusdelaparole.in2p3.fr

scientifiques : méthodologie de collecte, opérations d'annotations et impacts sur les données.

Ce programme doit permettre, non seulement le développement d'une base de données patrimoniales sur l'oral, mais aussi le développement d'outils de traitement automatique des langues et d'ingénierie linguistique rendant possible l'interopérabilité des bases de données de grands corpus.

La DGLFLF, entre 2006 et 2008, a soumis une proposition à la MRT dans le cadre du « plan numérisation » du ministère de la Culture et de la Communication et a ainsi pu bénéficier d'une aide financière. Cette aide a permis la définition d'une organisation, de méthodes, la construction d'un prototype et son alimentation initiale en données ainsi que la numérisation de milliers de documents en différentes langues. La proposition de la DGLFLF était la constitution d'une base de données sur les langues de France constituée de ressources d'enregistrements audio ou vidéo récoltés au fil du temps par des scientifiques (linguistes, anthropologues...) pour leurs propres études. Le dossier soumis et accepté, montre une collaboration étroite entre la DGLFLF et le CNRS, ce dernier à travers deux acteurs principaux : le premier assurant un travail scientifique, le deuxième assurant un rôle de support technique.

La composante scientifique du projet est représentée par les deux fédérations de linguistique TUL[5] et l'ILF[6]. Son rôle consiste en l'identification des corpus existants et la constitution de nouveaux corpus sur des langues de France émanant de la communauté de recherche en linguistique. Ces corpus, dans la définition de départ du ministère de la Culture et de la Communication devant être composés de ressources existantes, anciennes, en danger et pour lesquelles la numérisation pouvait représenter une solution de sauvegarde, mais aussi de nouveaux corpus intégrant dès leur conception des nouvelles pratiques orientées vers la diffusion et la conservation. Pour ce projet, les fédérations de recherche en linguistique du CNRS exploitent leurs réseaux de contacts (laboratoires, projets, chercheurs) afin de faire émerger des propositions de collaboration, évaluent l'importance scientifique de ces propositions de numérisation avec l'aide d'un conseil scientifique dédié à ce programme et assurent la gestion et le suivi des candidatures retenues.

L'organisation du support technique du projet a été définie au départ en 2006 par la Direction de l'information scientifique (DIS) du CNRS. Celle-ci a divisé les tâches à réaliser en deux lots, et a confié chaque lot à des structures distinctes : l'INIST[7] a été chargée de la conception d'un portail d'accès et de valorisation des ressources constituées dans ce cadre et le CRDO-Paris[8] a été chargé de la gestion de ces ressources au sein de son entrepôt de ressources. Une architecture fonctionnelle

5. Typologie et universaux linguistiques.
6. Institut de linguistique Française.
7. Institut national de l'information scientifique et technique.
8. Centre de ressources pour la description de l'oral.

fondée sur le modèle des archives ouvertes[9] devait permettre la communication entre les deux modules (portail et entrepôt). Le travail du CRDO-Paris a été principalement un travail de définition d'une collection sur les langues de France puis de contrôle des ressources entrant dans la collection au fur et à mesure de leur collecte. Le reste des tâches utiles (gestion du catalogue, exposition des métadonnées, stockage, conservation, gestion d'accès, etc.) étant des missions directement inscrites dans la définition du CRDO.

2. *Les langues de France*

La base de données qui prendra finalement le nom de *Corpus de la parole* en écho aux « Archives de la parole » de Ferdinand Brunot[10], comporte des enregistrements de parole en français et en « langues de France[11] ». Il existe plus de 75 langues relevant de l'appellation « langues de France », les langues régionales (alsacien, breton...), les langues d'outre-mer (arawak, futunien...), les langues non territoriales (berbère, judéo-espagnol...) et la langue des signes française. Au fur et à mesure du temps la collection s'est enrichie par des langues entretenant un rapport étroit avec ces premières langues et qui faisaient l'objet de recherches au sein des laboratoires de linguistique (français parlé hors du territoire national, langues des signes émergentes...).

Tous les enregistrements de la collection sont issus de corpus récoltés par des linguistes dans le cadre de projets scientifiques.

La plus grande partie des enregistrements est audio, mais pour certaines langues (les langues des signes en particulier) ou pour certains usages (étude des interactions, étude du langage enfantin) la vidéo a également été utilisée par les chercheurs.

Les plus anciens enregistrements audio ont été faits avec des techniques analogiques sur des supports de type bandes magnétiques, puis, par la suite, sur des cassettes magnétiques. Les enregistrements les plus récents ont été faits avec des techniques numériques sur des supports de type disques optiques, disques magnéto-optiques, disques durs, mémoires flash. Parmi ces enregistrements, priorité a été donnée aux supports analogiques en fin de vie qui, en raison de leur ancienneté et de leur fragilité, étaient généralement les plus en danger. Pour autant, certains supports numériques ont des durées de vie très courtes, souvent inférieures aux supports analogiques et l'obsolescence technologique y est aussi plus présente et plus fréquente, ce qui a pu conduire parfois à récupérer en urgence le contenu de ces supports.

9. *Open Archives Initiative* (OAI).

10. Le centenaire a été marqué en 2011 à la BnF par une journée d'étude, le 17 juin 2011.

11. Les langues de France sont les langues régionales ou minoritaires parlées traditionnellement par des citoyens français sur le territoire de la République, et qui ne sont la langue officielle d'aucun État.

2.1. *Les projets existants similaires ou en relation*

Il existe bien sûr d'autres initiatives qui recoupent au moins en partie les préoccupations du projet *Corpus de la parole*. À une échelle locale on peut, bien sûr, citer des projets de laboratoires de recherche ou des projets trans-laboratoires (financés en particulier par l'ANR). Nous mettons dans ces catégories des projets sur le français tels que PFC (Phonologie du français contemporain), CFPP2000 (Corpus de français parlé parisien des années 2000), CRFP (Corpus de référence du français parlé), etc[12]. À une échelle plus large, il existe aussi des projets nationaux et internationaux portés par des institutions ou des communautés. Par exemple le projet CHILDES (*Child Language Data Exchange System*) est un projet créé dans le but de faciliter les échanges de corpus oraux au sein de la communauté qui étudie l'acquisition du langage par les enfants. Ce projet, pionnier à son époque, a défini des conventions pour la transcription, des outils pour la création et l'interrogation de corpus ainsi qu'une organisation du partage avec une banque de données des corpus partagés. Ce projet comporte aussi des données sur le français et les langues de France. D'autres projets internationaux plus récents ont vu le jour suite à une prise de conscience de l'aspect « en danger » de certaines langues ou de la difficulté de partage et de conservation du patrimoine scientifique et culturel que représentent ces données. Dans les projets sur les langues en danger nous pouvons citer par exemple le projet HRELP[13], le programme de l'UNESCO, le programme Sorosoro[14] de la fondation Chirac. Dans la mesure où il existe de nombreuses langues de France qui ne sont pas écrites ni enseignées, notamment à l'outre-mer en Nouvelle-Calédonie ou en Guyane, les interactions entre ces projets et le projet *Corpus de la parole* peuvent être étroites et les chercheurs qui travaillent sur ces langues participent parfois aussi à ces projets. Enfin, dans les projets institutionnels qui défendent des objectifs patrimoniaux, scientifiques et/ou culturels, nous pouvons citer les projets européens CLARIN (qui vise à la mise en place d'une infrastructure sur les ressources et technologies de la langue) ou FLaReNet (*Fostering Language Resources Network*). Le réseau public des archives en France collecte lui aussi des ressources linguistiques et cette communauté, en collaboration avec les autres institutions équivalentes dans les autres pays, participe également à des projets européens, en particulier pour favoriser l'accès à ces contenus (projets APEnet ou Europeana).

En résumé il existe de nombreux projets et organisations qui, pour des raisons diverses, financent et constituent des bases de données sur des ressources proches de celles collectées dans le cadre de *Corpus de la parole*. L'interaction avec ces projets et organisations se fait en général à deux niveaux :

12. Pour une liste plus complète des corpus existants on se reportera utilement à l'inventaire effectué en 2005 par Paul Cappeau et Magali Seijido « Les corpus oraux en français » (http://www.dglf.culture.gouv.fr/recherche/corpus_parole/Presentation_Inventaire.pdf).
13. « *Hans Rausing Endangered Languages Project* » est un projet de la SOAS (*School of Oriental and African Studies*) de l'université de Londre.
14. Sorosoro signifie « souffle, parole, langue » en araki (langue du Vanuatu).

- au niveau du chercheur qui contribue en apportant au programme des ressources nouvelles, mais qui, pour d'autres raisons, participe aussi à d'autres projets aux objectifs variés ;
- entre les organisations ce sont plutôt les méthodes et les bonnes pratiques, qui font l'objet d'un partage et d'une discussion. Ainsi, depuis le départ du projet *Corpus de la parole*, les choix d'organisation, de formats, de normes de qualités ont été discutés avec des institutions (BnF, Archives de France, TGE-Adonis), qui pour des raisons extérieures au projet lui-même, participent à des groupes de standardisation et de normalisation et sont également présentes dans d'autres projets du même type.

2.2 *Les projets d'archivage et les théories*

Les développements parfois parallèles, parfois sécants, de la sociolinguistique et de la linguistique de corpus à partir des années soixante ont renouvelé les questions méthodologiques et théoriques de constitution et d'exploitation des données orales en linguistique. Cet article n'est pas le lieu d'une présentation de l'ensemble des points théoriques abordés ces cinquante dernières années, mais il convient de noter qu'il reste un clivage important entre un courant linguistique scindant les données et l'analyse des données, et un courant intégrant l'analyse aux conditions mêmes de productions des données. Selon ce second courant, les questions rencontrées lors des étapes de constitution, de conservation et de diffusion relèvent de positions théoriques. Ainsi, les choix de catalogage et de codage, les niveaux d'annotations et la description des éléments constitutifs d'un corpus sont autant de partis pris de théories trop souvent non explicites.

L'explicitation de la démarche de constitution et de ressources devient un exercice nécessitant une réflexivité placée au cœur même du travail scientifique. Les paragraphes suivants décrivent les lieux possibles et nécessaires de cette réflexivité et de cette explicitation.

3. Les préconisations

Afin d'effectuer le travail de préparation des ressources par leurs détenteurs, un certain nombre de préconisations ont été définies, portant tant sur les moyens à mettre en œuvre, que sur les résultats attendus.

3.1. *Les préconisations pour la numérisation des enregistrements*

Pour la numérisation des anciens supports analogiques, le CRDO-Paris a défini des critères de qualité minimaux. Ces critères, inspirés de ceux préconisés par

IASA[15] (IASA 2009) ont été validés par le conseil scientifique du programme *Corpus de la parole* en accord avec le département des archives sonores de la BnF et communiqués, *via* les fédérations de linguistique, aux chercheurs et laboratoires qui pratiquaient eux-mêmes la numérisation. Il s'agissait d'une préconisation contractuelle qui a donné lieu à l'élaboration d'une annexe technique, systématiquement présente dans les conventions de la DGLFLF. Le CRDO-Paris, qui pilotait également une partie des numérisations pour le compte des chercheurs et laboratoires qui le souhaitaient, appliquait aussi obligatoirement ces préconisations lors des opérations de numérisation à l'aide des équipements d'un laboratoire qui s'en était doté pour ses besoins propres (le LACITO[16]). Pour les enregistrements audio ces préconisations étaient les suivantes : échantillonnage 44,1 kHz au minimum (96 kHz au LACITO) ; quantification : 16 bits au minimum (24 bits au LACITO) ; copie droite sans retouche ; format WAV ; encodage : PCM[17].

3.2. *Les préconisations pour l'écriture des transcriptions*

Pour les annotations pouvant accompagner les enregistrements[18], le CRDO-Paris a défini, toujours après validation du conseil scientifique, des recommandations allant jusqu'à un modèle cible en XML. Ce modèle, exprimé dans une DTD XML, définit une structure minimale permettant :

- de coder la transcription d'un enregistrement ;
- d'ajouter une traduction en français (ce qui été demandé pour les langues autres que le français ou pour des transcriptions non orthographiques du français) ;
- de découper la transcription en segments (phrase ou groupe de souffle) ;
- de noter les repères temporels de début et de fin des segments.

Quelques raffinements du modèle permettent également d'indiquer le locuteur (utile pour les dialogues), le type de transcription (orthographique, phonétique, phonologique).

15. *International Association of Sound and Audiovisual Archives*.
16. Laboratoire de langues et civilisations à tradition orale.
17. Pulse-code modulation. Il s'agit d'un codage sans compression.
18. Le programme de la DGLFLF prévoit également une phase de valorisation des enregistrements à l'aide d'une ou plusieurs couches d'annotations (transcription, glose, traduction, annotations morphosyntaxiques, syntaxiques ou autres).

```
<S id="ESTAQUEs1" who="INT">
    <FORM kindOf="ortho">Se akodra de antes de la gera, kómo, kómo era la vida en el Estaque ?
        Ké aziya la djente...</FORM>
    <TRANSL xml:lang="fr">Vous vous rappelez, avant la guerre, comment, comment était la vie à
        l'Estaque ? Que faisaient les gens...</TRANSL>
    <AUDIO start="0.0000" end="9.5063"/>
</S>
<S id="ESTAQUEs2" who="LOC">
    <FORM kindOf="ortho">En Estaque bivimo bivimos komo los... la djente ke viviyan en el
        Estaque,</FORM>
    <TRANSL xml:lang="fr">A l'Estaque nous avons vécu comme les... les gens qui vivaient à
        l'Estaque,</TRANSL>
    <AUDIO start="9.5063" end="17.9336"/>
</S>
```

Figure 1. *Transcription extraite du corpus* « Judeo-Spanish in France Archive »

Ce modèle minimal peut être atteint soit directement, soit en passant par des formats et outils qui permettent de faire une annotation plus riche et plus fine. En particulier, les formats en sortie des outils Transcriber[19] et ITE[20] peuvent être directement exploités dans le cadre du projet, la transformation vers le format cible étant alors complètement automatisée. D'autres formats, tels que ceux utilisés par les outils CLAN ou ELAN, doivent faire l'objet d'une normalisation afin d'être transformés de manière souvent *ad hoc* dans le format cible. Dans ce dernier cas, les deux formats sont conservés : le format d'origine et le format cible.

Ces recommandations ne portent que sur la forme à utiliser pour exprimer les transcriptions. Aucune indication ni directive n'est donnée pour expliquer aux chercheurs comment ils doivent transcrire leurs enregistrements et en donner une traduction en français. Les linguistes ont parfois établi des conventions accompagnées de manuels[21], mais d'une langue à l'autre (du français à la langue des signes française, par exemple), ou d'un domaine linguistique à un autre (de la phonétique à la dialectologie, par exemple), ces conventions sont peu partagées. Il est en revanche conseillé d'identifier, sous forme d'une référence dans les métadonnées, la ou les ressources qui décrivent de manière explicite l'ensemble des conventions utilisées.

3.3. *Les consignes pour la description des ressources*

La description des ressources (les enregistrements et les transcriptions) a également donné lieu à l'élaboration de préconisations. Cette description repose sur un jeu de métadonnées qui doivent suivre le schéma XML défini par OLAC[22]. Ce

19. Transcriber (http://trans.sourceforge.net/).
20. ITE Interlinear Text Editor (http://michel.jacobson.free.fr/ITE/).
21. Conventions pour ESLO (http://eslo.in2p3.fr/).
22. *Open Language Archives Comunity*.

schéma reprend ceux du Dublin-core et du Dublin-core qualifié[23] auxquels sont ajoutés cinq attributs associés à des vocabulaires contrôlés (*role*, *language*, *linguistic-field*, *linguistic-type* et *discourse-type*). Les recommandations précisent la manière d'utiliser ce schéma OLAC pour décrire les ressources (les éléments obligatoires, facultatifs, des explications, des exemples, etc.).

4. Organisation de la production

La première opération consiste pour les fédérations de linguistique à identifier dans leurs réseaux de contact, ou par l'intermédiaire d'un appel à projets, les corpus existants qui pourraient entrer dans le cadre du projet. Puis elles prennent contact avec les responsables de ces corpus afin d'évaluer l'intérêt scientifique et la charge de travail (en temps, ressources humaines, budget) que représente l'entrée du corpus (tout ou partie) dans la collection. C'est aussi le moment où une expertise des aspects juridiques est nécessaire. Il s'agit tout d'abord d'évaluer si les ressources identifiées sont susceptibles de poser des problèmes en termes de propriété intellectuelle et de respect de la vie privé. Cette expertise nécessite la prise en compte d'informations précises qui ne peuvent être formulées que dans le cadre de l'explicitation de la démarche suivie par les chercheurs tout au cours de leur projet de recherche. Ainsi, la description de la méthodologie de collecte et de traitements des données ne peut se faire sans une présentation précise des cadres théoriques mobilisés, par exemple dans les opérations de catégorisation des participants et de leurs productions langagières.

En fonction de l'évaluation des risques, des mesures peuvent être envisagées telles que la recherche des autorisations ou l'anonymisation. Si aucune solution ne permet la libre accessibilité à la ressource, celle-ci ne pourra tout simplement pas entrer dans le cadre strict du projet. Cette contrainte juridique est en effet dépendante des objectifs du « plan numérisation » du ministère de la Culture et de la Communication, qui conditionne le financement de la numérisation à la mise à disposition des données publiques. Ceci n'interdit nullement d'archiver les ressources en question dans le respect des conditions d'accessibilité définies par le Code du patrimoine[24], mais elles ne pourront alimenter le réservoir C*orpus de la parole* qu'à l'issue d'une période plus ou moins longue. Si le corpus en question représente bien un intérêt, un accord précise que les responsables du corpus cèdent de manière non exclusive les droits de représentation et de reproduction des ressources qui seront numérisées dans ce cadre à la DGLFLF afin que cette dernière puisse alimenter le portail avec ces ressources.

23. Dublin-core ou norme ISO 15836.
24. « Que sont les archives ? » dans la lettre d'information de l'InSHS n°13 de septembre 2011 (http://www.cnrs.fr/inshs/Lettres-information-INSHS/lettre_infoINSHS_13.pdf).

4.1. *Numérisation*

Le travail de numérisation est effectué, suivant les cas, soit directement par le chercheur s'il dispose des outils et des compétences nécessaires, soit par un tiers qui souvent est le laboratoire du chercheur. Les consignes précisées plus haut sont transmises aux chercheurs et le CRDO apporte son expertise et un accompagnement à toutes les étapes du projet.

4.2. *Collecte*

La collecte des ressources est une tâche assurée par les fédérations de linguistique. Elle consiste en un suivi de planning pour les différents contributeurs avec des échanges et des relances jusqu'à obtenir la livraison complète des enregistrements numérisés correctement documentés et éventuellement, suivant les cas, accompagnés de fichiers de transcriptions et traductions. À partir de 2009, les fédérations ont recruté sur contrat un ingénieur pour assurer l'accompagnement des projets auprès des chercheurs et des laboratoires. Une fois cette collecte faite, les fédérations de linguistique effectuent une livraison du corpus au CRDO-Paris. La livraison doit respecter les préconisations techniques y compris le respect des consignes de nommage et de formats des fichiers. Elle s'effectue comme un simple dépôt de fichiers sur le serveur dans une zone réservée à cet effet.

La tâche d'accompagnement été particulièrement soignée dans le but de respecter l'ensemble des contraintes relevant de choix méthodologiques et théoriques des chercheurs. Il ressort de cette expérience le constat d'une très grande hétérogénéité des pratiques qui demande à être respectée afin de préserver la chaîne des opérations qui lie la collecte à la diffusion en passant par l'analyse et l'exploitation. Les problèmes rencontrés ont été systématiquement étudiés au sein du conseil scientifique.

4.3. *Contrôle*

Le contrôle de la livraison est effectué à réception des ressources par le CRDO-Paris. Une première analyse qualité est effectuée pour s'assurer de la complétude des informations : à chaque fichier livré doit correspondre un jeu de métadonnées et chaque jeu de métadonnées doit décrire un fichier unique et distinct. La présence d'informations obligatoires dans les métadonnées est également vérifiée : nom du déposant, identification de la langue... Enfin le CRDO-Paris contrôle la bonne formation des fichiers (enregistrements, annotations et métadonnées) et enrichit les métadonnées avec des informations de nature technique. Suivant la nature du contrôle ce dernier est effectué par un documentaliste ou par un informaticien.

La vérification des fichiers d'enregistrement est faite à l'aide d'un outil qui extrait les informations de formatage (format de l'enveloppe, codage des contenus,

fréquence d'échantillonnage, taille des échantillons, nombre de canaux). Ce programme émet des alertes lorsque les critères définis ne sont pas respectés, si des plages silencieuses dépassent une durée seuil[25] ou encore si des informations inattendues sont présentes dans le document (par exemple : des métadonnées, des jalons temporels).

La vérification des fichiers de métadonnées et des fichiers d'annotations passe systématiquement par des technologies de validation de schémas XML.

Parallèlement à ces contrôles, essentiellement techniques, une autre vérification est effectuée par un opérateur humain de formation documentaliste, dont le rôle est de vérifier la cohérence des métadonnées, leur complétude et leur exactitude et d'entretenir un catalogue de ressources homogène et compréhensible. Au besoin les métadonnées seront donc complétées et normalisées, éventuellement après des échanges auprès du déposant de la ressource pour récupérer l'information et la valider.

L'enrichissement automatique des métadonnées concerne principalement les informations techniques à propos du fichier lui-même (type mime, type DCMI, durée pour les enregistrements, liaison au schéma pour les annotations) ainsi que quelques informations de gestion telles que la date de dernière modification des métadonnées, le lien à la collection, les URLs d'accès.

4.4. *Versement*

Une fois la ressource contrôlée, elle est mise en ligne sur le site du CRDO-Paris avec une restriction d'accès qui n'autorise que les administrateurs et le déposant à la consulter. Le déposant est alors encouragé à vérifier si sa ressource et la description qui en est faite sont correctes avant de donner son accord pour publication. Une fois cet accord obtenu, la restriction d'accès est supprimée et la ressource est accessible par tous et notamment sur le portail *Corpus de la parole*.

5. Description de l'architecture technique

Depuis le début du projet, le portail d'accès aux ressources et l'entrepôt de ressources sont séparés logiquement et physiquement. Les liens entre ceux-ci se font à l'aide du protocole OAI-PMH[26] défini par les « archives ouvertes ».

La conception du portail a été confiée au départ du projet à l'INIST. Sur la base de choix validés par la DIS et par l'Observatoire des pratiques linguistiques, il a

25. La durée paramétrable a été fixé arbitrairement à 10 secondes et permet d'alerter le technicien afin qu'il puisse vérifier si ces plages silencieuses sont intentionnelles (anonymisation) ou accidentelles (mauvaise transmission, défaut à la numérisation, etc.).
26. *Open Archives Initiative - Protocol for Metadata Harvesting*.

développé un site fondé sur un gestionnaire de contenu largement répandu (SPIP[27]). Différents rédacteurs peuvent ainsi saisir des contenus avec éventuellement une politique de validation. L'INIST a développé et mis en œuvre une charte graphique pour le site et a été conduit à développer quelques scripts (écrit en langage PHP et utilisables sous forme de « plugin » dans SPIP) afin de permettre d'afficher dans les pages du site des moteurs de recherche et des animations multimédia (pour la consultation des ressources mélangeant audio, vidéo et annotations). Par la suite la charte graphique du site a été revue par l'atelier de création « des signes graphiques » (figure 2) et l'ensemble des moteurs de recherche et outils de consultation ont été réécrits par le CRDO-Paris. Enfin, le portail lui-même après avoir été hébergé les premières années sur le serveur du CRDO-Paris l'est aujourd'hui sur les serveurs du centre de calcul de l'IN2P3 au sein de la grille de services du TGE-Adonis.

Pour faciliter le travail et accroître l'indépendance du portail, une base de données a été conçue qui stocke les métadonnées moissonnées à l'aide du protocole OAI-PMH de sorte qu'une fois ce moissonnage effectué les moteurs de recherche du portail n'aient plus besoin d'avoir recours aux services du CRDO-Paris pour retrouver de l'information. La périodicité du moissonnage a été fixée à sept jours (après avoir été quotidienne dans les premiers temps) et la technique utilisée est différentielle, c'est-à-dire qu'on ne moissonne que les changements intervenus depuis la dernière moisson, plutôt que complète (ce qui était fait dans un premier temps quand les volumes à moissonner étaient plus petits).

Figure 2. *Page d'accueil du site* Corpus de la parole

27. Système de publication pour l'Internet.

5.1. *Les moteurs de recherche*

Trois moteurs de recherche ont été développés pour le portail. Les deux premiers permettent une recherche dans les métadonnées des ressources, le dernier dans les transcriptions elles-mêmes.

Le premier moteur de recherche (cf. figure 3) permet des requêtes assez classiques dans les métadonnées en utilisant comme critères de recherche les quinze catégories définies dans la norme Dublin-core (la langue en tant qu'objet d'étude de la ressource pouvant être surajoutée aux autres critères de recherche). Le résultat retourné par ce moteur de recherche est composé d'une liste de ressources faisant apparaître comme seuls critères : le titre, le nom du chercheur et la langue en tant qu'objet d'étude (cf. figure 3). Cette liste permettant elle-même de donner accès *via* des liens à l'ensemble des métadonnées ainsi qu'aux ressources elles-mêmes.

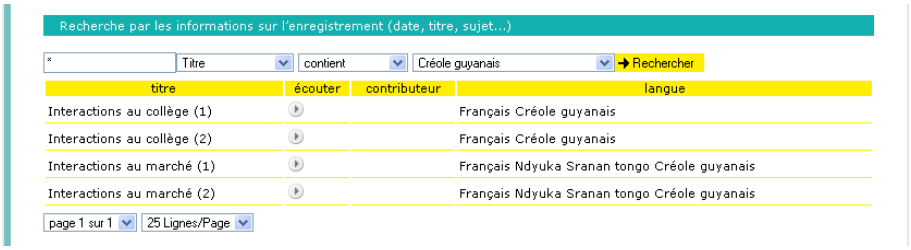


Figure 3. *Moteur de recherche dans les métadonnées*

Le deuxième moteur, mono-critère (la langue en tant qu'objet d'étude), retourne soit une liste de ressources comme le précédent moteur (cf. figure 3), soit une carte géographique sur laquelle les ressources qui sont géoréférencées sont indiquées (cf. figure 4). L'API Google Maps est utilisée pour l'affichage et la navigation dans la carte. De la même manière que pour le précédent moteur la liste tout comme la carte permettent de donner accès *via* des liens à l'ensemble des métadonnées ainsi qu'aux ressources elles-mêmes.

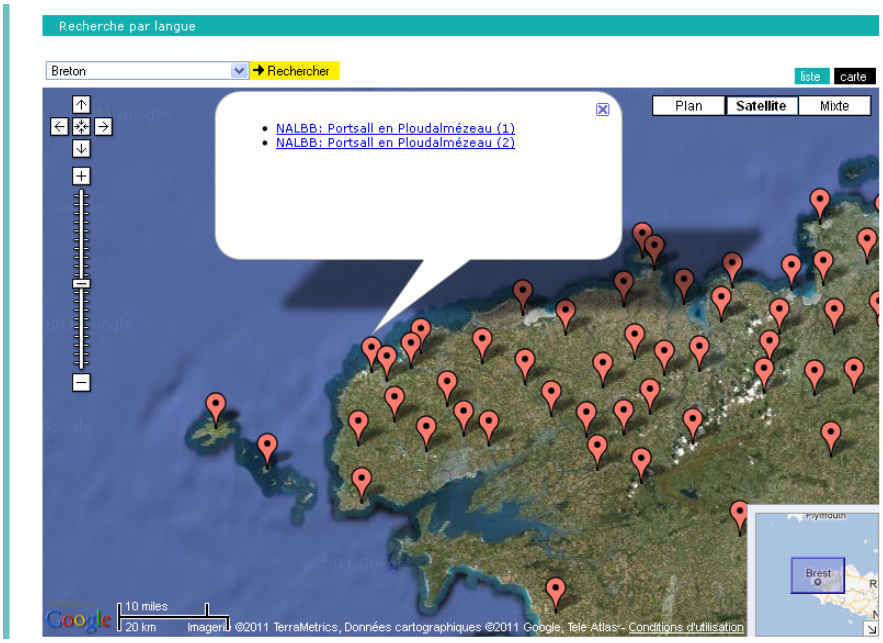


Figure 4. *Moteur de recherche géographique*

Le dernier moteur de recherche exploite cette fois l'ensemble des annotations et non plus les métadonnées, afin de pouvoir chercher un mot ou un motif soit dans la traduction soit dans la transcription. Le résultat retourné présente l'ensemble des segments qui contiennent ce mot ou ce motif (cf. figure 5).

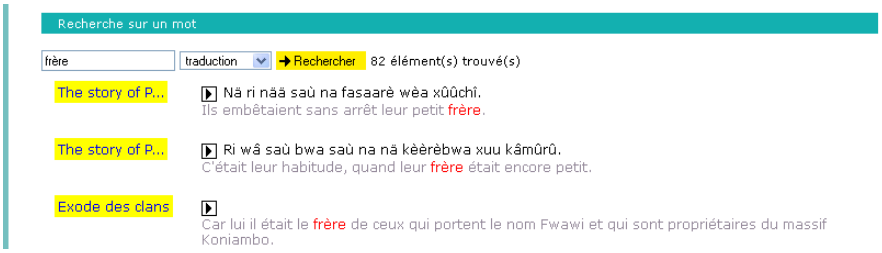


Figure 5. *Moteur de recherche dans les annotations*

5.2. *La consultation d'une ressource*

Lorsqu'une ressource est choisie par l'utilisateur, celui-ci peut la consulter par une interface multimédia (cf. figure 6). Dans tous les cas la page de consultation inclut un lecteur audiovisuel qui donne accès à l'enregistrement audio ou vidéo, ainsi qu'une fiche documentaire présentant une partie des métadonnées. Dans les cas où la ressource possède une transcription/traduction, celle-ci est également présentée dans la page et sa lecture est synchronisée avec le lecteur de sorte que cliquer sur le texte permet de démarrer la lecture de l'enregistrement à ce moment, et déplacer le curseur du lecteur positionne le texte à l'emplacement correspondant au moment choisi.

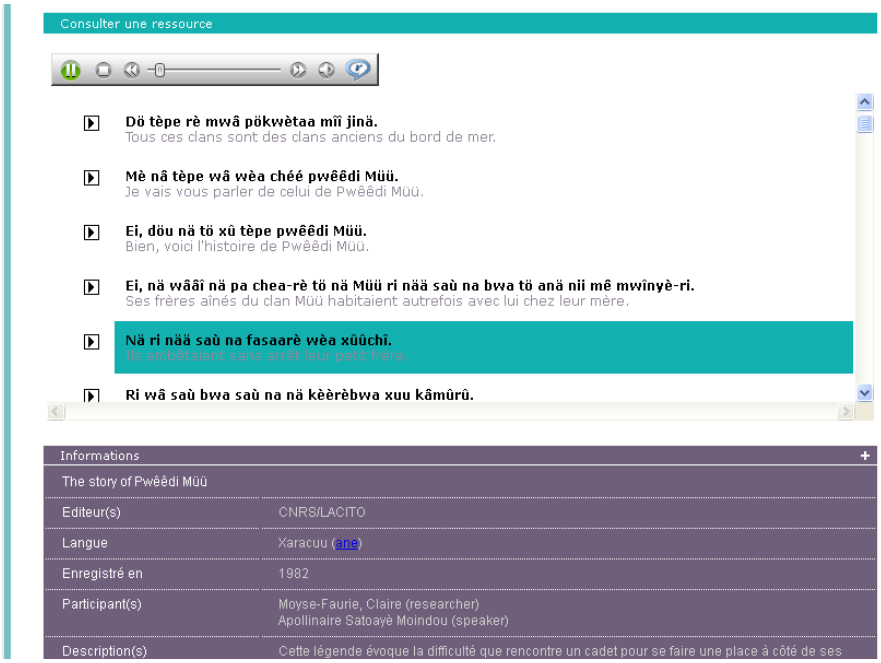


Figure 6. *Interface de consultation d'un document*

Les technologies qui ont été choisies pour réaliser ces fonctionnalités sont fondées sur le plugin Real, mais le nouveau HTML 5, qui offre la possibilité de coder directement en HTML des balises audio et vidéo, devrait permettre de s'affranchir de la technologie propriétaire de Real.

6. Prise en charge de l'archivage à long terme

Depuis mi-2008, le TGE-Adonis s'est engagé dans un programme d'archivage pérenne de données et documents numériques issus de la communauté des sciences humaines et sociales (SHS). L'organisation que le TGE-Adonis met en place dans le cadre de ce programme s'inspire du modèle de la norme OAIS[28]. Le modèle fonctionnel de cette norme distingue différentes briques illustrées en figure 7 et centrées autour :

- de l'entrée des archives : cette brique aborde le traitement des paquets d'informations versés par les producteurs d'archives. Elle comporte des mécanismes de préparation, de transmission, de contrôle, de rejet, de conversion de format, etc. Une fois le paquet d'informations validé et complété, cette brique le met à disposition des briques stockage et gestion de données ;
- de l'accès aux archives : cette brique traite des mécanismes d'accès, de consultation et de livraison des informations disponibles dans le système d'archivage (métadonnées et contenus). Elle comprend la mise à disposition d'un système de recherche dans les métadonnées, de sélection dans les résultats de la recherche, d'une interface de consultation et éventuellement un mécanisme de suivi des commandes jusqu'à leur livraison ;
- du stockage : cette brique traite de la conservation des informations à partir du moment où elles sont mises à sa disposition par la brique entrée et jusqu'à leur éventuelle destruction. C'est cette brique qui traite du choix des supports, de la gestion du contrôle de l'intégrité des données et de la gestion des migrations (rafraîchissement de supports, duplication et ré-empaquetage) ;
- de la gestion de données : cette brique assure la conservation, la mise à disposition et la mise à jour des informations descriptives (métadonnées) associées aux contenus d'informations conservés par la brique stockage ;
- de la planification de la pérennisation : cette brique assure une veille technologique et propose des recommandations, des évolutions et des stratégies pour prévenir l'obsolescence et garantir l'accès, sur le long terme, aux informations ;
- de l'administration : cette brique permet d'assurer l'exploitation de l'ensemble du système d'archivage électronique et traite en particulier de la gestion des utilisateurs au sens de leurs droits d'accès.

Le TGE-Adonis, sur la base d'une étude sur l'hébergement de services informatiques et de données numériques pour les SHS en France[29], a choisi d'adosser ce service à deux grands centres informatiques existants. Les centres choisis sont le

28. Modèle de référence pour un système ouvert d'archivage d'information (OAIS). Norme ISO 14721:2003

Centre informatique national de l'enseignement supérieur (CINES) et le centre de calcul de l'Institut national de physique nucléaire et de physique des particules (CC-IN2P3). Le CINES, qui avait déjà une expérience dans le domaine de la conservation avec les thèses et les données de numérisation des revues du portail Persée, s'est vu confier la brique d'entrée alors que le CC-IN2P3, qui avait déjà une forte expérience en hébergement et services, se voyait confier la brique d'accès correspondante. Les autres briques étant partagées entre eux avec une légère prédominance du CINES.

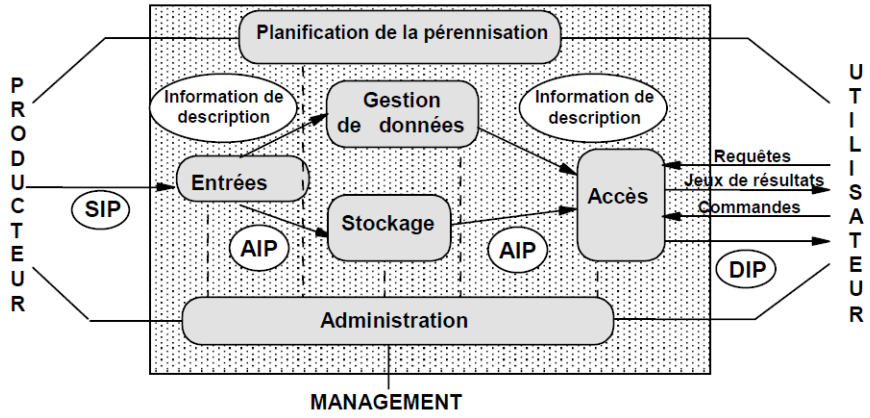


Figure 7. *Les fonctions du modèle OAIS*

Pour la mise en place de cette architecture, le TGE-Adonis a pris contact avec les Archives de France et a souhaité démarrer une première expérience sur les données orales gérées par le CRDO. Pour mener à bien cette expérience un groupe de travail a été constitué en rassemblant des membres des deux centres informatiques (CINES et CC-IN2P3), des membres des deux antennes du CRDO (Paris et Aix), un représentant des Archives de France, un chef de projet du TGE-Adonis et un consultant extérieur[30]. Après un peu plus d'un an de mise au point, l'architecture est entrée dans une phase de production réelle le 22 juin 2010 et offre un certain nombre de services au CRDO et donc au projet *Corpus de la parole*.

29. Étude Olof Bärring, « *Hosting of IT services and data for Human and Social sciences in France* », version 1.0, janvier 2008. Rapport final de l'étude commandée par le TGE Adonis au CERN. Document non public, communicable par le TGE Adonis sur demande.
30. Huc C, Habert B, Building together digital archives for research in social sciences and humanities, Social Science Information, vol. 49, n°3, septembre 2010, p. 415-443.

6.1. *Contrôle des entrées*

Dans l'organisation mise en place, les services servant (CRDO-Paris et CRDO-Aix) versent leurs archives au CINES *via* un protocole[31] qui prévoit dans leur ordre d'émission, les messages : bordereau de versement, accusé de réception, certificat d'archivage ou avis d'anomalie.

Entre le CRDO et le CINES, trois scénarios ont été définis : a) le transfert initial d'un objet, b) la modification de la description d'un objet et c) le transfert d'une nouvelle version d'un objet existant. Un dernier scénario dit de « restitution » a également été défini. Ce dernier ne doit intervenir qu'en cas de défaillance du CINES ou de fin de contrat entre le CNRS et le CINES afin de confier les archives dont le CINES a la responsabilité de la conservation à une autre organisation.

L'objet que l'on transfère une première fois et dont on peut par la suite modifier la description et/ou ajouter des versions, correspond à une unité d'archive dans laquelle peut se trouver un ensemble plus ou moins grand de fichiers. Chaque service versant doit donc définir de quoi est composé cet objet et le décrire dans un document[32] pour le CINES.

Les ressources du *Corpus de la parole* ne peuvent être que de l'un des trois types suivants :

- un enregistrement (audio ou vidéo) ;
- une annotation d'enregistrement (transcription, traduction, etc.). Cette annotation doit être un document structuré en XML et suivre le schéma défini dans les consignes du projet ou suivre l'une des deux DTD : celle du logiciel Transcriber ou du logiciel ITE ;
- une collection (rassemblement de plusieurs enregistrements et annotations).

L'identification du format de chaque fichier est une des informations techniques obligatoires à renseigner dans le bordereau de versement. En effet, le CINES ne peut engager sa responsabilité que sur des fichiers dont il connaît la structure et dont il peut s'assurer qu'ils respectent bien cette structure. Pour pouvoir prendre en charge les documents du CRDO, le CINES a été conduit à évaluer les formats et codages audiovisuels utilisés au CRDO et à faire évoluer sa plate-forme afin de pouvoir les ajouter à la liste des formats pris en charge. Cette étude[33] a identifié un certain nombre de formats de représentations acceptables tant pour l'audio que pour la vidéo. Par exemple pour l'audio, les formats acceptables par le CINES et utilisés dans *Corpus de la parole* sont le format WAV avec un encodage PCM ainsi que le

31. Le protocole s'inspire du « Standard d'échange de données pour l'archivage ».
32. *Project Preservation Description Information* (PPDI) terme de la terminologie OAIS.
33. « Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles ».

format FLAC[34]. Pour les vidéos, les formats conseillés par le CINES et que le programme *Corpus de la parole* a finalement retenus sont le format conteneur MKV avec le codec vidéo H264 et l'encodage FLAC pour l'audio ainsi que le format conteneur MPEG-4 avec le codec vidéo H264 et l'encodage AAC pour l'audio.

Tous les fichiers transmis au CINES, bien qu'ils aient déjà fait l'objet d'une validation par le CRDO-Paris avec ses propres outils, font l'objet à leur réception d'une deuxième validation par le CINES et des messages d'anomalies sont retournés s'ils ne suivent pas toutes les spécifications du format dont ils se réclament ou si les formats utilisés ne font pas partie de la liste des formats acceptables par le CINES.

Lors de la réception des fichiers le CINES effectue aussi un contrôle d'intégrité en comparant les empreintes des fichiers calculées par le service versant lors du transfert et renseignées dans le bordereau de versement avec les empreintes qu'il calcule lui-même sur les fichiers reçus. Ce contrôle est important afin de s'assurer que le fichier reçu correspond bien au fichier envoyé et qu'il n'y a pas eu par exemple d'erreur de communication lors du transfert.

À réception d'un paquet le CINES exécute deux opérations :

- il attribue au paquet un jeton d'horodatage calculé sur la base de source de temps extérieure et fiable. Ce jeton d'horodatage peut servir de garantie opposable d'antériorité en permettant de démontrer qu'une ressource existait bien à partir d'une date et d'une heure précises et certifiées ;
- il associe au paquet un identifiant pérenne de type ARK[35]. Cet identifiant est unique non seulement au sein de la plate-forme d'archivage du CINES mais également hors de ce contexte. Il permet ainsi de garantir une pérennité même à travers les différents organismes qui prendront par la suite en charge cette archive. C'est aussi un identifiant qui permet la « citabilité » de la ressource.

Les URL ARK sont composées de deux parties : la première partie appelée NMA[36] est une URL d'accès qui va permettre la localisation de la ressource. Cette partie est facultative, peut être modifiée (par exemple en cas de changement d'organisme en charge de cette gestion) ou multiple (gestion confiée à plusieurs organismes en même temps). La deuxième partie est l'identifiant à proprement parler de la ressource. Cet identifiant est obligatoire et ne peut ni changer ni être recyclé en cas de suppression de la ressource. Il se décompose en un identifiant de l'organisme qui affecte les identifiants aux ressources NAAN[37] suivi d'un identifiant de la

34. Ce format a été utilisé pour les enregistrements qui dépassent le volume accepté par le format WAV.
35. *Archival Resource Key*.
36. *Name Mapping Authority* ou autorité d'adressage.
37. *Name Assigning Authority Number* ou autorité d'assignement de noms. Par exemple, le NAAN du CINES est 87895, celui de la BNF est 12148. Une liste complète est maintenue à l'url http://www.cdlib.org/uc3/naan_registry.txt.

ressource, suivi éventuellement d'un « *qualifier* » permettant de gérer la granularité de la ressource, ses versions, ses formats, etc. Par exemple l'identifiant ARK attribué par le CINES pour un enregistrement en alsacien effectué à Aschbach dans le Bas-Rhin en 1980 dans le cadre des atlas linguistiques est ark:/87895/1.5-124712.

Dans l'état actuel du projet, l'autorité d'assignation n'étant pas encore définie, la citation d'une ressource se limite donc à sa seule identification (comme on le ferait avec le numéro ISBN d'un ouvrage) et non pas à sa localisation. Afin de permettre cette localisation le CRDO-Paris maintient des identifiants OAI. Ainsi la ressource du dernier exemple est localisable en interrogeant l'entrepôt du CRDO-Paris avec l'identifiant oai:crdo.vjf.cnrs.fr:crdo-HUD_0003_SOUND[38] par le protocole OAI-PMH.

6.2. *Transfert de responsabilité pour la conservation des informations*

Une fois tous les contrôles effectués par le CINES, si ce dernier juge acceptable le paquet d'informations versé, il retourne au service versant un certificat d'archivage afin de lui notifier le transfert de responsabilité et de lui communiquer l'identifiant permettant de récupérer au besoin l'archive. Le service versant (CRDO-Paris) peut alors se débarrasser des fichiers sur son système d'information puisqu'il n'a plus la responsabilité de leur conservation.

La responsabilité de la conservation par le CINES couvre toutes les tâches concourant à s'assurer que l'information transmise reste intègre et lisible à travers le temps, de s'assurer de son authenticité, de son accessibilité et éventuellement de sa confidentialité. Cela entraîne de nombreuses précautions à prendre en matière de stockage : l'information doit être copiée en plusieurs exemplaires, sur plusieurs sites distants, éventuellement sur des supports de différentes natures. Ces supports doivent être surveillés régulièrement et des migrations de supports devront être planifiées afin de lutter contre leur vieillissement et leur obsolescence. Des migrations de formats devront également être planifiées en cas d'obsolescence des formats d'origine.

De nombreuses autres tâches incombent au CINES du simple fait de prendre en charge la responsabilité de la conservation de ces informations. Afin de s'assurer que le CINES possède les compétences nécessaires en la matière et offre des garanties quant à leur mise en œuvre, il lui a été demandé, avant de pouvoir assurer cette mission, d'obtenir un agrément du ministère de la Culture et de la Communication. Cet agrément obtenu en décembre 2010 porte sur la conservation d'archives publiques numériques courantes et intermédiaires. Cet agrément permet en

38. La requête pour récupérer les métadonnées OLAC de la ressource en OAI-PHM est http://crdo.vjf.cnrs.fr/crdo_servlet/oai-pmh?verb=GetRecord&identifier=oai:crdo.vjf.cnrs.fr:crdo-HUD_0003_SOUND&metadataPrefix=olac

particulier de s'assurer que le service respecte bien les normes en usage dans la profession : à savoir la norme OAIS ainsi que la norme NF-42-013.

6.3. *Accès*

Une fois le versement accepté par le CINES, une copie est envoyée au CC-IN2P3 afin d'offrir un service d'accès plus large que celui que propose le CINES qui ne s'adresse qu'au seul service versant. Cela permet, par la même occasion, d'avoir une autre copie distante des archives (le CINES est situé à Montpellier alors que le CC-IN2P3 est situé à Villeurbanne près de Lyon).

Quand le CC-IN2P3 reçoit un paquet d'informations en provenance du CINES, celui-ci analyse son contenu et alimente son outil de gestion (Fedora-commons) en créant un nouvel objet. À cet objet sont associées les métadonnées fournies dans le paquet. Les liens de filiation (liens entre la transcription et l'enregistrement ou entre un enregistrement ou une transcription et une collection) sont reconstruits en rattachant les nouveaux objets aux objets parents précédemment archivés. Il en est de même pour les liens de version. Les relations inverses sont également ajoutées automatiquement. Ces relations sont exprimées dans l'outil Fedora-commons sous forme RDF avec des verbes du type isSonOf, isFatherOf, hasNextVersion, hasPreviousVersion, etc.

Chaque fichier (enregistrement ou transcription) est placé dans un système de fichiers et devient accessible par le Web *via* une URL. Pour certains types de fichiers, des formats de diffusion ont été définis de sorte que l'accès à une même information puisse se faire dans plusieurs formats correspondant à des usages différents. Pour le format d'archivage WAV/PCM nous avons ainsi défini au CRDO-Paris, trois formats de diffusion : 1) le format d'origine ; 2) une version dégradée en WAV/PCM à 22 kHz/16bits/mono ; 3) et une version dégradée au format MP3. Enfin une dernière version au format RealMedia devrait être prochainement disponible pour une diffusion par un serveur de streams.

Des mécanismes d'authentification associés à des mécanismes de restriction d'accès sont prévus afin de pouvoir assurer pendant le temps nécessaire la confidentialité des ressources qui le demandent. Ces mécanismes ne sont pas encore en production au moment de l'écriture de l'article, mais les ressources présentes sur le portail *Corpus de la parole* sont jusqu'à présent sans restriction d'accès. Au-delà même de cette liberté d'accès, les ressources du portail sont également toutes assorties d'une mention de licence Creative-commons[39] afin de préciser leurs conditions de réutilisation.

Une fois que l'ingestion des informations du paquet reçu dans Fedora-commons est terminée, l'objet devient accessible et le CRDO peut récupérer les informations

39. CreativeCommons est une organisation, qui un peu sur le modèle du mouvement des logiciels libre a défini des licences permettant la cession de certains droits aux utilisateurs pour des œuvres.

d'enrichissement par l'emploi des fonctions OAI-PHM de l'entrepôt OAI de Fedora-commons. En particulier, les informations de date d'archivage, d'identifiant ARK ainsi que les différentes URL d'accès aux fichiers sont récupérées pour enrichir et modifier les métadonnées présentes au CRDO-Paris et donc également sur le portail *Corpus de la parole*.

7. État du chantier

En 2012, le réservoir *Corpus de la parole* contient plus de mille heures d'enregistrements audio et plusieurs centaines de transcriptions ou traductions. Parmi les 78 langues de France, 42 sont présentes. Tous les documents présents sont accompagnés de leurs métadonnées et ils sont tous accessibles par le site portail selon les conditions présentées dans les paragraphes précédents.

Cette évaluation, purement quantitative, ne signifie rien sans la prise en compte des phases du projet consacrées à l'interopérabilité de données nativement hétérogènes. L'architecture mise en place a permis de tester les propositions de catalogage, de codage et d'archivage de corpus collectés et exploités par des communautés scientifiques. La diversité des laboratoires, des projets et des données assure la représentativité des pratiques actuelles des chercheurs.

8. Perspectives et conclusions

Cette expérience en cours devrait prochainement bénéficier d'une convention entre la BnF, la DGLFLF et le CNRS. La volonté d'une coopération entre la BnF et le CNRS n'est pas récente. Déjà en 1979, le CNRS et le département de la Phonothèque et de l'Audiovisuel de la Bibliothèque nationale avaient signé une convention. Celle-ci stipulait que les chercheurs œuvrant dans le cadre des Atlas linguistiques versaient des enregistrements et leur fiche descriptive à la BnF. La BnF en effectuait une copie destinée à la conservation et restituait aux chercheurs des bandes magnétiques vierges afin qu'ils puissent continuer leurs enquêtes. La convention précisait que les bandes magnétiques seraient consultables à la Bibliothèque nationale et dans une institution régionale afin qu'un public de chercheurs puisse avoir accès à un fonds patrimonial sonore normalisé et répertorié[40].

Les avancées technologiques et la modification des pratiques des chercheurs offrent aujourd'hui l'opportunité de construire une nouvelle convention qui s'appuierait sur l'expérience acquise dans le cadre du programme *Corpus de la parole*. L'architecture proposée permet notamment de conserver et d'exploiter des corpus polymorphes dont l'état est intrinsèquement non stabilisé, tout en facilitant le versement d'un état stabilisé du corpus vers la BnF. Charge ensuite à la BnF de

40. Cordereix 2005 p. 253-264.

conserver et de donner accès à ce fonds dans le cadre des missions de son service des archives sonores.

Le travail conjoint des chercheurs et des institutions de conservation offre l'opportunité de donner toute sa dimension aux travaux scientifiques fondés sur les données. Il est nécessaire que cette collaboration soit engagée sur l'ensemble d'une chaîne de traitement qui va de la collecte à la conservation en passant par la mise à disposition de ressources libres. Ainsi il convient de ne pas séparer le travail scientifique d'analyse de la gestion des données tant il est vrai que les questions théoriques se posent et s'interpellent à tous les niveaux. C'est le défi que doivent relever tous les acteurs engagés dans de telles recherches.

9. Bibliographie

Baude Olivier et al. Corpus oraux, guide des bonnes pratiques, CNRS et PUO, Paris, 2006.

Paul Cappeau et Magali Seijido, Les corpus oraux en français, DGLFLF, 2005.

CINES, Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles, deuxième édition, 2001.

Cordereix, Pascal, « Les fonds sonores du département de l'audiovisuel de la bibliothèque nationale de France », Le temps des médias n°5. Éditions du nouveau monde, p 253-264, 2005.

DAF/DGME, Standard d'échange de données pour l'archivage, 2006.

Huc Claude, Habert Benoit, Building together digital archives for research in social sciences and humanities, Social Science Information, vol. 49, n°3, p. 415-443, 2010.

IASA Technical Committee, Guidelines on the Production and Preservation of Digital Audio Objects, Kevin Bradley éditeur. Second edition, 2009.

ISO, Modèle de référence pour un Système ouvert d'archivage d'information (OAIS) - Norme ISO 14721:2003

Jacobson Michel, Corpus oraux en linguistique de terrain, Traitement automatique des langues. 45/2, p. 63-88, 2004.

Jacobson Michel, Que sont les archives ?, lettre d'information de l'InSHS, n°13, 2011.

Olof Bärring, Hosting of IT services and data for Human and Social sciences in France, version 1.0. Rapport final de l'étude commandée par le TGE Adonis au CERN. Document non public, 2008.

Titre	***Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France***
Type	Article
Editeur	
Année	2011
Référence	Jacobson M, **Baude O.**, (2011) « Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France », in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 47-69.

Un grand corpus oral « disponible » : le corpus d'Orléans[1] 1968-2012

Iris Eshkol-Taravella* — Olivier Baude* — Denis Maurel — Linda Hriba* — Céline Dugua*— Isabelle Tellier*****

** Université d'Orléans - Laboratoire Ligérien de Linguistique - UMR 7270*
{Iris.Eshkol, Olivier.Baude, Linda.Hriba, Celine.Dugua}@univ-orleans.fr
***Université François Rabelais Tours – Laboratoire d'Informatique*
{Denis.Maurel@univ-tours.fr}
****Université Paris 3 - Sorbonne Nouvelle, Lattice[2]*
{isabelle.tellier@univ-paris3.fr}

RÉSUMÉ. *Cet article présente la constitution et la mise à disposition du corpus oral ESLO. Notre objectif est de montrer qu'il ne s'agit pas seulement de recueillir et rendre disponible des données langagières mais aussi de rendre explicite l'ensemble de la chaîne de traitement qui permet d'élaborer un tel corpus. Après avoir présenté le projet et le corpus nous préciserons les problèmes juridiques et méthodologiques qui ont conditionné les opérations de traitement du corpus et notamment les procédures d'anonymisation indispensables à la libre diffusion de cette ressource. Dans une seconde partie, nous présenterons les différentes annotations effectuées sur les données brutes avec quelques exemples de leurs exploitations. Nous expliquerons la méthodologie suivie qui est toujours guidée par la nature des données et l'objectif final visé : constituer un grand corpus oral variationniste du français. Nous aborderons enfin les questions de mise à disposition du corpus en ligne.*

ABSTRACT. *This article presents the building and putting online the oral corpus ESLO. Our purpose is to show that it is important not only to collect and make available language data and metadata but also to make explicit the whole chain of treatments. In the first part, we will present the project and the corpus, then we will specify the legal and methodological problems which determined all corpus treatments, in particular the anonymisation procedures which are required to freely make available this kind of resource. In the second part, we present different annotations made on the raw data with some examples of their use. We will explain the followed methodology which is always guided by the nature of the data and by the final objective: build a large sociolinguistic variationist oral corpus of French. Finally, we will discuss the issues of putting the corpus online.*

MOTS-CLÉS : *corpus oral, corpus variationniste, mise à disposition, anonymisation, transcription, annotation, variations.*

KEYWORDS: *oral corpus, variationniste corpus, anonymisation, transcription, corpus annotation, variations.*

1. http://eslo.in2p3.fr/
2. *Ce travail a été réalisé à l'université d'Orléans (au LIFO).*

TAL. Volume 53 – n° 2/2011, pages 17 à 46

1. Introduction

L'apparition d'Internet et le développement des outils informatiques ont permis la mise à disposition et la consultation des différents corpus. Des corpus nationaux représentatifs de leur langue comme le BNC[3], le *Russian National Corpus*[4] ou encore le *National Corpus of Polish*[5] apparaissent sur la Toile. Un corpus national représente la langue dans son développement en essayant de tenir compte de toutes les variétés de genres, de styles, d'utilisation, de variantes territoriales, sociales, etc. Ainsi, tous ces corpus contiennent les textes écrits mais aussi des transcriptions d'échanges oraux. Il s'agit souvent de la langue parlée par des locuteurs variés et dans les situations différentes : des conversations spontanées, des entretiens ou des émissions radio. D'autres corpus consacrés exclusivement à la langue parlée sont mis en ligne. Le *Santa Barbara Corpus of Spoken American English*[6] contient des enregistrements d'interactions orales des locuteurs de différentes origines régionales, ethniques, sociales. La majorité du corpus contient des entretiens en face-à-face, mais le corpus comprend aussi des conversations téléphoniques, des parties de jeux de cartes, des conférences, des narrations, des assemblées publiques, etc. Un autre exemple est le projet CORPAFROAS[7]. Il s'agit du premier corpus oral de langues afro-asiatiques (chamito-sémitiques) dont l'objectif consiste en une « *typological comparability among languages: prosodic analysis, and morphosyntactic glossing* » (Mettouchi, A. et Chanard, 2010, p. 258). Nous finirons par mentionner le livre-DVD (Cresti *et al.,* 2005) qui regroupe des corpus comparables en langues romanes (français, italien, portugais et espagnol) de discours spontanés avec des exemples de quelques études comparatives et avec l'accès simultané au son et à la transcription.

La situation semble être différente en France. Force est de constater que l'oral a été longtemps marginalisé dans le champ de la linguistique française (Blanche-Benveniste et Jeanjean, 1983) comme dans celui de la linguistique de corpus. Faisant l'inventaire des corpus oraux en français, Cappeau et Gadet (2007) notent qu' « il n'y a pas eu en France de volonté institutionnelle qui aurait conduit à la constitution d'un grand corpus oral. C'est, en contraste, ce qui a été fait pour l'écrit ». Cependant les travaux sur « le français parlé » puis l'apport des nouvelles technologies ont permis un engouement récent pour ce domaine. Parmi les initiatives actuelles, nous pouvons citer la base CLAPI[8] constituée pour étudier les interactions orales, le corpus PFC[9] pour analyser certains phénomènes

3. British National Corpus, http://www.natcorp.ox.ac.uk/
4. http://www.ruscorpora.ru/en/index.html
5. http://nkjp.pl/index.php?page=0&lang=1
6. http://www.linguistics.ucsb.edu/research/sbcorpus_obtaining.html
7. http://corpafroas.tge-adonis.fr/index_fr.html
8. Corpus de langues parlées en interaction, http://clapi.univ-lyon2.fr/
9. Phonologie du français contemporain, http://www.projet-pfc.net/?accueil:intro

phonologiques, le corpus CRFP[10] pour la morphosyntaxe ou le cas du corpus de français spontané EPAC[11] composé des interviews et des débats d'émissions de télé.

Des initiatives institutionnelles (Centre de ressources numériques du CNRS, ANR Corpus, Programme corpus de la parole de la DGLFLF en partenariat avec les fédérations de recherche en linguistique du CNRS, la création du TGE-ADONIS et du TGIR CORPUS) n'ont pas encore permis une mise à disposition d'envergure des corpus oraux. En 2011, la création du consortium Corpus oraux et multimodaux au sein de l'IRCORPUS qui doit répondre à l'objectif de « fédérer les équipes, laboratoires, chercheurs et enseignants-chercheurs engagés dans la constitution de corpus oraux et multimodaux, afin de faire converger les pratiques et de les rendre conformes aux standards internationaux[12] » confirme l'intérêt important pour ce champ d'études.

À la différence de l'écrit qui n'utilise qu'un seul support, l'oral associe le plus souvent la parole enregistrée à une représentation écrite et/ou codée (transcriptions, traductions, annotations). Cette donnée « secondaire » permet son exploitation par les outils de la linguistique de corpus. Toutefois les spécificités de l'oral nécessitent une adaptation des outils. La superposition des voix, ou « chevauchement de parole », ainsi que les phénomènes de disfluences – hésitations, répétitions, reprises, fausses amorces – n'existent pas à l'écrit et rendent le traitement de l'oral compliqué.

Cet article présente un exemple de constitution et de mise à disposition d'un grand corpus oral de français. Il s'agit de porter un regard sur le travail effectué au sein du programme de recherche ESLO (*Enquête Sociolinguistique à Orléans*) et par-delà, sur les méthodes actuelles d'exploitation des corpus et des données sociolinguistiques[13].

ESLO est un corpus de référence du français parlé hier et aujourd'hui à Orléans. Sa caractéristique première est de permettre des analyses sur la variation dans le français. La seconde caractéristique réside dans la structure même de ce corpus, composé de deux enquêtes : ESLO1 en 1968-1971 (un corpus clos/stable) et ESLO2 en 2008-2012 (un corpus ouvert/évolutif).

Cette expérience est donc l'occasion d'aborder quelques problèmes liés aux opérations de conservation et de diffusion d'un tel corpus. Comment tenir compte de l'hétérogénéité des données ? Quelles sont les données à constituer et à traiter ? Comment les coder ? Quels outils utiliser ? Toutes ces questions sont indissociables

10. Corpus de référence du français parlé, http://www.up.univ-mrs.fr/delic/crfp
11. Exploration de masse de documents audio pour l'extraction et le traitement de la parole conversationnelle, http://projet-epac.univ-lemans.fr/doku.php?id=accueil.
12. http://www.corpus-ir.fr/index.php?page=ircom
13. Ce travail a été réalisé grâce au soutien de l'ANR (projet Variling) et du Feder Région Centre (projet Entités).

et doivent être posées : de la collecte jusqu'à la diffusion. Mettre à disposition ce corpus implique de maîtriser l'ensemble de la chaîne de traitement tout en veillant à une cohérence au sein des contraintes imposées par la nature des données et par la nécessité de conserver les moyens d'une comparaison entre les deux corpus.

Nous développerons plus précisément les travaux d'annotation et de structuration du corpus rendus nécessaires par l'anticipation des problèmes juridiques de diffusion d'un corpus oral.

2. Corpus

2.1. *Historique*

ESLO1, la première enquête sociolinguistique à Orléans, a été réalisée en 1968-1971 par des professeurs de français de l'University of Essex, Language Centre, Colchester (Royaume-Uni), en collaboration avec des membres du B.E.L.C. (Bureau pour l'étude de l'enseignement de la langue et de la civilisation françaises de Paris). L'objectif en était double. D'une part rendre disponible l'ensemble du corpus – « Des listes de transcriptions et enregistrements sont disponibles à ceux qui s'adressent à nous. » (Lonergan *et al.*, 1974) – et d'autre part, constituer un corpus « sociolinguistique » autour du concept de « portrait sonore d'une ville » afin de croiser représentativité et variations au sein d'une communauté d'auditeurs dans un espace géographique et socioéconomique clairement défini (Bergounioux *et al.*, 1992). Dans les années 1980-1990, une partie du corpus a été transcrite et étiquetée puis mise à disposition sur la Toile dans le cadre du projet ELILAP/LANCOM[14]. Entre 1993 et 2001, une partie du corpus a été reprise par des chercheurs de l'Université de Louvain (Debrock *et al.*, 2000) dans le cadre du projet ELICOP[15].

Quarante ans après le projet initial, le Laboratoire Ligérien de Linguistique (LLL) de l'université d'Orléans a entrepris un double projet : diffuser largement le corpus ESLO1 et réaliser un nouveau corpus représentatif du français parlé à Orléans dans les années 2010 (ESLO2), en prenant en compte l'expérience d'ESLO1 et l'évolution des cadres théoriques et méthodologiques de la constitution et de l'exploitation de grands corpus oraux à visée variationniste.

14. ELILAP 1980-83 puis LANCOM 1993-2001, voir Mertens (2002).
15. http://bach.arts.kuleuven.be/elicop/

2.2. *ESLO en chiffres*

Actuellement, ESLO1 est composé de 470 enregistrements, d'une durée totale de 317 heures et évalué à 4 500 000 mots. Plusieurs genres y sont représentés : la grande partie des enregistrements consiste en des entretiens en face-à-face (157 enregistrements comportant autant de profils sociologiques différents), mais ESLO1 compte également des enregistrements libres dans des situations privées ou professionnelles faites en l'absence des chercheurs (16 enregistrements), des interviews des personnalités de la ville (45 enregistrements), des reprises de contact complètement informelles (55 enregistrements), des communications téléphoniques (50 enregistrements), des conférences-débats ou même des discussions (26 enregistrements), des interviews au centre médico-psychopédagogique entre des parents d'élèves et des assistantes sociales (37 enregistrements), ainsi que des enregistrements divers dans les lieux publiques : magasins, marchés, visites d'ateliers, etc. (84 enregistrements).

ESLO2, débuté en 2008, est un corpus en cours. À terme, il comprendra plus de 350 heures d'enregistrement afin de former avec ESLO1 un corpus de plus de 700 heures et atteignant les dix millions de mots. Il s'agira alors d'un grand corpus oral réalisé selon des bonnes pratiques de constitution garantissant l'interopérabilité des données avec d'autres projets semblables.

La mise à disposition d'un tel corpus se heurte en premier lieu aux aspects juridiques. Nous verrons que l'anticipation de ceux-ci a eu une incidence sur l'ensemble de la chaîne de traitement.

3. Mettre à disposition : les aspects juridiques

À l'époque de la conservation pérenne et de la diffusion des archives numériques il est important de prendre en compte l'ensemble des aspects juridiques dès la conception du projet. Nous avons bénéficié des réflexions et recommandations émanant du groupe de travail du ministère de la Culture et du CNRS[16] : *Corpus oraux. Guide des bonnes pratiques 2006*.

Les problèmes rencontrés se concentrent sur deux grands domaines juridiques : le respect de la vie privée et la protection de la propriété intellectuelle.

16. Groupe de travail composé de linguistes, juristes, informaticiens et conservateurs.

3.1. *Respect de la vie privée*

3.1.1. *Complexité de la notion*

Selon le *Guide des bonnes pratiques 2006* pour gérer les droits liés au respect de la vie privée, il convient de suivre scrupuleusement le cadre légal de gestion des données personnelles, de s'assurer que les locuteurs ont exprimé leur consentement « éclairé » ou, à défaut de celui-ci, de procéder à l'anonymisation des données. Juridiquement, l'anonymisation consiste à rendre impossible l'identification d'une personne. Cette notion est assez complexe à mettre en place dans le cadre de recherches linguistiques sur l'oral (il n'est pas question de « brouiller » totalement la voix des locuteurs). On restreindra l'objectif au masquage ou à la suppression des éléments permettant une identification par un large public qui utiliserait des moyens classiques de requêtes. Il faut donc repérer et traiter les indices permettant d'identifier directement ou indirectement la personne, ainsi que les éléments qui peuvent lui porter préjudice. Il peut s'agir des formes nominatives, des professions, statuts, ou titres, des activités sociales, des liens de parenté, des réseaux, des références à des lieux et/ou des références à des caractéristiques de la personne ou encore des propos légalement répréhensibles.

Les données traitées concernent les fichiers audio, les données primaires textuelles ainsi que les données secondaires comme les transcriptions, les métadonnées ou les analyses effectuées sur ces données.

La diffusion des données anonymisées présuppose aussi la préservation et la conservation des données originales non anonymisées ainsi que l'accès restreint à ces données.

3.1.2. *Respect de la vie privée dans ESLO : anonymisation*

Dans le cas du corpus ESLO1, le recueil de consentement pose deux problèmes. Premièrement, il n'existe aucun document rempli par les locuteurs qui permettrait d'exprimer ce consentement ; deuxièmement, il serait illusoire de penser que les locuteurs de la fin des années soixante imaginaient le type d'exploitation et notamment la diffusion instantanée par Internet. Toutefois le contenu des enregistrements et certains documents annexes prouvent que les locuteurs étaient conscients d'être enregistrés à des fins d'exploitation scientifique et didactique. Pour ESLO2, le recueil d'un consentement éclairé est systématique et permet une diffusion de l'ensemble des données brutes.

Le choix de l'équipe a néanmoins été d'anonymiser l'ensemble des données d'ESLO1 et d'ESLO2. L'idée est de repérer dans les enregistrements des éléments sensibles pouvant donner une information personnelle sur le locuteur.

Nous nous sommes tout d'abord intéressés aux noms propres. Il faut considérer que tous les noms propres ne sont pas à anonymiser : la *Loire* et *Jeanne d'Arc* ne sont pas à inclure dans l'effacement, ainsi que les toponymes se trouvant dans la

réponse à la question « *Où parle-t-on bien le français ?* » ou encore le nom des animateurs célèbres de l'époque, dans les réponses sur les questions concernant les émissions télévisées ou radiophoniques. En revanche, dans la phrase « *Je travaille au collège de Saint-Jean-de-Braye* », l'entité *collège de Saint-Jean-de-Braye* n'est plus seulement un établissement scolaire, mais également un lieu de travail du locuteur. Les noms communs désignant les métiers, par exemple, peuvent aussi à leur tour donner une information personnelle. En observant le corpus, nous avons constaté que c'est souvent le regroupement de plusieurs indices qui peut renvoyer vers l'identité du locuteur. Être un *professeur* ne permet pas d'identification, mais il n'en va pas de même s'il est précisé par ailleurs que c'est un *professeur d'université spécialisé en électronique* et, ailleurs encore, que c'est une *femme*, auquel cas on peut arriver à un singleton. À cela s'ajoutent les exemples comme « *mon père a fondé le plus grand cabinet d'ophtalmologiste de la ville* » qui sont rarement présents dans le corpus et permettent, en revanche, l'identification directe du locuteur.

Si au début, nous avons pensé automatiser complètement le processus d'anonymisation en nous fondant sur des couches d'annotations automatiques de ce type d'information, nous avons vite été confrontés à l'impossibilité d'effectuer cette tâche, d'où le travail manuel de la validation selon le contexte des entités repérées automatiquement[17]. Celles qui identifient le plus le locuteur ont été remplacées par leur hyperonyme et le son a été masqué. Pour cette raison, il a été décidé de simplifier le processus d'anonymisation pour ESLO2 : la phase de remplacement par un hyperonyme s'y fait dès la transcription.

3.2. *Propriété intellectuelle*

À la différence du corpus constitué en 1968-1971, en 2012, la question de la propriété intellectuelle de corpus contenant des paroles de locuteurs enregistrés, des enrichissements de ces paroles (transcriptions, annotations), constitués en base de données dans le cadre de projets financés par les institutions de l'État et réalisés par des enseignants-chercheurs en fonction est éminemment complexe.

L'objectif annoncé du projet ESLO de mise à disposition de l'ensemble du corpus a permis à l'équipe de se positionner dans une démarche de ressources libres et accessibles. Ainsi ESLO est un corpus mis à disposition librement sous licence *Creative Commons*[18]. L'usage de ce type de licences permet de gérer les droits d'exploitation en spécifiant la paternité et les conditions d'utilisation. Nous verrons

17. Dans la section 4.4., nous présenterons une couche d'annotations automatiques dont nous nous sommes servis pour l'anonymisation des entretiens d'ESLO1.
18. Les licences Creative Commons sont des contrats-types pour la mise à disposition d'œuvres en ligne. Il s'agit d'autorisations non exclusives données par les titulaires des droits au public. Ces autorisations spécifient les conditions d'utilisation des œuvres.

également l'impact de ce choix sur la structuration de la base de données et sur les opérations de traitement du corpus.

Les sections suivantes sont consacrées à la description des opérations de traitement du corpus entrant en jeu dans la préparation de ressources destinées à une libre diffusion. Comme nous le verrons il ne s'agit pas simplement de problèmes techniques qui pourraient apparaître neutres sur le plan théorique.

4. Annotation

Pour exploiter un corpus oral il est nécessaire de le transcrire et certaines tâches d'annotation deviennent utiles et/ou indispensables. « Mais c'est certainement une erreur que d'imaginer que le modèle suivi pour l'écrit pourrait être transféré à l'oral. En effet, les corpus oraux sont liés à des exploitations extrêmement diversifiées (analyse prosodique, analyse de discours, analyse syntaxique, approches pragmatiques ou sociolinguistiques, etc.) qui nécessitent des informations par nature très disparates. » (Cappeau et Gadet, 2007). Les choix d'annotation diffèrent d'un projet à l'autre suivant des objectifs variés. Ainsi, dans le cadre du projet OTIM[19], le travail d'annotation a porté sur un grand nombre de domaines : phonétique, prosodie, phonologie, syntaxe, discours et gestes. Le corpus EPAC que nous avons mentionné avant a été annoté en prenant en compte divers phénomènes : bruits, musiques, inspirations, prononciations particulières ou erronées, mots étrangers, néologismes... Le projet Rhapsodie[20], quant à lui, met au centre de ses activités les annotations prosodique et syntaxique des données orales existantes.

Les outils d'annotation varient également selon la nature de l'annotation, c'est-à-dire selon les phénomènes que l'on veut distinguer. Ainsi, l'annotation automatique des coréférences, par exemple, pose de nombreux problèmes et nécessite le recours à l'intervention humaine. Toutefois, il existe des outils d'aide à l'annotation manuelle comme Transcriber[21], Praat[22], ANVIL[23], ELAN[24] pour la transcription, et Glozz[25], Gate[26], etc. pour d'autres niveaux d'annotation. L'annotation automatique ou semi-automatique peut se faire avec des méthodes à base de règles linguistiques décrivant le contexte d'emploi de phénomènes à annoter sous forme de grammaires locales ou avec des méthodes d'apprentissage automatique à partir d'un corpus de référence

19. Outils pour le traitement de l'information multimodale, http://www.lpl-aix.fr/˜otim.
20. http://rhapsodie.risc.cnrs.fr/fr/index.html
21. http://trans.sourceforge.net/en/presentation.php. Une nouvelle version de ce dernier est disponible depuis juillet 2011.
22. http://www.fon.hum.uva.nl/praat/
23. http://www.anvil-software.de/
24. http://icar.univ-lyon2.fr/projets/corinte/confection/elan.htm
25. http://www.glozz.org/
26. http://gate.ac.uk/

annoté manuellement. C'est la nature des données qui dicte le choix de la méthodologie à adopter. Le corpus ESLO est un exemple de cette démarche. Nous montrerons dans cette partie, comment nous avons adopté les outils et techniques existants à chaque type d'annotation.

D'après Leech (1997), l'annotation est une « valeur ajoutée » aux données brutes, c'est-à-dire un apport d'informations. Toujours selon cet auteur, la transcription possède un statut ambigu car la frontière entre les données brutes, neutres et leur annotation n'est pas clairement délimitée. Tout commentaire (balisage des bruits, notes du transcripteur) appartient également au domaine de l'annotation et peut donc être considéré comme de l'interprétation.

Nous considérons le processus d'annotation comme *porteur d'une interprétation*. Ainsi, il n'y a pas qu'une version de corpus annoté mais plusieurs versions –existantes ou potentielles – du même corpus. Les différents annotateurs humains peuvent interpréter et percevoir différemment les données. Même dans le cas de l'annotation automatique, les annotations diffèrent par les conventions, les techniques, etc. ESLO est un corpus de variations : variations entre le français d'hier et d'aujourd'hui, entre les différents locuteurs, entre les différentes situations d'enregistrement, mais aussi entre les différentes annotations.

Afin de rendre disponible et exploitable notre corpus, nous avons suivi les principes d'annotation suivants pour la transcription :

– lisibilité ;
– conservation des spécificités de l'oral ;
– volonté d'un maximum d'interopérabilité ;
– codage non ambigu ;
– contraintes de comparabilité (d'ESLO1 à ESLO2).

Pour nous, la transcription doit être considérée comme un premier niveau d'annotation : le son étant enrichi d'une information orthographique.

Dans le cadre d'un corpus sociolinguistique, on souhaite également intégrer des descripteurs de la situation d'interaction, notamment les données sociologiques sur les locuteurs et la description de la situation d'enregistrement. Ces données peuvent être décrites dans les métadonnées mais peuvent aussi être contenues directement dans le corpus, comme par exemple dans un entretien ou un récit de vie. L'exploitation sociolinguistique nécessite évidemment la disponibilité de ces informations.

Nous montrerons, dans la partie qui suit, les différentes annotations effectuées sur le corpus afin de répondre à cet objectif.

4.1. *Annotation du niveau zéro : transcription*

4.1.1. *Contraintes et choix de transcription*

La transcription qui est le premier degré d'annotation de l'oral est une étape primordiale dans la constitution du corpus puisque c'est sur ce premier niveau que vont s'ajouter d'autres annotations. Les choix faits à ce stade influencent donc tout le traitement postérieur. La tâche a été d'autant plus difficile qu'il n'y a pas de conventions de transcription admises par la communauté scientifique.

Plusieurs contraintes ont influencé nos choix. Notre volonté première était de mettre à disposition des chercheurs une grande quantité de données transcrites (700 heures d'enregistrement). Le processus de transcription devait donc être effectué rapidement mais avec une bonne efficacité. Il n'existe pas aujourd'hui d'outils de transcription automatique disponible, il s'agit donc de transcriptions manuelles. Ceux qui ont travaillé sur l'annotation manuelle savent que moins on annote d'informations, plus on gagne dans la quantité et la qualité car l'annotateur est moins dispersé et donc plus concentré sur sa tâche. Nous sommes allés dans la même direction et nous avons choisi l'annotation minimale. Il s'agit de la transcription orthographique qui conserve les spécificités de l'oral (amorces, disfluences, répétitions, etc.). Les conventions de transcription ont été réduites ainsi au minimum. Pour éviter l'anticipation de l'interprétation (Blanche-Benveniste et Jeanjean, 1987), les marques typographiques comme le point, la virgule, le point d'exclamation ou encore la majuscule en début d'énoncé sont absentes. La segmentation a été faite soit sur une unité intuitive de type « groupe de souffle » repérée par le transcripteur humain, soit sur un « tour de parole », défini uniquement par le changement de locuteurs.

La synchronisation avec le son était une autre contrainte. On devait pouvoir naviguer dans la transcription et le son en parallèle. L'objectif a été défini de transcrire et rendre disponible l'intégralité du corpus en y associant des jalons temporels.

La mise à disposition des ressources implique l'utilisation de normes et même si les conventions de transcription ne sont pas normalisées, l'exigence sur le format standardisé des fichiers s'impose. Ce format interopérable devait permettre un traitement plus facile du corpus par les outils du TAL.

Notre choix s'est arrêté sur le logiciel de transcription Transcriber qui répondait complètement à nos attentes : outil facile d'utilisation permettant la synchronisation entre la transcription et le signal sonore et ayant un format de sortie XML, un format normé, facilement exportable et largement utilisé, gage d'interopérabilité. À l'aide de ce logiciel, l'ensemble des données dont l'acoustique était acceptable a été transcrit, ce qui nous permet de disposer d'un panorama varié des différents types d'enregistrements ESLO1 (Baude et Dugua, 2011).

Pour nous, la phase de transcription relève et révèle systématiquement des variations linguistiques. Cette conception de la transcription nous a donc guidés dans l'élaboration de notre méthodologie. Chaque enregistrement est transcrit en trois étapes successives, donnant lieu à trois versions différentes :

– transcription (A), première transcription rapide ;
– transcription (B) qui est la transcription (A) relue et corrigée par un deuxième transcripteur ;
– transcription (C), la transcription (B) relue et corrigée par un troisième transcripteur.

Avec cette méthode, « nous évaluons le temps de transcription à 10 fois pour une première version brute, 5 fois pour une deuxième et autant pour une troisième. » (Baude et Dugua, 2011). Cette méthodologie nous a permis de constater et de confirmer notre hypothèse sur la variation dans l'annotation ; ce sont trois perceptions différentes de l'écoute, qui manifestent toutes trois des types de variations et des opérations particulières.

4.1.2. *Transcriptions et variations*

La comparaison[27] des trois versions de transcriptions a montré d'importantes divergences. Trois cent trente différences ont été relevées en moyenne entre les trois versions. Dans le cadre d'un travail de thèse (Hriba, en cours), leur description, leur analyse et leur catégorisation, à partir d'un corpus constitué de vingt enregistrements et de soixante fichiers de transcription (225 173 mots), ont permis la mise en évidence de trois types de variations :

– des variations graphiques qui regroupent l'ensemble des erreurs qui, d'une part, correspondent aux fautes induites par les outils de saisie (clavier) et d'aide à la transcription (Transcriber) et, d'autre part, qui correspondent au non-respect d'une norme orthographique ou de codage (cf. conventions de transcription) ;

  ESLO1_110B/C : les g**ra**çons/les g**ar**çons
– des variations de segmentation qui concernent des différences d'alignement temporel, la segmentation en sections, en tours de parole et les pauses ;
– des variations de perception manifestant des divergences d'écoute ;

  ESLO1_062A/B : **on** le mettait tout à fait **en bas** à gauche / **il** le mettait tout à fait **en haut** à gauche.

L'étude des variations a montré qu'aux trois types de variations correspondaient des opérations communes, d'une part, et des opérations spécifiques à chacune d'elles, d'autre part. Les opérations correspondent aux interventions que les relecteurs font sur la version précédente (de la B sur la A et de la C sur la B). Nous

27. Comparaison réalisée à l'aide d'un outil spécialisé Beyond Compare 3 : http://www.scootersoftware.com/

présentons ici les résultats généraux en nous limitant aux opérations communes, au nombre de trois : des modifications, des suppressions et des rajouts. Le tableau 1 fait apparaître la proportion de chaque type d'opération pour les deux relectures.

Types d'interventions	**Interventions de la version B sur la version A**	**Interventions de la version C sur la version B**
Modifications	56 %	49 %
Suppressions	11 %	9 %
Ajouts	33 %	42 %

Tableau 1. *Interventions des relecteurs*

Globalement, nous constatons que les types d'opérations présentent une répartition semblable que ce soit pour la relecture B ou pour la validation C.

Parmi les trois types de variations, les variations de perception sont celles qui ont fait l'objet d'une étude particulière. En effet, à travers l'analyse de 130 exemples et des interférences qui ont conduit aux divergences constatées, Hriba a pu relever des mécanismes qui démontrent que la variation est inhérente au système, y compris dans un traitement de description du signal sonore.

ESLO est donc un corpus qui en imbrique d'autres. Nous le considérons plutôt comme une « archive » qui permet l'extraction de « corpus d'études ». Ainsi, les trois versions des transcriptions constituent à leur tour un corpus à part entière qui demande à être analysé. Le choix de garder les trois versions correspond à notre conception d'un corpus dont les traitements sont révélateurs et même porteurs de variations.

4.2. *Annotations et métadonnées*

À toutes les étapes et donc dès la collecte des données, se pose la question des éléments descripteurs de la ressource afin de faciliter son exploitation, sa réutilisation et son archivage. Il est important de prévoir l'annotation des métadonnées qui décrivent et enrichissent le corpus dans la chaîne de traitement dès le début. Cette tâche devient encore plus compliquée pour le corpus oral qui met en jeu les différents types de fichiers (sonores et écrits), les différentes situations d'enregistrement : entretiens en face-à-face, conversations spontanées, réunions de travail, etc., les locuteurs qui se distinguent par leurs âge, sexe, profession, lieu de naissance, milieu social, etc. Les contributeurs du corpus varient également : ceux

qui font les enregistrements ne sont pas toujours ceux qui les transcrivent. Il s'agit donc de décrire d'une manière homogène toute cette variété de données.

Dans le cadre du projet ESLO, nous pouvons distinguer les métadonnées décrivant les enregistrements et les transcriptions et les métadonnées enrichissant la situation de production linguistique et notamment le profil sociologique du locuteur.

Pour les fichiers de transcription, une partie des métadonnées est contenue dans les balises XML proposées par Trancriber au cours de l'étape de transcription. Cette information permet de décrire le fichier de transcription.

Chaque fichier de transcription correspondant à un enregistrement est caractérisé, en premier lieu, par le nom du transcripteur, le numéro de l'enregistrement, la version ainsi que la date de la transcription :

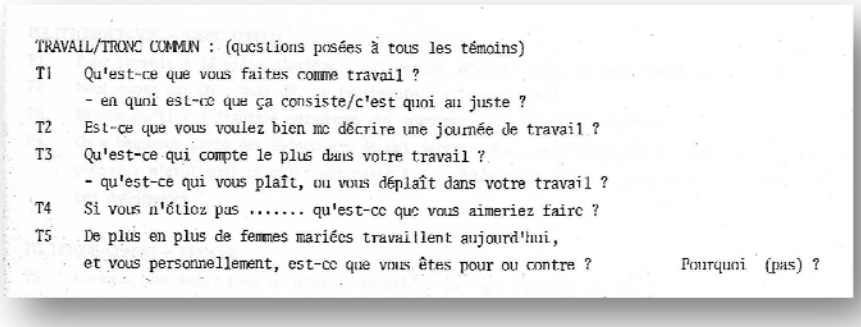
```
<Trans scribe="Panot" audio_filename="001" version="15" version_date="091210">
```

L'attribut *version* marque le nombre d'interventions au cours de la transcription. Cette information peut s'avérer très intéressante si l'on veut analyser la durée de transcription manuelle d'un enregistrement et le nombre d'interruptions que le transcripteur humain fait pendant ce processus.

Dans le cas des entretiens, on décrit des différentes thématiques de l'interview :

```
<Topics>
<Topic id="to1" desc="QP1"/>
<Topic id="to2" desc="QP2"/>
```

Les codes décrits font référence à une trame d'entretien établie lors de la constitution du plan expérimental. Un document annexe dont voici un extrait du catalogue tapuscrit original de 1974 (figure 1) détaille les thématiques.



TRAVAIL/TRONC COMMUN : (questions posées à tous les témoins)
T1 Qu'est-ce que vous faites comme travail ?
- en quoi est-ce que ça consiste/c'est quoi au juste ?
T2 Est-ce que vous voulez bien me décrire une journée de travail ?
T3 Qu'est-ce qui compte le plus dans votre travail ?
- qu'est-ce qui vous plaît, ou vous déplaît dans votre travail ?
T4 Si vous n'étiez pas ..... qu'est-ce que vous aimeriez faire ?
T5 De plus en plus de femmes mariées travaillent aujourd'hui,
et vous personnellement, est-ce que vous êtes pour ou contre ? Pourquoi (pas) ?

Figure 1. *Extrait des thématiques de l'interview dans ESLO1*

Chaque question est incluse dans une thématique plus vaste traitant du travail, des loisirs, de la politique, de l'éducation, de la culture ou encore de la langue elle-

même. L'extraction de cette information permet de constituer des sous-corpus selon un sujet traité.

L'information sur les participants de l'enregistrement est intégrée dans la balise <Speakers> :

<Speaker id="spk1" name="HM" check="no" dialect="native" accent="" scope="local"/>

Les métadonnées sont décrites aussi dans la base de données connectée à l'interface Web ce qui donne accès à la fois à l'enregistrement (fichier sonore), à sa/ses transcription(s) (fichier de Transcriber) et aux informations relatives au locuteur principal de chaque entretien : date et lieu de naissance, sexe, profession et appartenance sociale. Ces critères ont été renseignés au moment de l'enquête.

Le corpus des ESLO est également déposé au CRDO[28] avec les métadonnées requises par la procédure d'archivage de celui-ci (quinze étiquettes extraites du standard DUBLIN-CORE OLAC). Voici un exemple extrait du corpus ESLO au CRDO et une description des métadonnées utilisant les étiquettes du DUBLIN CORE OLAC[29] :

dc:title. [1, 1]. Titre.
dc:subject. [1, n]. Description du sujet du contenu de la ressource.
dc:type. [0, n]. Nature ou genre du contenu de la ressource.
dc:source. [0, n]. Référence à une ressource à partir de laquelle la ressource actuelle a été dérivée.
dcterms:spatial. [0, n]. Couverture spatiale. En général le point d'enquête, sauf s'il ne représente absolument pas la couverture spatiale visée
dcterms:temporal. [0, n]. Couverture temporelle de la ressource. A ne pas confondre avec les dates d'enregistrement,
dc:creator. [0, n]. Entité responsable de l'élaboration du contenu de la ressource (individu, institution, organisation)
dc:publisher. [0, n]. Entité responsable de la mise à disposition de la ressource, dans sa forme actuelle.
dc:contributor. [0, n]. Entités ayant contribué à la création du contenu de la ressource. Préciser leurs rôles par choix
dc:rights. [0, 1]. Indiquer la mention de copyright
dcterms:license. [1, 1]. URL de la licence creative-commons choisie
dcterms:created. [1, 1]. Date de création de la ressource.
dcterms:modified. [1, 1]. Date de dernière modification de la ressource.
dc:language. [0, n]. Langue du contenu intellectuel de la ressource. Langue de l'enquêteur dans un document sonore.

Ces métadonnées correspondent à des normes internationales qui permettent non seulement de décrire une situation de production mais aussi de cataloguer les ressources électroniques afin d'en faciliter l'accès ultérieurement.

Enfin, dans le cadre du projet ESLO, nous avons développé notre propre jeu complémentaire de métadonnées structuré dans une base de données dédiée à celui-ci.

28. Centre de ressources pour la description de l'oral – CNRS.
29. http://www.language-archives.org/REC/role.html.

4.3. *Annotation du discours*

L'étape suivante n'est plus descriptive mais a comme objectif la segmentation du flux de paroles, qui se manifeste par les changements de locuteurs, les pauses, les événements comme la toux, les rires, etc. Les tours de parole, le temps d'énoncé, les pauses, les différents événements interrompant le flux de paroles, doivent être marqués pour faciliter le traitement du corpus par les outils informatiques et permettre l'alignement avec le fichier sonore. Ce processus d'annotation a été effectué sous Transcriber.

La transcription est découpée d'abord en sections (ou *Report*) qui correspondent dans les entretiens, par exemple, aux questions de la trame du questionnaire. La deuxième segmentation se fait par le transcripteur soit intuitivement selon le groupe de souffle ou s'il y a une pause dans le discours du locuteur, soit selon le tour de parole, défini uniquement par les changements de locuteurs :

```
<Turn speaker="spk1" startTime="0.449" endTime="3.114">
<Sync time="0.449"/>
vous savez euh
<Sync time="1.317"/>
<Sync time="2.814"/>
enfin
</Turn>
```

Les attributs « *startTime* » et « *endTime* » indiquent le temps de début et de fin des segments de parole. Un des phénomènes de l'oral qui nécessite une segmentation particulière est la pause, elle est notée par un segment vide. Cette segmentation permettra d'avoir précisément la durée de la pause.

Les divers événements au sein du discours oral annotés avec Transcriber concernent les différents bruits : rire, micro, passages non transcrits, bruits de respiration (inspiration, soupir, respiration) ou encore des clics ou bruits de bouche :

```
<Event desc="rire" type="noise" extent="instantaneous"/>
```

Nous ajouterons à ces événements les phénomènes de prononciation :

```
petite <Event desc="pi" type="pronounce" extent="instantaneous"/> moyenne
les techniciens ils <Event desc="i" type="pronounce" extent="previous"/> font
```

Les informations annotées sont un autre exemple de variation. La variation dans le même enregistrement peut être observée entre les différentes questions posées, entre la pause et le discours, entre le discours et l'événement. Les différents enregistrements d'interviews mettent en évidence la variation entre les différents locuteurs. Ainsi, dans (Dupont *et al.,* 2012), les auteurs se sont intéressés au sous-corpus composé des réponses à une question posée aux différents locuteurs sur les événements de mai 1968. Certaines informations comme la durée de pause après la question, la durée de pause par section ou la durée de section ainsi que les renseignements sur le locuteur comme son sexe, son âge, son niveau d'études, etc. ont été analysées avec des méthodes de statistiques descriptives. Il s'agissait

d'étudier des variations et des corrélations entre d'une part, des annotations de temps dans les fichiers de transcription, et, d'autre part, des valeurs sociologiques.

4.4. *Annotations des informations personnelles concernant le locuteur*

L'annotation est toujours liée à la nature des données à annoter et à l'objectif visé. Les enquêtes sociolinguistiques à Orléans contiennent beaucoup d'informations personnelles sur les locuteurs interviewés. Repérer ces données est d'autant plus nécessaire qu'elles peuvent être utilisées dans le processus d'anonymisation.

Ce type d'annotation a été réalisé sur une partie du corpus ESLO1[30]. Nous avons sélectionné 112 entretiens en face-à-face. Les entretiens contiennent beaucoup de questions du type : « *Depuis combien de temps habitez-vous Orléans ?* » « *Quel âge avez-vous ?* » « *Qu'est-ce que vous faites comme métier ?* » « *Où travaillez-vous ?* » « *Qu'est-ce que fait votre époux(se) ?* », etc. Les réponses des locuteurs montrent comment les Orléanais parlent d'eux-mêmes et représentent une masse de données mettant en valeur une variation intéressante à analyser. L'intérêt des entretiens pour annoter automatiquement ces informations est qu'ils présentent des données riches et homogènes. Dans les discours spontanés, les énoncés contenant des informations personnelles permettant une identification sont plus rares et surtout moins structurés, ce qui rend plus difficile une annotation automatique. C'est la raison pour laquelle nous nous sommes limités à ce sous-corpus.

4.4.1. *Méthodologie adoptée*

L'annotation s'est faite en deux étapes. Nous avons repéré et annoté, en premier lieu, les entités nommées comme le nom de la personne, son âge ou son lieu de travail. Nous avons recherché ensuite les éléments plus personnels concernant le locuteur comme son métier, le nombre d'enfants qu'il avait, le métier de son conjoint, etc. que nous avons appelés « entités dénommantes » (Eshkol, 2010). Le processus de reconnaissance de ces informations s'est effectué sur le corpus annoté en entités nommées.

L'annotation des entités nommées et dénommantes a été décrite dans (Maurel *et al.,* 2011). Nous nous contenterons dans cet article de présenter une synthèse de ce travail.

Pour repérer et annoter les entités nommées et dénommantes, nous avons choisi l'approche en surface permettant de construire les grammaires locales selon le contexte en utilisant le système CasSys (Friburger, 2002) intégré à la plate-forme Unitex (Paumier, 2003).

---

30. À l'origine de cette étude les transcriptions ESLO2 n'étaient pas disponibles.

Notre choix a été guidé d'une part, par le corpus déjà constitué des entretiens dans lesquels les mêmes questions avaient été posées aux différents locuteurs. L'analyse de ce corpus a permis de créer des règles d'extraction (patrons) fondées sur ces questions et sur les structures répétées dans les réponses. Le système CasSys, d'autre part, nous a permis de ne pas développer un nouvel outil. Il s'agissait d'adapter CasSys à nos données, ce qui nous a semblé être plus économique et approprié à notre tâche.

L'annotation a été réalisée sur cent douze fichiers Transcriber, soit un total de 35,75 Mo. Six fichiers ont été réservés pour les tests et neuf fichiers pour l'évaluation.

L'adaptation de l'outil à ESLO1 s'est faite sur plusieurs niveaux :

– prétraitement du corpus : le découpage en phrases d'Unitex a été remplacé par un découpage en fonction des balises Transcriber, c'est-à-dire en général par un découpage en tours de parole[31] ;
– enrichissement des cascades par des dictionnaires et des graphes spécifiques ;
– élaboration de la typologie des entités adéquates à notre besoin, à partir de la typologie de la campagne d'évaluation Ester2 (*campagne d'évaluation des systèmes de transcription enrichie d'émissions radiophoniques*)[32].

Il était indispensable de prévoir aussi la présence éventuelle de disfluences et de reprises syntaxiques, ainsi que d'insertions et d'amorces, comme par exemple dans :

– *je m'appelle euh Patrick Mallon*

4.4.2. *Jeu d'étiquettes choisi*

Le contenu d'annotation a été guidé par les questions posées au locuteur sur lui ou sa famille, questions qui portaient sur :

– l'identité (date de naissance, date d'arrivée à Orléans, âge, origine, date de mariage, etc.) ;
– le travail (métier, secteur d'activité, lieu ou nom d'entreprise, etc.) ;
– l'engagement (associations, syndicats, etc.) ;
– les études (diplômes, lieux ou établissements) ;
– les voyages.

La typologie ainsi définie concerne les informations sur la personne interrogée (*pers.speaker*), son conjoint (*pers.spouse*), ses enfants (*pers.child*) et d'autres membres de la famille (*pers.parent*) : voir le tableau 2[33].

---

31. Méthode recommandée par Anne Dister (2007).
32. http://www.afcp-parole.org/ester/docs/Conventions_EN_ESTER2_v01.pdf

Personne (+pers)	la personne interrogée (+speaker)
	son conjoint (+spouse)
	ses enfants (+child)
	les autres membres de la famille (+parent)
Identité (+identity)	le nom (+name)
	l'adresse (+addr)
	l'âge (+age)
	le mariage (+wedding)
	l'origine (+origin)
	la naissance (+birth)
	l'arrivée à Orléans (+arrival)
	le nombre d'enfants (+children)
Travail (+work)	métiers (+occupation)
	secteur d'activité (+field)
	lieu de travail (+location)
	entreprise (+business)
Engagement (+involvement)	association (+voluntary)
	militaire (+military)
	scolaire (+school)
	syndical (+tradeunion)
Voyage (+trip)	études (+study)
	vacances (+holiday)
	professionnel (+work)
Etudes (+study)	lieu (+location)
	diplôme (+degree)
	établissement (+edu)

Tableau 2. *Typologie des entités dénommantes*

33. Pour plus de détails, voir (Maurel *et al.*, 2011).

Ainsi, nous annotons tout d'abord le sujet sur qui porte l'information : le locuteur ou les autres membres de sa famille ; nous précisons ensuite la nature de cette information : l'identité, le travail, les études, l'engagement associatif ou syndicale, les vacances.

Voici quelques exemples d'annotations, tout d'abord pour les entités nommées[34] :

– chez moi<ENT type="pers.hum"> Bérénice Nutal</ENT>
– moi je suis native de<ENT type="loc.admi"> Pithiviers</ENT> j'aime mieux <ENT type="loc.admi">Orléans</ENT>

Puis pour les entités dénommantes :

– <DE type="pers.child"> il est parti <DE type="work.location"> à <ENT type="loc.admi"> Paris </ENT></DE> il travaille dans les <Sync time="1526.195"/> <DE type="work.field"> dans les assurances </DE></DE>
– alors <DE type="pers.speaker"> <DE type="identity.name"> je suis <ENT type="pers.hum"> monsieur Gabrion </ENT></DE></DE> <DE type="pers.speaker"> je suis <DE type="work.occupation"> ingénieur chimiste </DE> </DE>
– <DE type="pers.speaker"> <DE type="identity"> je m'appelle euh <ENT type="pers.hum"> Patrick Mallon </ENT></DE></DE>

Le corpus annoté a été vérifié ensuite manuellement. L'évaluation des résultats a été détaillée dans (Maurel *et al.*, 2009). Les entités dénommantes ont été reconnues avec la précision estimée à 94,2 % et le rappel de 84,4 %.

4.4.3. *Exemple d'utilisation des données annotées : l'anonymisation*

Nous avons déjà mentionné les enjeux importants de l'anonymisation des données orales lorsqu'elles sont mises en libre distribution. Pour anonymiser ESLO1, les entités dénommantes qui renvoient vers les informations personnelles concernant le locuteur et sa famille et qui peuvent éventuellement permettre sa reconnaissance, ont été repérées et étiquetées. On procède ensuite à l'analyse manuelle consistant en la validation des éléments annotés dans un contexte. Ceux qui identifient directement le locuteur sont remplacés par un hyperonyme.

Les premières formes de remplacement choisies sont : *NPERS*[35] pour un nom de personne, *NLIEU* pour un nom de lieu, et *NPROF* pour un nom de profession. En deuxième lieu, nous les remplacerons par des identifiants uniques et anonymes ce

34. La cascade utilisée pour les entités nommées est disponible sous licence LGPL-LR à l'URL : http://tln.li.univ-tours.fr/Tln_CasEN.html

35. L'anonymisation du fichier texte a été réalisée sur la plus petite partie possible (par exemple, le nom mais pas le prénom).

qui devrait permettre de traiter les phénomènes tels que les coréférences. Pour le fichier son, nous avons utilisé le logiciel Praat[36] et un script développé par D. Hirst[37]. La procédure consiste à segmenter l'enregistrement (création d'intervalles correspondant exactement à la partie du signal qui doit être brouillée), puis nous avons annoté ces intervalles avec un code, « buzz », et le script opère automatiquement un traitement du signal.

Comme il a été mentionné au début de l'article, la méthodologie a été modifiée pour ESLO2. L'impossibilité de rendre le processus totalement automatique nous a conduits à la simplification de l'anonymisation pour ESLO2 qui consiste dans le remplacement manuel par un hyperonyme d'un élément identifieur et se fait dès la transcription (figure 2).

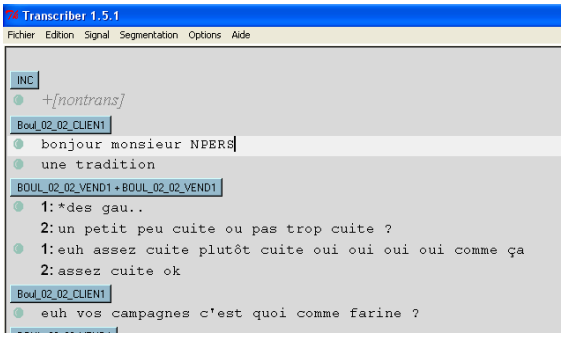


Figure 2. *L'anonymisation du corpus ESLO2*

4.4.4. *Information annotée versus information contenue dans la base de données*

ESLO contient de nombreuses métadonnées permettant une catégorisation sociologique des locuteurs et de la situation. Comme nous l'avons évoqué, celles-ci sont stockées dans une base de données et sont donc accessibles pour un traitement informatique, notamment sous la forme de requêtes croisées.

La base de données répertorie les différentes informations sur le locuteur, comme le montre la figure 3.

Ces informations sont remplies manuellement par la personne qui a fait un enregistrement, qu'on appelle un contributeur chercheur. Certains champs comme le sexe, la catégorie professionnelle INSEE, la situation de famille, etc. sont normalisés

36. http://www.fon.hum.uva.nl/praat/
37. http://uk.groups.yahoo.com/group/praat-users/files/Daniel_Hirst/anonymise_long_sounds.praat

et permettent leur interrogation dans les requêtes par les utilisateurs[38]. D'autres champs sont libres comme les remarques, les informations sur les enfants, la profession en termes propres, etc. et ne peuvent être interrogés.

L'annotation effectuée automatiquement par CasSys reprend certaines informations contenues dans la base de données, comme l'année et le lieu de naissance, par exemple, mais, contrairement à celle-ci, l'information annotée pourrait être interrogée. L'annotation réalisée apporte des informations complémentaires non contenues dans la base de données comme celles sur les loisirs, les vacances ou encore la vie associative du locuteur. Ces informations sont parfois indiquées dans la base de données dans un champ « remarques diverses », mais d'une façon non systématique et ne permettant pas leur étude.

Fiche locuteur

Identifiant locuteur : BA725

Anonyme:	OUI
Année de naissance:	1912
Tranche d'âge:	55/65
Lieu de naissance:	loiret
Sexe:	Homme
Niveau d'études:	CEP
Commentaire:	Enseignement primaire à Orléans
Age de fin d'études:	14
Catégorie Professionnelle (INSEE):	Artisans, commerçants et chefs d'entreprise
Profession en termes propres:	boucher, gérant boucherie supermarché
Langue(s):	Français
Commentaire niveau langue:	
Situation de famille:	Marié
Année d'arrivée:	1912
Domicile:	Orléans centre
Nombre d'enfants:	2
Information sur les enfants:	fils 1 : coiffeur, fils 2 ?
Remarques diverses:	Famille : femme sans activitéfils, brevet, coiffeurfilsEnseignement : primaire à orléansdiplôme : CEPProblème : non-renseigné
Fiche modifiée par:	obaude
Enregistrements et transcriptions:	▪ Enregistrement ESLO1_ENT_001 • Transcription ESLO1_ENT_001_A • Transcription ESLO1_ENT_001_B • Transcription ESLO1_ENT_001_C

Figure 3. *Fiche du locuteur de l'enregistrement 008*

Le corpus annoté des entités nommées et dénommantes n'étant pas anonymisé, il ne peut pas être mis à disposition librement sur le Web, l'accès en est restreint à la communauté scientifique sous réserve de signature d'une convention.

38. Les quatre champs : nom, prénom, nom de jeune fille et adresse, sont stockés dans une base de données séparée et ne sont plus accessibles pour le public (voir la section 3 de cet article).

Nous envisageons ultérieurement de relier les deux processus, pour que l'information annotée automatiquement soit stockée directement dans la base de données. Cette tâche est prévue dans la suite des travaux sur ESLO.

Parallèlement à l'annotation des informations personnelles concernant le locuteur, nous avons pu procéder aux premiers tests de la phase de traitement consacrée à l'annotation morphosyntaxique.

4.5. *Annotation morphosyntaxique*

Le travail développé dans cette partie est encore en cours de réalisation. Il s'agit de l'annotation morphosyntaxique, qui consiste à attribuer à chaque unité lexicale du corpus une étiquette apportant certaines informations (sa catégorie syntaxique, ses éventuels genre, nombre, temps verbal, etc.) dans le contexte où elle apparaît. Cette étape est précédée par une phase de segmentation dans laquelle le logiciel doit reconnaître et séparer les unités lexicales les unes des autres.

L'étiquetage morphosyntaxique est important pour la mise à disposition et la consultation du corpus car il permet de faire des requêtes précises. On pourrait ainsi vouloir extraire tous les noms communs employés par un certain locuteur, tous les verbes d'un autre, ou encore analyser les différentes prépositions utilisées après un certain verbe, etc. Comme le travail décrit ici n'est pas fini, la consultation d'ESLO en utilisant ce type de critères n'est pas encore possible, mais est envisagée ultérieurement.

Il existe des outils variés pour l'étiquetage morphosyntaxique : libres (TreeTagger[39], Sem[40], Melt[41], LGTagger[42]) ou payants (Cordial[43]). Le problème majeur est que ces outils ne sont pas adaptés à l'oral, caractérisé par les phénomènes de disfluences : répétitions, autocorrections, amorces de mots, etc. Cette tâche est d'autant plus difficile que les transcriptions ne sont pas ponctuées (voir la section 4.2).

Pour étiqueter ESLO, nous avons choisi de développer notre propre étiqueteur en utilisant la technique d'apprentissage automatique la plus adaptée pour cela à l'heure actuelle : les CRF (Lafferty *et al.*, 2001 ; Sutton et McCallum, 2006 ; Tellier et Tommasi, 2011), qui permettent de construire un modèle statistique à partir de données étiquetées fournies en exemple. L'objectif est de développer un étiqueteur libre et gratuit adapté au jeu d'étiquettes établi, en tenant compte des spécificités du corpus traité.

---

39. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
40. http://www.univ-orleans.fr/lifo/Members/Isabelle.Tellier/SEM.html
41. (Denis et Sagot, 2010).
42. http ://igm.univ-mlv.fr/˜mconstan/research/software
43. http://www.synapse-fr.com/Cordial_Analyseur/Presention_Cordial_Analyseur.htm

En 2010, nous avons réalisé des expériences préliminaires exploitant les CRF, en partant d'un corpus d'apprentissage déjà segmenté et annoté manuellement. Le programme développé permettait d'attribuer une étiquette morphosyntaxique à une unité lexicale selon trois niveaux : POS[44], informations morphologiques (genre, nombre, etc.), informations sémantiques et/ou syntaxiques, comme dans les exemples suivants :

oui	ADV	ADV	ADV
en_effet	ADV	ADV	ADV
on	P	P3I	P3IPER
peut	V	V3SINDP	V3SINDP
commencer	V	VINF	VINF
rire	V	VINF	VINF

Cette structure en trois niveaux autorise une certaine souplesse, suivant la nature et la qualité de l'information attendue : le premier niveau est plus simple à étiqueter et donc plus fiable, le troisième niveau inclut des informations linguistiques plus riches mais est potentiellement plus sujet à l'erreur d'étiquetage. On pourrait ainsi faire des requêtes plus ou moins précises, faisant appel spécifiquement à un certain niveau. Les premières expériences ont été décrites dans (Eshkol *et al.*, 2010), la phase de segmentation étant supposée être déjà réalisée. Elles utilisaient un ensemble d'entraînement assez réduit de 1 723 « énoncés » (tours de parole ou séquences entourées de « blancs ») étiquetés comportant 18 424 unités lexicales. Elles ont montré qu'il était possible d'apprendre un étiqueteur atteignant de 89 % d'exactitude (pour le troisième niveau) à 94 % (pour le premier niveau). Ce qui est moins bon que les performances annoncées des meilleurs étiqueteurs actuels du français, qui obtiennent entre 97 et 98 % d'exactitude (Denis et Sagot 2010, Constant *et al.*, 2011) avec un jeu d'étiquettes comparable à notre premier niveau. Mais ces derniers étiqueteurs ont été appris et testées sur le French Treebank, avec plus de 300 000 unités lexicales en entraînement. Ce corpus extrait d'articles du journal *Le Monde* est rédigé dans une langue beaucoup plus normée que le nôtre. L'oral présentant plus d'irrégularités, il est à prévoir que les modèles statistiques entraînés sur ESLO requièrent encore plus d'exemples pour parvenir à des résultats équivalents.

Dans la suite du travail, nous avons réfléchi à quelques modifications dans le jeu d'étiquettes, en prenant encore plus en compte les spécificités de l'oral. Certaines nouvelles étiquettes telles que « marqueurs discursifs » (MD) incluant trois sous-classes : MD (marqueurs discursifs propres), MDEUH (*euh* d'hésitation) et MDINT (interjections) et « présentatif » (PRES) pour les structures comme « c'est », « voici », etc. ont été ajoutées. Pour essayer d'éviter l'ambiguïté comme, par exemple, celle du participe passé dans « je suis prise/je suis partie »[45], nous avons aussi introduit une étiquette pour le passif.

---

44. *Part of speech*.
45. Dans les deux cas, le participe passé du verbe est précédé du verbe auxiliaire « être ».

Nous prévoyons de procéder à la segmentation préliminaire du texte par un segmenteur qui sera également appris avec un CRF. Il est aussi possible d'apprendre conjointement à segmenter et à étiqueter un texte, comme cela a été fait dans (Constant *et al.*, 2011), mais nous préférons pour l'instant réaliser les deux étapes indépendamment, parce que notre jeu d'étiquettes morphosyntaxiques est plus riche que le leur, ce qui risque d'augmenter les erreurs d'étiquetage. Pour apprendre à segmenter, le plus simple est de découper au maximum les unités du texte et de « recoller les morceaux » des mots composés par une phase d'annotation. Pour cela, nous segmentons tout d'abord le texte au maximum sur une base formelle, en prenant l'espace, l'apostrophe et les ponctuations comme séparateurs. Puis le texte est mis en format tsv[46] et des connaissances externes (comme l'étiquette fournie par un autre étiqueteur) sont ajoutées. L'étiquetage en B (pour « *Begin* », désignant le début d'une unité lexicale) et I (pour « *In* », désignant la suite d'une unité lexicale commencée précédemment) marquant les frontières du mot est suivi par leur fusion dans le cas des mots composés.

```
Pomme    N       B
De       PREP    I    =>    pomme_de_terre    N_PREP_N
terre    N       I
```

Comme connaissances externes (la colonne intermédiaire dans notre exemple), nous utilisons les résultats de l'étiqueteur libre SEM et des ressources linguistiques libres comme le lexique du Lefff et les tables de verbes et de noms prédicatifs du Lexique-Grammaire, passées au format alexina du Lefff. Cette étape est en cours de finalisation.

La deuxième phase de l'étiquetage morphosyntaxique s'appuiera sur cette segmentation. Les expériences d'apprentissage tenant compte des nouvelles étiquettes sont en cours. Le nouveau modèle d'annotation sera appris à partir d'un corpus de référence annoté manuellement ainsi que de connaissances externes (les mêmes que pour la segmentation).

Nous travaillons aussi sur la compatibilité de l'étiquetage avec les fichiers XML de Transcriber pour permettre la synchronisation avec les fichiers sonores. Le fichier de transcription, est prétraité pour donner un fichier en texte brut qui sera segmenté et étiqueté. Le résultat sera ensuite fusionné avec le fichier de départ pour donner un fichier étiqueté compatible avec le format original. On conservera ainsi les performances de notre étiqueteur et le même procédé pourra être applicable à n'importe quel autre format d'entrée et de sortie.

```
<Turn speaker="spk2" startTime="5.0" endTime="7.533">
<Sync time="5.0"/>
<Sync time="5.03"/>
et qu'est-ce qui vous a amené à vivre à Orléans
</Turn>
=>
```

46. *Tab separated values.*

```
<Turn speaker="spk2" startTime="5.0" endTime="7.533">
<Sync time="5.0"/>
<Sync time="5.03"/>
<w total="CONJCOO"> et </w>
<w total="PIINT"> qu'est-ce qui </w>
<w total="P2PPERCOMPL"> vous </w>
<w total="V3SINDPAUX"> a </w>
<w total="VMSPP"> amené </w>
<w total="PREP"> à </w>
<w total="VINF"> vivre </w>
<w total="PREP"> à </w>
<w total="NP"> Orléans </w>
</Turn>[47]
```

Les étiquettes de cet exemple sont celles du troisième niveau, et donc les plus précises. Elles intègrent, en quelque sorte, les niveaux précédents, permettant de retrouver si besoin tous les verbes (dont les étiquettes sont de la forme V*) indépendamment de leurs flexions. Mais si les étiquetages des différents niveaux sont de précisions notablement différentes, on peut aussi garder trois attributs différents pour chaque mot, correspondant à chacun de ces trois niveaux. Dans ce cas, pour rechercher les verbes, il suffira d'interroger uniquement l'attribut de premier niveau.

5. Consultation

Cette dernière partie est consacrée à la consultation d'ESLO. L'objectif du projet est de rendre le corpus disponible pour une large communauté scientifique mais aussi pour le grand public. L'application Web est disponible sous licence Creatives Commons[48].

La mise à disposition du corpus pose des questions sur sa consultation par les différents profils d'utilisateurs : chercheurs, contributeurs ou grand public. L'accès au corpus ne peut pas être le même pour chacune de ces catégories. Un corpus oral comme ESLO rend la tâche encore plus compliquée en raison de la masse et de la diversité des données et des métadonnées pouvant être consultées.

L'application Web de consultation des corpus ESLO se décompose en trois parties : la partie institutionnelle destinée à publier des informations sur le projet, la partie publique destinée à un large public et la partie administration.

Les quatre types d'utilisateurs sont pris en compte : administrateur, contributeur qui peut en outre ajouter ou modifier des informations d'enregistrement ou de

47.CONJCOO = Conjonction de coordination, PIINT = Pronom invariable interrogatif, P2PPERCOMPL = Pronom 2e personne pluriel personnel complément, V3SINDPAUX = Verbe 3e personne singulier indicatif présent auxiliaire, VMSPP = Verbe masculin singulier participe passé, PREP = Préposition, VINF = Verbe infinitif, NP = Nom propre.
48. Application réalisée par le prestataire ARES-GFI.

transcription, chercheur invité qui aura les mêmes accès qu'un utilisateur non authentifié avec en plus la possibilité de voir les données non anonymisées et toutes les versions de transcription et enfin l'utilisateur grand public.

On peut consulter le catalogue selon trois axes indépendants : les enregistrements, les transcriptions et les locuteurs auxquels sont associées les métadonnées (pour les enregistrements : type, durée, lieu, etc. ; pour les transcriptions : nom du transcripteur, problèmes et remarques, etc. ; et pour les locuteurs, leurs caractéristiques sociologiques : date et lieu de naissance, sexe, profession, etc.).

Des recherches simples et des recherches avancées sont possibles en précisant le corpus (ESLO1 ou ESLO2), les descripteurs de métadonnées de l'axe considéré ou encore selon l'occurrence (figure 4). La sélection de la version des transcriptions (brute, relue ou validée) est possible pour les profils administrateur, contributeur et chercheur invité. L'utilisateur peut faire des requêtes sur une occurrence ou en utilisant quelques expressions régulières. L'exploration du lexique d'une transcription choisie est affichable également, il s'agit d'une liste de mots avec leur fréquence. Parmi les autres actions réalisables sur le site, on peut naviguer synchroniquement dans la transcription et le fichier son, ou bien télécharger un enregistrement ou une transcription.

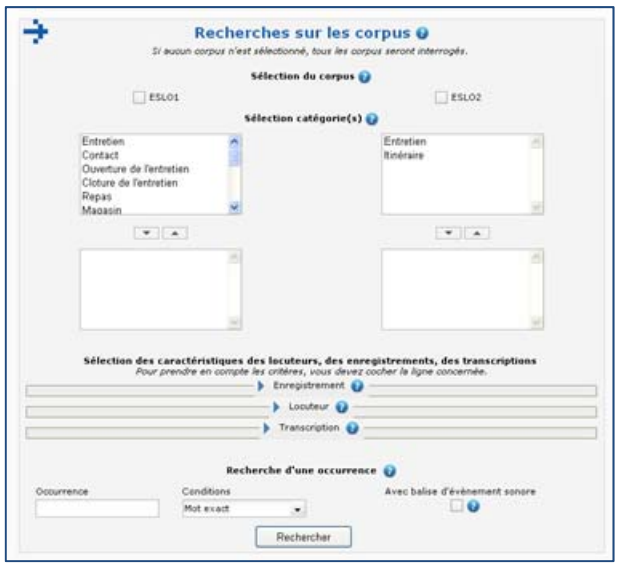


Figure 4. *Interface Web*

Pour des raisons juridiques, une partie non anonymisée des données n'est accessible qu'aux chercheurs et ce, après identification et convention (non diffusion et respect de la confidentialité) ce qui est le cas du corpus annoté des entités nommées et dénommantes. Les transcriptions rendues disponibles au grand public seront les troisièmes versions. Les utilisateurs auront d'ailleurs la possibilité de proposer des transcriptions alternatives avec des corrections.

Les données anonymisées sont également disponibles par l'entrepôt du CRDO-TGE Adonis et par le site Corpus de la parole de la DGLFLF – ministère de la Culture.

Dans le travail sur la consultation du corpus, nous avons été guidés par les différents profils d'utilisateurs potentiels avec la volonté de leur donner accès à des données variées et riches constituant ESLO.

6. Conclusion

En conclusion, nous rappelons quelques objectifs du projet ESLO qui ont guidé notre équipe tout au long du travail.

ESLO est un travail de longue haleine. Il s'agit tout d'abord de constituer un grand corpus de français parlé de quelque 700 heures en préservant prioritairement l'hétérogénéité maximale des données observées et contenant de nombreuses informations complémentaires (profil du locuteur, caractéristiques de la situation, etc.).

La mise à disposition de ce corpus a nécessité d'anticiper les contraintes juridiques et techniques qui ne peuvent être dissociées des cadres théoriques du projet de recherche. Les différentes opérations de traitement des données ont un impact fort sur la constitution de l'objet scientifique et ne peuvent être considérées comme des opérations de prétraitements des données en préalable au travail d'analyse. Ainsi ce projet aborde résolument une démarche réflexive. Il ne s'agit pas seulement de recueillir et de rendre disponibles des données et métadonnées langagières mais aussi de rendre explicite l'ensemble de la chaîne qui permet d'y arriver, de la collecte à l'analyse, en passant par la transcription et les autres opérations d'annotation. On se rend alors compte que toutes les opérations sont liées les unes aux autres, et que les choix qui sont faits doivent l'être en prenant en compte l'ensemble des objectifs et des contraintes, à tous les niveaux.

Le projet ESLO n'est pas achevé. Le corpus ESLO2 est en cours puisque des enregistrements d'Orléanais d'aujourd'hui dans des situations variées continuent. Les travaux sur l'annotation se poursuivent. Comme nous l'avons mentionné, nous sommes en train de travailler sur l'étiquetage morphosyntaxique par apprentissage automatique avec les CRF. Le nouveau modèle d'annotation sera appris à partir d'un corpus de référence annoté manuellement ainsi que de connaissances externes comme les résultats de l'étiqueteur libre SEM et des ressources linguistiques libres

comme le lexique du Lefff et les tables de verbes et de noms prédicatifs du Lexique-Grammaire, passées au format alexina du Lefff. L'étiquetage sera compatible avec les fichiers XML de Transcriber pour permettre la synchronisation avec les fichiers sonores. D'autres couches d'annotations sont programmées sur ESLO, comme l'annotation syntaxique en *chunks*[49], par exemple. On citera également le projet ANCOR[50] décrit dans (Schang *et al.,* 2011) consistant dans l'annotation et l'étude des coréférences à l'oral et se fondant majoritairement sur le corpus ESLO.

Cent vingt et un entretiens d'ESLO1 annotés des entités nommées et dénommantes ou un corpus de 1 723 « énoncés » (tours de parole ou séquences entourées de « blancs ») comportant 18 424 unités lexicales étiquetées morphosyntaxiquement représentent des ressources riches et très utiles pour les techniques d'apprentissage.

Le corpus ESLO peut être aujourd'hui exploité de manières diverses. Nous avons mentionné un exemple du travail qui a croisé l'information annotée avec les métadonnées. Il s'agit de l'analyse de la variation des valeurs numériques de temps dans les fichiers de transcription en fonction de différents profils sociologiques des locuteurs à l'aide des méthodes statistiques (pour plus de détails voir Dupont *et al.*, 2012). La richesse des données que permet d'étudier le corpus ESLO est importante. Sa mise à disposition permettra ainsi leur meilleure exploitation.

7. Bibliographie

Abney S., Parsing by chunks. In Berwick R., Abney R. et Tenny C., éditeurs : Principlebased Parsing. Kluwer Academic Publisher, 1991.

Baude O., *Corpus oraux : guide des bonnes pratiques*, CNRS-Editions et Presses universitaires d'Orléans, 2006.

Baude O., Dugua C. « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? », vol. 10, *Corpus, Varia,* 2011.

Bergounioux G., Baraduc J., Dumont C., « L'étude socio-linguistique sur Orléans (1966-1991): 25 ans d'histoire d'un corpus », n°93, *Langue française*, 1992, p. 74-93.

Blanche-Benveniste C., Jeanjean C., *Le français parlé, transcription et édition*, Paris, Didier érudition, 1987.

Cappeau P., Gadet F., « Où en sont les corpus sur les français parlés ? » *Revue Française de Linguistique Appliquée*, Vol. XII, 2007, p. 129-133.

49. Les *chunks* sont des constituants continus et non récursifs (Abney, 1991) qui définissent la structure syntaxique des phrases ou énoncés.

50. http://tln.li.univ-tours.fr/Tln_Ancor.html

Constant M., Tellier I., Duchier D., Dupont Y., Sigogne A., Billot S., « Intégrer des connaissances linguistiques dans un CRF : application à l'apprentissage d'un segmenteur-étiqueteur du français », *TALN2011,* Montpellier, 2011.

Cresti E., Moneglia M. *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*, Studies in Corpus Linguistics 15. Amsterdam : Benjamins, 2005.

Debrock M., Mertens P., Truyen F., Brosens V., *ELICOP, Étude LInguistique de la COmmunication Parlée: Constitution et exploitation d'un corpus de français parlé automatisé*, K.U.Leuven: Departement Linguïstiek, 2000.

Denis P., Sagot B. « Exploitation d'une ressource lexicale pour la construction d'un étiqueteur morpho-syntaxique état-de-l'art du français », *TALN2010*, Montréal, 2010.

Dister A,. De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données textuelles orales VALIBEL, Thèse de doctorat, Université catholique de Louvain, 2007.

Dupont A., Eshkol-Taravella I., Delsol L., « Étude d'application des méthodes et des outils statistiques sur les données du corpus ESLO : cas de la question sur mai 68 », *11es Journées Internationales d'analyse statistique des données textuelles JADT 2012,* Liège, 13-15 juin, 2012 (à paraître).

Eshkol I., « Entrer dans l'anonymat. Etude des "entités dénommantes" dans un corpus oral », *Eigennamen in der gesprochenen Sprache*, Narr Francke Attempto Verlag GmbH, Germany, 2010, p. 245-266.

Eshkol I., Tellier I, Taalab S., Billot S., « Étiqueter un corpus oral par apprentissage automatique à l'aide de connaissances linguistiques », *10es Journées Internationales d'analyse statistique des données textuelles JADT 2010*, Rome, 9-11 juin, 2010.

Friburger N., Reconnaissance automatique des noms propres ; application à la classification automatique de textes journalistiques, Thèse de doctorat, Université François Rabelais Tours, 2002.

Hriba L., Identification automatique des locus de variation dans un corpus de français parlé, Thèse de doctorat. Université d'Orléans, Orléans, en cours.

Kaufmann A., « The Santa Barbara Corpus of Spoken American English. Part 1 », *Journal of Pragmatics* 34, 2002, p. 1309-1316.

Lafferty J., McCallum A., Pereira F., « Conditional random fields : Probabilistic models for segmenting and labeling sequence data », *Proceedings of ICML'01*, 2001, p. 282–289.

Leech G., « Introduction corpus annotation ». In Garside R., Leech G., McEnery A., (Eds.), *Corpus annotation: Linguistic information from computer text corpora*. London: Longman, 1 :18, 1997.

Lonergan J., Kay J., Ross J., *Etude sociolinguistique sur Orléans, catalogue des enregistrements*, Colchester: Multigraphié, 1974.

Maurel D., Friburger N., Antoine J.-Y., Eshkol-Taravella I., Nouvel D., Cascades autour de la reconnaissance des entités nommées, *TAL* 52-1, 2011.

Maurel D., Friburger N., Eshkol I., « Who are you, you who speak? Transducer cascades for information retrieval », *4th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland, 6-8 novembre, 2009, p. 220-223.

Mertens, P., « Les corpus de français parlé ELICOP : consultation et exploitation », in Binon, J.*, et al.* (éd.) *Tableaux Vivants. Opstellen over taal-en-onderwijs aangeboden aan Mark Debrock*. Leuven: Universitaire Pers., 2002.

Mettouchi, A., C. Chanard, « From Fieldwork to Annotated Corpora: the CorpAfroAs Project », *Faits de Langue-Les Cahiers*, 2 : 255-265, 2010.

Nazarenko, A., « Le point sur l'état actuel des connaissances en traitement automatique du langage (TAL) », *Compréhension des langues et interaction*, Lavoisier, 2006, p. 31-70.

Paumier S., De la reconnaissance de formes linguistiques à l'analyse syntaxique, Thèse de Doctorat, Université de Marne-la-Vallée, 2003.

Schang E., Boyer A., Muzerelle J., Antoine J-Y., Eshkol I., Maurel, D. « Coreference and Anaphoric Annotations for Spontaneous Speech Corpora In French », *The 8th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC2011)*, Faro, Algarve, Portugal, 6-7 October, 2011.

Sutton C., McCallum A., « An Introduction to Conditional Random Fields for Relational Learning », In L. Getoor and B. Taskar, Eds., *Introduction to Statistical Relational Learning*. MIT Press, 2006.

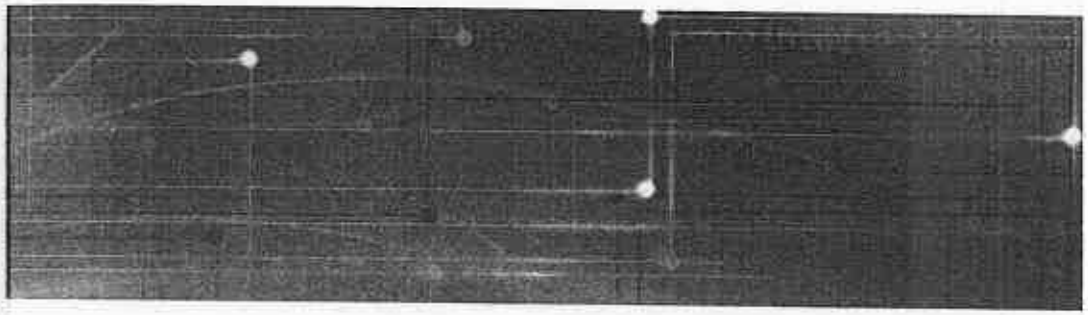
Tellier I., Tommasi M., « Champs Markoviens Conditionnels pour l'extraction d'information », dans *Modèles probabilistes pour l'accès à l'information textuelle*, Hermès, 2011, p. 223-267.

Titre	***Corpus oraux, Guide des bonnes pratiques 2006 (édition coréenne)***
Type	Livre
Editeur	Korea Pagijong Press
Année	2012
Référence	Baude, O., coord. (2012) *Corpus oraux, guide des bonnes pratiques*, Korea Pagijong Press. Version coréenne du guide Baude, O., coord. (2006) *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

인문자식기반총서 2

구어 말뭉치
실용 안내서

CORPUS ORAUX
Guide des bonnes pratiques



올리비에 보드 엮음
손현정 · 손희연 옮김

프랑스의 언어들 및 프랑스어 심의위원회(DGLFLF)
http://www.dglflf.culture.gouv.fr

ISBN 2-271-06425-2(프랑스 국립연구센터 발행)
ISBN 2-913454-30-5(PUO)
EAN 9782271064 257(프랑스 국립연구센터 발행)
EAN 9782913454 309(PUO)

오를레앙 대학 출판부/프랑스 국립연구센터 발행

이 책은 **이자벨 드 랑베르트리(Isabelle de Lamberterie)**가 주도하는 연구 그룹의 공동 창작물이다. 내용을 모아 엮는 역할은 **올리비에 보드(Olivier Baude)**가 맡았다.

올리비에 보드(Olivier Baude)(DGLFLF, CORAL-오를레앙 대학)
클레르 블랑쉬 벤베니스트(Claire Blanche-Benveniste)(EPHE, 프로방스 대학)
마리 프랑스 칼라스(Marie-France Calas)(DMF)
폴 카포(Paul CAPPEAU)(프와티에 대학)
파스칼 코르드레스(Pascal Cordereix)(BnF)
로랑스 구리(Laurence Goury)(CNRS-CALIA)
미셸 자콥슨(Michel Jacobson)(CNRS-LATICO)
이자벨 드 랑베르트리(Isabelle de Lamberterie)(CNRS-CECOJI)
크리스티안 마르셸로 니지아(Christiane Marchello-Nizia)(CNRS-ILF, ENS-LSH-리옹)
로렌자 몬다다(Lorenza Mondada)(ICAR, CNRS, 리옹2대학)

그 외의 참여자:
질 아다(Gilles ADDAS)(COPTE LIMSI-CNRS), **미셸 알레지오(Michel ALESSIO)**(DGLFLF), **알랭 카루(Alain CAROU)**(BnF), **이브라힘 쿨리발리(Ibrahim COULIBALY)**(CDF-그르노블 대학), **발레리 감므(Valérie GAME)**(BnF), **파브리스 몰로(Fabrice MOLLO)**(CNRS-CECOJI), **미셸 레이날(Michel RAYNAL)**(INA), **장 시빌(Jean SIBILLE)**(DGLFLF), **도미니크 테롱(Dominique THERON)**(BnF), **뤽 베리에(Luc VERRIER)**(BnF)

저자 소개

❑ 올리비에 보드(Olivier Baude)

오를레앙 대학교 언어학과 교수. 인류학과 언어학 연구센터(Centre Orléanais de Recherche en Anthropologie et Linguistique)(EA-3850) 연구원. 언어 행태 관측소(Observatoire des pratiques linguistiques) 학술 이사.

❑ 클레르 블랑쉬 벤베니스트(Claire Blanche-Benveniste)

프랑스고등연구원(EPHE, École Pratique des Hautes Études à Paris)과 프로방스 대학교 명예교수. 연구 분야: 구어와 문어, 통사론, 형태론, 구어 말뭉치 구축.

❑ 마리 프랑스 칼라스(Marie-France Calas)

문화재청장. 박물관 감사위원회 위원장. 프랑스 박물관 위원회 위원장. 역사, 경영, 구어 자료, 음악 자료, 자연의 소리 자료, 즉 오늘날 무형 유산에 속하는 대상들에 대한 보존과 평가를 포함한 학제간 영역의 소리 분야 전문가.

❑ 파스칼 코르드렉스(Pascal Cordereix)

도서관 정사서. 프랑스 국립 도서관의 시청각과 소리 자료 서비스 책임자. 음향 및 시청각 자료 소지자 프랑스 협회(AFAS, Association française des détenteurs de documents audiovisuels et sonores)의 부회장. 이 단체의 주요 활동은 소리의 기록 보관에 관한 것이다.

❑ 로랑스 구리(Laurence Goury)

발달 연구소(IRD, Institut de Recherche pour le Développement)의 책임연구원. 아메리카 인디언 언어 연구센터(CELIA, Centre d'Etude des Langues Indigènes d'Amérique) 연구원, 연구 분야: 현장언어학과 유형론(크레올 언어 전공).

❑ 미셸 자콥슨(Michel Jacobson)

프랑스 국립연구센터(CNRS, Centre National de la Recherche Scientifique)의 구술 전통 언어 문화 연구소(LACITO, Laboratoire des langues et civilisation à tradition orale) 소속 컴퓨터 엔지니어. '아카이브' 프로그램의 공동 책임자. 구어 말뭉치 운영 전문가.

❑ 이자벨 드 랑베르트리(Isabelle de Lamberterie)

프랑스 국립 과학원 책임연구원. 국제 사법 협력 연구센터(CECOJI, Centre d'études sur la coopération juridique internationale)(UMP 6224)의 '정보사회와 규범성' 연구팀 책임자. 프랑스 국립과학연구센터 윤리위원회 위원.

❑ 크리스티안 마르셀로 니지아(Christiane Marchello-Nizia)

리옹 고등사범학교(ENS-LSH, École Normale Supérieure) 언어학과 교수, 프랑스 국립연구센터 소속 프랑스 언어학 연구원(Institut de Linguistique Française) 원장. 연구 분야: 언어사, 프랑스어의 역사, 언어 발달 이론.

❑ 로렌자 몬다다(Lorenza Mondada)

리옹 2대학 언어학과 교수. ICAR(UMP CNRS 5191) 연구원. 연구 분야: 상호작용 구어말뭉치에서 언어적 상호작용 연구. 영상 말뭉치에서 다면적 분석 연구.

서문

자비에 노스(XAVIER NORTH)
프랑스의 언어들 및 프랑스어 심의위원회 위원

문화 정책과 과학의 역사에서 가공되지 않은 모든 정보와 확신할 수 없는 실체들이 지식의 대상이 되었던 순간은 그리 많지 않다. 이러한 정보와 실체에서 비롯된 본 안내서가 출판되면서, 이제 모든 연구자들은 유용한 도구, 즉 '실용 안내서'를 지니고 '전환'을 이끌어낼 시점에 이르게 된다. '전환'은 곧 사람들의 말을 구어 말뭉치로 변화시키는 것이다. 말은 이제 분석되고 보존되며 나아가 한 나라의 문화유산으로 자리할 수 있다.

'글'의 형태로 등장하는 문학 작품, 사료와 같은 언어적 생산물들은 고정되어 있고 결정적인 것으로서, 도서나 아카이브와 관련된 문화부 정책의 중심에 있어 왔다. 아주 최근에 이르러서야, 바로 터져 나오는 말, 일상적이고 보편적인 발화, 놀라우리만치 다양한 말투 등에 녹아 있는 살아 있는 언어활동에 관심을 두기 시작한 것이다. 안정된 기반을 가지고 진정한 '구어 아카이브'가 구축될 가능성은 이제 처음으로 열리기 시작했다.

구어 말뭉치는 인류의 구어를 기록하여 모아둔 단순한 소장품을 넘어선다. '건설'을 필요로 하는 것인데, 다시 말해 가공(디지털화, 전사, 색인 목록화)이 요구된다. 가공을 통해 말뭉치가 구축되면, 이제 구어는 보존되고, 연구의 대상이 되거나 새로운 가치를 창출하는, 말 그대로 새로운 지위를 가지게 되는 것이다. 이는 물론 사용하기에 쉽고 일관성 있는 방법을 구상하여 이를 바탕으로 이루어져야 한다.

본 '실용 안내서' 덕분에 연구자들에게는 드디어 새롭고 거대한 연구 분야가 펼쳐지게 되었다. 또한 '언어 행태 관측소'의 참여를 바탕으로 '프랑스의 언어들 및 프랑스어 심의위원회'도 출발의 동력을 제공할 수 있었고, 나아가 본 안내서 편찬을 위해 문화부의 다른 행정 부서와 관련 연구 분야에서 동원되었던 인력과 자원들을 재분배하고 조정할 수 있었다.

구어 말뭉치의 발전, 배포 및 보존을 확실히 실행하는 것은 곧 다양하고 풍부하며 생명력이 넘치는 프랑스어 문화 유산를 개방하는 것이고, 알리는 것이다. 또한 언어 정책 및 사회, 교육 정책을 수립하는 데에 필요한 실제적 언어 지식이라는 소중한 도구를 갖추는 것이기도 하다.

수개월 동안 법률가, 언어학자, 자료 관리자, 전산 기술자 등이 모여 법을 존중하는 테두리 안에서 피어나는 문화와 연구의 새로운 여정을 공유하고 개척해 나갔다. 본 안내서는 이렇게 수고로운 공동 사유의 결과물이다. 이제 본 안내서가 풍부한 연구들이 생산될 수 있는 토대가 되기를 바란다.

서문

베르나르 뫼니에(BERNARD MEUNIER)
프랑스 국립연구센터 원장

구어와 문어, 두 단어는 굉장히 많은 것들을 떠오르게 한다. 문화는 우선 구어적 실제로 구성된다. 문자를 통해 구성되는 것은 그 이후의 일이다. 문자는 주어진 공간과 시간 속에서 효과적인 전승, 전달을 가능하도록 하기 위해 창조된 것이다.

연구자로서 학술 지식의 전파를 위한 구어와 문어 각각의 역할 모두를 염두에 두고 있다. 그러나 학술적 생각들을 전파하고 설득시키고 또 공유하는 데에 있어, 글의 역할 또한 중요하지만, 동료나 대중들 앞에서 이루어지는 구두 발표가 글을 넘어서는 본질적인 것이라는 사실을 잊지 않고 있다. 구어는 설득력을 지닌다. 현재의 음성영상 기술을 통해 녹음되고 전파된다면 구어의 설득력은 배가될 것이다.

문어 말뭉치가 그러하듯이 구어 말뭉치의 구축과 활용에도 실용적 안내가 필요하다. 하나의 문장이 맥락을 떠나서 아무렇게나 전파될 경우, 말하는 사람, 단체, 공동체 등을 얼마나 위협할 수 있는 것인지, 우리는 잘 알고 있다.

본 안내서의 저자들은 문어 말뭉치를 사용하고 구축하는 데에 관련되는 법적인 문제들을 심도 있게 다루어 왔다. 이 책이 구어 말뭉치의 주체이자 사용자들, 우리 스스로도 언제든 이러한 주체이자 사용자일 수 있는데, 그들에게 최상의 조건에서 보급되기를 바란다.

서문

장노엘 장느네(JEAN-NOEL JEANNENEY)
프랑스 국립 도서관 관장

프랑스 국립 도서관이 본 안내서를 만드는 데에 도움이 될 수 있어서 기쁘다. 프랑스 국립 도서관은 구두적 사용 언어들의 보존과 전파에 힘쓰면서 이러한 구어와 오래되고 밀접한 관계를 유지해 온 것이 사실이다. 도서관 내의 멀티미디어과는 1911년 페르디낭 브뤼노(Ferdinand Brunot)가 만든 '구어 기록 보관소'의 전통을 이어 오고 있다. 1911년 이후로 프랑스 국립 도서관은 모든 종류의 구어적 표현물들을 수집하고 보존하는, 그리고 가장 많은 대중들에게 이를 공개하기 위한 최상의 조건을 구비하기 위해 끊임없이 노력해 왔다.

오늘날 디지털 기술은 역사와 학술의 결속을 더욱 강화하고 있다. 자료 보존의 측면에서, 본 도서관의 소장 자료를 디지털화하는 야심찬 계획이 추진 중이고 특히 음향 및 음성영상 자료들이 주요한 대상이 되고 있다. 한편, 온라인 디지털 도서관 시스템 '갈리카(Gallica)'의 비약적인 발전의 결과, 사용자들은 본 도서관의 풍부한 소장 자료를 도서관 안팎에서 검색할 수 있다. '갈리카(Gallica)'는 인터넷으로 접속한 모든 사용자들이 어디에 있건, 또 무엇을 찾거나 궁금해하건, 본 도서관이 제공하는 중요한 지식 원천에 접근할 수 있도록 해 준다.

신뢰를 바탕으로 한 공동 작업의 결과로서, 본 안내서는 언어학자, 법률가, 자료 관리자, 전산기술자, 음성 및 영상 기술자들이 서로의 지식을 보완하여 채워나가는 모습을 보여준다. 다시 한 번, 프랑스 국립 도서관이 이러한 혁신적이고 풍부한 작업에 동참할 수 있어 기쁘게 생각하는 바이다.

목차

3. 연구의 전개 과정

4. 구어 말뭉치, 문화유산으로서의 성격

5. 부록

1
서론

1.1 구어 말뭉치 실용 안내서의 편찬 목적

최근 '구어 말뭉치' 즉 구두적 음성 혹은 이미지, 몸짓 등을 포함한 다면적 경로로 생산된 언어를 녹음하거나 녹화하여 편집한 자료를 활용한 기초 연구나 응용 연구가 대량으로 생산되고 있다. '실용 안내서'의 기능을 하는 이 책은 언어 자료의 원천을 보존하면서도 그러한 원천에서 생산된 구두적 자료에 다양한 방식으로 접근하려는 여러 언어학자들의 문제의식에 바탕을 두고 있기 때문에 무엇보다도 언어학자들에 의해 그리고 그들을 위해 만들어지고 활용되는 '구어 말뭉치'를 다룬다. 그러나 구어 말뭉치를 구축하고 활용하는 것은 여러 학문 분야에 걸쳐 제기되는 다양한 의문들을 동반할 수밖에 없다. 그 학문 분야들은 예를 들어 민속지학, 인류학, 사회학, 심리학, 인구통계학 등이 있고 또한 일상 구술, 인터뷰, 증언, 구두적 조사 등의 방법을 주로 사용하는 구전 역사와도 관련된다. 따라서 이 안내서는 언어학의 방법론을 바탕에 두고 있으면서도 구어 말뭉치를 사용하는 다른 분야의 학자들, 예를 들어 발화 인식과 합성에 구어 말뭉치를 사용하는 학자들이 가지는 관심사에 접근해 있기도 하다. 그러나 이러한 학자들이 지니고 있는 특정 학문적 요구가 이 책에서 전면적으로 다루어질 수 없는 것은 사실이다.

Titre	***Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ?***
Type	Actes
Editeur	Peter Lang
Année	2015
Référence	Baude, O. Dugua,C. (2015) « Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ? » in Dostie, G., Hadermann, P ., *La dia-variation en français actuel*, collection "Sciences pour la communication", Peter Lang ed, Berlin, pp 349-372

Usages de la liaison dans le corpus des ESLOs : vers de nouveaux (z)ouvrages de référence ?

Olivier Baude
Université d'Orléans, France

Céline Dugua
Université d'Orléans, France

1. Introduction

Les ESLOs (Enquêtes sociolinguistiques à Orléans) constituent un corpus exceptionnel pour observer la variation à la fois en synchronie et en diachronie. Nous nous intéresserons ici à un phénomène classiquement utilisé dans les études variationnistes et largement présent dans les ouvrages de référence : la liaison. L'étude de l'usage des liaisons au sein d'un corpus de français parlé est alors l'occasion de questionner les conditions d'analyses fondées sur des corpus « de référence » et de les confronter aux descriptions présentes dans les grammaires et les méthodes de langue.

Dans cet article, nous présentons tout d'abord les principales caractéristiques de la liaison telles qu'elles sont décrites mais aussi prescrites dans différents ouvrages de références et telles qu'elles apparaissent dans les études récentes réalisées sur corpus.

Nous proposons ensuite une première série d'analyses reposant sur l'étude des variations de réalisation des liaisons au sein d'un ensemble d'extraits du corpus des ESLOs. En confrontant des analyses qualitatives d'extraits significatifs abordés sous différents angles théoriques aux connaissances actuelles contenues dans les ouvrages de référence et les travaux récents, nous souhaitons décrire et contrôler les effets

méthodologiques d'une linguistique qui se fonde sur l'interprétation de données situées.

2. La liaison dans les ouvrages de référence et les corpus

2.1 La liaison : repères théoriques

La liaison est une forme particulière d'enchainement qui se manifeste à la frontière de deux mots. Dans sa description traditionnelle « [Elle] consiste à faire entendre devant un mot commençant par une voyelle une consonne finale normalement muette » (Chevalier *et al.* 1964). Plus récemment selon Côté (2005 : 66) « La liaison correspond à la prononciation entre deux mots d'une consonne qui n'apparait dans aucun de ces mots prononcé dans d'autres contextes ». Des contraintes formelles interviennent dans la définition des contextes de liaison : le premier mot doit se terminer par une consonne graphique muette, c'est-à-dire non présente lorsque le mot est produit en isolation (*petit* [pti]) ou devant un mot à initiale consonantique (*des tables* [detabl]), le deuxième mot doit commencer par une voyelle. La définition des contextes dans lesquels les liaisons peuvent ou doivent se réaliser, ainsi que l'étude des facteurs de leur réalisation font intervenir plusieurs niveaux linguistiques (morphosyntaxe, lexique, orthographe) ainsi que des paramètres sociaux et contextuels.

De nombreux travaux s'attachent au phénomène de liaison en français ; ils se répartissent dans des domaines de recherche divers. De manière non exhaustive, nous pouvons citer certains de ces domaines. La liaison est un processus généralement décrit comme principalement phonologique ; elle a d'ailleurs très souvent servi de phénomène test pour l'expression des représentations phonologiques (Schane 1967 ; Encrevé 1988 ; Tranel 1996 ; Côté 2005). Son évolution diachronique en lien avec les évolutions de l'orthographe trouve place dans les travaux de linguistique historique (Clédat 1917 ; Fouché 1952 ; Bourciez et

Bourciez 1971). Au-delà des traitements purement phonologiques, des auteurs ont montré, par exemple, que la liaison en /z/ peut être traitée comme un marqueur morphologique du pluriel (Morin et Kaye 1982 ; Morin 2003 [1998]). Des critères de groupes syntaxiques et sémantiques (Laks 2005), comme des critères de groupes prosodiques (Grammont 1914 ; De Jong 1990) sont utilisés pour délimiter des contextes de liaisons possibles. Les chercheurs dans le domaine du traitement cognitif du langage s'appuient également sur la liaison pour cerner des phénomènes qui lui sont liés, tels la segmentation du lexique et le traitement *on-line* de l'accès lexical (Spinelli, Cutler et McQueen 2002 ; Spinelli et Meunier 2005). Depuis les années 2000, les recherches en acquisition s'efforcent de décrire et formaliser les étapes de mise en place de ce phénomène chez les jeunes enfants en langue première (Chevrot, Dugua et Fayol 2009 ; Nardy et Dugua 2011 ; Wauquier et Braud 2005) et en langue seconde (Harnois-Delpiano *et al.* 2012 ; Wauquier 2009). Certains travaux de sociolinguistes (Encrevé 1988 ; Gadet 2003) utilisent la liaison en tant que phénomène de variation pour préciser, par exemple, les facteurs extralinguistiques influant sur la réalisation ou la non réalisation de cette unité phonologique variable.

2.2 Traitement de la liaison par les ouvrages de référence

Outre ces descriptions linguistiques, les ouvrages de référence (grammaires, dictionnaires) évoquent systématiquement la question de la liaison, généralement en deux temps : une définition factuelle et une présentation des classements des contextes de liaisons en trois catégories : liaison obligatoire (invariante), facultative (variable, recommandée), interdite (impossible, inusitée, erratique, abusive). Deux remarques méritent d'être soulignées quant aux classements proposés. Premièrement, ces classements ne sont pas absolus, certains contextes prennent place dans des catégories différentes d'un ouvrage de référence à l'autre. Un exemple typique, le contexte entre l'auxiliaire *être* ou *avoir* et le participe passé ; considéré par Fouché (1959) comme un contexte dans lequel « on fait la liaison », il appartient aux liaisons obligatoires chez

des auteurs s'inscrivant dans des approches normatives (Dupré 1972), comme dans des grammaires descriptives (Riegel, Pellat et Rioul 1994) et enfin, il intègre la catégorie des liaisons facultatives pour Grevisse (1988) et Delattre (1966). Notons pour ce contexte un intérêt tout particulier de la part de Dupré (1972) qui souligne à plusieurs reprises dans son passage sur la liaison, la tendance à faire évoluer ce contexte dans l'usage, et en substance son regret d'un tel mouvement.

> La liaison des auxiliaires terminés par *t* avec le participe passé qui suit était régulière et conforme à la logique. Or par un phénomène curieux, cet usage traditionnel est en train de se perdre. On entend constamment des hiatus comme ceux-ci : *deux guerres mondiales nous on (t-) appris…*, *ils on (t) été renvoyés au dépôt*. (Dupré 1972 : 1469)

Cet auteur conclut l'article en revenant sur cette observation pour en souligner l'importance « […] Insistons sur la nécessité de la liaison entre auxiliaire et participe : *ils étaient-t-allés*. ». Cette hétérogénéité observée peut s'expliquer par des différences d'interprétations des critères classiquement utilisés pour catégoriser les contextes de liaison, tels que la cohérence syntaxique entre les deux mots faisant liaison et l'intuition sur l'usage qui en est fait ou sur les règles à mettre en œuvre. Soulignons à ce propos que les ouvrages de référence ne fournissent pas les sources sur lesquelles ils s'appuient pour organiser leurs classements. C'est notamment le signe d'un manque de relation entre ouvrages de référence et théories linguistiques.

Une deuxième raison permettant d'expliquer la perméabilité des classements est la prise en compte des variations sociales et stylistiques, comme le souligne d'ailleurs Delattre (1966 : 43).

> Il est évident que les distinctions « obligatoires, facultatives, interdites » ne sont pas absolues. Elles varient selon le style. Dans les exemples qui suivent, elles se rapportent, autant que possible, au style de la conversation soignée courante – style encore variable selon les individus et le milieu où ils se trouvent.

Se pose évidemment la question de savoir ce que l'on entend par conversation soignée courante. Nous comprenons l'intention de Delattre, mais nous devons aujourd'hui affiner la définition des sources de variations

pour décrire et analyser l'usage de la liaison. Les corpus recueillis actuellement sont un des outils qui devrait permettre de mieux décrire et cerner ces sources de variations.

2.3 Corpus et liaison

Les études sur la liaison nécessitent le recours à l'observation des pratiques et donc a fortiori à la méthodologie d'enquête. En effet, la liaison est un phénomène linguistique repéré comme porteur d'une forte identité de marqueur social et comme le lieu de variations régulières. Comme le souligne Encrevé (1983 : 42) :

> … la liaison est un indicateur social explicite, un des rares lieux de la langue où les plus anti-variationnistes des linguistes ont été amenés à reconnaitre la variation sociale et l'hétérogénéité linguistique.

Cette reconnaissance de la variation se retrouve également dans les ouvrages de référence les plus normatifs qui proposent systématiquement une catégorie de liaisons facultatives sans toutefois donner les clefs de l'usage social de ces variations.

De fait, bien que soumise à de grandes régularités, l'analyse des variations de la liaison nécessite une méthodologie d'enquête quantitative. L'apport des corpus pour étudier l'usage de la liaison est alors évident. Comment, sans ceux-là, apporter des réponses à des questions cruciales, comme celles posées par Laks (à paraitre) qui font écho aux classements avancés dans les ouvrages de référence ?

> la question centrale concernant la liaison en français reste celle de la description des différents usages attestés : quelles sont les liaisons systémiquement catégoriques, pour quels locuteurs et dans quelles circonstances ? Quelles sont les liaisons systémiquement variables, selon quelles fréquences et dans quelle organisation diastratique et diaphasique.

Dès le début des années 1970, Ågren (1973), à partir de 40 heures d'enregistrements radiophoniques impliquant des journalistes, des hommes politiques et des écrivains, fournit un premier travail sur la liaison à

partir d'un corpus de parole spontanée. Laks (1980) dans son travail de thèse a constitué un corpus d'adolescents de banlieue parisienne et a ainsi marqué un tournant dans le champ de la sociolinguistique en France. A la même période, Encrevé (1988) a rassemblé les paroles d'hommes politiques afin de rendre compte, notamment, d'un nouvel usage de la liaison : la liaison sans enchainement. Par la suite, d'autres corpus ont été constitués avec pour objectif premier l'étude de la liaison, citons évidemment PFC[1] (Durand *et al.* 2002 ; Durand *et al.* 2011), mais également plus récemment le projet ALIPE[2] (Liégeois *et al.* 2011) ou HPOL[3] (Laks 2007). Dans ces corpus, les contextes de liaisons sont repérés et les réalisations / non réalisations codées. Plus précisément, dans le corpus PFC par exemple, il est possible d'obtenir des résultats généraux facilement, mais aussi d'affiner ces derniers en sélectionnant des critères sociodémographiques (âge, sexe des locuteurs), situationnels (lecture, discussion guidée, discussion informelle), géographiques et de type morphosyntaxique. Les choix de PFC sont clairement énoncés dans leurs publications de présentation du projet : le codage est systématique et ne s'appuie sur aucune classification préalable telle que celles définies dans les ouvrages de références antérieurs. La méthodologie de corpus permet ici de s'affranchir de cadres théoriques afin de réaliser une première étape purement descriptive de données attestées. Par la suite, l'analyse de ces données met explicitement en question ces classifications (Durand et Lyche 2008).

Le projet ALIPE, quant à lui, fournit un contexte particulier, celui des interactions parents-enfants, et permet d'observer à la fois l'usage de la liaison chez les parents, lorsqu'ils s'adressent à l'enfant ou se parlent entre eux, et chez les enfants (âgés entre 2 et 4 ans).

Les autres corpus de français parlé disponibles – citons par exemple CFPP2000, Valibel, CFPQ, OFROM, Clapi, CoLaJE, ESLO – constituent des bases de données exploitables pour étudier la liaison mais ne

1 Phonologie du Français Contemporain : <http ://www.projet-pfc.net/>.

2 Acquisition de la liaison dans des interactions parents-enfants : <http://lrlweb.univ-bpclermont.fr/spip.php?article282>.

3 Corpus d'hommes politiques, en situations de paroles publiques accessible à partir du site PFC.

fournissent pas d'outil de repérage et d'exportation des contextes de liaison. La mise en perspective des différentes bases de données devrait permettre d'une part, de disposer d'une masse de données importante et d'effectuer des analyses quantitatives et, d'autre part, d'avoir accès à des situations de communication et des locuteurs variés.

Toutefois, l'annotation des corpus disponibles ne répond pas à une méthodologie commune et partagée qui fournirait par exemple un codage interopérable entre les différents projets. Pour une part, cette limite est due aux outils utilisés pour la transcription/annotation (Praat, TEI, CLAN, Transcriber, etc.) et aux conventions adoptées qui ne sont pas nécessairement interopérables. D'un autre côté, il pourrait s'avérer compliqué de généraliser un codage sur la liaison à l'ensemble des corpus, sachant qu'un nombre restreint d'entre eux s'intéresse vraiment à ce phénomène.

3. Corpus ESLO

3.1 ESLO : un « nouveau » corpus pour la liaison ?

Dans le cadre de ce travail, nous nous appuyons sur le corpus ESLO, élaboré au sein du laboratoire ligérien de linguistique (LLL) de l'Université d'Orléans. Ce corpus comprend deux enquêtes.

ESLO1, réalisée par une équipe britannique (plus exactement franco-britannique) à la fin des années 1960, avait pour objectif de fournir un corpus représentatif du français tel qu'il est parlé, dans le but premier de constituer une méthode de langue fondée sur des documents authentiques. Ce projet a permis la réalisation du portrait sonore de la ville d'Orléans sur la base de 470 enregistrements variés correspondant à plus de 300 heures.

Quarante ans plus tard, le LLL a entrepris un double projet : diffuser largement le corpus ESLO1 dans un format correspondant aux outils de traitement des données actualisés et réaliser un nouveau corpus

représentatif du français parlé à Orléans dans les années 2010 (désormais ESLO2), en prenant en compte l'expérience d'ESLO1 et l'évolution des cadres théoriques et méthodologiques de la constitution et de l'exploitation de grands corpus oraux à visée variationniste. D'autres enquêtes font partie du corpus ESLO, citons LCO (Langues en Contact à Orléans) qui consiste à décrire les langues parlées dans l'agglomération, et ESLO-DIA(chronie) qui correspond à des entretiens menés auprès de sept locuteurs enregistrés à quarante années d'intervalle, enquête sur laquelle nous reviendrons par la suite.

A terme, ESLO comprendra quelque 700 heures d'enregistrements (approximativement 10 millions de mots), transcrits en trois versions, et rendus disponibles sur le site <http://eslo.tge-adonis.fr/>. Au-delà de ces caractéristiques quantitatives, la spécificité du projet consiste en la volonté de maitriser l'ensemble des étapes repérées dans la constitution d'un corpus, depuis les collectes et les prises de contact à l'analyse. Cette maitrise de l'ensemble des étapes s'entend comme la volonté de mettre en perspective les différents choix qui doivent être faits au fil de l'avancée du projet. Ainsi, par exemple, les transcriptions proposées reposent sur des conventions minimales et peu porteuses d'enjeux théoriques et d'interprétations.

3.2 ESLO : du réservoir de données au corpus d'études

ESLO est donc conçu comme grand corpus au sens de réservoir de données (Habert 2000). Il fournit une masse de données langagières, à partir de laquelle tout type d'étude linguistique est possible, directement, ou nécessitant sélection, transcription, annotation, codage, de la part du chercheur. La richesse du corpus des ESLO permet de sélectionner des sous-corpus d'études selon différents critères. En effet, dès ESLO1, les auteurs du corpus avaient apporté une grande attention à la documentation et au catalogage des documents afin de fournir des données « situées ». Tous les enregistrements sont donc documentés par des métadonnées précises qui concernent le locuteur, le contexte d'enregistrement et l'exploitation scientifique des données sonores et des transcriptions.

Le parti pris « variationniste » des auteurs d'ESLO1 et d'ESLO2 se concrétise donc dans l'architecture du corpus afin de permettre la prise en compte de différents types de variations.

La méthodologie d'échantillonnage d'un panel de locuteurs et le recours à des entretiens biographiques rendent possibles des analyses fines de la variation diastratique. Dans ESLO1, les entretiens concernent 147 locuteurs sélectionnés selon les critères classiques de la sociolinguistique (sexe, âge, catégorie socio-professionnelle) mais aussi selon une méthodologie novatrice (cf. infra partie *variations et trajectoires*). C'est sensiblement la même méthodologie qui a été utilisée pour 120 locuteurs dans ESLO2.

La collecte de diverses situations discursives représentant une typologie de genres de situations communicatives et sociales apporte les conditions d'analyse de la variation diaphasique. ESLO1 est composé majoritairement d'entretiens, mais dès l'origine du projet, il a été prévu de couvrir d'autres situations[4]. ESLO2, prenant en compte l'apport des théories sur les genres et le développement de l'analyse conversationnelle (CA) est élaboré sur une plus grande diversité[5].

Enfin, la réalisation de deux corpus à quarante ans d'intervalle et l'existence d'un sous-corpus de sept locuteurs présents dans les deux corpus offrent des données inédites pour l'étude de la variation diachronique.

Ces différents aspects ont été conservés dans la structure du corpus numérique ainsi que dans les outils qui permettent l'exploitation des données. Ainsi le croisement des données et des métadonnées permet l'extraction de sous-corpus d'études selon des critères très précis (sur le locuteur et sur le type d'enregistrement) afin de permettre la manipulation de données massives mais finement situées.

4 Au total, ESLO1 comprend : 157 entretiens, 79 situations informelles, 51 communications téléphoniques, 46 interviews de personnalités, 29 conférences débat, 84 enregistrements divers dont certains réalisés en micro caché et 41 consultations au centre médico-psychopédagogique.

5 22 situations prévues pour la première phase.

3.3 Méthodologies pour une étude exploratoire

Notre démarche de construction de corpus est résolument réflexive. Aussi il nous apparaît essentiel de contrôler les effets induits voire construits par la méthodologie développée lors de la constitution d'un corpus.

Si l'architecture même du corpus (*corpus design*) des ESLO2 a été conçue dès l'origine pour être comparable avec celle des ESLO1, il convenait de s'appuyer sur des analyses fines du premier corpus afin de mieux maîtriser les effets de transformation de l'objet que peut entraîner une telle méthodologie quantitative.

Nous avons donc souhaité, dans une démarche exploratoire et préparatoire, procéder à des micro-analyses à partir de différents axes d'approche du corpus afin de confronter des résultats partiels à de grandes tendances. Nous avons interrogé le corpus sur plusieurs extraits, délimités par des critères variés selon une méthodologie de « carottage ». Le terme carottage est utilisé dans un sens métaphorique pour définir une méthodologie empruntée à d'autres disciplines et qui consiste à analyser des échantillons du corpus à travers les différentes strates qui le constituent. Nous considérons en effet chacune de ces micro-analyses fondée sur des échantillons (carottes), comme une possibilité de sonder un grand corpus non pas à différents « endroits » mais plus exactement selon différents angles et par là, différentes perspectives théoriques.

Il s'agit ici de ponctionner quelques carottes à partir de critères issus de postulats variationnistes selon lesquels les « données » ne sont jamais « données » (Encrevé, 1976 : 13) mais relèvent à la fois de contraintes méthodologiques et des outils d'observation et d'analyse mobilisés pour la recherche. Nous présenterons ici trois carottes, autour de trois approches de la variation diachronique, diastratique et diaphasique.

Au sein de ces différentes micro-analyses, nous nous limiterons à l'observation de deux contextes particuliers de liaison *c'est* + *X* et *il est* + *X*[6]. Pourquoi ceux-là ? Ces contextes ne semblent pas susciter

6 Nous avons fait le choix dans cette étude de ne pas intégrer dans nos données la forme particulière *c'est-à-dire* qui peut être considérée comme une expression

un accord unanime entre les différentes sources de classements : ils sont en effet généralement considérés comme obligatoires (Delattre 1966 ; Grevisse 1988 ; Riegel *et al.* 1994), classement invalidé par des études de corpus. Citons par exemple De Jong (1994) qui, à partir de 45 entretiens extraits du corpus ESLO1, a entrepris une première étude sur l'usage de la liaison. Sans distinguer les formes *c'est* et *il est,* l'auteur trouve qu'après la forme *est* la liaison est réalisée à hauteur de 69 %, taux suffisamment en deçà de 100 % pour considérer ce contexte comme non systématique. Quant au corpus PFC, en limitant la recherche aux situations d'entretiens guidés et libres, on obtient respectivement 26.20 % et 42.83 % de liaison réalisée après *c'est* et après *est*.

L'un des enjeux de cet article consiste alors à apporter un éclairage nouveau à ces différentes données en intégrant et en contrôlant la prise en compte de différentes sources de variations.

4. Résultats et analyses

4.1 Variations diachroniques

Le sous-corpus DIAchronie sur lequel porte notre étude correspond à un module du corpus ESLO qui se singularise par le fait que les participants de ce module ont été interviewés à deux reprises : une première fois dans les années 1970 lors de la collecte d'ESLO1, et une deuxième fois dans les années 2000 (Vaslin-Chesneau 2008). Ce module compte sept locuteurs, sommairement présentés dans le tableau ci-dessous :

figée, fixant donc un usage de la liaison. Notons simplement que les formes produites ne semblent pas unifiées, on retrouve par exemple, des formes où seules les sons vocaliques subsistent, telles : [seai], la liaison est de fait absente.

Tableau 1 : Informations concernant les locuteurs du module DIAchronie.

Code locuteur	*Sexe*	*Année et lieu de naissance*	*Age fin d'étude*	*Age**	*Profession**	*Echelle AM**
DJ39	H	1932 Amiens	27 ans	37 ans	Médecin ophtalmologiste	A
				73 ans	Retraité (médecin)	A
QB100	F	1945 Orléans	25 ans	24 ans	Elève infirmière	B
				60 ans	Cadre hospitalier	A
CF4	H	1943 Tillay le Peneu	17 ans	26 ans	Ajusteur	C
				64 ans	Retraité (ajusteur)	C
PY94	F	1943 NR	14 ans	26 ans	Gardienne d'immeuble	D
				64 ans	Retraité (contrôleur PTT)	C
RF211	H	1947 Entre Deux Guiers	25 ans	22 ans	Elève professeur	A
				58 ans	Professeur de physique	A
YR399	H	1942 Limoges	17 ans	27 ans	Rectificateur	D
				65 ans	Retraité (ouvrier qualifié)	C
YT387	H	1928 Saumur	18 ans	41 ans	Contre maitre outilleur	C
				77 ans	Retraité (Contre maitre outilleur)	C

* Pour les colonnes Age, Profession, Echelle AM, les deux lignes par locuteur correspondent aux informations au moment d'ESLO1 et au moment d'ESLO2.

On dispose donc de quatorze enregistrements, d'une durée totale d'environ 17 heures (10 heures pour ESLO1 et 7 heures pour ESLO2). La trame d'entretien utilisée pour ESLO2 a été élaborée pour permettre une comparaison précise. Elle calque en l'ajustant le questionnaire d'ESLO1. Il s'agit d'un entretien guidé, autour de grandes thématiques telles que la famille, le travail, les activités dans la ville, et les pratiques sociales et culturelles.

Une première « carotte » nous permet d'appréhender l'usage de la liaison en mettant l'accent sur l'aspect diachronique. En effet, le corpus

DIAchronie rend compte de l'évolution des pratiques, chez sept locuteurs, à quarante années d'intervalle. Nous avons choisi d'étudier dans les deux contextes choisis –après *c'est* et après *est* – le taux de liaison facultative réalisée. Un premier résultat global (Figure 1 ci-dessous) fait apparaitre une baisse globale de l'usage de la liaison dans ce contexte : de 51.21 % dans ESLO1 à 37.76 % dans ESLO2. Notons toutefois que les deux contextes ne présentent pas le même comportement : la baisse est nettement plus forte après *c'est* qu'après *est* ; dans ce dernier contexte, on peut même parler de stabilité. Cette première série de résultats globaux va dans le sens des constats systématiques d'un moindre usage de la liaison et nuance en même temps cette tendance générale en faisant apparaitre que la baisse n'affecte pas dans les mêmes mesures tous les contextes.

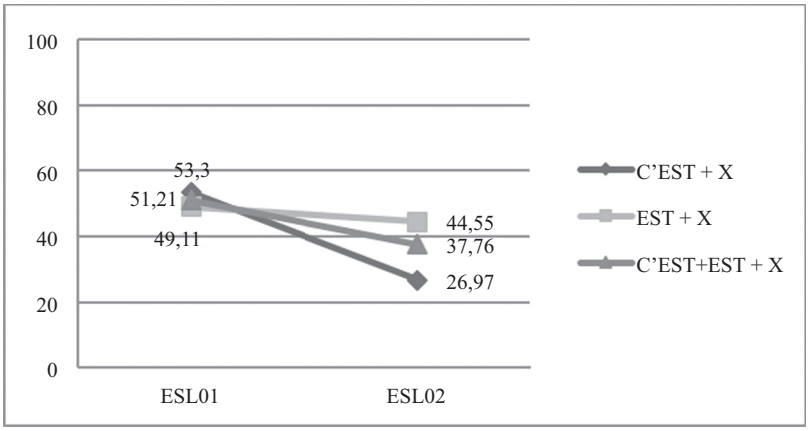


Figure 1 : Evolution des taux de liaisons réalisées après *c'est* et *est* entre ESLO1 et ESLO2.

L'observation fine des productions des sept locuteurs permet de nuancer ce premier résultat. La figure 2 ci-dessous montre des taux de réalisations dans ESLO1 variant entre 0 % et 93 % et dans ESLO2 entre 0 % et 74 %, respectivement pour les locuteurs 048 et 003. Entre ces deux extrêmes, on constate une répartition relativement homogène pour les cinq autres locuteurs. Outre des variations dans les taux de réalisation,

on peut observer des courbes d'évolutions variées. Une pente forte pour le locuteur 150, des pentes comparables (d'environ 20 points) pour les 003, 017, 149, une stabilité pour les 115, 048.

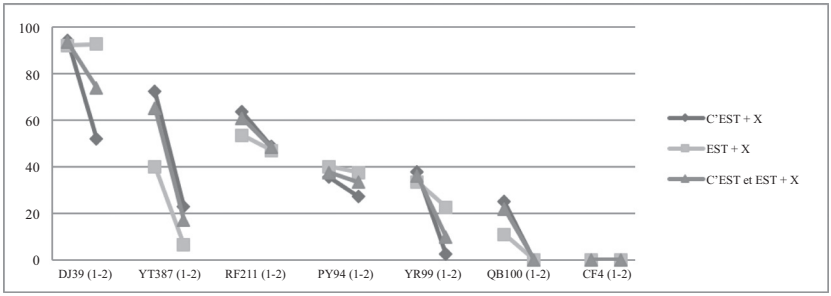


Figure 2 : Evolution des taux de liaisons réalisées après *c'est et est* chez les sept locuteurs du module DIAchronie.

4.2 Variations et trajectoires

La variation diastratique était une des caractéristiques premières à l'origine de la constitution des enquêtes ESLO1 et il en est de même dans ESLO2. Cette caractéristique se retrouve dans les objectifs des enquêtes : disposer de données linguistiques représentatives d'une communauté socialement stratifiée. Cet objectif a une incidence sur la technique d'enquête et sur les procédés de constitution du panel des témoins. Mais comment cet objectif se concrétise-t-il lors de la phase d'exploitation des données ? C'est cette question que nous souhaitons aborder empiriquement en délimitant un sous-corpus déterminé par les données diastratiques disponibles.

La méthodologie d'échantillonnage d'ESLO1 est très révélatrice d'une période où la sociologie quantitative se fondait principalement sur une catégorisation des témoins selon les caractéristiques d'âge, de sexe et de profession. La rigueur scientifique recherchée découle de cette approche (Mullineaux et Blanc 1982). Il était prévu un tirage au sort par les services de l'INSEE d'un échantillon de six cents témoins selon ces trois critères : sexe (deux catégories), âge (trois tranches) et

catégorie socio-professionnelle (cinq catégories). Le croisement de ces critères équilibrés donnait trente sous-groupes de vingt témoins. Les difficultés rencontrées par les chercheurs ont considérablement affaibli cette démarche. Le taux de refus était beaucoup plus important dans certaines catégories sociales[7]. Au terme de l'enquête, seul un quart du panel a pu être réalisé.

Au-delà de ces critères de classifications classiques, ESLO1 présente une tentative particulière de catégorisation menée par Alix Mullineaux (Mullineaux et Blanc 1982). Aux critères précédemment cités, elle a ajouté, dans un premier temps, le niveau et l'âge de fin d'études puis établi la perspective de compléter ces critères par une évaluation du capital culturel (repéré à l'issue des entretiens et notamment en prenant en compte les questions sur les goûts et pratiques culturelles) de chaque locuteur. L'objectif était alors de diviser le corpus en cinq groupes (échelle « AM » – du nom de l'auteur – de A à E). C'est ce classement que nous avons utilisé pour la suite de l'étude. Notons que l'ensemble de ces éléments est intégralement disponible dans la version numérique du corpus, sous la forme de métadonnées constituées en base de données.

C'est ainsi que, dans le sous-corpus DIAchronie, nous repérons trois locuteurs qui présentent, entre les périodes d'enregistrement d'ESLO1 et d'ESLO2, une trajectoire sociale ascendante, passant pour deux d'entre eux (PY94 et YR399) de D à C et pour la troisième (QB100) de B à A. C'est essentiellement le critère de la profession qui est à l'origine de ces trajectoires (voir tableau 1).

Une combinaison des critères diachroniques et diastratiques permet de repérer des différences de comportement significatifs. En effet si la baisse du taux de liaison est bien présente chez les trois locuteurs conformément aux prévisions des statistiques sur l'ensemble des corpus, nous constatons (voir figure 2) que le locuteur YR399 effectue une baisse de 27 points, le locuteur QB100 de 22 points et le locuteur PY94 de 4 points seulement. Par ailleurs, ces différences ne compensent pas des variations importantes, puisque les trois locuteurs conservent un taux de liaison très différent (respectivement 9,84 %, 0% et 33,33 %). Enfin, nos présentes

7 Voir Beaud et Weber (1997) pour une illustration de la difficulté à recueillir les paroles de personnes qui ne se sentent pas légitimes.

données vont à l'encontre des tendances générales sur l'usage de la liaison en fonction des données sociales (Ashby 1981, De Jong 1991) : c'est ici la locutrice qui se situe dans l'échelle la plus élevée qui a le taux de liaison le plus bas. Il faudrait prendre en compte d'autres critères pour cerner cette variation et s'interroger sur les conditions d'un éventuel changement linguistique. Ainsi, Lyche et Otsby (2009 : 225), dans leur article consacré au français de la haute bourgeoisie parisienne, relèvent que « *(…) toutes choses égales par ailleurs, certains locuteurs sont plus susceptibles que d'autres de produire des liaisons et les locuteurs les plus âgés exhibent un taux de liaisons plus conséquent* ».

Ces premiers chiffres reposent sur un nombre d'occurrences nettement plus faible que celui d'un grand corpus, ils sont évidemment beaucoup plus difficiles à interpréter. Toutefois, les variations qu'ils révèlent par rapport à d'autres travaux nous incitent à développer une analyse qui établit systématiquement un croisement entre les données de masse et une granularité plus fine.

La prise en compte du troisième critère de variation (diaphasique) nous confortera dans notre démarche de mise en perspective de données macrosociologiques et microsociologiques.

4.3 DIA-variations

Si le souci de prendre en compte la variation diaphasique est clairement exprimé dans les objectifs d'ESLO1, force est de constater qu'un déficit de cadre théorique (l'analyse de conversation n'a pas encore porté ses fruits) et des difficultés techniques (encombrement du dispositif d'enregistrement) n'ont pas permis une pleine exploitation de cet aspect du corpus. Néanmoins presque la moitié du corpus est consacré à un éventail de situations discursives variées (de la conversation lors de repas à des conférences en passant par des appels téléphoniques ou des discussions sur le marché). Cette perspective diaphasique sera au cœur d'ESLO2, dont l'architecture du corpus a été élaborée en conséquence (Baude et Dugua 2011).

Afin de tester l'impact de ces choix méthodologiques d'élaboration d'un corpus variationniste, nous avons procédé à l'extraction d'un

sous-corpus particulier permettant l'analyse d'une carotte relevant de variations diachroniques, diastratiques et diaphasiques.

Ce sous-corpus présente la particularité de regrouper quatre locuteurs (deux hommes, deux femmes), d'échelles AM différentes (A, B et D) et enregistrés dans des situations variées (entretiens, repas, appels téléphoniques). Le total représente vingt-quatre enregistrements, de durées variables (de 1 à 89 minutes) pour un total de 11h08 minutes (environ 100 000 mots).

Un premier comptage conforte les analyses antérieures sur le taux de liaison. Ainsi après *(c)est*, nous constatons un taux de réalisation de 65 %, comparable à celui décrit par De Jong (1994) (69 %). Toutefois une analyse qui combine les différents critères de variations tout en s'intéressant à des comportements décelables à un niveau beaucoup plus fin est riche d'enseignements et va quelque peu bouleverser cette impression d'homogénéité.

Nous nous sommes intéressés plus précisément à deux locuteurs (Gilbert « BA725 » et Georges « 1134 »[8]) de la même catégorie AM : D (hommes de 50/60 ans, vendeurs, sans diplôme, fin de scolarité à 13/14 ans) qui ont un taux de liaison différent tant dans la situation d'entretien (respectivement 75 % et 100 %) que dans l'ensemble des autres situations (respectivement 69 % et 33 %) (figure 3 ci-dessous). Ainsi si Gilbert réalise un taux de liaison conforme à la moyenne, nous constatons que celui-ci est peu sensible à la variation diaphasique. Le comportement de Georges est très différent puisqu'on constate une variation diaphasique forte (de 100 à 33 %).

Locuteur	Entretien	Situations hors entretien
Gilbert (code BA725, échelle AM=D)	**75 %**	**69 %**
Georges (code 1134, échelle AM=D)	**100 %**	**33 %**

Figure 3 : Taux de liaisons chez Gilbert et Georges, en situation d'entretien et hors situation d'entretien.

8 Les prénoms ont été modifiés.

Une étude du discours produit lors des entretiens apporte un éclairage très intéressant sur la trajectoire sociale que vont suivre les deux témoins. En voici quelques extraits :

> **Gilbert** souhaite devenir « *gérant de plusieurs boucheries* », il aime « *la lecture et la musique* », compte « *visiter des musées quand [il sera] à la retraite* ». Il deviendra gérant de boucheries.
>
> **Georges** « *rêvait d'être boulanger, ne prend pas de vacances sauf la pêche* », a été « *une fois au cinéma en 17 ans* » et ne connait pas « *le dictionnaire utilisé par [son] enfant* ». Il restera vendeur.

Ces informations, croisées avec l'analyse des taux de liaison, permettent de déceler des stratégies linguistiques bien plus prononcées chez Georges qui fait un effort important pour passer de 33 % à 100 % de liaisons réalisées en situation formelle (entretiens). L'habitus linguistique de Gilbert est différent et son taux de liaison est moins sensible à la variation diaphasique mais aussi plus élevé d'une manière générale. On est ici au cœur de la variation linguistique, que seules l'enquête sociolinguistique et l'exploitation maitrisée du corpus permettent d'atteindre.

5. Conclusion

Ces premières analyses reposent sur des comptages très partiels. Elles sont néanmoins significatives sur différents points. Si elles confirment la diminution du taux de liaison entre ESLO1 et ESLO2, et si les variations diastratiques et diaphasiques prises isolément correspondent globalement aux thèses classiques de la co-variation, seule une analyse qui croise avec précisions l'ensemble des critères de variations offre l'opportunité d'une analyse linguistique qui ne refoule pas la véritable nature sociale de la langue et permet d'appréhender un système linguistique dont la variation est au cœur.

Il faut souligner que ces analyses ont été rendues possibles par la nature même de la constitution des enquêtes ESLO et des choix retenus pour celles-ci. La linguistique de corpus ne peut se réduire à la manipulation de données de masse dont la nécessaire uniformisation risque de construire des artefacts bien éloignés de l'objet de la linguistique. Les corpus doivent être élaborés et exploités afin de donner accès à la variation linguistique. Pour cela il est nécessaire qu'outre le long travail d'élaboration d'un corpus selon des critères maitrisés, il soit toujours possible (car nécessaire) tout au long de l'exploitation et l'analyse des données, de répondre aux critères de l'adéquation observationnelle : savoir qui parle, à qui et dans quelle situation sociale ? Pour pouvoir répondre à ces questions il faut que la linguistique de corpus s'oriente définitivement vers une exploration maitrisée de données situées, documentées, disponibles, sur lesquelles un retour est systématiquement possible.

Il est également temps que les corpus soient rendus interopérables, tout au moins à un niveau qui permette de confronter les données de différentes analyses. C'est dans cette voie que s'est engagé ESLO dans sa méthodologie de mise à disposition des enregistrements, des transcriptions (formats), des métadonnées (normes et standards) et une documentation d'accompagnement (conventions, manuels, guides) (Eshkol *et al.* 2009). Il ne s'agit pas de penser naïvement qu'une standardisation des données est simple et intéressante. Mais plus le lien entre données et analyses est fort plus les choix du chercheur doivent ne faire aucune concession à des contraintes technologiques d'interopérabilité. Cependant les conditions d'une confrontation des données par un retour possible sur celles-ci ou par leur dialogue avec d'autres données relèvent de la responsabilité des chercheurs qui fondent leurs analyses sur des enquêtes et des corpus.

C'est alors, au bout d'un long chemin d'exploration, qu'on peut appréhender la variation linguistique dans toute sa complexité : une étape nécessaire pour des ouvrages de référence qui prendraient en compte l'apport de la linguistique variationniste fondée sur corpus.

Références

ESLO 1 et 2, <http://eslo.huma-num.fr/>.

PFC, <http://www.projet-pfc.net/>.

Ågren, J., 1973, *Etude sur quelques liaisons facultatives dans le français de conversation radiophonique : fréquences et facteurs*, Uppsala ; Stockholm : Almqvist och Wiksell.

Ashby, W., 1981, « French liaison as a sociolinguistic phenomenon », in : Cressey, W. W. et D. J. Napoli (éds.), *Linguistics Symposium on Romance Languages (9th)*, Washington, DC : Georgetown University Press, p. 46–57.

Baude, O. et C. Dugua, 2011, « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf linguiste ? », *Corpus*, 10, *Varia*, p. 99–118.

Beaud, S. et F. Weber, 1997, *Guide de l'enquête de terrain : produire et analyser des données ethnographiques*, Paris : La Découverte.

Bourciez, E. et J. Bourciez, 1971, *Phonétique française – Etude historique*, Paris : Klincksieck.

Chevalier, J.-Cl., Blanche-Benveniste, Cl., Arrivé, M. et Peytard, J., 1964, *Grammaire Larousse du français contemporain*, Paris : Larousse.

Chevrot, J.-P., C. Dugua et M. Fayol, 2009, « Liaison, word segmentation and construction in French : a usage-based account », *Journal of Child Language*, 36 (3), p. 557–596.

Clédat, L., 1917, *Manuel de phonétique et de morphologie historique du français*, Paris : Librairie Hachette et Cie.

Côté, M.-H., 2005, « Le statut lexical des consonnes de liaison ». *Langages*, 158, p. 66–78.

De Jong, D., 1990, « The syntax-phonology interface and French liaison », *Linguistics*, 28, (1), p. 57–88.

De Jong, D., 1991, « La liaison à Orléans (France) et à Montréal (Québec) », *Actes du XIIe Congrès International des Sciences Phonétiques,* Aix-en Provence, France, p. 198–201.

De Jong, D., 1994, « La sociophonologie de la liaison orléanaise », in : C. Lyche (éd.), *French Generative Phonology : Retrospective and Perspectives*, Salford : ESRI, p. 95–129.

Delattre, P., 1966, *Studies in french and comparative phonetics : selected papers in French and English*, The Hague, London, Paris : Mouton & Co.

Dupré, P., 1972, *Encyclopédie du bon français dans l'usage contemporain*, Paris : Edition de Trévise.

Durand, J. et C. Lyche, 2008, « French Liaison in the Light of Corpus Data », *Journal of French Language Studies*, 18 (1), p. 33–66.

Durand, J., B. Laks et C. Lyche, 2002, « La phonologie du français contemporain : usages, variétés et structure », in : C. D. Pusch et W. Raible. (éds.), *Romanistische Korpuslinguistik – Korpora und esprochene Sprache / Romance Corpus Linguistics – Corpora and Spoken Language*. Tübingen : Gunter Narr Verlag, p. 93–106.

Durand, J., B. Laks, B. Calderone et A. Tchobanov, 2011, « Que savons-nous de la liaison aujourd'hui ? », *Langue française*, 169, p. 1 03–135.

Encrevé, P., 1976, « Labov, linguistique, sociolinguistique », in : W. Labov, *Sociolinguistique*, Paris : Les éditions de Minuit, p. 9–35.

Encrevé, P., 1983, « La liaison sans enchainement » in *Actes de la recherche en sciences sociales*, vol. 46, p. 39–66.

Encrevé, P., 1988, *La liaison avec et sans enchaînement, phonologie tridimensionnelle et usage du français*, Paris : Edition du Seuil.

Eshkol-Taravella, I., O. Baude, D. Maurel, L. Hriba, C. Dugua et I. Tellier, 2011, « Un grand corpus oral « disponible » : le corpus d'Orléans 1968–2012 », *Traitement Automatique des Langues,* vol. 52 (3), p. 17–46.

Fouché, P., 1952, *Phonétique historique du français – Volume III : les consonnes et index général*, Paris : Klincksieck.

Fouché, P., 1959, *Traité de prononciation française*, Paris : Klincksieck.

Gadet, F., 2003, *La variation sociale en français*, Gap, Paris : Ophrys.

Grammont, M., 1914, *Traité pratique de prononciation française*, Paris : Delagrave.

Grevisse, M., 1988, *Le bon usage – Grammaire française*, Douzième édition, refondue par André Goosse, Paris : Duculot.

Habert, B., 2000, « Des corpus représentatifs : de quoi, pour quoi, comment ? », in : M. Bilger (éd.), *Linguistique sur corpus. Etudes et réflexions*, Perpignan : Presses Universitaires de Perpignan, p. 11–58.

Harnois-Delpiano, M., C. Cavalla et J.-P. Chevrot, 2012, « L'acquisition de la liaison en L2 : étude longitudinale chez des apprenants coréens de FLE et comparaison avec des enfants francophones natifs », in : F. Neveu *et al.* (éds), *Actes du 3ème Congrès Mondial de Linguistique Française*, Lyon, France, 4–7 Juillet 2012, p. 1575–1589.

Laks, B., 1980, *Différentiation linguistique et différentiation sociale : quelques problèmes de sociolinguistique française*, Thèse de doctorat, Université Paris VII-Vincennes.

Laks, B., 2005, « La liaison et l'illusion », *Langages*, 158, p. 101–125.

Laks, B., 2007, « Les hommes politiques français et la liaison (1908–1999) », in : L. Baronian et F. Martineau (éds), *Modéliser le changement : Les voies du français*, Montréal : Presses de l'Université de Montréal, p. 237–269.

Laks, B., à paraitre, « Diachronie de la liaison 1999–2011 : le cas de la parole publique », in : J. Durand *et al.* (éds.), *La phonologie du français : normes, périphéries, modélisation*, Presses Universitaires de Paris Ouest.

Liégeois, L., D. Chabanal et T. Chanier, 2011, « La liaison en discours adressé à l'enfant, spécificités et impacts sur l'acquisition », Colloque du Réseau Français de Phonologie, 1–3 Juillet 2011, Tours.

Lyche, C. et K. A. Østby, 2009, « Le français de la haute bourgeoisie parisienne : une variété conservatrice ? », in : J. Durand, B. Laks et C. Lyche (éds.), *Phonologie, variation et accents du français*, Paris : Hermès, p. 203–230.

Morin, Y.-C., 2003 [1998], « Remarks on prenominal liaison consonant in French », in : S. Ploch (éd.), *Living on the Edge – 28 Papers in Honour of Jonathan Kaye*, Berlin : Mouton de Gruyter, p. 385–400.

Morin, Y.-C. et Kaye, J. D., 1982, « The syntactic bases for French liaison », *Journal of Linguistics*, 18, p. 291–330.

Mullineaux A. et M. Blanc, 1982, « The problems of classifying the population sample in the socio-linguistic survey of Orléans (1969) in terms of socio-economic, social and educational categories », *Review of Applied Linguistics*, 55, p. 3–37.

Nardy, A. et C. Dugua, 2011, « Le rôle de l'usage sur le développement des constructions nominales chez les enfants pré-lecteurs », *Travaux de linguistique*, 162, p. 129–148.

Riegel, M., J.-Ch. Pellat et R. Rioul, 1994, *Grammaire méthodique du français*, PUF, Paris

Schane, S. A., 1967, « L'élision et la liaison en français », *Langages*, 8, p. 37–59.

Spinelli, E., A. Cutler et J. M. McQueen, 2002, « Resolution of liaison for lexical access in French », *Revue Française de Linguistique Appliquée*, VII-1, p. 83–96.

Spinelli, E. et F. Meunier, 2005, « Le traitement cognitif de la liaison dans la reconnaissance de la parole enchaînée », *Langages*, 158, p. 79–88.

Tranel, B., 1996, « Exceptionality in optimality theory and final consonants in French », in : K. Zagona (éd.), *Grammatical theory and Romance languages – Selected papers from the 25th linguistic symposium on Romance languages* (LSRL XXV), Amsterdam/Philadelphia : John Benjamins Publishing Company, p. 275–291.

Vaslin-Chesneau, A., 2008, *Analyse diachronique de la variation sociolinguistique à partir de deux corpus orléanais*, Thèse de doctorat, Université d'Orléans.

Wauquier-Gravelines, S. et V. Braud, 2005, « Proto-déterminant et acquisition de la liaison obligatoire en français », *Langages*, 158, p. 53–65.

Wauquier-Gravelines, S., P. Encrevé T. et Scheer, 2005, « Liaison in French, towards an unified explanation of variation », Colloque Phonologie du Français Contemporain, Phonological Variation, the case of French, Tromsø, Norway, 25–27 August 2005.

Wauquier, S., 2009, « Acquisition de la liaison en L1 et L2 : stratégies phonologiques ou lexicales ? » *Aile… Lia* 2, p. 93–130.

Titre	***Les ESLO, du portrait sonore au paysage digital***
Type	Article
Editeur	
Année	2015 à paraitre, accepté pour publication
Référence	Baude, O., Dugua, C., (2015 à paraitre, accepté pour publication) « Les ESLO, du portrait sonore au paysage digital », *Corpus*

Les ESLO, du portrait sonore au paysage digital

Baude Olivier, Céline Dugua
Laboratoire Ligérien de Linguistique, UMR 7270

Résumé : Cet article souhaite porter un regard réflexif sur un projet scientifique de constitution et d'exploitation d'un grand corpus de français parlé, les *Enquêtes sociolinguistiques à Orléans*, né à l'aube de la sociolinguistique et qui se développe au tournant méthodologique et épistémologique des *digital humanities*. Quels objectifs ? Quelles données ? Quels traitements ? Ce sont les questions qui guident la réflexion proposée ici afin d'apporter une contribution à l'élaboration de nouvelles pratiques scientifiques dans une perspective variationniste contemporaine.

Abstract : This article is an analysis of the constitution and the exploitation of a large corpus of spoken French: *Les Enquêtes sociolinguistiques à Orléans* (ESLO). This corpus has been created from the beginnings of sociolinguistics and now it evolves with digital humanities, methodological and epistemological specificities. Which objectives? Which data? Which analysis? These are the questions that guide our thinking in order to contribute to the elaboration of new scientific practices in a variationnist perspective.

Mots clés : Sociolinguistique, corpus, linguistique variationniste, digital humanities.
Key words : sociolinguistic, corpora, variationnist linguistic, digital humanities.

Les Enquêtes sociolinguistiques à Orléans (dorénavant ESLO) forment un grand corpus oral de plusieurs millions de mots. Ces corpus ont été réalisés à deux époques importantes de la linguistique contemporaine. La première (ESLO1), élaborée à la fin des années soixante, accompagne la naissance d'une sociolinguistique urbaine fondée sur un grand corpus d'enquêtes, et la seconde (ESLO2), commencée au début des années 2000, a profité du tournant numérique produit par les *Digital Humanities* en sciences humaines et sociales. Résolument ancrées dans le courant de la sociolinguistique et de la linguistique variationniste, les ESLO forment le socle d'études sur le français parlé à Orléans dans une perspective qui place les données au coeur d'études sur la nature sociale de la langue.

Cet article vise à décrire le travail réalisé depuis une dizaine d'années par l'équipe du projet des ESLO en le confrontant à ses cadres théoriques et méthodologiques. Après avoir abordé brièvement l'ancrage sociolinguistique du statut des données et le périmètre du français parlé, nous présenterons le travail réalisé afin de faire de ces corpus un « objet scientifique disponible » et situé.

1. Sociolinguistique et corpus

La notion de corpus croise différentes approches parfois relativement éloignées selon qu'on se situe dans une perspective de linguistique de terrain ou de linguistique informatisée. Elle prend néanmoins un sens bien plus défini dans le cadre du programme de la sociolinguistique tel qu'il a été établi dans la seconde moitié du vingtième siècle.

1.1. Nature sociale de la langue

La sociolinguistique s'est fondée sur une relecture pertinente de définition même de l'objet de la linguistique et sur la volonté de couvrir l'ensemble du domaine.

> *Pour Labov, la sociolinguistique n'est pas une des branches de la linguistique, et pas davantage une discipline interdisciplinaire : c'est d'abord*

> *la linguistique, toute la linguistique - mais la linguistique remise sur ses pieds. Elle se fonde sur l'ambition de remplir dans sa totalité le programme que la linguistique se donne dans sa définition moderne – et de l'outrepasser du seul fait de ne pas réduire son objet*. (Encrevé, 1976 : 9)

Dans cette perspective, la sociolinguistique définit la langue comme étant *partie prise* et *partie prenante* d'un social qui ne peut se réduire à un trésor collectif. Si le social est divisé et lieu de luttes et d'enjeux qui le structurent, la langue en porte, dans sa nature même, les caractéristiques qui font de la variation le principe même de celle-ci :

> *Une partie fondamentale des variations présentées par les paroles individuelles est elle aussi « instituée socialement », et par là même gouvernée par des règles : elle fait partie du système de la langue. Elle trouve normalement sa place dans la « linguistique interne » telle que la définit le CLG : « Est interne tout ce qui concerne le système et les règles (…) est interne tout ce qui change le système à un degré quelconque ».* (Encrevé 1976 :11-12)

Cette conception de la variation comme composante inhérente de la langue a une incidence directe sur la définition de l'objet d'étude sur lequel les linguistes doivent se pencher. Si les variations linguistiques sont à étudier au sein du domaine de la linguistique interne, la langue est bien le lieu où productions linguistiques et marché linguistique sont étroitement liés selon une « grammaire de la réception » qui situe la langue, comme le faisait déjà Saussure, dans le circuit de la parole :

> *Ainsi la langue d'un sujet, contrairement au sujet commun, ce n'est pas la langue qu'il parle, c'est la langue qu'il entend. Or que reçoit l'oreille d'un sujet parlant : très précisément ce que la sociolinguistique veut enregistrer et que la linguistique actuelle refuse d'écouter, les multiples paroles dont l'ensemble hétérogène*

> *arrivera à former la langue de la communauté* (Encrevé 1976 : 7).

Ainsi la communauté linguistique doit être saisie en tant qu'organisation concrète structurée et structurante des dynamiques sociales. C'est bien au cœur de celles-ci plutôt que dans une recherche illusoire d'une langue stabilisée au sein d'une communauté homogène, qu'il faut aller observer la langue afin d'obtenir l'adéquation observationnelle première que Chomsky lui-même réclamait.

> *Au total, c'est dans le caractère intrinsèquement social de la langue, dans l'intimité du lien entre langue et communauté linguistique socialement qualifiée que Weinreich, Labov et Herzog (1968) voient la source première et le moteur du changement linguistique. La communauté linguistique rappellent-ils, est une organisation sociale concrète. Elle est donc, ex definitio, profondément hétérogène, divisée, hiérarchisée, structurée par des dynamiques sociales antagoniques. La variation et l'hétérogénéité linguistique d'une part, la variation et l'hétérogénéité sociale de l'autre, ne sont alors que les deux aspects du même réel social. C'est ainsi parce qu'il n'existe jamais de communauté homogène parfaitement stable qu'il n'existe jamais de langue homogène parfaitement invariante et stable.* (Laks 2013 : 41)

Là encore, la langue ne peut se définir en dehors d'un réel social qu'il convient d'appréhender pour toute étude sur la langue. Selon Bourdieu, l'expression linguistique résultait d'une production émanant d'un habitus linguistique confronté à un marché linguistique (Bourdieu 1984 : 121). Il en résulte que l'acquisition du langage met en jeu des intériorisations socialement réglées. Ainsi comme le souligne Encrevé :

> *Aussi la grammaticalité est-elle toujours de nature sociale quant à son origine concrète pour un sujet : elle est toujours reçue et acquise assortie de sanctions sociales, dont la nature et*

> *l'importance varient avec le marché de la langue en cause – corrections, reprises, réprimandes dans la famille ; rire, moquerie de la part des égaux pour les dialectes dominés ; sanctions du marché scolaire, du marché matrimonial, du marché du travail pour les dialectes dominants.* (Encrevé 1976 : 7-8)

Il est alors aisé de concevoir le changement linguistique comme un processus résultant d'une lutte au sein de l'hétérogénéité des pratiques linguistiques évaluées socialement. La boucle est bouclée, de l'acquisition du langage au changement linguistique, la sociolinguistique offre un cadre théorique où la nature sociale de la langue est maintenant clairement définie. Cette définition de l'objet de la linguistique par la sociolinguistique se concrétise en premier lieu, et de manière centrale, autour de la question des données.

1.2 Sociolinguistique et données

En effet, définir la langue comme un fait social, nécessite de l'observer comme une pratique socialement située. C'est donc au sein même de l'activité sociale qu'elle devient appréhendable :

> *Partie structurée d'un tout qu'elle structure, la langue, en effet, n'est jamais « donnée ». Les « données » de la langue dans son usage quotidien, telle que veut l'étudier Labov, ne sont « produites » qu'au terme d'un long chemin d'aveuglette où se construit pas à pas une science de l'enquête linguistique qui est la première conquête de la sociolinguistique* (Encrevé 1976 : 13).

Pour la sociolinguistique, il ne s'agit pas d'une simple question méthodologique qui déterminerait l'observation des données comme une étape préliminaire à l'analyse scientifique, bien au contraire la définition même des données et des conditions de leur production sont au cœur du travail du linguiste. La

première incidence concerne le périmètre des données linguistiques. Comme le souligne Laks (2013), on ne peut concevoir d'analyser des données linguistiques orphelines de l'habitus du locuteur et du marché qui structure ses productions :

> *Observer la variation dans sa systématicité et rendre compte de l'hétérogénéité comme étant structurée impose évidemment d'adopter une méthodologie adéquate. On sait en effet que décontextualisée, l'observation détruit la systématicité des phénomènes variables et les fait paraitre erratiques. Observer les faits linguistiques hors de l'écosystème social qui les conditionne détruit en effet tout ce que la pratique doit précisément à son caractère pratique. C'est la raison pour laquelle l'analyse de la variation systémique commence nécessairement par une réflexion critique sur les observables.* (Laks 2013 : 36)

Dans les années soixante-dix, la réflexion sur la place des données a entrainé une véritable science de l'enquête linguistique pour laquelle les avancées de la sociologie à la même époque, depuis Bourdieu, Chamboredon et Passeron en 1968 jusqu'à Beaud et Weber en 1997, ont été déterminantes en ce domaine. Parallèlement et parfois simultanément à l'apport de la sociologie de l'enquête, la naissance du domaine de l'analyse de conversations et les études sur les données « naturelles » ou plus justement sur les données issues de « situations non provoquées par le chercheur » sont également des éléments essentiels du développement de la science de l'enquête linguistique.

Enfin, le troisième domaine constitutif de cette démarche méthodologique et théorique provient de la linguistique de corpus dans son versant « informatique et traitement automatique du langage ».

1.3 Données et posture du chercheur

Dans cette perspective la place des données devient prédominante et le travail du linguiste ne peut s'affranchir d'une démarche réflexive sur la méthodologie de constitution et d'exploitation des données. Il lui revient alors de rendre explicite ses motivations scientifiques, sa méthodologie de collecte, la description des données et le traitement de celles-ci (Habert, 2005). C'est alors une véritable posture qui se profile sur la base d'une confrontation scientifique qui doit rendre possible la disponibilité des données, y compris pour un retour évaluatif ou contrastif, leur interopérabilité et leur description fine. En outre cette posture ne peut s'affranchir d'une réflexion éthique et juridique (Baude, 2006) sur les données, les locuteurs et le terrain non exempts d'enjeux sociaux.

Il s'agit donc de définir une conception de la sociolinguistique et par-delà de la linguistique, à partir de la relation de cette discipline aux données, nécessairement variationnistes et situées. Ceci nécessite que le linguiste sache ce qu'il fait (Gadet 2007) dans la continuité d'une évolution méthodologique et théorique d'une science de l'enquête à une science du corpus.
Les Enquêtes sociolinguistiques à Orléans, qui se concrétisent par un ensemble de deux corpus réalisés à quarante années d'intervalle, offrent l'opportunité d'évaluer, à partir de projets concrets, le cadre de ce positionnement.

2. Le français ordinaire

2.1 La recherche du français parlé

ESLO1 a pour origine un projet à finalité didactique. L'équipe constituée à la fin des années soixante autour de Michel Blanc avait comme objectif de réaliser une méthode d'enseignement audiovisuelle du français langue seconde à partir de documents authentiques. Celui-ci est clairement défini dans un court article paru en 1971 (Blanc & Biggs). A « une époque où le rôle essentiel de la langue parlée dans l'enseignement d'une langue étrangère » venait d'être acquis, il a fallu « constituer un

ensemble cohérent de matériaux vivants, rassemblés de manière systématique » valable « à la fois pour l'application pédagogique et pour la recherche sur la langue parlée ». Partant du constat qu'une collection ordonnée de documents de ce type n'était disponible, l'équipe a entrepris de collecter un vaste corpus représentatif du français parlé à partir d'une enquête ciblée sur une ville « moyenne » française exempte de caractéristiques trop marquées.
La démarche a d'emblée été résolument ancrée dans le champ de la sociolinguistique et la variation fut au cœur du travail de définition de la représentativité du corpus :

> *Selon nous une recherche sociolinguistique impliquait une étude de la langue dans sa diversité plutôt que comme un tout homogène et figé. En effet, même si on étudie un état de langue à un moment précis de l'histoire, il n'empêche qu'il offre une variété à plusieurs niveaux : différences entre les générations, différences dialectales entre communautés, différences entre les milieux sociaux, différences liées aux conditions de production du discours.* (Blanc & Biggs 1971 :16).

Cette prise en compte de la diversité n'exclut pas, bien au contraire, la recherche d'une langue partagée par une communauté linguistique. C'est ainsi que le projet s'est orienté vers la réalisation du portrait sonore de la ville d'Orléans. Il s'agissait d'observer et de capter à un moment précis, dans un lieu restreint, la dynamique des pratiques linguistiques partagées par les habitants d'une cité. Le corpus est donc constitué d'une collection d'entretiens de locuteurs socialement situés et catégorisés, mais aussi d'enregistrements variés donnant accès au « français parlé dans une ville moyenne par la population de la ville à une époque précise » (Blanc & Biggs 1971).

2.2 La découverte du français entendu

La grande originalité pour l'époque et le parti pris très fort choisi par l'équipe a été de définir les pratiques linguistiques communes non pas par les productions de locuteurs types mais par l'hétérogénéité des pratiques linguistiques entendues dans la ville. Comme le soulignent Blanc & Biggs « C'est une communauté d'auditeurs qui est construite, autant qu'une communauté de locuteurs, à notre connaissance pour la première fois en France (…) On ne cherche pas « cet individu mythique, l'orléanais moyen » (Blanc & Biggs 1971 : 23). On est ici dans la même perspective de la sociolinguistique que celle défendue par Encrevé quelques années plus tard quand il reprend l'affirmation de Saussure selon laquelle la langue comme objet de la linguistique se situe dans le circuit de la parole, pour préciser immédiatement que

> *pour Saussure la langue est entièrement, et exclusivement, du côté de l'audition, de la réception : on peut la (la langue) localiser dans la partie déterminée du circuit (de parole) où une image auditive vient s'associer à un concept ; c'est par le fonctionnement des facultés réceptives et coordinatives que se forment chez les sujets parlants des empreintes qui arrivent à être sensiblement les mêmes pour tous. Ces deux points sont manifestement reliés : seule l'audition met le sujet en contact avec la masse parlante. Ainsi la langue d'un sujet, contrairement au jugement commun, ce n'est pas la langue qu'il parle, c'est la langue qu'il entend* (Encrevé 1977 : 6).

Nous le verrons dans le chapitre consacré à l'architecture des corpus des ESLO, ce cadre théorique et ses incidences méthodologiques apportent une très forte identité à l'ensemble du projet.

2.3 La linguistique du français parlé d'ESLO1 à ESLO2

Entre les deux enquêtes ESLO1 et ESLO2, la linguistique française a bénéficié des très précieux travaux de Blanche-Benveniste et de l'école du GARS sur la description du français

parlé. Ces études, principalement grammaticales, ont incontestablement marqué le champ de la discipline. Or, comme ces travaux du GARS reposent essentiellement sur l'analyse de corpus, on peut s'attendre à une avancée importante sur la description du français parlé et simultanément sur la méthodologie de corpus entre les années soixante et les années deux mille dix. Si l'avancée a été majeure et déterminante pour les travaux sur la syntaxe du français, elle n'a apporté qu'une contribution très faible à la linguistique de corpus ou plus exactement à la linguistique *sur* corpus. La relation relativement distante entretenue entre les travaux du GARS et la sociolinguistique explique ce rendez-vous manqué.
Quatre disciplines vont avoir une incidence plus forte dans la même période sur les corpus de français parlé. Discipline compagne, la sociologie va opérer un lourd travail sur le recueil des données et sur la méthodologie d'entretien qui reste une part importante des corpus oraux. Parallèlement, la linguistique de l'interaction et plus particulièrement l'Analyse de Conversations va se développer très fortement et proposer une nouvelle approche du recueil de données « non provoquées par le chercheur ». Ensuite, le domaine de l'acquisition du langage fournira une méthodologie très rigoureuse de grandes bases de données partagées (volet français du programme CHILDES, notamment pour ce qui concerne l'adoption d'un format et d'un codage communs[1]) de corpus de productions d'enfants.
Enfin la recherche en technologies de la parole, de la reconnaissance à la synthèse en passant par la traduction repose sur le traitement de données orales massives.

La reprise du projet ESLO1 par l'équipe du CORAL (devenue LLL) en 2004 avec comme perspective de rendre disponible l'intégralité du corpus[2] et d'en constituer un nouveau devait

[1] MacWhinney B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ : Lawrence Erlbaum Associates.

[2] Un travail remarquable avait déjà été réalisé dans le cadre du projet ELILAP-ELICOP : ELILAP 1980-83 puis LANCOM 1993-2001, voir Mertens (2002)

nécessairement tenir compte des avancées apportées par ces disciplines.

Un bref bilan de l'impact de celles-ci révèle la qualité du travail précurseur des auteurs d'ESLO1 et facilite la reprise du projet avec une forte continuité même si plusieurs choix sont caractéristiques de l'évolution d'ESLO2.
Outre le soin apporté à la technique de conduite d'entretiens, les principales évolutions concernent l'intérêt accru pour assurer une représentation de l'hétérogénéité du panel de locuteur et des situations enregistrées (cf chapitre sur l'architecture du corpus en infra) et pour la description des langues en contact avec le français.

2.4 Conserver et diffuser le français ordinaire

Le bouleversement le plus fort concerne un élément peu fréquent jusqu'à très récemment dans les projets sur les corpus de français parlé : celui de la conservation et de la diffusion.
Pourtant sur ce point aussi ESLO1 était totalement précurseur.
Alors que dix ans auparavant les responsables du *Français fondamental* effaçaient les enregistrements réalisés dans le cadre de ce projet d'ampleur internationale (Abouda & Baude 2006), les auteurs d'ESLO1 décidaient d'apporter un soin particulier au catalogage de leurs enregistrements afin d'en assurer la meilleure diffusion. Ainsi un des six objectifs d'ESLO1 était de :

> *préparer et publier un catalogue descriptif et analytique des documents sonores et écrits, afin de les rendre disponibles aux chercheurs, notamment dans les domaines de la linguistique, de la sociologie et de la pédagogie des langues* (Lonergan, 1974 :2).

Cette volonté affichée dès l'origine du projet aura une forte incidence sur son développement. Elle porte la marque d'une relation particulière aux données et au rôle de leur exploitation partagée dans la constitution d'un savoir collectif. C'est également une reconnaissance de la légitimité de la langue

parlée comme objet scientifique et patrimonial. L'ESLO deviendra alors une référence sous le nom du *Corpus d'Orléans* et voyagera de la France à l'Angleterre, des Pays-Bas à la Belgique au gré des nombreux travaux de chercheurs dans une discipline en plein développement.

3. Le corpus des ESLO

3.1 Un très grand corpus

Le corpus des ESLO[3] a comme objectif d'être un très grand corpus de français parlé constitué de plusieurs centaines d'heures d'enregistrements afin d'atteindre une masse de 10 millions de mots.

Il est composé du corpus ESLO1, qui est un corpus clos, réalisé entre 1968 et 1971 et qui comprend 470 enregistrements d'une durée totale de 318 heures ce qui représenterait, selon l'estimation de l'époque, 4.5 millions de mots[4].

Le corpus ESLO2, en cours de réalisation, affiche un objectif de plus de six millions de mots pour 450 heures d'enregistrements.

Réunis dans une même base de données comprenant les enregistrements, leurs transcriptions orthographiques et les métadonnées décrivant les documents, le contexte d'enregistrement et les locuteurs, le corpus des ESLO est actuellement le plus grand corpus de français parlé disponible pour la recherche en linguistique.

L'objectif du projet n'est pas de produire un corpus représentatif, mais d'offrir un réservoir de corpus conçu dans un souci de représentativité des pratiques linguistiques d'une communauté d'auditeurs dans une ville donnée à des moments distincts. La sélection d'un sous-corpus d'études à partir de ces données reste à la charge du chercheur dans une démarche où la sélection des données est une étape fondamentale de l'analyse.

[3] Cf. Baude & Dugua 2011

[4] Environ 70 % du corpus présente une qualité acoustique suffisante pour une transcription.

Il revient alors aux auteurs des ESLO de rendre disponibles les données tout en les situant à la fois dans le cadre de leur contexte de production par les locuteurs et de celui de production par l'équipe scientifique y compris dans ses aspects et contraintes technologiques.
Il ne s'agit donc pas de produire un corpus de masse de données sans en préciser l'architecture et les cadres théoriques qui la conditionnent.

3.2 Architecture du corpus

La composition du corpus a subi une évolution sensible entre ESLO1 et ESLO2.
Comme nous l'avons indiqué, le corpus ESLO1 correspond déjà à une prise en charge des variations linguistiques selon différents axes. Cette recherche de la variation s'est concrétisée par une architecture qui en donnant une place centrale aux entretiens en face à face a néanmoins intégré sept autres modules dédiés à la diversité des situations de production de discours :
- Interviews sur questionnaires (interviews en face-à-face sur des questionnaires standardisés, avec un échantillon statistique aléatoire choisi d'après la liste INSEE du recensement de la population 1968). 157 enregistrements, 182,5 heures.
- Opérations sur le vif : contacts (prises de contact, reprises de contact, ouverture et clôture des entretiens enregistrés à l'insu du témoin). 55 enregistrements, 12,5 heures.
- Opérations sur le vif : témoins en situations sociales ou professionnelles (enregistrements de témoins INSEE dans des situations sociales ou professionnelles, faits en l'absence des chercheurs). 16 enregistrements, 14,5 heures.
- Communications téléphoniques. 50 enregistrements, 2,15 heures.
- Interviews sur mesure (entretiens avec des individus choisis selon leur rôle dans la "microsociété" orléanaise). 45 enregistrements, 48,33 heures.

- Conférences-débats (conférences-débats ou discussions à plusieurs participants, les dernières comportant souvent des témoins INSEE). 26 enregistrements, 34,15 heures.
- Enregistrements divers (enregistrements divers comportant des témoins inconnus, visite d'atelier, marchés, magasins, etc.). 84 enregistrements, 14,33 heures.
- CMPP (interviews au Centre Medico-Psychopédagogique, parents d'élèves et assistante sociale). 37 enregistrements, 10 heures.

L'ensemble de ces modules sont décrits dans le catalogue original (Lonergan, 1974 : 1) et présentés sur le site de diffusion du corpus ESLO[5].

L'architecture va considérablement évoluer dans le cadre du corpus ESLO2[6] afin de prendre en compte l'avancée méthodologique et théorique réalisée entre 1968 et 2008. D'une part l'évolution technologique a une forte incidence sur la collecte des corpus oraux. Si les auteurs d'ESLO1 se félicitaient de disposer de matériel d'enregistrement peu volumineux (de la taille d'une petite valise), et léger (à peine 7 kilos), l'équipe d'ESLO2 dispose d'un matériel numérique offrant les possibilités d'équiper des locuteurs de micro cravates HF pour une qualité d'enregistrement de tout premier ordre. Ainsi pour l'un des modules qui consiste à enregistrer l'intégralité de ce qu'une personne entend pendant 24 heures, les locuteurs sont équipés d'un micro les accompagnant dans toutes les activités de la vie quotidienne, de la toilette à la soirée entre amis en passant par l'activité professionnelle et les conversations familiales.

Cette évolution technologique s'accompagne d'un engouement fort pour la captation d'enregistrements les plus diversifiés dans des situations non provoquées par le chercheur selon les objectifs de l'Analyse de conversations.

[5] http://eslo.huma-num.fr/

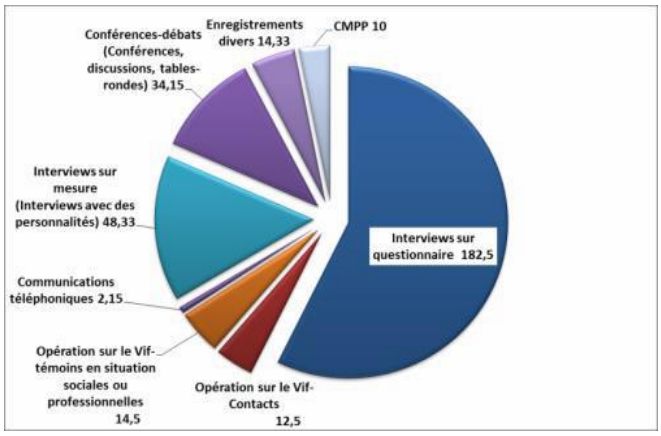
[6] http://eslo.huma-num.fr/index.php/pagecorpus/pagepresentationcorpus

L'objectif de dresser un portrait sonore ne peut donc se résumer à la collecte d'entretiens selon un échantillonnage sociologique. Il convient également d'élaborer une architecture de corpus qui permet de rendre compte de la diversité des situations de production et d'audition. Force est de constater qu'ESLO1 était balbutiant sur cet aspect. Si les entretiens ont été réalisés avec beaucoup de rigueur, les autres types d'enregistrements sont très souvent de très mauvaise qualité et correspondent à des objectifs peu maitrisés. La tentative d'enregistrer la même personne dans diverses situations s'est réduite à de simples tests sur quelques locuteurs. ESLO2 a donc comme ambition de présenter une forte évolution de la méthodologie de collecte de situations variées et représentatives des pratiques d'une communauté.

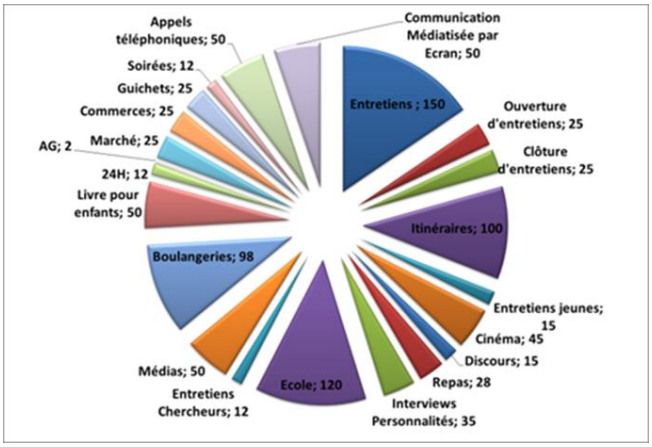
C'est toute l'architecture du corpus qui doit être modifiée afin de prendre en compte une grande diversité de situations de productions linguistiques tout en les situant au sein d'un marché linguistique plus général.

Le premier effet de ce changement est de pondérer la place des entretiens par rapport à d'autres types d'enregistrements. Les graphiques suivants qui expriment en nombre d'heures et en pourcentage la place de chacun des modules pour les deux corpus, rendent compte de ce changement.

ESLO1 :



ESLO2 :



3.3 Catégorisation des modules

L'architecture d'un corpus ne peut se résumer au pourcentage des genres, styles ou situations représentées. Elle nécessite également une réflexion sur la pertinence de ces catégories au sein d'une structure globale.

Ainsi, assurer la collecte de la diversité des pratiques linguistiques répond à un objectif d'enquête sociolinguistique et de description linguistique. Le conditionnement en corpus numérique du résultat de cette collecte nécessite un travail de catégorisation des modules constituant l'architecture du corpus. Cette catégorisation se doit d'être explicite et disponible à des fins de traitement des données. La classification habituelle dans les corpus de français parlé repose sur une opposition simpliste entre discours public et discours privé décrivant le niveau de formalité des énoncés.

Ainsi, *Le Corpus de référence du français parlé*, réalisé par Claire Blanche Benveniste et l'équipe DELIC à partir de 1998, repose sur une structure en trois modules : parole privée, parole professionnelle et parole publique. Cette distinction est assez rudimentaire si on se réfère aux travaux de l'analyse de

conversations ou même à la description des registres de langue (Koch & Oesterreicher 2001).

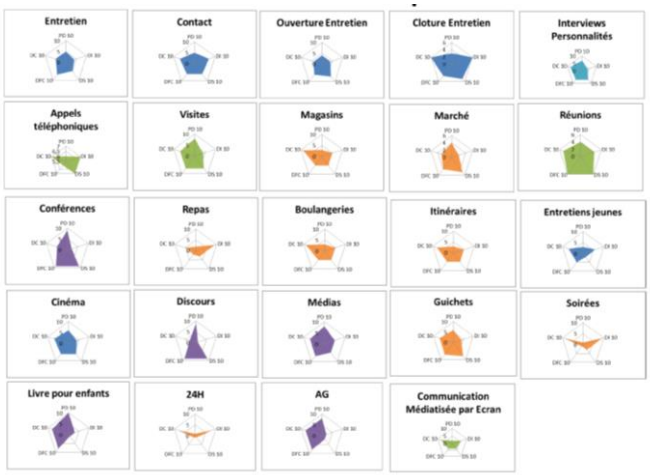
Le corpus ESLO2 est l'occasion de tenter une description des registres, styles ou types de situations en partant des caractéristiques apriori et, apostériori, des différents modules.

Chaque module est décrit a priori, c'est-à-dire avant la collecte et non sur la base d'une analyse du contenu, selon les critères suivants :

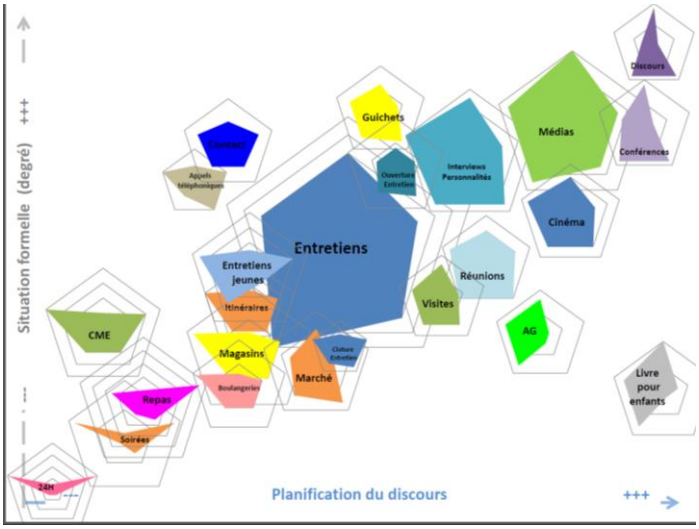
- Degré de planification du discours (en opposant le registre « spontané » de la conversation ordinaire à celui de conférences ou le discours est écrit),
- Degré d'interactivité (du monologue au dialogue et autres conversations relevant d'un travail conséquent d'interaction),
- Degré de distance sociale entre les interactants (à partir des critères traditionnels de la sociologie : âge, sexe, niveau d'études, profession),
- Degré de convergence (de la polémique au consensus),
- Degré de formalité du cadre (au sens de Goffman, chaque situation pouvant se définir selon un cadre social impliquant des statuts, rôles et comportements langagiers).

Chacun de ces critères est évalué sur une échelle de 0 à 10 et le module peut être visualisé selon la forme obtenue par un graphique en radar :

Les différents modules constitutifs de l'arcitecture ESLO2 :



Cette démarche permet de décrire l'architecture du corpus en raffinant une prise en compte des axes traditionnels qui situent un contexte de production de discours selon le degré de formalisme de la situation sociale d'une part et le degré de planification de l'énoncé d'autre part.



Cette représentation de l'architecture du corpus répond à deux objectifs. Premièrement, il s'agit de définir avec précisions les différents modules qui composent le corpus complet en situant les situations enregistrées selon les critères de la sociologie et de la pragmatique. Cela répond à une conception des pratiques linguistiques comme relevant systématiquement d'un *contexte*, qui n'est autre qu'un marché linguistique au sein duquel les locuteurs mobilisent des comportements langagiers dans un but d'interaction.
Deuxièmement, l'évaluation des modules selon différents critères permet un travail réflexif sur une définition apriori et un constat apostériori à partir des données précises de la situation enregistrée. Ainsi, si le module entretien répond globalement à une définition selon les critères présentés, celle-ci va être pondérée pour chaque entretien. L'évaluation de la distance sociale et du degré d'interactivité peuvent par exemple être très différents d'un entretien à l'autre et déboucher sur une représentation proche d'une conversation ordinaire dans un cas ou d'un discours public ou médiatique dans un autre.
In fine, cette réflexion sur l'architecture du corpus permet de concevoir ESLO2 comme un corpus ouvert sans pour autant le réduire à un empilement, opportuniste et sans fin, d'enregistrements variés.

3.4 État du corpus

L'ensemble des enregistrements est maintenant numérique. L'intégralité des enregistrements ESLO1 a été numérisée dans le cadre du dépôt du fonds à la Bibliothèque Nationale de France. ESLO2 est nativement collecté en numérique à l'aide de différents matériels selon les contraintes des modules[7]. Si ESLO1 est un corpus clos, la collecte d'ESLO2 continue à la date de la rédaction de cet article.

Tous les enregistrements sont catalogués et indexés (cf. chapitre suivant) et la transcription de l'intégralité des corpus est en cours.

[7] Principalement : enregistreurs Marantz PMD 661 MKII + microcravates AKG C417L, TASCAM DR100, Edirol R09 : http://eslo.huma-num.fr/index.php/pagemethodologie?id=70

Les opérations de formatage, catalogage et transcription sont excessivement lourdes ce qui explique le peu de corpus d'envergure disponibles. Face à cette difficulté, les chercheurs se replient souvent vers un usage du corpus restreint à leur recherche. La particularité forte du projet des ESLO est au contraire de maintenir un objectif scientifique clairement identifié tout en attribuant au corpus une valeur patrimoniale et scientifique qui dépasse le cadre du projet initial. Il en résulte un vaste chantier de traitement du corpus qui sera détaillé dans la dernière partie de cet article. Nous pouvons néanmoins faire état de l'avancement de ces opérations. Ainsi, au premier mai 2015 le corpus des ESLO est composé de :

	Enregistrements		**Transcrits**	
	Nbre	Heures	Nbre	Heures
ESLO1	468	318	336	274
ESLO2	590	266	583	259
TOTAL	**1058**	**584**	**919**	**533**

4. Un corpus pour les *Humanités numériques*

4.1 Le temps des humanités numériques

Le projet de diffusion des ESLO au début des années 2000 est contemporain de la mutation des sciences humaines et sociales dans ce qu'on appelle dorénavant le tournant des *Digitals Humanities* ou *Humanités numériques* voire *Humanités digitales* (Le Deuff, 2014)[8]. Les discussions sur ce que sont les *humanités numériques* sont très vives et la définition reste très ouverte. Il ne s'agit pas de rentrer ici dans une vaste discussion sur la pertinence d'une approche en terme de naissance d'une discipline, d'une trans-discipline ou d'une appropriation d'outils numériques par des disciplines traditionnelles, nous nous contenterons de constater que la linguistique est en

[8] Le Deuff O. dir. (2014). *Le temps des humanités digitales,* la mutation des sciences humaines et sociales.

première ligne d'un questionnement sur les conditions de constitution, de diffusion et de partage d'un savoir transformé par le croisement de l'informatique, du numérique et des arts et lettres au sein des sciences humaines et sociales. Ces grands principes ont été définis dans le *Manifeste des Digital humanities*[9].

D'une manière plus concrète encore nous présentons ici les principales caractéristiques qui inscrivent le projet des ESLO dans cette approche des corpus en sciences humaines et sociales. Le soin apporté à la diffusion d'ESLO1 en 1974 en réalisant un « *catalogue descriptif et analytique des documents sonores et écrits, afin de les rendre disponibles aux chercheurs (Lonergan 1974 : 2)* » peut être interprété comme la première pierre posée dans l'édifice d'un corpus qui dépasse les enjeux de l'étude des auteurs. La seconde pierre viendra de l'équipe de Piet Mertens et du projet ELICOP quelque trente ans plus tard en rendant accessible une partie du corpus après un lourd travail de normalisation des conventions de transcription et même d'annotations morphosyntaxiques contenues dans des balises au format SGML. Ce travail s'appuie sur les perspectives dressées par la linguistique de corpus telle qu'elle est définie par Habert, Nazarenko et Salemn en 1997, mais n'est pas encore directement orienté vers un traitement d'ensemble.

C'est à partir de 2004 avec la numérisation d'ESLO1 et le souhait de rendre le corpus intégralement disponible pour des usages scientifiques mais aussi culturels que l'édifice s'ancrera définitivement dans les humanités numériques.

4.2 L'interopérabilité et l'archivage

La question de la réutilisation d'un corpus n'est pas anodine et ne va pas de soi. Il ne s'agit pas ici d'affirmer que toute recherche linguistique doit s'appuyer sur un corpus et que tout corpus peut être réutilisé pour d'autres recherches. Rien n'est moins sûr mais dans le cas des ESLO c'est un parti pris affirmé par les différents auteurs du projet. Le périmètre du projet est de

[9] http://tcp.hypotheses.org/318

fait vaste, il s'agit de produire un portrait sonore d'une ville en faisant l'hypothèse que le corpus produit peut être utile à diverses recherches en linguistique, sociologie, histoire, didactique et acquiert ainsi une dimension patrimoniale qui a également pour effet de légitimer le français tel qu'il est parlé dans sa très grande diversité.

L'objectif affirmé est donc de disposer de données répondant à un critère d'interopérabilité. Celui-ci se concrétise à différents niveaux.

Premièrement, les enregistrements sont conservés dans un format numérique selon les recommandations d'une structure internationale, l'*International Association of Sound and Audiovisual Archive*[10].

Deuxièmement, les documents sont systématiquement accompagnés de métadonnées descriptives. Le choix retenu est celui du format *DUBLIN-CORE Open Language Archives Community*[11]. Il s'agit d'un choix minimal qui a été repris dans le cas de diffusions liées à d'autres objectifs. Ainsi le format CMDI[12] est celui utilisé dans la perspective européenne CLARIN, le format EAD[13] par la BnF pour l'intégration à son catalogue *Archives et manuscrits*, et l'EDM dans le cadre de la bibliothèque européenne *Europeana*[14].

Troisièmement, les enregistrements sont transcrits et synchronisés avec le signal sonore selon des conventions minimales[15] répondant à un format interopérable. Le format choisi est un format XML qui est ensuite repris pour un enrichissement en TEI (TEIML[16]). Les transcriptions sont segmentées en unités prosodiquement, syntaxiquement et

[10] http://www.iasa-web.org/ : Wave, stéréo, 16 bits, 44100 Hz.

[11] http://www.language-archives.org/OLAC/metadata.html

[12] http://www.clarin.eu/content/component-metadata

[13] http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_ead.html

[14] http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation

[15] http://eslo.huma-num.fr/index.php/pagemethodologie?id=71

[16] Norme ISO/CD 24624 en cours d'élaboration

sémantiquement cohérentes afin d'assurer une synchronisation à l'aide de jalons temporels fréquents. La transcription proposée repose sur des conventions minimales. A ce stade, il s'agit de répondre à un simple objectif de navigation dans le corpus. Pour toute analyse ultérieure une reprise de la transcription avec des conventions répondant aux cadres théoriques du chercheur est indispensable.

L'ensemble de ces choix permet l'utilisation d'un service d'archivage. Expérimenté dans le cadre du projet pilote sur l'archivage de l'oral par le TGE ADONIS puis poursuivi par la TGIR HUM-NUM, les données (enregistrements, transcriptions et métadonnées) sont confiées à la plateforme Cocoon[17] qui en assure le stockage sécurisé sur la grille Huma-Num hébergée au centre de calcul de l'IN2P3. Pendant cette phase, Cocoon assure des services de contrôle de la qualité des données puis verse les données au Centre Informatique National de l'Enseignement Supérieur pour une conservation intermédiaire avant de rejoindre les Archives Nationales pour un archivage définitif. Parallèlement, les bandes magnétiques originales ont été confiées au service sonore du département de l'audio-visuel de la BnF.
Les opérations d'archivage sont également l'occasion d'attribuer un identifiant unique et pérenne à tous les documents constitutifs du corpus.

4.3 Les aspects juridiques

La diffusion du corpus est bien évidemment liée à des aspects juridiques. Sur ce point, le projet a bénéficié du travail diffusé par le *Guide des bonnes pratiques 2006*[18].
Le choix de l'équipe a été d'apporter beaucoup d'attention à une démarche éthique en recueillant le consentement éclairé de

[17] http://cocoon.huma-num.fr/exist/crdo/

[18] Baude et al., (2006).

toutes les personnes enregistrées[19]. Les enregistrements et les transcriptions sont également anonymisés et les données personnelles conservées dans une base de données séparée.
Les données sont diffusées sous licence creatives commons[20] (BY NC SA : Attribution, pas d'utilisation commerciale et partage dans les mêmes conditions) : le titulaire des droits autorise l'exploitation de l'œuvre originale à des fins non commerciales, ainsi que la création d'œuvres dérivées, à condition qu'elles soient distribuées sous une licence identique à celle qui régit l'œuvre originale.

4.4 Le signalement et la diffusion

La conservation des données étant assurée à différents niveaux (stockage sécurisé, conservation intermédiaire et archivage pérenne) et les aspects juridiques ouverts à une large diffusion, il faut en assurer l'accès pour différents usages.
Sur ce point, le soin apporté à l'interopérabilité devient crucial.
Les données ESLO sont accessibles sur un site dédié au projet[21], géré par l'équipe du Laboratoire Ligérien de Linguistique et hébergé sur la grille Huma-Num.
Le site, réalisé à l'aide du CMS Joomla et intégrant une application, a été conçu en trois parties :
- Une interface « back office » qui permet la gestion du corpus. Cette interface permet, à l'aide de formulaires, de renseigner les métadonnées et dispose de fonctionnalités pour attribuer aléatoirement les identifiants anonymes, transférer les fichiers sonores et les transcriptions sur la plateforme Cocoon et pour accéder à une base de données mysql qui contient les transcriptions et les métadonnées.
- Une interface d'accès aux corpus avec des outils spécifiques. L'accès aux corpus se fait par une recherche des documents dans leur intégralité sous la forme d'un catalogue ou par la recherche d'une chaine de caractères au sein des transcriptions.

[19] http://eslo.huma-num.fr/index.php/pagemethodologie?id=69

[20] http://creativecommons.fr/licences/les-6-licences/

[21] http://eslo.huma-num.fr/

Un outil de requête permet de croiser les critères de recherche sur les transcriptions avec les informations sur les documents et les locuteurs.
Un second outil offre la possibilité d'écouter l'enregistrement synchronisé sur le signal.
Enfin, l'ensemble des documents sont téléchargeables directement soit pour tout utilisateur du site soit pour un utilisateur ayant signé une convention lorsqu'il y a des restrictions juridiques.
- La dernière fonctionnalité du site est d'offrir un contenu éditorial principalement orienté vers les documents méthodologiques : conventions et guides de transcriptions, documents techniques et juridiques, documents scientifiques.

Cette diffusion du corpus par un site spécifique répond principalement aux objectifs du Laboratoire Ligérien de Linguistique. La gestion des données selon des bonnes pratiques d'interopérabilité et d'archivage permet un signalement et une diffusion beaucoup plus large.

Ainsi, la plateforme Cocoon propose un entrepôt exposant les métadonnées en Open Archive Initiative. Le corpus des ESLO est donc signalé par tout instrument reposant sur un moissonnage en OAI. C'est notamment le cas de la plateforme ISIDORE[22] qui permet la recherche et l'accès aux données numériques en sciences humaines et sociales. Au premier mai 2015, une recherche sur ESLO dans le moteur d'ISIDORE apporte 2001 réponses, soit l'ensemble des documents disponibles à ce moment-là dans la collection ESLO de l'entrepôt Cocoon.

Comme ESLO existe également sous la forme de bandes magnétiques originales conservées et décrites par la BnF, le corpus est également signalé dans ses catalogues.

[22] http://www.rechercheisidore.fr/

Enfin le corpus des ESLO a été naturellement intégré à l'EQUIPEX ORTOLANG[23] dont l'objectif est de gérer une « *infrastructure en réseau offrant un réservoir de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement clairement disponibles et documentés* ».

4.5 Le web de données

Le travail sur la structuration des données et des métadonnées et la gestion de la diffusion du corpus des ESLO permet un travail exploratoire dans le cadre du web de données (ou web sémantique). Cette étape concrétise la volonté de construire un corpus réutilisable pour une grande variété d'usages. Le web de données vise à publier des données structurées sur le Web, afin de les relier entre elles et donc d'enrichir un réseau d'informations. Elle nécessite l'utilisation, dans un format spécifique, de vocabulaires, référentiels et ontologies facilitant le liage des données.

Nous pouvons citer quelques exemples d'expérimentations en cours auxquelles participe ESLO :

- la plateforme ISIDORE qui repose sur les principes du Web de données,
- data.bnf.fr, le projet qui donne accès aux données contenues dans ses catalogues et dans Gallica,
- le programme *Sémantisation du Corpus de la parole* du Ministère de la Culture,
- le projet « Cabinet de curiosités des langues de France » réalisé dans le cadre de l'appel à propositions « services culturels innovants du Ministère de la culture ».

Ces différents projets sont trop récents pour en tirer un premier bilan. Un seul exemple peut néanmoins démontrer l'intérêt de rendre un corpus disponible selon les pratiques en vigueur dans le domaine du web de données. Une recherche sur le terme « abattoirs » permet, par l'outil data.bnf.fr de signaler, d'écouter et de télécharger l'enregistrement d'ESLO consacré à l'entretien d'un boucher d'Orléans, et la même requête sur

[23] https://www.ortolang.fr/

ISIDORE permet de trouver une correspondance entre cet enregistrement et un entretien sur le même thème réalisé par des sociologues à Toulouse dans les années 1960.

Conclusion

Le corpus des ESLO a été réalisé par des linguistes et il a donné lieu à de très nombreux travaux en linguistique. Après les différents travaux en phonologie, syntaxe, prosodie, lexique et autres domaines engendrés par ESLO1, l'équipe d'ESLO2 réalise différentes études directement issues d'une analyse du corpus ou fondées sur une comparaison avec d'autres corpus[24].
A partir d'ESLO1 une méthode d'apprentissage des langues particulièrement innovante[25] a été réalisée et des travaux sont en cours de réflexion dans le cadre d'un usage didactique du corpus ESLO2.
On peut donc considérer que l'objectif d'obtenir un portrait sonore d'une communauté d'auditeurs d'une même ville est une source importante d'études linguistiques et d'applications liées.
Il convient néanmoins d'être prudent, ce portrait sonore ne peut se résumer à des enregistrements divers et variés sans un cadre théorique qui fait de la linguistique de corpus une discipline qui doit entendre autant si ce n'est plus, la sociolinguistique que la linguistique outillée par l'informatique.
Le tournant des *humanités numériques* est l'occasion de repenser cette définition de la linguistique sur corpus afin de définir une véritable science des données linguistiques. Face à ce défi, le linguiste doit maitriser l'ensemble de la chaine qui le conduit à travailler, exploiter et diffuser ces données collectées qui ne lui sont jamais « données ». Il est aussi important qu'il prenne conscience que cette science relève d'un domaine au sein duquel il n'est pas le seul acteur.

[24] Comme par exemple les travaux sur la liaison dans ESLO, PFC et d'autres corpus (Dugua et Baude, à paraitre).

[25] Biggs P. & Dalwood M. 1976, *Les Orléanais ont la parole.*

Références bibliographiques

Site ESLO : http://eslo.huma-num.fr

Abouda L. & Baude O. (2009). « Du Français Fondamental aux ESLO », *in* Bruxelles, Mondada, Simon, Traverso (ed.) *Grand corpus de français parlé, Bilan historique et perspectives de recherche*. Cahiers de Linguistique Revue de sociolinguistique et de sociologie de la langue française 33/2, EME, Louvain, 131-146.

Abouda L. & Baude O. (2007). « Constituer et exploiter un grand corpus oral, choix et enjeux théoriques : le cas des ESLO », in actes du colloque Corpus en lettres et sciences sociales, *Des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification juin 2006, Presses universitaires de Toulouse: 161-168.

Baude O. & Bergounioux G. (à paraitre 2015). « L'ESLO : une enquête en son temps » *in Linguistique de corpus : une étude de cas La recette de l'omelette*, Champion.

Baude O. & Lacheret A. (à paraitre 2015). "The collection of data for the Rhapsodie Treebank: typological criteria and ethical issues" in A. Lacheret, S. Kahane & P. Pietrandrea (ed.) *Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French, col Studies in Corpus Linguistics*. Amsterdam, Benjamins.

Baude O. & Dugua C. (2011). « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? » Corpus 10, *Varia*, 99-118.

Baude O. coord. (2006). *Corpus oraux, guide des bonnes pratiques*, Paris et Orléans, Editions du CNRS et Presses Universitaires d'Orléans.

Bergounioux G., Baraduc J. & Dumont C. (1992). « L'étude socio-linguistique sur Orléans (1966-1991) : 25 ans d'histoire d'un corpus », *Langue française* 93 : 74-93.

Biggs P. & Dalwood M. (1976). *Les Orléanais ont la parole : Teaching Guide and Tapescript*, Londres, Longman (Livre du maître).

Biggs P. & Dalwood M. (1976). *Les Orléanais ont la parole*, Londres, Longman (Livre de l'élève).

Blanc M. & Biggs P. (1971). « L'enquête socio-linguistique sur le français parlé à Orléans », *Le français dans le monde* 85 : 16-25.

Blanche-Benveniste C. et alii (1990). *Français parlé. Etudes grammaticales*. Paris : CNRS.

Bourdieu P., Chamboredon J.-C. & Passeron J.-C. (1968). *Le métier de sociologue*. Paris : Mouton de Gruyter/Bordas.

De Jong D. (1988). *Sociolinguistic aspects of French liaison, Academisch proefschrift*. Amsterdam : Vrije Universiteit Amsterdam.

Bourdieu P. (1984). « Le marché linguistique », *Questions de sociologie*, éditions de Minuit, Paris.

Equipe DELIC (2004). Recherches sur le français parlé n° 18, Publications de l'université de Provence, 265 p.

Encrevé P. (1976). « Présentation », *in* W. Labov, *Sociolinguistique*, éditions de Minuit, Paris.

Eshkol-Taravella I., Baude O., Maurel D., Hriba L., Dugua C. & Tellier I. (2012). « Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012 », in *Ressources linguistiques libres, TAL*. Volume 52 – n° 3/2011, 17-46.

Habert B., Nazarenko A. & Salem A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin.

Jacobson M. & Baude O. (2012). « Corpus de la parole : collecte, catalogage, conservation et diffusion des ressources orales sur le français et les langues de France », in *Ressources linguistiques libres, TAL*. Volume 52 – n° 3/2011, 47-69.

Koch P. & Oesterreicher W. (2001). « Langage oral et langage écrit », *Lexicon der Romanistischen Linguistik*, tome 1-2, Tübingen, Max Niemeyer, 584-627.

Laks B. (2013). « Why is there variation instead of nothing », *Language Sciences* 39: 31-53

Labov W. (1976). *Sociolinguistique*, Paris : Editions de Minuit.

Le Deuff O. dir (2014). *Le temps des humanités digitales*, Limoges : FYP éditions.

Lonergan J., Kay J. & Ross J. (1974). *Etude sociolinguistique sur Orléans, catalogue des enregistrements*. Colchester : Multigraphié.

Mertens P. (2002). « Les corpus de français parlé ELICOP : consultation et exploitation », in J. Binon et al. (eds) *Tableaux Vivants*. Opstellen over taal-en-onderwijs aangeboden aan Mark Debrock. Leuven : Universitaire Pers.

Mullineaux A. & Blanc M. (1982). « The problems of classifying the population sample in the socio-linguistic survey of Orléans (1969) in terms of socio-economic, social and educational categories », *Review of Applied Linguistics* 55 : 3-37.

Titre	***L'ESLO : une enquête en son temps***
Type	Chapitre
Editeur	Champion
Année	A paraitre
Référence	Baude, O., Bergouniuox, G., (à paraitre), chapitre « L'ESLO : une enquête en son temps » in *Linguistique de corpus : une étude de cas La recette de l'omelette*, Champion

LINGUISTIQUE DE CORPUS
UNE ÉTUDE DE CAS

LA RECETTE DE L'OMELETTE DANS L'ENQUÊTE SOCIO-LINGUISTIQUE À ORLÉANS (ESLO)

Gabriel Bergounioux

&

Olivier Baude, Annie Chesneau, Gilles Cloiseau, Céline Dugua, Iris Eshkol

2. ESLO, UNE ENQUÊTE EN SON TEMPS : ENJEUX, MÉTHODES ET RÉSULTATS

2.1 INTRODUCTION

ESLO est apparue à la fin des années 1960 pour combler un manque : l'absence de corpus fiable concernant le français parlé (Bergounioux, 1992). A ce titre, cette enquête a constitué un exemple isolé pendant des années ; elle est devenue, depuis 1970, une œuvre pionnière que d'autres ont prolongé avec, pour s'en tenir au cas du français parlé, les travaux entrepris au Canada, en Belgique et, en France à Aix, à Lyon, à Paris et Nanterre ainsi que dans la constellation des équipes impliquées dans le programme de « Phonologie du Français Contemporain » (PFC). Couramment désignée aujourd'hui comme ESLO1 pour la distinguer de la nouvelle enquête entreprise sur le même terrain (ESLO2), elle continuera à figurer dans ce livre sous la forme ESLO sans risque de confusion.

La création d'ADONIS et d'un organisme tel que Corpus avec, en particulier pour les corpus oraux et multimodaux, le consortium IRCOM adossé à la fédération de recherche Typologie et Universaux Linguistiques (TUL), a conféré à présent une meilleure visibilité aux données numériques collectées par les laboratoires.

En préambule à l'exploitation d'une toute petite partie du corpus (moins de 1 %), on restituera ESLO dans son contexte en marquant sa place dans l'histoire des études sur le français parlé, en présentant les motivations des chercheurs qui en sont à l'origine et en rappelant quelques-unes des perspectives ouvertes par cette initiative.

2.2 PARLER ET ÉCRIRE (LIRE ET ENTENDRE)

L'histoire des recherches sur les langues est inséparable de leur écriture. Le nom même de *grammaire*, qui contient en lui le nom grec de *lettre*, d'*inscription*, le dit. L'avènement de la linguistique, issue d'une tradition plurimillénaire, est d'une certaine façon le récit de son arrachement à l'écrit, la reconnaissance que les langues sont parlées, qu'elles ont une forme sonore.

L'écriture advient à un peuple comme un événement *historique* dans toute l'acception du terme puisque le fait qu'il y ait de l'écrit produit de la connaissance historique comme savoir réflexif et comme objectivation en même temps qu'il divise le cours des temps entre une « préhistoire » et l'Histoire. L'écriture assure la pérennisation, la continuité dans le temps d'une formation sociale. Elle induit en même temps une transformation radicale des cadres de représentation collective qui déplace autant les modes de conservation des savoirs et leur apprentissage que la reproduction de l'ordre social et le fonctionnement économique ou le jugement esthétique. On a insisté sur l'accumulation des connaissances que permet l'écriture, sur la mémorisation qu'elle assure, mais elle est aussi une leçon d'amnésie. On se repose sur elle pour n'avoir plus à apprendre par cœur des épopées et des généalogies, des secrets de fabrication et des classements botaniques. A. Lord (1960) et M. Parry (1987) ont donné un aperçu de la poésie orale qui s'effaçait dans le monde slave au fur et à mesure que l'alphabétisation y progressait. Ce qu'efface en premier l'usage de l'écriture auprès de ceux qui s'en rendent possesseurs, c'est un rapport à la langue qui ferait abstraction de cette compétence, qui percevrait la parole sans recourir à l'image qu'en procure la transcription.

La situation actuelle du chercheur emprunte ses caractéristiques à la structure du mythe, au renvoi vers un passé irréversiblement perdu d'une relation immédiate des hommes à un environnement qui leur aurait été transparent, motivé. Au moment où elle accède à un instrument irremplaçable de représentation de sa langue et de son passé, de ses connaissances et de sa poésie, une formation sociale perdrait, par le fait même, sa capacité à préserver la solidarité de ses ressources et des moyens qui les expriment, désormais contraints d'emprunter, dans tous les emplois sanctionnés et valorisés, le truchement d'un support et d'un tracé qui sont l'apanage d'une classe de professionnels.

Que l'histoire des écritures soit orientée téléologiquement vers la phonétisation ne rend pas obligatoire l'usage de l'alphabet. Le développement de systèmes d'écritures idéographiques ou syllabaires montre qu'il existe d'autres solutions. Néanmoins, ce sont les alphabets qui, en asservissant l'écrit à l'oral, comme une copie imagée des paroles, ont contribué à la confusion de la langue et de sa notation et à un délaissement des investigations sur la parole réduites aux mises en voix (l'art du discours, l'éloquence…). Ceux qui ont en charge l'enseignement, la défense de la tradition et de la doxa, ont partie liée avec le privilège accordé à la graphie sur les usages parlés, un rôle qui s'accommode d'une posture conservatrice. Les maîtres de la langue légitime sont les premiers à oublier que les langues n'ont aucun besoin d'être écrites, que ce sont les sociétés qui éprouvent la nécessité de recourir à cet artifice.

C'est aussi par l'écriture que la révision du jugement sur l'étude scientifique des langues s'est opérée, au terme d'un détour inattendu par le sacré. Alors qu'autour de la Méditerranée les dogmes religieux s'étaient fondés sur un Livre souvent établi à partir de transcriptions – parole prophétique, évangile, dictée coranique – et sur une conception du salut, en Inde, au contraire, étaient consignées non des propos tenus mais des paroles à dire, des prières destinées à s'attirer la bienveillance de divinités pointilleuses sur l'exécution des rites (Malamoud, 1989), en particulier sur la prononciation des adresses. De l'obligation, dans le védisme, d'articuler les mots d'une façon immuable et conforme pour que la sollicitation soit recevable par les dieux, ont résulté des descriptions phonétiques d'une qualité inégalée en Europe (Pinault, 1989). Le sanskrit, déchiffré et transmis vers l'Occident, a joué un rôle décisif dans la nouvelle façon d'appréhender les langues qui a préludé à la linguistique comparée. Paradoxalement, cette connaissance de l'oral portait sur des langues qui n'étaient plus d'usage ; elle visait leur état le plus ancien, adoptant le modèle proposé par l'étude du sanskrit jusqu'à se donner pour objectif la reconstruction d'une forme si archaïque qu'il ne peut s'en découvrir aucun témoignage : le proto-indo-européen.

Comme l'ont souligné en leur temps H. Schuchardt (2011) et F. de Saussure, dans leur relation critique aux Junggrammatikers, une conception mieux ajustée à la pratique des langues s'est imposée à travers l'étude des langues romanes. La disponibilité des sources, la proximité géographique et la continuité des transformations en synchronie appelaient une observation moins distante et mieux documentée. Les études dialectologiques ont permis de restituer les variétés d'évolution du latin dans la Romania et les circonstances pour en rendre compte impliquaient un affranchissement des formes écrites. En particulier, la dialectologie est par nature associée à la phonétisation puisqu'il s'agissait, en première approche, de déterminer des écarts de réalisation faibles par rapport à une langue standard, officielle, de restituer les« nuances » entre les différents « patois », comme le dira Gaston Paris (1888), de reporter sur la carte des différences sonores que l'orthographe annule.

Dès avant, comme le relevait Michel Bréal préfaçant la traduction de la *Grammaire comparée* (Bopp, 1866), le différend entre Grimm et Bopp sur le vocalisme des langues indo-européennes et le rôle de l'apophonie trouvait sa source dans leur approche respective. Grimm s'intéressait à la culture orale, aux contes, au folklore, aux formes parlées, aux dialectes de l'allemand. Bopp concentrait son attention sur la restitution approximative de langues mortes qu'il s'efforçait de déduire de leur écriture vernaculaire.

2.3 REPRÉSENTER LES LANGUES

Pour surmonter la difficulté que présente l'objectivation des langues, ou plutôt des discours par quoi elles se manifestent sans que d'autre trace ne s'en conserve que le souvenir labile des auditeurs, il y a eu l'écriture. Pour en assurer une préservation qui ne dépende pas des conventions orthographiques d'une langue donnée mais tente d'en conserver l'expression, trois techniques interfèrent qui ont été exploitées parallèlement : (i) la production d'une image du signal, (ii) l'enregistrement et (iii) la transcription des timbres et des articulations.

L'enregistrement commence avec le phonautographe qui reproduit seulement l'image du signal. C'est le point de départ d'une histoire des techniques aujourd'hui bien décrite qui, à partir de la découverte d'Edison (1877), aboutit à des appareils utilisables pour la prise du son et sa restitution. L'usage des enregistrements magnétiques, de l'utilisation professionnelle en studio dans les années 1940 à la production de masse dans les années 60, a eu des conséquences importantes dans le travail d'enquête et dans les applications didactiques. ESLO en est l'illustration.

La production d'images du signal est liée en France au nom de Marey et surtout de l'abbé Rousselot qui fonde un laboratoire, transféré de l'Institut Catholique de Paris au Collège de France sous l'autorité de M. Bréal (1897). Dans ses *Principes de phonétique expérimentale* (1897-1901), Rousselot recense l'ensemble des moyens mécaniques qui permettent d'analyser et de fixer une représentation du langage en s'inspirant des principes posés par Helmholtz. Bien qu'il se soit consacré au traitement des aspects articulatoire et acoustique pendant une quarantaine d'années, Rousselot a commencé sa carrière de linguiste par une thèse de dialectologie (1891) et c'est à lui qu'il a été fait appel pour fixer l'alphabet de la *Revue des Patois Gallo-Romans* qu'utilisera Gilliéron dans l'*Atlas Linguistique de la France* et qui est resté d'usage courant chez les romanistes français.

La notation des sons du langage a été et demeure le premier souci d'un linguiste de terrain. C'était en particulier le cas pour ceux qui s'astreignaient à transcrire des langues menacées dans leur existence même par le déploiement d'une colonisation quand des immigrants européens s'appropriaient des territoires dont ils refoulaient les autochtones. Le travail accompli par F. Boas et ses élèves, en particulier E. Sapir, mais aussi, toujours aux Etats-Unis, par L. Bloomfield, recoupait les préoccupations des linguistes russes qui parcouraient le Caucase ou la Sibérie. À la suite, la présentation faite par le Cercle de Prague devant le Premier Congrès International des Linguistes à La Haye (1928) venait répondre au questionnaire adressé aux participants sur la meilleure façon de produire des transcriptions.

Parallèlement, une autre demande émergeait, liée au développement des relations internationales et à l'extension de la scolarité dans les pays industrialisés. Encouragés par les demandes des professeurs de langues vivantes, en rupture avec la domination des études classiques, trois phonéticiens s'engageaient dans la réalisation d'un système de notation. L'initiative de W. Viëtor, sous le pseudonyme de *Quousque tandem*, relayée en Angleterre par H. Sweet et en France par P. Passy s'est conclue par l'adoption de l'Alphabet Phonétique International (API) dans les années 1880 afin de promouvoir un enseignement oral des langues vivantes.

Pas plus que l'écriture alphabétique, ni l'enregistrement, ni l'image instrumentale, ni l'API ne sont des instruments suffisants pour étudier les langues. Les méthodes d'enregistrement reproduisent ad libitum l'émission sonore, en font durer la trace et assurent la décomposition du signal ou la catégorisation de ses paramètres mais elles n'en permettent pas directement la connaissance en tant que forme symbolique. Les tracés sonores ne donnent pas accès aux phonèmes, aux coupes syllabiques, encore moins aux variations morphologiques. Enfin, l'API, couramment utilisé aujourd'hui, peine à différencier ce qui est d'ordre phonétique et ce qui est phonologique, entre norme et usage, partagé entre la forme conventionnelle et la reproduction acoustique, entre les contraintes articulatoires du classement et les variations intrinsèques. Si l'enseignement secondaire a eu recours de façon croissante à l'API, c'est à proportion de la distance entre les représentations graphiques et les productions vocales, pour compenser les inconséquences de l'orthographe anglaise notamment. Adopté pour des raisons de commodité, pédagogique et économique, le recours en API à l'alphabet roman a brouillé un peu plus la démarcation de l'écrit et de l'oral.

2.4 LE FRANÇAIS PARLÉ

Le français parlé, exclu des emplois administratifs, est attesté à l'écrit, qu'il apparaisse à des fins carnavalesques (Rabelais, les écrits poissards), politiques (les mazarinades, *Le Père Duchêne*), littéraires (Molière, Hugo) ou didactiques (la stigmatisation des formes non standard). Quelque soit la

façon dont il est consigné, l'oral est exploité en vue d'une fin qui lui est extérieure. Ce n'est pas pour livrer un témoignage verbal que les textes sont publiés mais pour énoncer une critique politique, pour faire rire les spectateurs des maladresses des classes dominées ou pour inculquer aux élèves les bonnes façons de s'exprimer. Aussi, la rédaction en est-elle souvent stylisée, dans ses thèmes, ses tours, son lexique et jusque dans les artifices typographiques qui la démarquent. Le constat valait déjà en ancien français (Cerquiglini, 1981).

Avec la généralisation de l'alphabétisation et la constitution d'une science du langage, un regard différent est porté sur une paysannerie que le développement de l'industrie marginalise et sur les « classes dangereuses » qui se concentrent dans les villes. L'idée de recueillir du français populaire s'est confondue, au XIX^{e} siècle, avec la composition de textes et de lexiques consacrés à l'argot, un intérêt précédé par une exploitation littéraire du procédé, exemplifié dans *Les Mystères de Paris* d'E. Sue (1842-1843), et avec l'étude des « patois » appréhendés dans leur état ancien (Raynouard, 1816-1821) et rédimés en proportion inverse de la dévaluation relative de la rente foncière (Bergounioux, 1994).

Dans l'attention accordée aux parlures populaires, argots et patois, le français ordinaire, ravalé par la norme littéraire inculquée à l'école, est d'emblée écarté. Gilliéron et Edmont laissent en blanc sur leur atlas Paris et sa région et F. Brunot, au moment de la création des Archives de la parole, entreprend un inventaire dans les Ardennes, reprenant l'étude dialectale de Charles Bruneau (1913), puis se dirige ensuite vers le Centre et le Limousin, sans graver sur les cylindres Pathé des échanges quotidiens en français à l'exception d'un très court dialogue avec un menuisier (Cordereix, 2006).

La demande sociale aurait pu provenir de l'enseignement du français langue étrangère en un temps où les voyages et les séjours linguistiques ne facilitaient pas la connaissance des langues. Les cours phonographiques semblent n'avoir pas produit plus de résultats que les cours radiodiffusés. La demande concernait préférentiellement l'usage d'un registre très soutenu, très artificiel aussi, comme on l'entend dans les enregistrements de pièces de théâtre ou d'émissions de la TSF avant 1960.

L'intérêt pour le français ordinaire emprunte une voix frayée aux Etats-Unis trente ans auparavant pour l'établissement des fréquences verbales. La collecte du Français Fondamental, motivée par le succès du BASIC English, centre son attention sur le lexique mais innove en abandonnant le dépouillement des textes à quoi sont substituées des enregistrements commandés pour l'occasion. Réalisée par le CREDIF au cours des années 50 pour établir la liste des mots les plus courants dans des échanges quotidiens (Gougenheim *et al.,* 1956), l'enquête témoigne rétrospectivement d'une certaine naïveté sociologique et la technique de capture du son a été, pour des raisons économiques, très sommaire. Comme l'objectif était de constituer un vocabulaire minimal, à la rigueur une syntaxe élémentaire, aucune attention n'a été portée aux données sonores et les enregistrements, effectués sur un support de piètre qualité, ont été effacés au fur et à mesure de l'exécution de transcriptions qui ne s'embarrassaient pas de relever des variations. La modestie des consignes données contraste avec l'importance des recommandations que l'équipe d'Aix-en-Provence prodiguera (Claire Blanche-Benveniste, 1987). L'ampleur des moyens engagés, en dépit de ses prolongements dans la diffusion du FLE, aura peu de répercussions sur la description du français vivant et sur les méthodes des linguistes en France, passé une polémique vite éteinte (Cohen, 1955).

Tel est le constat de carence que dressent les universitaires anglais qui, à la fin des années 60, se proposent de confectionner un manuel du français à l'usage de l'enseignement secondaire en se fondant sur des enregistrements. Ils ont tôt fait de constater qu'il leur était impossible de se procurer les matériaux dont ils avaient besoin dans les fonds d'archives alors que leur préoccupation n'était plus l'établissement de statistiques lexicales mais l'accès à la diversité des réalisations du français parlé auxquelles seront exposés les lycéens britanniques. Au nombre des dissymétries entre l'enquête du Français Fondamental et ESLO, on peut relever l'inégale légitimité des acteurs : un parrainage par des linguistes parisiens reconnus (A. Sauvageot, G. Gougenheim) avec le soutien de l'Education nationale d'un côté, des professeurs de langue étrangère issus de facultés qui ne sont pas les plus prestigieuses.

2.5 L'ÉQUIPE D'ESLO ET SON ENQUÊTE

L'Angleterre des années 60 est traversée par un courant de rénovation dans la vie politique porté par l'aile gauche du Labour Party qui, avec la fin de l'empire colonial, tourne ses regards vers l'Europe. Cette orientation a des répercussions sur le système d'enseignement. La scolarisation secondaire, dominée traditionnellement par le modèle des *public schools*, s'ouvre à tous et des allocations d'études sont proposées aux étudiants pour lutter contre la sélection sociale. Il s'ensuit une demande pour une pédagogie rénovée de l'apprentissage des langues qui, refusant le dilemme entre des applications immédiates au domaine commercial et l'acquisition d'un signe électif d'appartenance aux classes dominantes, devient un élément central de la culture transmise par le système éducatif. A l'opposé de la prédilection pour l'écrit et les auteurs classiques au principe des cours dispensés jusqu'alors, la connaissance d'au moins une langue étrangère à destination de publics moins sensibles au prestige de la tradition doit se faire avec des méthodes modernes et des contenus modernes selon le jugement de quelques enseignants du supérieur impliqués dans la formation des professeurs.

2.5.1 « PORTRAIT SONORE D'UNE VILLE »

La forme que prend l'Enquête Socio-Linguistique à Orléans s'est décidée dans l'interaction d'un mouvement d'aggiornamento pédagogique et d'une nouvelle conception sociologique. L'ambition d'intégrer à la fois une nouvelle technologie (le magnétophone), de nouveaux contenus (le français ordinaire sous sa forme orale) et de nouveaux objectifs (le développement d'une compétence communicative en situation réelle) est portée par des universitaires qui entendent accompagner dans leur pratique leurs collègues en poste dans les établissements du second degré. Au-delà de la réalisation de produits pédagogiques, l'importance accordée à la recherche concerne essentiellement le versant sociologique. Les conceptions linguistiques sont assez traditionnelles, dans la construction du questionnaire comme dans la présentation de l'ouvrage. En revanche, les promoteurs sont réceptifs aux idées de Bernstein (1975) sur la construction des énoncés, les formes de raisonnement, l'opposition entre un « code restreint » et un « code élaboré ». L'orientation progressiste du projet est sensible au travers de questions très ciblées (sur l'enseignement du latin, l'école privée, Mai 68 par exemple) comme dans les objectifs généraux ou le fonctionnement collectif de l'équipe.

Le titre « Portrait sonore d'une ville » résume plusieurs intentions. La ville se situe à l'opposé de l'enquête dialectale sur un terroir par le choix d'une agglomération en pleine croissance, où le brassage des populations dilue la transmission endogène. Avec le son, on s'affranchit au moins partiellement de l'écrit, de l'effacement de la variation à quoi avaient abouti les transcriptions. Avec le« portrait », on renonce à mettre l'accent sur un état de langue pour concentrer l'attention sur les locuteurs, leurs interactions. Pourquoi avoir fait choix d'Orléans ? Les raisons ont été explicitées par les auteurs eux-mêmes. Il s'agissait de recenser une réalisation du français qui ne soit pas identifiée à un accent régional marqué, dans une ville d'une certaine importance qui venait de rouvrir son université, bien reliée à Paris et qui n'en soit pas trop distante pour des raisons logistiques.

Les témoignages pour obtenir une photographie des différentes compétences langagières ont été sollicités par le biais de l'INSEE qui a fourni à l'équipe un échantillon de la population urbaine constitué par tirage au sort à partir de ses fichiers, combinant des critères d'âge, de sexe et de catégorie socio-professionnelle (CSP). La désignation aveugle des témoins aura pour effet une distorsion des réponses : l'inégale légitimité à se concevoir comme un représentant attitré de l'usage du français conduira la plupart des personnes sollicitées dans les milieux modestes à se récuser.

Pour compléter les entretiens en face à face et varier les situations d'échange, d'autres témoignages sont sollicités (ou obtenus par micro caché) qui font intervenir des situations plus formelles (interviews de personnalités) ou plus spontanées (visite, transaction commerciale) et d'autres locuteurs. La série des enregistrements répertoriés joue sur la variété des usages en exploitant

différents paramètres : (i) les protocoles d'enregistrement, (ii) les situations d'entretien (débats, dîner de famille, appel téléphonique…), (iii) la familiarité relative avec les témoins (dans les reprises de contact) et bien sûr (iv) les indicateurs sociologiques, qu'ils aient fait l'objet de la requête, comme la CSP, ou non, comme l'origine géographique des témoins qui n'est pourtant pas neutre en ce domaine.

Les enquêteurs ont concentré l'essentiel de leur collecte sur le premier semestre 1969, associant à leur démarche des assistants et des doctorants français et anglais. L'inadéquation d'une partie du questionnaire les a conduits à apporter dans l'urgence des réponses. On relève leur satisfaction à découvrir, dans le panel, des locuteurs originaires de régions périphériques (Lorraine, Pyrénées, rapatrié d'Afrique du nord…) mais aucune n'est originaire de l'outre-mer ou d'un pays étranger.

Le développement le plus intéressant concerne le glissement dans les critères sociologiques où est substituée, de façon hésitante mais cohérente, aux cadres déterministes de Bernstein, une différenciation entre ce que P. Bourdieu, rapidement consulté, caractérisera ultérieurement comme la distinction que produit la composition relative du capital économique et du capital culturel dans la définition sociale d'un agent. L'échelle AM (pour Alix Mullineaux, une chercheuse associée au groupe dont les propositions figurent en appendice dans le *Catalogue*) pondère le classement INSEE en prenant en considération l'âge de fin d'étude, correspondant dans le capital culturel à la part généralement déterminante du capital scolaire.

Si ESLO n'a pas joué un rôle déterminant dans les principes de la sociolinguistique, une révision venue plutôt des Etats-Unis avec W. Labov dont les publications fondatrices sont contemporaines (1966), l'ensemble constitue plus qu'un document d'une qualité scientifique exceptionnelle, une nouvelle pratique méthodologique, la en linguistique de corpus.

2.5.2 LE QUESTIONNAIRE

Le questionnaire est long (il faut en moyenne une heure pour l'administrer), hétérogène avec ses diverses « branches », si maladroit parfois que certaines questions ont dû être abandonnées (par exemple sur l'utilisation d'un guide-âne dans la rédaction de la correspondance), intrusif aussi quand il s'aventure sans ménagement excessif sur le terrain politique. Néanmoins il constitue la meilleure source d'information, et la plus fiable, sur le français parlé il y a un demi-siècle. C'est en tout cas le premier explicitement conçu à des fins linguistiques qui ait pris en compte la variation et qui ait eu le souci de préserver les métadonnées tout en protégeant l'anonymat des témoins.

Le questionnaire préparé pour les entretiens en face à face se compose de trois séquences :

- une déclinaison des identifiants hors micro (dit « questionnaire fermé »),
- une série de questions posées à tous les témoins, dites « tronc commun »et
- des questions facultatives, dites « branches », sollicitées en cas d'intérêt supputé de l'interviewé pour un des thèmes abordés dans les séquences.

Dans la branche dénommée « langue et culture », la dernière est formulée ainsi :

> Comment est-ce qu'on fait une omelette ? Pourriez-vous m'expliquer comment on fait ? / Pouvez-vous me donner la recette de l'omelette ?

C'est la seule question qu'on trouve parfois précédée d'un commentaire de l'enquêteur annonçant qu'il ne s'agit pas forcément d'une question « sérieuse ». Ce volume est consacré en entier aux réponses à cette question, dans la conviction qu'un observatoire très précis des pratiques linguistiques constitue un apport à la connaissance des langues dans leur usage.

2.6 CONCLUSION

Le plus remarquable dans ESLO est sa singularité. L'enquête, saluée en son temps pour sa nouveauté et son ambition, a connu des exploitations qui lui ont valu, surtout en Angleterre, une réputation certaine. Pourtant, elle n'a eu aucune filiation directe. On peut en chercher l'explication dans l'absence de lien direct entre l'équipe anglaise et des chercheurs français qui s'inscrivaient dans une tradition plus marquée par les orientations de Marcel Cohen ou avec l'équipe du GARS (Groupe Aixois de Recherche en Syntaxe) qui, autour de Claire Blanche-Benveniste, a choisi de se consacrer plutôt aux questions de transcription et à une réflexion sur la grammaire.

Les corpus oraux sur le français n'ont pas suivi les leçons de cette première initiative qui apparaît sans continuité jusqu'au lancement d'ESLO2.

Gabriel BERGOUNIOUX & Olivier BAUDE

Titre	***The collection of data for the Rhapsodie Treebank: typological criteria and ethical issues***
Type	Chapitre
Editeur	Benjamins
Année	2015
Référence	A. Lacheret, P. Pietrandrea, O. Baude, A. C. Simon (accepté pour publication) "The collection of data for the Rhapsodie Treebank: typological criteria and ethical issues" in Lacheret A., Kahane S., Pietrandrea P., *Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French*, col Studies in Corpus Linguistics, Amsterdam, Benjamins.

The collection of data for the Rhapsodie Treebank: typological criteria and ethical issues

1. Introduction

The *Rhapsodie* Treebank was created with a main objective: to propose and to test on the widest range of structures as possible new instruments of annotation and analysis for modeling the prosodic-syntactic interface in spoken French.
Such a research objective had an immediate consequence for the design of our corpus. We aimed at gathering a collection of samples of spoken French sufficiently diversified as for textual typology.
Under this respect we could only poorly build on previous experiences of spoken Treebank construction. The majority of existing Treebanks are not diversified as textual typology: the Verbmobil Treebanks of English, German and Japanese (Hinrichs et al. 2000), for example, only comprise samples of spontaneous conversations, the Spoken French Ester Treebank (Cerisara et al. 2010) is only based on manual transcripts for French radio news; the Venice Treebank (Delmonte et al 2007) only includes (regionally diversified) Italian conversations and the Switchboard corpus (Meteer et al. 1995), only telephone conversation.
The design of the CHRISTINE Treebank (Sampson 2000) aimed at providing a more diversified and balanced sampling based on extracts from the "demographically-sampled" speech section of the British National Corpus. Such design choice guaranteed a sociolinguistic diversification of the corpus, rather than a textual typology diversification.
Particular attention to structural variety was paid by the creators of the British component of the International Corpus of English (Nelson/Wallis/Aarts 2002), of the Diachronic Corpus of Present-Day Spoken English. (Wallis et al 2006) and of the CNG Spoken Dutch Corpus (Schuurman et al. 2004). These three large Treebanks are all based on the sampling principles established by the International Corpus of English, which recommend to include in the corpus an equal number of monologues and dialogues, public and private speeches, scripted and unscripted speeches.
Our work was largely inspired by the ICE-based sampling experiences, but we met a further difficulty. While most spoken Treebanks were extracted from existing representative corpora, or at least built within the frame of large projects dealing with corpus construction, the creation of the Rhapsodie repository could not rely on these conditions. No representativecorpusok spoken French is available to present which could function as a basis for the selection of samples to be annotated in the Treebank. Neither we could , in the frame of our project, counton enough time and money to ensure the collection of a new corpus.
This peculiar situation urged us to construct our textually diversified, well-balanced repository of samples by extracting excerpts from a number of relatively small and specific pre-existing spoken corpora of French.

To do that we had to answer a number of challenging questions, which led us to elaborate what we could consider as an innovative sampling strategy. First of all, we had to decide how to sample our corpus in order to ensure that a sufficient variety of textual typologies and a sufficient number of speakers be represented (section 2). Secondly, we had to define a number of linguistic criteria and linguistic hypotheses on which to base the selection of the source corpora, the source samples, and the particular excerpts that alimented our repository (section 3). Thirdly, we had to define a procedure to acknowledge the reuse of source corpora, to refer to them and to ensure the possibility of retrieving the original samples (section 4). Finally we had to choose a metadata standard that allowed us to provide an exhaustive textual characterization of each sample, to provide complete information about source corpora and to precisely describe the annotations of each sample, which are at the heart of the Rhapsodie project (section 5).

2. Genres and speakers: how to maximize the diversity?

As mentioned above, our first concern in the creation of the corpus thus was to create a coherent repository of non-elicited[1] data including the largest number of structures as possible.

In order to achieve this goal we followed, in the collection of our data, Biber's et al (1999:4) hint suggesting that "the vocabulary and grammar that we use to communicate are influenced by a number of factors, such as the reason for the communication, the context, the people with whom we are communicating, and whether we are speaking or writing". In other words we assumed that a strict relation exists between textual typologies - i.e. textual patterns defined in intrinsically linguistic terms - and genres - i.e. socio-communicative patterns- in that a particular linguistic structure reflects a particular genre, and the other way around, a given genre engenders a limited number of linguistic structures (Bakhtine 1984, Swales 1990, Maingeneau 1996, Adam 1999). As for the specific objectives of our research we assumed that a correlation exists between discourse genres and the distribution of intonosyntactic structures, and by consequence that a corpus comprising a high variety of discourse genres also comprises a high variety of intonosyntactic structures. On these grounds, in the design of our corpus we pursued the objective of maximizing the variety of genres.

In order to maximize the variety of genres represented in our corpus we took into account the multifactorial nature of discourse genres. It is well known that discourse genres can be described as multifactorial phenomena involving a number of socio-communicative variables independent from one another (Biber et al 1999, Koch, P. and W. Oesterreicher 2001). A given discourse genre, for example, can be described in terms of the nature of the speech situation (its location, its goals, the degree of formality); in terms of the physical constraints the speech situation undergoes (in particular the type of channel of communication), or in terms of the topic, i.e. the semantic content of the exchange.

The design of the Rhapsodie corpus (table 1) was first of all based on the balance between:

(i) monologues (M), i.e., discourses produced by a single speaker addressed to interlocutors who could not freely take the turn of speech (whether a large audience or a single interviewer);

(ii) dialogues (D), i.e., produced by two or more speakers in a situation of either low or high interactivity.

Having included both monologues and dialogues in our corpus, we could take into account in the development of our annotations a number of phenomena typically related to interaction such as overlaps, turn taking or interactional discourse markers.

Secondly, following Bilger (2007), we slightly modified the traditional distinction between public and private speech, by introducing a tripartite distinction between private, public and professional speech.

(i) Private speech (Rhap-M0 or Rhap-D0) is composed of samples extracted either from face to face interviews between the linguist researcher and one or more French speakers, or from everyday life interactions. Private speech may cover any topic except business which is recorded in Professional speech.

(ii) Professional speech (Rhap-M1 or Rhap-D1) has the same format as private

[1] Following the Imdi Metadata initiative vocabulary, speech is elicited when investigator asks speaker(s) to produce isolated phonemes/ words/ utterances / grammatical structures. It may also be possible to *elicit* semi-spontaneous speech (planning type) if the consultant is asked to respond "as fast as possible without thinking".

speech (interviews or spontaneous conversations) but is focused on business activity: the speaker is recorded while he/she is working (teaching activity for exemple) or is speaking about his/her work or professional plans.

(iii) In Public speech (Rhap-M2 or Rhap-D2), the speaker addresses an audience (conferences, radio or televisions talks: political speech or debates, talk shows, scientific press, reportage, forecast, etc).

	Type of speech	**Number of samples**	**Speakers genre**		**Lenght (in seconds)**	**Number of words**
MONOLOGUES	**Private speech**	22	H	F	1246 s	3384
			10	13		
	Professional Speech	2	0	2	335 s	1180
	Public Speech	6	5	1	2506 s	4993
	Total monologues	30	15	16	4087 s	9557
DIALOGUES	**Private speech**	11	12	7	2258 s	8558
	Professional Speech	3	5	1	824 s	2027
	Public Speech	13	22	8	4132 s	14219
	Total dialogues	27	39	16	7214 s	24804
	TOTAL	57	52	35	11301s	34361

Table 1. A first view of the composition of the Rhapsodie corpus

Having collected private, professional and public speech samples was particularly useful to refine the prosodic annotation schema, since the distribution of prominences is quite sensitive to the degree of formality and affectivity of the exchange, which changes according to the type of speech.

Finally, we also took into account the following situational variables:

(i) the degree of planning of the speech (more or less spontaneous;

(ii) the degree of interactivity;

(iii) the channel of communication;

(iv) the type of discursive sequence characterizing the speech (description, vs. argumentations, vs. narrations, vs. procedures – see Adam 1999.

(v) The type of task: interviews, sermons, sportscasting, movie scene description, travel planning, etc.

We hypothesized that each of these variables could have an influence on the nature and complexity of syntactic and prosodic constructions, the presence of overlaps and dysfluences, the presence and distribution of phatic and deictic markers, as well as the mechanisms of thematic progression.

Table 2 summarizes the types of situational variables that have been taken into account in the construction of the Rhapsodie corpus

Event structure	Monologues, dialogues
Types of speech	Private, Public, Professional
planning	Spontaneous, semi-spontaneous, planned
interactivity	Interactive, semi-interactive, non-interactive
chanel	Broadcast ; face-to-face
Discursive sequence	Argumentation, narration, description…
Task	Info-kiosk, semon, lesson, professional project description...

Table 2. Situational variables in Rhapsodie

With the same objective of maximization of diversity, we decided to include in our corpus a large number of short samples containing speeches uttered by 89 Central French adult native speakers (males and females) from the early eighties to nowadays, for a total of 57 short audio samples (5 minutes long on average), amounting to 3 hours of speech and a 33 000 word corpus.

3. Gathering data: external and internal sources

Gathering new data with the aim of getting a diversified collection of textual samples would have been too costly and time-consuming for a project whose first aim was to develop instruments of annotation and analysis for spoken French. We preferred to build the bulk of our corpus (32 samples) by selecting data from 7 corpora of spoken French created in the last years for various scientific projects. Table 3 describes in detail the 7 external source corpora that alimented the *Rhapsodie*'s reservoir: CFPP2000, C-Prom, ESLO, PFC, the Avanzi Corpus, the Lacheret Corpus, the Mertens Corpus.

Source	**Description**	**Number of samples**	**Samples**
CFPP2000	The CFPP2000 (Le Corpus de Français Parlé Parisien) is made up of interviews about Paris districts and suburbs. It provides data to study Parisian French as is used in real communication (Branca-Rosoff et al 2012) http://cfpp2000.univ-paris3.fr/	4	D0001, D0002, D0004, D0006
Corpus Avanzi	The Avanzi Corpus was gathered for the intonosyntactic study of macrosyntactic phenomena conducted by Mathieu Avanzi for his Ph.D. dissertation at the University of Neuchatel, 2011 (Avanzi 2012)	19	M0001, M0003-M0017, D0007, D0008, D0020
Corpus Lacheret	The Lacheret Corpus was gathered by Anne Lacheret within the frame of her habilitation work, which focused on the continuous and functional modeling of French prosody (Lacheret 2003).	2	D2004, D2005
Corpus Mertens	The Mertens Corpus was gathered by Piet Mertens for his doctoral dissertation. Mertens's dissertation constitutes the first approach to the intonosyntactic modeling of French elaborated with the aim of developping an automatic system of identification of intonational units (Mertens 1987).	2	D2001, D2009
C-Prom	C-PROM is an aligned and annotated corpus developed with the aim of studying syllable prominences in French. It comprises 24	1	M2001

	recordings representing 7 different genres produced by French, Belgian and Swiss native speakers of French (Avanzi et al 2010) http://sites.google.com/site/corpusprom/		
ESLO	ESLO, L'Enquête Sociolinguistique à Orléans is a 300 hours, about 4500000 words corpus of spoken French gathered in Orleans, France in 1969-71 with a sociolinguistic aim. It includes 157 interviewes and more than 200 recordings of spontaneous private and professional conversations, telephonic exchanges, public meetings, and commercial trades http://eslo.tge-adonis.fr/ (Eshkol-Taravella et al 2010)	1	D1001
PFC	The international project PFC, directed by Marie-Hélène Côté (University of'Ottawa), Jacques Durand (ERSS, University de Toulouse-Le Mirail), Bernard Laks (MoDyCo, Université de Paris Ouest Nanterre la Défense) & Ch. Lyche (Oslo University) aims to obtain the more accurate picture of both the similarity and diversity of phonetic varieties of contemporary spoken French. The elements of the PFC database which were used for Rhapsodie sampling are directed conversations between a subject and an interviewer and informal conversations between two persons belonging to a dense social network. http://www.projet-pfc.net/ Durand et al (2009)	3	D0003, D0005, D0009

Table 3. External source corpora used in Rhapsodie

It should be noted that most of the *Rhapsodie* samples are shorter (between 1 to 10 minutes) than the original source files. We had to make the choice of selecting only a portion of the source file because of our objective of maximization of genre and speaker diversity (see § 2): in order to pursue this objective, indeed, we had to organize our reservoir in a collection of many shorts samples produced by many speakers. We are aware of the fact that this choice, which allows for fine and complete intonosyntactic analyses of each sample, is not perfectly suitable for complex semantic textual analyses. However, as will be shown in § 4, users can retrieve the whole original source files through an identifier that is given in the metadata associated to each sample.

In order to guarantee a balanced representativeness of all the situational variables listed in §2, we also collected 27 original samples. Table 4 describes in details the 3 original subcorpora collected by three PhD students, Anja Kaglik, Djamila Cherbal and Nicolas Obin for the Rhapsodie reservoir.

Subcorpora	**Description**	**Number of samples**	**Samples**
Movie description Corpus	The Movie description corpus gathers 7 monologues in which 7 different speakers are invited, in an informal setting, to describe a short scene of a Charlie Chaplin movie	7	M0002, M0022,M019, M018, M021, M023
Professional Corpus	The Professional corpus consists of 4 monologic and dialogic professional speeches	4	M1001, M1003, D1002, D1003
Broadcast corpus	The Broadcast corpus consists of a 14 broadcasted monologues, dialogues and conversations gathered on the Internet for the Rhapsodie project	14	M2004, D2003, M2002, M2006, M2003, D2007, D2008, D2010, D2006, M2005, D2012, D2011, D2013, D2002

Table 4. Internal sources created for the Rhapsodie project

describes in details the 3 original subcorpora collected by three PhD students, Anja Kaglik, Djamila Cherbal and Nicolas Obin for the Rhapsodie reservoir.

A synoptic view of the criteria adopted for gathering and sampling the Rhapsodie corpus is in figure 1.

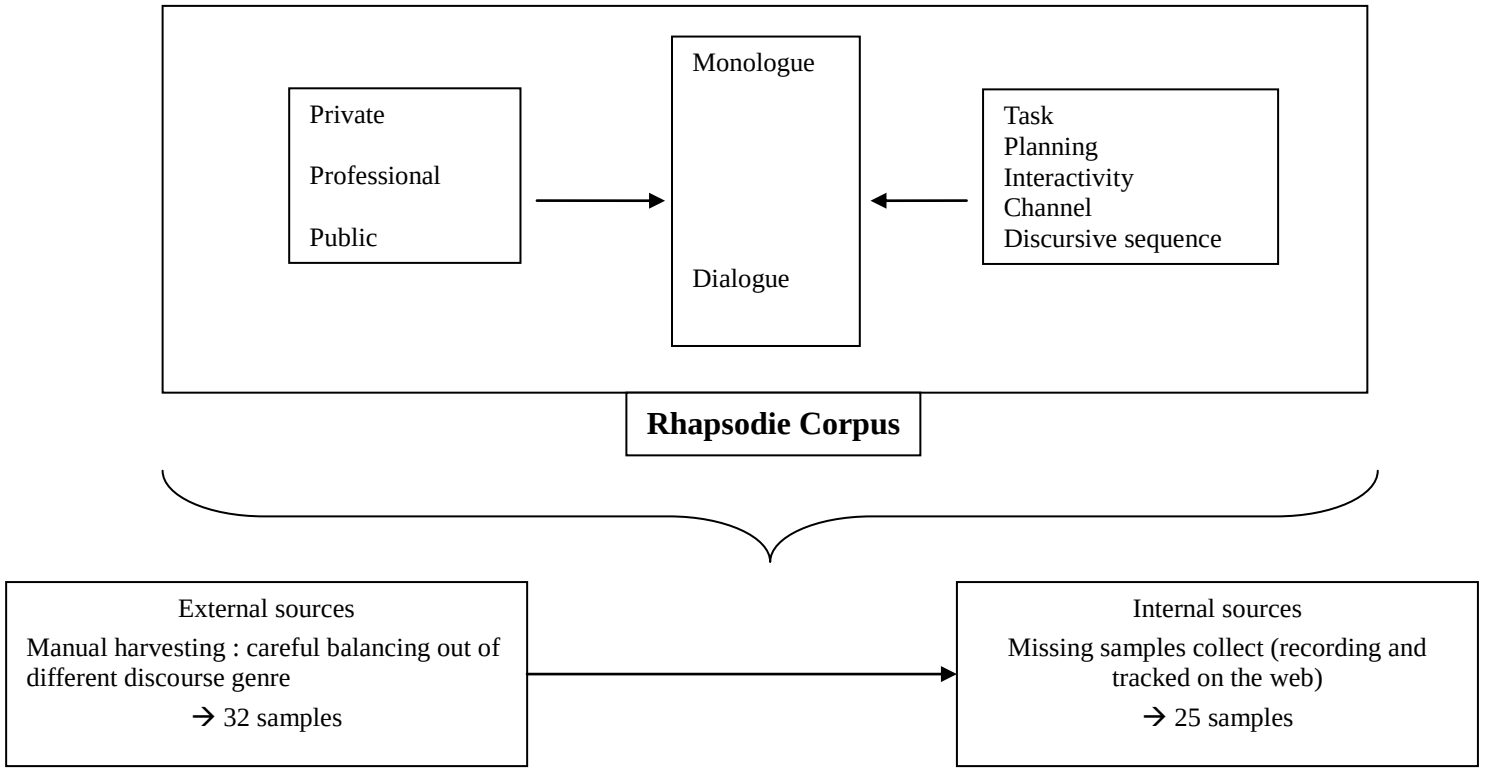


Figure 1.

Corpus design and data gathering in the *Rhapsodie* Project

4. Ethical issues

Digital formats along with the development of the Internet have greatly facilitated the dissemination of linguistics resources in the academic and scientific world. Ethical and legal points concerning recorded people and rights of authors are crucial issues linked to the development of open access projects. Namely, the collection of the corpus raised three important ethical issues concerning (i) the respect of the intellectual property of the creators of source corpora: How to manage between the right of the scientist to have a free access to sources and the protection of authorship? How to value the scientific reutilisation of existing sources? Which deontological measures to adopt when citing a second-hand sample in a publication? (ii) the respect of the privacy of participants. What measures to adopt in order to avoid that the open publication of recordings cause damage to recorded speakers? (iii) the open access to data. How to guarantee a copyright on data, allowing at the same time for a free enrichment of them? In France, a reflexive study on good practices in the community is barely beginning (Baude et al, 2006). In the absence of shared practices, we adopted specific strategies to achieve a balance between these three competing priorities.

In order to guarantee the respect of the intellectual property we endowed our corpus with CMDI metadata (see next section). The CMDI format allowed for a full description of the source corpora and their related publications.

We also adopted a well-defined citation format for the examples drawn from the Rhapsodie reservoir and based on external sources: these are cited by specifying the name of the Rhapsodie sample, preceded by the prefix Rhaps and followed by the name of the source corpus. Ex.

(1) *j'accorde une puissance énorme à l'acte d'écrire* [Rhap-D2009, corpus Mertens]
(2) *c'était ils préféraient rigoler que de travailler* [Rhap-D0002, CFPP2000]
(3) *je suis heureux de me retrouver ce soir parmi vous* [Rhap-M2001, C-PROM]
(4) *et puis finalement bah on a choisi de rester* [Rhap-D0003, PFC]

Besides, we asked colleagues that use the Rhapsodie data based on external sources to make reference to the bibliography relative at each corpus, as is shown in table 2.

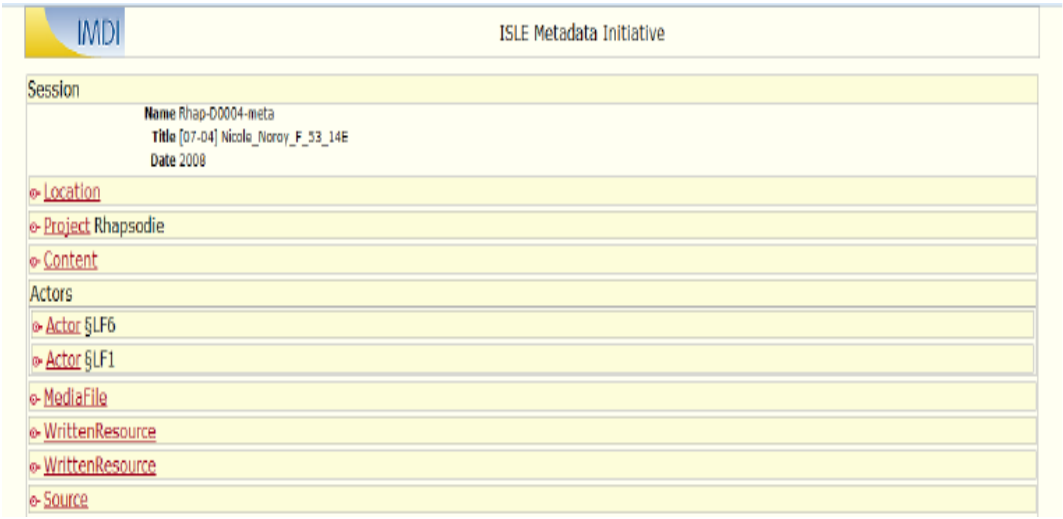
In order to guarantee the respect of the privacy of speakers we mostly selected recordings for which an informed consent was obtained at the outset and we anonymized the proper nouns, included the toponyms. It should also be noted that the short-sampling strategy used in the constitution of the corpus limited the amount of information provided on speakers, which are not easily identifiable and therefore better protected.

In order to freely distribute our Treebank and to protect at the same time our copyright, as well as the copyright of the source corpora, we adopt the Attribution, No Commercial, Share Alike Common Free Licence.

5. Metadata

We chose to encode our metadata in the CMDI format developed at the Max Planck Institute for Psycholinguistics in Nijmegen (CMDI, http://www.clarin.eu/cmdi) (Broeder et al 2011, 2012). Such a format is flexible enough to be adapted to the metadata encoding needs of different projects. In Rhapsodie, we did not need a fine-grained description either speakers sociolinguistic characteristics. Neither we needed to detail the modalities of gathering of sources, since the collection of data was for basically conducted in the frame of the source

projects (see § 3). Rather we needed: (i) to explicitly mention and describe the source corpora; (ii) to guarantee the access to the source files; (iii) to finely describe the speech situation; (iv) to provide for each sample a thorough description of the annotations; (v) to acknowledge the intellectual property of annotators.
Using CMDI we could define the profile of metadata we needed. Generally speaking, we chose to provide information on the sample, speakers, discourse situation, corpus sources, written and media resources associated to each sample. And the use of these components easily allowed us to meet our needs. In particular, we used the CMDI "source" component to finely describe the source corpora. In the "session" component we could endow our samples with a unique identifier that allows for a quick retrieval of the source sample and its metadata. With only a few modifications to the CMDI "discourse" component, we could provide a complete description of the situational variable characterizing each sample. In the "written source" component both we provided detailed information concerning the annotation of each sample and we acknowledged the intellectual property of annotators.
In order to manipulate the metadata we used the Arbil tool associated to the CMDI format (http://tla.mpi.nl/tools/tla-tools/arbil/description/) which allowed us to easily edit XMLs for our metadata and to convert them in HTML. The following pictures show the general format of metadata (figure 2); a zoom on the description of the speech situation (figure 3) and a zoom on the description of the source corpora (figure 4).



- Figure 2 – The general format of *Rhapsodie* metadata

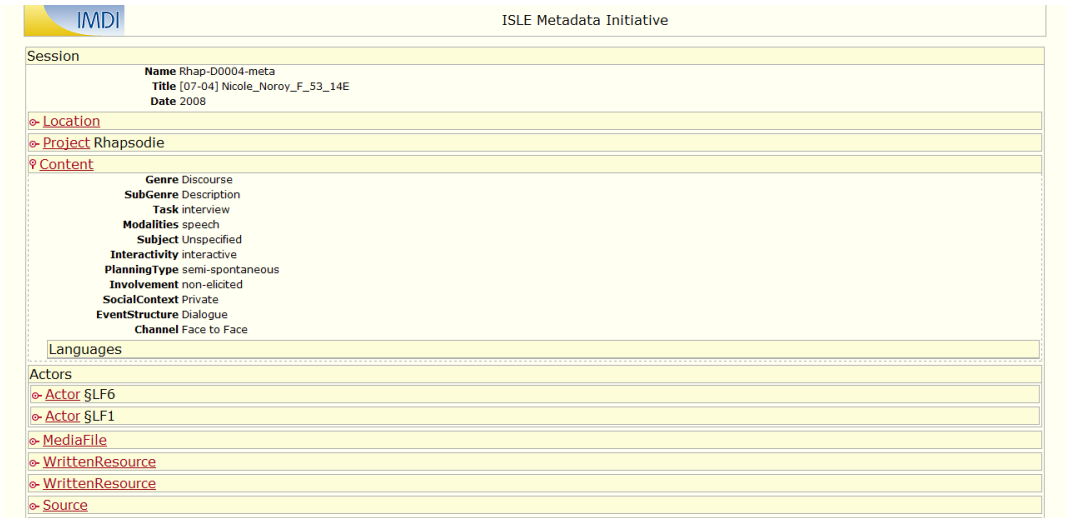
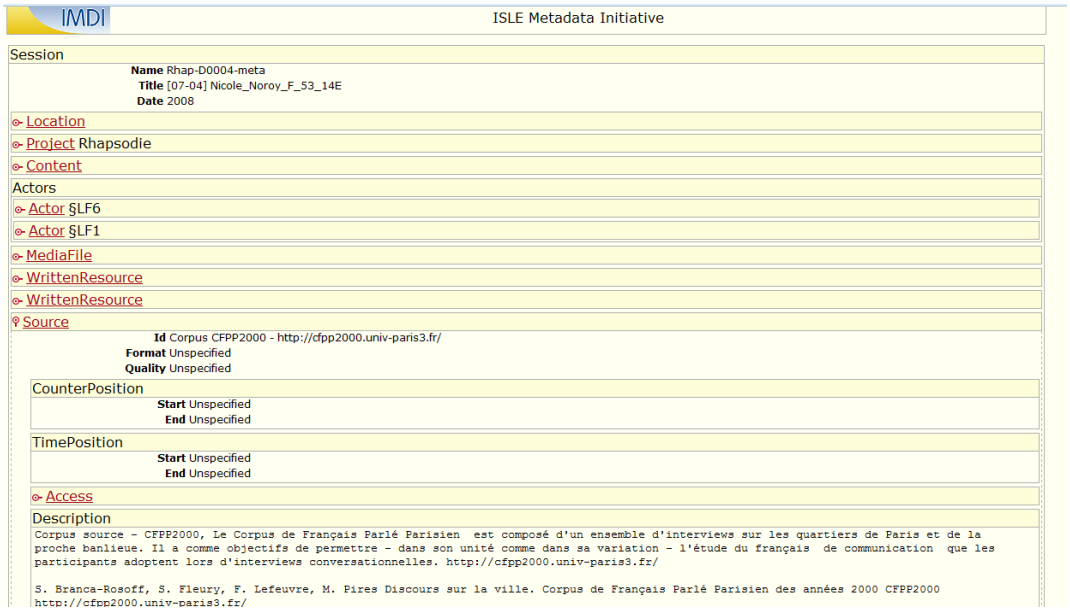


Fig. 3 The encoding of speech situation



- Figure 4. The description of a source corpus

Conclusions

In designing the Rhapsodie corpus we had a clear objective: gathering a sufficient variety of textual typologies to test and improve the flexibility of our annotation schemata. We also had a significant constraint: we could not rely on a pre-existing representative corpus of spoken French from which extracting our reservoir.

This limitation obliged us to select and extract excerpts from a number of different heterogeneous corpora and to collect new data wherever a given textual typology was not

represented. This *bouquet de corpus* approach had never been adopted before, at least in France, which raised a number of challenging new theoretical, ethical, and juridical questions.

From a theoretical point of view, we could not count on a unified model of textual diversity. We decided therefore to presuppose a correlation between the heterogeneity of speech situations and the variety of textual typologies. We hypothesized in other words that each particular speech situation engenders a number of specific linguistic constructions. We gathered data along a number of axes of situational variation: monologues vs. dialogues, private vs. public vs. professional speech, interactive vs. non interactive speeches, face-to-face vs. broadcast samples; more or less formal register. As we will see in the next chapters, the completeness of the *Rhapsodie* annotation schemata is exactly due to the variety of texts that have been annotated.

From an ethical and juridical point of view, a number of questions arose due to the fact that we actually made a public use of public resources. This only apparently trivial task led us to define a Good Practice with respect to the acknowledgment of the intellectual property of both authors and annotators as well as of the privacy of speakers. This led us to propose a short sampling strategy for the optimization of the anonymization; to define a standard for the citation of second-hand data and to choose a flexible and detailed format of metadata such as CMDI to guarantee a fine and complete description of source corpora, annotations, and speech situations.

REFERENCES

ADAM J.M. (1999), *Linguistique textuelle : des genres de discours au texte*, Paris, Nathan.

AVANZI, M. (2012), *L'interface prosodie/syntaxe en français. Dislocations, incises et asyndètes*, Bruxelles, Peter Lang.

AVANZI, M., SIMON, A.C., GOLDMAN, J.-P. & A. AUCHLIN. (2010), C-PROM. Un corpus de français parlé annoté pour l'étude des proéminences, *Actes des 23èmes journées d'étude sur la parole* (Mons, Belgique, 25-28 mai 2010).

BAKHTINE M. (1984), *Esthétique de la création verbale*, Paris, Gallimard.

BAUDE O. *coor., (2006), Corpus oraux, guide des bonnes pratiques*, Orléans et Paris, PUO et CNRS Editions.

BIBER D., JOHANSSON S., LEECH G., CONRAD S., FINNEGAN E. (1999) : *Longman Grammar of Spoken and Written English*, Harlow, Longman.

BILGER M. (2007), Réflexions sur un obscur objet de désir ; le corpus, *Les Cahiers de l'Association for French Language Studies*, 13-1, (http://www.afls.

BRANCA-ROSOFF S., FLEURY S., LEFEUVRE Fl., PIRES M (2012), *Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*, http://cfpp2000.univ-paris3.fr/

BROEDER, D., Schonefeld, O., Trippel, T., Van Uytvanck, D., and Witt, A. (2011). A pragmatic approach to XML interoperability — the component metadata infrastructure (CMDI). I

BROEDER, D., Van UYTVANCK, D., GAVRILIDOU, M., TRIPPEL, T., & WINDHOUWER, M.(2012). Standardizing a component metadata infrastructure. In N. Calzolari (Ed.), Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, May 23rd-25th, 2012 (pp. 1387-1390). European Language Resources Association (ELRA)

Durand, J., Laks, B. & Lyche, C. (2009), Le projet PFC (phonologie du français contemporain): une source de données primaires structurées. In: J. Durand, B. Laks & C. Lyche (eds.), *Phonologie, variation et accents du français*. Hermès, Paris, pp. 19-61.

Eshkol-Taravella I., Baude O., Maurel D., Hriba L., Dugua C., Tellier I., (2012), Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012, in *Ressources linguistiques libres*, TAL. Volume 52 – n° 3/2011, 17-46.

KOCH, P. and W. OESTERREICHER (2001). Langage parlé et langage écrit, in Lexicon der Romanistischen Linguistik, T1-2, Tübingen, Max Niemeyer Verlag, 584-627.

LACHERET A. (2003), *La prosodie des circonstants*, Leuven Peeters.

MAINGUENEAU D. (1996), *Les termes clés de l'analyse du discours*, Paris, Seuil.

MERTENS P. (1987), *L'intonation du français : de la description linguistique à la reconnaissance automatique*, Thèse de Doctorat, Université de Louvain.

SWALES J.M. (1990), *Genre Analysis*. Cambridge: Cambridge University Press.

Travaux :

Corpus des ESLO dans Cocoon

[fr] Corpus d'Orléans

Baude, Olivier (compiler)

(archivage: 2010-06-26T11:02:43+02:00; mise à disposition: 2010-06-26; dernière modification de la notice: 2015-02-26)

Editeur(s):	Centre Orléanais de Recherche en Anthropologie et Linguistique
Description(s):	[fr] Corpus linguistique composé d'enregistrements sonores et de leurs transcriptions réalisés à Orléans entre 1968 et 1974 (ESLO1) et à partir de 2008 (ESLO2). Entre 1969 et 1974, des universitaires britanniques ont réalisé un premier portrait sonore de la ville en enregistrant plusieurs centaines d'Orléanais dans la vie de tous les jours. Il s'agit du plus important témoignage sur le français des années soixante-dix. En 2014, quarante ans après cette première étude, l'université d'Orléans, en partenariat avec le CNRS, le Ministère de la Culture et la Région Centre, renouvelle l'expérience en procédant à des enregistrements avec des habitants de toute l'agglomération.
Type(s):	Collection
Sujet(s):	Français (code ISO-639: fra)
Droits:	Freely available for non-commercial use
Pour citer la ressource:	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO ou ark:/87895/1.5-124201
Membres:	1. Français / Corpus d'Orléans: ESLO1 2. Français / Corpus d'Orléans: ESLO1: la recette de l'omelette 3. Français / Corpus d'Orléans: ESLO2

1. Complément au guide du transcripteur: Lexique (metadata)
2. Guide du transcripteur/relecteur des ESLOs V1 (metadata)
3. Guide du transcripteur/relecteur des ESLOs V2 (metadata)
4. Guide du transcripteur/relecteur des ESLOs V3 (metadata)
5. Guide du transcripteur/relecteur des ESLOs V4 (metadata)

[fr] Corpus d'Orléans: ESLO1

Baude, Olivier (compiler)

(mise à disposition: 2014-03-21; création: start=1968;end=1971; archivage: 2015-07-09T09:52:50 02:00; dernière modification de la notice: 2015-07-09)

Editeur(s): Laboratoire Ligérien de Linguistique

Autres documents en relation:

Description(s): [fr] ESLO1 est la forme numérique de la première enquête ESLO. L'Enquête Socio-Linguistique à Orléans est une initiative prise en 1968 par un groupe d'universitaires anglais qui avait entrepris de collecter des documents sonores à Orléans avec une visée didactique : l'enseignement du français langue étrangère dans le système public d'éducation anglais. Cette enquête comprend environ 200 interviews, toutes référencées et au total plus de 300 heures de parole incluant une gamme d'enregistrements variés (conversations téléphoniques, réunions publiques, transactions commerciales, repas de famille, entretiens médico-pédagogiques, etc.). ESLO couvre l'ensemble des catégories socio-professionnelles, hommes et femmes, et présente un échantillon des formats de la communication, des tâches linguistiques et des types de discours selon une approche dialogique. Ce corpus représente, par son ampleur, sa rigueur et sa cohérence, le plus important témoignage disponible sur le français parlé avant 1980 (corpus de 4 500 000 mots environ). En 2004 le CORAL (devenu LLL en 2012) a entrepris la numérisation et la transcription de la totalité du fonds afin d'en assurer la diffusion dans le cadre de l'accès aux données scientifiques numériques. Les documents originaux d'ESLO, rebaptisé ESLO1, ont été déposés à la BnF et le corpus numérique archivé et rendu disponible selon les pratiques en vigueur au Ministère de l'enseignement supérieur et de la Recherche.
[fr] Enregistrements originaux conservés à la Bibliothèque nationale de France: description consultable en ligne sur le catalogue "Archives et manuscrits" (http://archivesetmanuscrits.bnf.fr/)

Source(s): http://archivesetmanuscrits.bnf.fr/ead.html?id=FRBNFEAD00009593

4

Type(s):	Collection
Sujet(s):	Français (code ISO-639: fra) [fr] Archives sonores [fr] Enquêtes linguistiques [fr] Analyse linguistique [fr] Description (linguistique) [fr] Parole spontanée [fr] Questionnaires [fr] Enquêtes par téléphone [fr] Enquêtes à domicile [fr] Congrès et conférences [fr] Entretiens [fr] France -- 20e siècle [fr] Orléans (Loiret)
Format(s):	[fr] 474 bandes magnétiques audio
Droits:	Freely available for non-commercial use
Citation bibliographique(s) :	[fr] Eshkol-Taravella I., Baude O., Maurel D., Hriba L., Dugua C., Tellier I., (2012) Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012. in Ressources linguistiques libres, TAL. Volume 52 – n° 3/2011, 17-46. [fr] ESLO, Enquête Sociolinguistique à Orléans lien: http://www.univ-orleans.fr/eslo/ [fr] Site du Laboratoire Ligérien de Linguistique lien: http://www.lll.cnrs.fr/eslo-1 [fr] Site du Portrait sonore de la ville d'Orléans lien: http://eslo.huma-num.fr/index.php/pagecorpus/pagepresentationcorpus
Pour citer la ressource:	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO1 ou ark:/87895/1.17-509086
Membres:	1. Français / ESLO1: consultation au centre médico-psycho-pedagogique 701 2. Français / ESLO1: consultation au centre médico-psycho-pedagogique 702 3. Français / ESLO1: consultation au centre médico-psycho-pedagogique 703 4. Français / ESLO1: consultation au centre médico-psycho-pedagogique 704

5. Français / ESLO1: consultation au centre médico-psycho-pedagogique 705
6. Français / ESLO1: consultation au centre médico-psycho-pedagogique 707
7. Français / ESLO1: consultation au centre médico-psycho-pedagogique 709
8. Français / ESLO1: consultation au centre médico-psycho-pedagogique 711
9. Français / ESLO1: consultation au centre médico-psycho-pedagogique 712
10. Français / ESLO1: consultation au centre médico-psycho-pedagogique 713
11. Français / ESLO1: consultation au centre médico-psycho-pedagogique 717
12. Français / ESLO1: consultation au centre médico-psycho-pedagogique 725
13. Français / ESLO1: consultation au centre médico-psycho-pedagogique 743
14. Français / ESLO1: consultation au centre médico-psycho-pedagogique 751
15. Français / ESLO1: consultation au centre médico-psycho-pedagogique 752
16. Français / ESLO1: contact avent entretien 201
17. Français / ESLO1: contact avent entretien 202
18. Français / ESLO1: contact avent entretien 203
19. Français / ESLO1: contact avent entretien 204
20. Français / ESLO1: contact avent entretien 205
21. Français / ESLO1: contact avent entretien 206
22. Français / ESLO1: contact avent entretien 208
23. Français / ESLO1: contact avent entretien 209
24. Français / ESLO1: contact avent entretien 210
25. Français / ESLO1: contact avent entretien 211
26. Français / ESLO1: contact avent entretien 212
27. Français / ESLO1: contact avent entretien 213
28. Français / ESLO1: contact avent entretien 214
29. Français / ESLO1: contact avent entretien 215
30. Français / ESLO1: contact avent entretien 216
31. Français / ESLO1: contact avent entretien 217
32. Français / ESLO1: contact avent entretien 218
33. Français / ESLO1: contact avent entretien 219
34. Français / ESLO1: contact avent entretien 222
35. Français / ESLO1: contact avent entretien 223
36. Français / ESLO1: contact avent entretien 224
37. Français / ESLO1: contact avent entretien 227
38. Français / ESLO1: contact avent entretien 228
39. Français / ESLO1: contact avent entretien 229
40. Français / ESLO1: contact avent entretien 233
41. Français / ESLO1: contact avent entretien 234
42. Français / ESLO1: contact avent entretien 241

43. Français / ESLO1: contact avent entretien 242
44. Français / ESLO1: contact avent entretien 243
45. Français / ESLO1: conversation téléphonique 302
46. Français / ESLO1: conversation téléphonique 303
47. Français / ESLO1: conversation téléphonique 304
48. Français / ESLO1: conversation téléphonique 305
49. Français / ESLO1: conversation téléphonique 306
50. Français / ESLO1: conversation téléphonique 307
51. Français / ESLO1: conversation téléphonique 308
52. Français / ESLO1: conversation téléphonique 309
53. Français / ESLO1: conversation téléphonique 310
54. Français / ESLO1: conversation téléphonique 311
55. Français / ESLO1: conversation téléphonique 312
56. Français / ESLO1: conversation téléphonique 316
57. Français / ESLO1: conversation téléphonique 317
58. Français / ESLO1: conversation téléphonique 318
59. Français / ESLO1: conversation téléphonique 319
60. Français / ESLO1: conversation téléphonique 331
61. Français / ESLO1: conversation téléphonique 333
62. Français / ESLO1: conversation téléphonique 335
63. Français / ESLO1: conversation téléphonique 337
64. Français / ESLO1: conversation téléphonique 339
65. Français / ESLO1: conversation téléphonique 340
66. Français / ESLO1: conversation téléphonique 342
67. Français / ESLO1: conversation téléphonique 346
68. Français / ESLO1: conversation téléphonique 350
69. Français / ESLO1: discussion en clôture de la séquence d'entretien 251
70. Français / ESLO1: discussion en clôture de la séquence d'entretien 255
71. Français / ESLO1: discussion en clôture de la séquence d'entretien 259
72. Français / ESLO1: discussion en clôture de la séquence d'entretien 264
73. Français / ESLO1: discussion en clôture de la séquence d'entretien 268
74. Français / ESLO1: discussion en ouverture de la séquence d'entretien 250
75. Français / ESLO1: discussion en ouverture de la séquence d'entretien 254
76. Français / ESLO1: discussion en ouverture de la séquence d'entretien 257
77. Français / ESLO1: discussion en ouverture de la séquence d'entretien 258
78. Français / ESLO1: discussion en ouverture de la séquence d'entretien 260
79. Français / ESLO1: discussion en ouverture de la séquence d'entretien 262
80. Français / ESLO1: discussion en ouverture de la séquence

d'entretien 263
81. Français / ESLO1: discussion en ouverture de la séquence d'entretien 265
82. Français / ESLO1: discussion en ouverture de la séquence d'entretien 267
83. Français / ESLO1: enregistrements divers 292
84. Français / ESLO1: enregistrements divers 619
85. Français / ESLO1: enregistrements divers 676
86. Français / ESLO1: enregistrements divers 680
87. Français / ESLO1: enregistrements divers 681
88. Français / ESLO1: enregistrements divers 682
89. Français / ESLO1: enregistrements divers 683
90. Français / ESLO1: enregistrements divers 684
91. Français / ESLO1: enregistrements divers 685
92. Français / ESLO1: enregistrements divers 686
93. Français / ESLO1: enregistrements divers 687
94. Français / ESLO1: enregistrements divers 688
95. Français / ESLO1: enregistrements divers 689
96. Français / ESLO1: enregistrements divers 690
97. Français / ESLO1: enregistrements divers 691
98. Français / ESLO1: enregistrements divers 692
99. Français / ESLO1: enregistrements divers 693
100. Français / ESLO1: enregistrements divers 695
101. Français / ESLO1: enregistrements divers 697
102. Français / ESLO1: enregistrements divers 699
103. Français / ESLO1: entretien 001
104. Français / ESLO1: entretien 002
105. Français / ESLO1: entretien 003
106. Français / ESLO1: entretien 004
107. Français / ESLO1: entretien 005
108. Français / ESLO1: entretien 006
109. Français / ESLO1: entretien 007
110. Français / ESLO1: entretien 008
111. Français / ESLO1: entretien 009
112. Français / ESLO1: entretien 010
113. Français / ESLO1: entretien 011
114. Français / ESLO1: entretien 012
115. Français / ESLO1: entretien 013
116. Français / ESLO1: entretien 014
117. Français / ESLO1: entretien 015
118. Français / ESLO1: entretien 016
119. Français / ESLO1: entretien 017
120. Français / ESLO1: entretien 018
121. Français / ESLO1: entretien 019
122. Français / ESLO1: entretien 020
123. Français / ESLO1: entretien 021
124. Français / ESLO1: entretien 022
125. Français / ESLO1: entretien 023
126. Français / ESLO1: entretien 024

127. Français / ESLO1: entretien 025
128. Français / ESLO1: entretien 026
129. Français / ESLO1: entretien 027
130. Français / ESLO1: entretien 028
131. Français / ESLO1: entretien 029
132. Français / ESLO1: entretien 030
133. Français / ESLO1: entretien 041
134. Français / ESLO1: entretien 042
135. Français / ESLO1: entretien 043
136. Français / ESLO1: entretien 044
137. Français / ESLO1: entretien 045
138. Français / ESLO1: entretien 046
139. Français / ESLO1: entretien 047
140. Français / ESLO1: entretien 048
141. Français / ESLO1: entretien 049
142. Français / ESLO1: entretien 050
143. Français / ESLO1: entretien 051
144. Français / ESLO1: entretien 052
145. Français / ESLO1: entretien 053
146. Français / ESLO1: entretien 054
147. Français / ESLO1: entretien 055
148. Français / ESLO1: entretien 056
149. Français / ESLO1: entretien 057
150. Français / ESLO1: entretien 058
151. Français / ESLO1: entretien 059
152. Français / ESLO1: entretien 060
153. Français / ESLO1: entretien 061
154. Français / ESLO1: entretien 062
155. Français / ESLO1: entretien 063
156. Français / ESLO1: entretien 064
157. Français / ESLO1: entretien 065
158. Français / ESLO1: entretien 066
159. Français / ESLO1: entretien 067
160. Français / ESLO1: entretien 068
161. Français / ESLO1: entretien 069
162. Français / ESLO1: entretien 070
163. Français / ESLO1: entretien 071
164. Français / ESLO1: entretien 072
165. Français / ESLO1: entretien 073
166. Français / ESLO1: entretien 075
167. Français / ESLO1: entretien 076
168. Français / ESLO1: entretien 078
169. Français / ESLO1: entretien 079
170. Français / ESLO1: entretien 080
171. Français / ESLO1: entretien 081
172. Français / ESLO1: entretien 082
173. Français / ESLO1: entretien 083
174. Français / ESLO1: entretien 084
175. Français / ESLO1: entretien 085

176. Français / ESLO1: entretien 086
177. Français / ESLO1: entretien 087
178. Français / ESLO1: entretien 088
179. Français / ESLO1: entretien 089
180. Français / ESLO1: entretien 090
181. Français / ESLO1: entretien 091
182. Français / ESLO1: entretien 092
183. Français / ESLO1: entretien 093
184. Français / ESLO1: entretien 095
185. Français / ESLO1: entretien 096
186. Français / ESLO1: entretien 097
187. Français / ESLO1: entretien 098
188. Français / ESLO1: entretien 100
189. Français / ESLO1: entretien 101
190. Français / ESLO1: entretien 102
191. Français / ESLO1: entretien 103
192. Français / ESLO1: entretien 105
193. Français / ESLO1: entretien 106
194. Français / ESLO1: entretien 107
195. Français / ESLO1: entretien 108
196. Français / ESLO1: entretien 109
197. Français / ESLO1: entretien 110
198. Français / ESLO1: entretien 111
199. Français / ESLO1: entretien 112
200. Français / ESLO1: entretien 113
201. Français / ESLO1: entretien 114
202. Français / ESLO1: entretien 115
203. Français / ESLO1: entretien 117
204. Français / ESLO1: entretien 118
205. Français / ESLO1: entretien 119
206. Français / ESLO1: entretien 120
207. Français / ESLO1: entretien 121
208. Français / ESLO1: entretien 122
209. Français / ESLO1: entretien 123
210. Français / ESLO1: entretien 124
211. Français / ESLO1: entretien 125
212. Français / ESLO1: entretien 126
213. Français / ESLO1: entretien 127
214. Français / ESLO1: entretien 129
215. Français / ESLO1: entretien 131
216. Français / ESLO1: entretien 132
217. Français / ESLO1: entretien 133
218. Français / ESLO1: entretien 139
219. Français / ESLO1: entretien 140
220. Français / ESLO1: entretien 141
221. Français / ESLO1: entretien 142
222. Français / ESLO1: entretien 149
223. Français / ESLO1: entretien 150
224. Français / ESLO1: entretien 160

225. Français / ESLO1: entretien 166
226. Français / ESLO1: entretien 167
227. Français / ESLO1: entretien 169
228. Français / ESLO1: entretien 170
229. Français / ESLO1: entretien 172
230. Français / ESLO1: entretien 173
231. Français / ESLO1: interaction dans un magasin 290
232. Français / ESLO1: interaction dans un magasin 623
233. Français / ESLO1: interaction dans un magasin 624
234. Français / ESLO1: interaction dans un magasin 630
235. Français / ESLO1: interaction dans un magasin 633
236. Français / ESLO1: interaction dans un magasin 635
237. Français / ESLO1: interaction dans un magasin 640
238. Français / ESLO1: interaction dans un magasin 641
239. Français / ESLO1: interaction dans un magasin 642
240. Français / ESLO1: interaction dans un magasin 643
241. Français / ESLO1: interaction dans un magasin 644
242. Français / ESLO1: interaction dans un magasin 646
243. Français / ESLO1: interaction dans un magasin 647
244. Français / ESLO1: interaction dans un magasin 648
245. Français / ESLO1: interaction dans un magasin 649
246. Français / ESLO1: interaction dans un magasin 651
247. Français / ESLO1: interaction dans un magasin 652
248. Français / ESLO1: interaction dans un magasin 653
249. Français / ESLO1: interaction dans un magasin 654
250. Français / ESLO1: interaction dans un magasin 655
251. Français / ESLO1: interaction dans un magasin 656
252. Français / ESLO1: interaction dans un magasin 657
253. Français / ESLO1: interaction dans un magasin 658
254. Français / ESLO1: interaction dans un magasin 659
255. Français / ESLO1: interaction dans un magasin 660
256. Français / ESLO1: interaction dans un magasin 661
257. Français / ESLO1: interaction dans un magasin 663
258. Français / ESLO1: interaction dans un magasin 664
259. Français / ESLO1: interaction dans un magasin 665
260. Français / ESLO1: interaction dans un magasin 666
261. Français / ESLO1: interaction dans un magasin 667
262. Français / ESLO1: interaction dans un magasin 670
263. Français / ESLO1: interaction dans un magasin 671
264. Français / ESLO1: interaction dans un magasin 672
265. Français / ESLO1: interaction dans un magasin 673
266. Français / ESLO1: interaction dans un magasin 674
267. Français / ESLO1: interaction lors d'un marché 609
268. Français / ESLO1: interaction lors d'un marché 610
269. Français / ESLO1: interaction lors d'un marché 611
270. Français / ESLO1: interaction lors d'un marché 612
271. Français / ESLO1: interaction lors d'un marché 613
272. Français / ESLO1: interaction lors d'un marché 616
273. Français / ESLO1: interaction lors d'un marché 618

274. Français / ESLO1: interview de personnalité 401
275. Français / ESLO1: interview de personnalité 402
276. Français / ESLO1: interview de personnalité 406
277. Français / ESLO1: interview de personnalité 407
278. Français / ESLO1: interview de personnalité 409
279. Français / ESLO1: interview de personnalité 414
280. Français / ESLO1: interview de personnalité 415
281. Français / ESLO1: interview de personnalité 417
282. Français / ESLO1: interview de personnalité 418
283. Français / ESLO1: interview de personnalité 419
284. Français / ESLO1: interview de personnalité 420
285. Français / ESLO1: interview de personnalité 421
286. Français / ESLO1: interview de personnalité 422
287. Français / ESLO1: interview de personnalité 423
288. Français / ESLO1: interview de personnalité 424
289. Français / ESLO1: interview de personnalité 425
290. Français / ESLO1: interview de personnalité 426
291. Français / ESLO1: interview de personnalité 427
292. Français / ESLO1: interview de personnalité 429
293. Français / ESLO1: interview de personnalité 430
294. Français / ESLO1: interview de personnalité 431
295. Français / ESLO1: interview de personnalité 436
296. Français / ESLO1: interview de personnalité 437
297. Français / ESLO1: interview de personnalité 438
298. Français / ESLO1: interview de personnalité 439
299. Français / ESLO1: interview de personnalité 441
300. Français / ESLO1: interview de personnalité 454
301. Français / ESLO1: interview de personnalité 460
302. Français / ESLO1: repas 272
303. Français / ESLO1: repas 273
304. Français / ESLO1: repas 276
305. Français / ESLO1: repas 277
306. Français / ESLO1: repas 281
307. Français / ESLO1: repas 555
308. Français / ESLO1: réunion 295
309. Français / ESLO1: réunion 297
310. Français / ESLO1: réunion 511
311. Français / ESLO1: réunion 512
312. Français / ESLO1: réunion 514
313. Français / ESLO1: réunion 515
314. Français / ESLO1: réunion 516
315. Français / ESLO1: réunion 517
316. Français / ESLO1: réunion 518
317. Français / ESLO1: réunion 519
318. Français / ESLO1: réunion 520
319. Français / ESLO1: réunion 521
320. Français / ESLO1: réunion 522
321. Français / ESLO1: réunion 523
322. Français / ESLO1: réunion 541

323. Français / ESLO1: réunion 542
324. Français / ESLO1: réunion 543
325. Français / ESLO1: réunion 551
326. Français / ESLO1: réunion 552
327. Français / ESLO1: réunion 553
328. Français / ESLO1: visite d'un lieu de travail 603
329. Français / ESLO1: visite d'un lieu de travail 604
330. Français / ESLO1: visite d'un lieu de travail 605
331. Français / ESLO1: visite d'un lieu de travail 607

[fr] Corpus d'Orléans: ESLO2

Baude, Olivier (compiler)

(mise à disposition: 2014-03-28; archivage: 2015-07-09T09:52:59 02:00; dernière modification de la notice: 2015-07-09)

Editeur(s):	Laboratoire Ligérien de Linguistique
Description(s):	[fr] En 2014, quarante ans après la première étude "Enquête SocioLinguistique d'Orléans" (ESLO1) réalisée par des universitaires britanniques sur un portrait sonore de la ville d'Orléans, l'université d'Orléans, en partenariat avec le CNRS, le Ministère de la Culture et la Région Centre, renouvelle l'expérience en procédant à des enregistrements avec des habitants de toute l'agglomération.
Type(s):	Collection
Sujet(s):	Français (code ISO-639: fra)
Droits:	Freely available for non-commercial use
Pour citer la ressource:	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO2 ou ark:/87895/1.17-509087

Role	Type	Title	ID
compiler	Collection	Corpus d'Orléans	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO
compiler	Collection	Corpus d'Orléans: ESLO1	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO1
compiler	Collection	Corpus d'Orléans: ESLO2	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO2
compiler	Collection	Corpus d'Orléans: ESLO1: la recette de l'omelette	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_ESLO_OMELETTE
compiler	Collection	Corpus de la parole	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_LANGUESDEFRANCE
editor	Sound	ESLO1: entretien 001	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-FRA_ESLO1_1_SOUND
editor	Sound	ESLO1: entretien 002	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_002
editor	Sound	ESLO1: entretien 003	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_003
editor	Sound	ESLO1: entretien 004	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_004
editor	Sound	ESLO1: entretien 005	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_005
editor	Sound	ESLO1: entretien 006	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_006
editor	Sound	ESLO1: entretien 007	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_007
editor	Sound	ESLO1: entretien 009	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_009
editor	Sound	ESLO1: entretien 010	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_010
editor	Sound	ESLO1: entretien 008	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_008
editor	Sound	ESLO1: entretien 011	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_011
editor	Sound	ESLO1: entretien 012	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_012
editor	Sound	ESLO1: entretien 013	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_013
editor	Sound	ESLO1: entretien 014	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_014
editor	Sound	ESLO1: entretien 015	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_015
editor	Sound	ESLO1: entretien 017	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_017
editor	Sound	ESLO1: entretien 018	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_018
editor	Sound	ESLO1: entretien 019	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_019
editor	Sound	ESLO1: entretien 020	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_020
editor	Sound	ESLO1: entretien 021	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_021
editor	Sound	ESLO1: entretien 022	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_022
editor	Sound	ESLO1: entretien 023	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_023
editor	Sound	ESLO1: entretien 024	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_024
editor	Sound	ESLO1: entretien 025	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_025
editor	Sound	ESLO1: entretien 026	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_026
editor	Sound	ESLO1: entretien 027	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_027
editor	Sound	ESLO1: entretien 028	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_028
editor	Sound	ESLO1: entretien 029	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_029
editor	Sound	ESLO1: entretien 030	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_030
editor	Sound	ESLO1: entretien 041	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_041
editor	Sound	ESLO1: entretien 042	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_042
editor	Sound	ESLO1: entretien 043	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_043
editor	Sound	ESLO1: entretien 044	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_044
editor	Sound	ESLO1: entretien 045	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_045
editor	Sound	ESLO1: entretien 046	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_046
editor	Sound	ESLO1: entretien 047	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_047
editor	Sound	ESLO1: entretien 048	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_048
editor	Sound	ESLO1: entretien 050	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_050
editor	Sound	ESLO1: entretien 051	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_051

editor	Sound	ESLO1: entretien 052	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_052
editor	Sound	ESLO1: entretien 053	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_053
editor	Sound	ESLO1: entretien 054	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_054
editor	Sound	ESLO1: entretien 055	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_055
editor	Sound	ESLO1: entretien 056	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_056
editor	Sound	ESLO1: entretien 057	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_057
editor	Sound	ESLO1: entretien 058	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_058
editor	Sound	ESLO1: entretien 059	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_059
editor	Sound	ESLO1: entretien 060	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_060
editor	Sound	ESLO1: entretien 061	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_061
editor	Sound	ESLO1: entretien 062	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_062
editor	Sound	ESLO1: entretien 063	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_063
editor	Sound	ESLO1: entretien 064	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_064
editor	Sound	ESLO1: entretien 065	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_065
editor	Sound	ESLO1: entretien 066	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_066
editor	Sound	ESLO1: entretien 067	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_067
editor	Sound	ESLO1: entretien 068	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_068
editor	Sound	ESLO1: entretien 069	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_069
editor	Sound	ESLO1: entretien 070	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_070
editor	Sound	ESLO1: entretien 071	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_071
editor	Sound	ESLO1: entretien 072	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_072
editor	Sound	ESLO1: entretien 073	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_073
editor	Sound	ESLO1: entretien 075	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_075
editor	Sound	ESLO1: entretien 076	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_076
editor	Sound	ESLO1: entretien 078	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_078
editor	Sound	ESLO1: entretien 079	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_079
editor	Sound	ESLO1: entretien 080	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_080
editor	Sound	ESLO1: entretien 081	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_081
editor	Sound	ESLO1: entretien 082	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_082
editor	Sound	ESLO1: entretien 083	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_083
editor	Sound	ESLO1: entretien 084	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_084
editor	Sound	ESLO1: entretien 085	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_085
editor	Sound	ESLO1: entretien 086	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_086
editor	Sound	ESLO1: entretien 087	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_087
editor	Sound	ESLO1: entretien 088	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_088
editor	Sound	ESLO1: entretien 089	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_089
editor	Sound	ESLO1: entretien 090	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_090
editor	Sound	ESLO1: entretien 091	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_091
editor	Sound	ESLO1: entretien 092	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_092
editor	Sound	ESLO1: entretien 093	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_093
editor	Sound	ESLO1: entretien 095	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_095
editor	Sound	ESLO1: entretien 096	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_096
editor	Sound	ESLO1: entretien 097	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_097
editor	Sound	ESLO1: entretien 100	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_100
editor	Sound	ESLO1: entretien 101	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_101
editor	Sound	ESLO1: entretien 102	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_102
editor	Sound	ESLO1: entretien 103	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_103
editor	Sound	ESLO1: entretien 105	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_105
editor	Sound	ESLO1: entretien 106	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_106

editor	Sound	ESLO1: entretien 107	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_107
editor	Sound	ESLO1: entretien 108	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_108
editor	Sound	ESLO1: entretien 109	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_109
editor	Sound	ESLO1: entretien 110	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_110
editor	Sound	ESLO1: entretien 111	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_111
editor	Sound	ESLO1: entretien 112	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_112
editor	Sound	ESLO1: entretien 113	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_113
editor	Sound	ESLO1: entretien 114	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_114
editor	Sound	ESLO1: entretien 115	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_115
editor	Sound	ESLO1: entretien 117	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_117
editor	Sound	ESLO1: entretien 118	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_118
editor	Sound	ESLO1: entretien 119	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_119
editor	Sound	ESLO1: entretien 120	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_120
editor	Sound	ESLO1: entretien 121	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_121
editor	Sound	ESLO1: entretien 122	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_122
editor	Sound	ESLO1: entretien 123	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_123
editor	Sound	ESLO1: entretien 124	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_124
editor	Sound	ESLO1: entretien 125	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_125
editor	Sound	ESLO1: entretien 126	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_126
editor	Sound	ESLO1: entretien 127	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_127
editor	Sound	ESLO1: entretien 131	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_131
editor	Sound	ESLO1: entretien 132	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_132
editor	Sound	ESLO1: entretien 133	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_133
editor	Sound	ESLO1: entretien 139	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_139
editor	Sound	ESLO1: entretien 140	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_140
editor	Sound	ESLO1: entretien 141	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_141
editor	Sound	ESLO1: entretien 142	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_142
editor	Sound	ESLO1: entretien 129	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_129
editor	Sound	ESLO1: entretien 149	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_149
editor	Sound	ESLO1: entretien 150	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_150
editor	Sound	ESLO1: entretien 160	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_160
editor	Sound	ESLO1: entretien 166	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_166
editor	Sound	ESLO1: entretien 167	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_167
editor	Sound	ESLO1: entretien 169	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_169
editor	Sound	ESLO1: entretien 170	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_170
editor	Sound	ESLO1: entretien 172	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_172
editor	Sound	ESLO1: entretien 173	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_173
editor	Sound	ESLO1: contact avent entretien 201	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_201
editor	Sound	ESLO1: contact avent entretien 202	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_202
editor	Sound	ESLO1: contact avent entretien 203	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_203
editor	Sound	ESLO1: contact avent entretien 204	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_204
editor	Sound	ESLO1: contact avent entretien 205	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_205
editor	Sound	ESLO1: contact avent entretien 206	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_206

editor	Sound	ESLO1: contact avent entretien 208	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_208
editor	Sound	ESLO1: contact avent entretien 211	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_211
editor	Sound	ESLO1: contact avent entretien 212	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_212
editor	Sound	ESLO1: contact avent entretien 214	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_214
editor	Sound	ESLO1: contact avent entretien 215	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_215
editor	Sound	ESLO1: contact avent entretien 216	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_216
editor	Sound	ESLO1: contact avent entretien 217	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_217
editor	Sound	ESLO1: contact avent entretien 223	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_223
editor	Sound	ESLO1: contact avent entretien 224	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_224
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 254	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_254
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 257	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_257
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 258	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_258
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 260	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_260
editor	Sound	ESLO1: discussion en clôture de la séquence d'entretien 255	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_255
editor	Sound	ESLO1: discussion en clôture de la séquence d'entretien 259	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_259
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 263	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_263
editor	Sound	ESLO1: repas 272	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_272
editor	Sound	ESLO1: repas 273	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_273
editor	Sound	ESLO1: interaction dans un magasin 290	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_290
editor	Sound	ESLO1: conversation téléphonique 302	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_302
editor	Sound	ESLO1: conversation téléphonique 307	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_307
editor	Sound	ESLO1: conversation téléphonique 308	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_308

editor	Sound	ESLO1: conversation téléphonique 312	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_312
editor	Sound	ESLO1: conversation téléphonique 316	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_316
editor	Sound	ESLO1: conversation téléphonique 318	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_318
editor	Sound	ESLO1: conversation téléphonique 319	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_319
editor	Sound	ESLO1: interview de personnalité 401	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_401
editor	Sound	ESLO1: interview de personnalité 402	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_402
editor	Sound	ESLO1: interview de personnalité 406	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_406
editor	Sound	ESLO1: interview de personnalité 407	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_407
editor	Sound	ESLO1: interview de personnalité 409	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_409
editor	Sound	ESLO1: interview de personnalité 414	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_414
editor	Sound	ESLO1: interview de personnalité 415	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_415
editor	Sound	ESLO1: interview de personnalité 417	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_417
editor	Sound	ESLO1: interview de personnalité 418	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_418
editor	Sound	ESLO1: interview de personnalité 419	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_419
editor	Sound	ESLO1: interview de personnalité 420	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_420
editor	Sound	ESLO1: interview de personnalité 421	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_421
editor	Sound	ESLO1: interview de personnalité 422	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_422
editor	Sound	ESLO1: interview de personnalité 423	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_423
editor	Sound	ESLO1: interview de personnalité 424	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_424
editor	Sound	ESLO1: interview de personnalité 425	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_425
editor	Sound	ESLO1: interview de personnalité 426	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_426
editor	Sound	ESLO1: interview de personnalité 427	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_427
editor	Sound	ESLO1: interview de personnalité 429	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_429
editor	Sound	ESLO1: interview de personnalité 431	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_431
editor	Sound	ESLO1: interview de personnalité 436	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_436

editor	Sound	ESLO1: interview de personnalité 437	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_437
editor	Sound	ESLO1: interview de personnalité 438	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_438
editor	Sound	ESLO1: interview de personnalité 439	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_439
editor	Sound	ESLO1: interview de personnalité 441	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_441
editor	Sound	ESLO1: réunion 517	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_517
editor	Sound	ESLO1: interaction dans un magasin 630	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_630
editor	Sound	ESLO1: réunion 519	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_519
editor	Sound	ESLO1: réunion 542	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_542
editor	Sound	ESLO1: interaction dans un magasin 673	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_673
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 701	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_701
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 702	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_702
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 703	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_703
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 704	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_704
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 705	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_705
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 707	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_707
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 709	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_709
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 712	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_712
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 713	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_713
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 717	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_717
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 743	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_743

editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 751	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_751
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 752	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_752
editor	Sound	ESLO1: enregistrements divers 292	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_292
editor	Sound	ESLO1: discussion en clôture de la séquence d'entretien 264	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_264
editor	Sound	ESLO1: entretien 098	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_098
editor	Sound	ESLO1: discussion en clôture de la séquence d'entretien 268	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_268
editor	Sound	ESLO1: contact avent entretien 209	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_209
editor	Sound	ESLO1: contact avent entretien 210	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_210
editor	Sound	ESLO1: contact avent entretien 222	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_222
editor	Sound	ESLO1: contact avent entretien 227	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_227
editor	Sound	ESLO1: contact avent entretien 233	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_233
editor	Sound	ESLO1: contact avent entretien 234	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_234
editor	Sound	ESLO1: contact avent entretien 241	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_241
editor	Sound	ESLO1: contact avent entretien 242	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_242
editor	Sound	ESLO1: contact avent entretien 243	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_243
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 250	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_250
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 262	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_262
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 265	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_265
editor	Sound	ESLO1: discussion en ouverture de la séquence d'entretien 267	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_267
editor	Sound	ESLO1: entretien 016	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_016
editor	Sound	ESLO1: entretien 049	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_049
editor	Sound	ESLO1: repas 276	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_276
editor	Sound	ESLO1: repas 277	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_277
editor	Sound	ESLO1: repas 281	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-

			ESLO1_REPAS_281
editor	Sound	ESLO1: réunion 295	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_295
editor	Sound	ESLO1: réunion 297	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_297
editor	Sound	ESLO1: interview de personnalité 454	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_454
editor	Sound	ESLO1: interview de personnalité 460	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_460
editor	Sound	ESLO1: repas 555	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_555
editor	Sound	ESLO1: réunion 511	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_511
editor	Sound	ESLO1: réunion 512	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_512
editor	Sound	ESLO1: réunion 514	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_514
editor	Sound	ESLO1: réunion 515	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_515
editor	Sound	ESLO1: réunion 516	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_516
editor	Sound	ESLO1: réunion 518	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_518
editor	Sound	ESLO1: réunion 521	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_521
editor	Sound	ESLO1: réunion 522	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_522
editor	Sound	ESLO1: réunion 523	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_523
editor	Sound	ESLO1: réunion 541	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_541
editor	Sound	ESLO1: réunion 543	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_543
editor	Sound	ESLO1: réunion 552	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_552
editor	Sound	ESLO1: réunion 553	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_553
editor	Sound	ESLO1: conversation téléphonique 303	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_303
editor	Sound	ESLO1: conversation téléphonique 304	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_304
editor	Sound	ESLO1: conversation téléphonique 305	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_305
editor	Sound	ESLO1: conversation téléphonique 306	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_306
editor	Sound	ESLO1: conversation téléphonique 309	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_309
editor	Sound	ESLO1: conversation téléphonique 310	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_310
editor	Sound	ESLO1: conversation téléphonique 311	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_311
editor	Sound	ESLO1: conversation téléphonique 317	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_317
editor	Sound	ESLO1: conversation téléphonique 331	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_331
editor	Sound	ESLO1: conversation téléphonique 333	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_333
editor	Sound	ESLO1: conversation téléphonique 335	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_335
editor	Sound	ESLO1: conversation téléphonique 337	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_337
editor	Sound	ESLO1: conversation téléphonique 339	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_339
editor	Sound	ESLO1: conversation téléphonique 340	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_340

editor	Sound	ESLO1: conversation téléphonique 342	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_342
editor	Sound	ESLO1: conversation téléphonique 346	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_346
editor	Sound	ESLO1: enregistrements divers 619	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_619
editor	Sound	ESLO1: interaction dans un magasin 633	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_633
editor	Sound	ESLO1: interaction dans un magasin 641	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_641
editor	Sound	ESLO1: interaction dans un magasin 642	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_642
editor	Sound	ESLO1: interaction lors d'un marché 609	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_609
editor	Sound	ESLO1: interaction lors d'un marché 610	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_610
editor	Sound	ESLO1: interaction lors d'un marché 611	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_611
editor	Sound	ESLO1: interaction lors d'un marché 613	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_613
editor	Sound	ESLO1: interaction lors d'un marché 616	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_616
editor	Sound	ESLO1: interaction lors d'un marché 618	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_618
editor	Sound	ESLO1: visite d'un lieu de travail 604	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_VISIT_604
editor	Sound	ESLO1: visite d'un lieu de travail 605	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_VISIT_605
editor	Sound	ESLO1: visite d'un lieu de travail 607	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_VISIT_607
editor	Sound	ESLO1: interaction dans un magasin 623	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_623
editor	Sound	ESLO1: interaction dans un magasin 624	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_624
editor	Sound	ESLO1: interaction dans un magasin 635	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_635
editor	Sound	ESLO1: interaction dans un magasin 640	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_640
editor	Sound	ESLO1: interaction dans un magasin 643	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_643
editor	Sound	ESLO1: interaction dans un magasin 644	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_644
editor	Sound	ESLO1: interaction dans un magasin 646	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_646
editor	Sound	ESLO1: interaction dans un magasin 647	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_647
editor	Sound	ESLO1: interaction dans un magasin 648	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_648
editor	Sound	ESLO1: interaction dans un magasin 649	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_649

editor	Sound	ESLO1: interaction dans un magasin 651	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_651
editor	Sound	ESLO1: interaction dans un magasin 652	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_652
editor	Sound	ESLO1: interaction dans un magasin 653	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_653
editor	Sound	ESLO1: interaction dans un magasin 654	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_654
editor	Sound	ESLO1: interaction dans un magasin 655	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_655
editor	Sound	ESLO1: interaction dans un magasin 656	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_656
editor	Sound	ESLO1: interaction dans un magasin 657	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_657
editor	Sound	ESLO1: interaction dans un magasin 658	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_658
editor	Sound	ESLO1: interaction dans un magasin 659	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_659
editor	Sound	ESLO1: interaction dans un magasin 660	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_660
editor	Sound	ESLO1: interaction dans un magasin 661	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_661
editor	Sound	ESLO1: interaction dans un magasin 663	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_663
editor	Sound	ESLO1: interaction dans un magasin 664	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_664
editor	Sound	ESLO1: interaction dans un magasin 665	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_665
editor	Sound	ESLO1: interaction dans un magasin 666	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_666
editor	Sound	ESLO1: interaction dans un magasin 667	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_667
editor	Sound	ESLO1: enregistrements divers 676	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_676
editor	Sound	ESLO1: enregistrements divers 680	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_680
editor	Sound	ESLO1: enregistrements divers 681	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_681
editor	Sound	ESLO1: enregistrements divers 682	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_682
editor	Sound	ESLO1: enregistrements divers 683	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_683
editor	Sound	ESLO1: enregistrements divers 684	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_684
editor	Sound	ESLO1: enregistrements divers 685	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_685
editor	Sound	ESLO1: enregistrements divers 686	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_686
editor	Sound	ESLO1: enregistrements divers 687	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_687

editor	Sound	ESLO1: interaction dans un magasin 670	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_670
editor	Sound	ESLO1: interaction dans un magasin 671	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_671
editor	Sound	ESLO1: interaction dans un magasin 672	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_672
editor	Sound	ESLO1: interaction dans un magasin 674	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_674
editor	Sound	ESLO1: enregistrements divers 688	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_688
editor	Sound	ESLO1: enregistrements divers 689	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_689
editor	Sound	ESLO1: enregistrements divers 690	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_690
editor	Sound	ESLO1: enregistrements divers 691	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_691
editor	Sound	ESLO1: enregistrements divers 692	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_692
editor	Sound	ESLO1: enregistrements divers 693	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_693
editor	Sound	ESLO1: enregistrements divers 695	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_695
editor	Sound	ESLO1: enregistrements divers 697	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_697
editor	Sound	ESLO1: enregistrements divers 699	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_DIV_699
editor	Sound	ESLO1: discussion en clôture de la séquence d'entretien 251	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_251
editor	Sound	ESLO1: réunion 520	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_520
editor	Sound	ESLO1: réunion 551	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_551
editor	Sound	ESLO1: visite d'un lieu de travail 603	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_VISIT_603
editor	Sound	ESLO1: conversation téléphonique 350	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_350
editor	Sound	ESLO1: contact avent entretien 213	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_213
editor	Sound	ESLO1: contact avent entretien 218	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_218
editor	Sound	ESLO1: contact avent entretien 219	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_219
editor	Sound	ESLO1: contact avent entretien 228	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_228
editor	Sound	ESLO1: contact avent entretien 229	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_229
editor	Sound	ESLO1: interview de personnalité 430	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_430
editor	Sound	ESLO1: interaction lors d'un marché 612	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAR_612

editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 711	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_711
editor	Sound	ESLO1: consultation au centre médico-psycho-pedagogique 725	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_725
researcher	Text	ESLO1: entretien 001_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-FRA_ESLO1_1B
researcher	Text	ESLO1: Entretien 001_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-FRA_ESLO1_1C
researcher	Text	ESLO1: Entretien 002_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_002_C
researcher	Text	ESLO1: entretien 003_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_003_C
researcher	Text	ESLO1: entretien 004_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_004_C
researcher	Text	ESLO1: entretien 005_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_005_C
researcher	Text	ESLO1: entretien 009_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_009_C
researcher	Text	ESLO1: entretien 008_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_008_C
researcher	Text	ESLO1: entretien 006_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_006_C
researcher	Text	ESLO1: entretien 010_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_010_C
researcher	Text	ESLO1: entretien 007_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_007_C
researcher	Text	ESLO1: entretien 015_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_015_C
researcher	Text	ESLO1: entretien 014_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_014_C
researcher	Text	ESLO1: entretien 013_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_013_C
researcher	Text	ESLO1: entretien 012_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_012_C
researcher	Text	ESLO1: entretien 011_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_011_C
researcher	Text	ESLO1: entretien 030_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_030_C
researcher	Text	ESLO1: entretien 029_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_029_C
researcher	Text	ESLO1: entretien 028_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_028_C
researcher	Text	ESLO1: entretien 027_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_027_C
researcher	Text	ESLO1: entretien 026_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_026_C
researcher	Text	ESLO1: entretien 025_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_025_C

researcher	Text	ESLO1: entretien 023_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_023_C
researcher	Text	ESLO1: entretien 022_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_022_C
researcher	Text	ESLO1: entretien 021_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_021_C
researcher	Text	ESLO1: entretien 020_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_020_C
researcher	Text	ESLO1: entretien 019_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_019_C
researcher	Text	ESLO1: entretien 018_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_018_C
researcher	Text	ESLO1: entretien 017_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_017_C
researcher	Text	ESLO1: entretien 024_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_024_C
researcher	Text	ESLO1: entretien 060_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_060_C
researcher	Text	ESLO1: entretien 059_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_059_C
researcher	Text	ESLO1: entretien 058_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_058_C
researcher	Text	ESLO1: entretien 057_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_057_C
researcher	Text	ESLO1: entretien 056_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_056_C
researcher	Text	ESLO1: entretien 055_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_055_C
researcher	Text	ESLO1: entretien 054_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_054_C
researcher	Text	ESLO1: entretien 053_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_053_C
researcher	Text	ESLO1: entretien 052_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_052_C
researcher	Text	ESLO1: entretien 051_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_051_C
researcher	Text	ESLO1: entretien 050_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_050_C
researcher	Text	ESLO1: entretien 048_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_048_C
researcher	Text	ESLO1: entretien 047_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_047_C
researcher	Text	ESLO1: entretien 046_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_046_C
researcher	Text	ESLO1: entretien 045_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_045_C
researcher	Text	ESLO1: entretien 044_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_044_C
researcher	Text	ESLO1: entretien 043_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_043_C

researcher	Text	ESLO1: entretien 042_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_042_C
researcher	Text	ESLO1: entretien 041_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_041_C
researcher	Text	ESLO1: entretien 093_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_093_C
researcher	Text	ESLO1: entretien 092_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_092_C
researcher	Text	ESLO1: entretien 091_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_091_C
researcher	Text	ESLO1: entretien 090_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_090_C
researcher	Text	ESLO1: entretien 089_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_089_C
researcher	Text	ESLO1: entretien 088_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_088_C
researcher	Text	ESLO1: entretien 087_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_087_C
researcher	Text	ESLO1: entretien 086_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_086_C
researcher	Text	ESLO1: entretien 085_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_085_C
researcher	Text	ESLO1: entretien 084_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_084_C
researcher	Text	ESLO1: entretien 083_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_083_C
researcher	Text	ESLO1: entretien 082_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_082_C
researcher	Text	ESLO1: entretien 081_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_081_C
researcher	Text	ESLO1: entretien 080_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_080_C
researcher	Text	ESLO1: entretien 079_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_079_C
researcher	Text	ESLO1: entretien 078_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_078_C
researcher	Text	ESLO1: entretien 076_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_076_C
researcher	Text	ESLO1: entretien 075_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_075_C
researcher	Text	ESLO1: entretien 073_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_073_C
researcher	Text	ESLO1: entretien 072_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_072_C
researcher	Text	ESLO1: entretien 071_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_071_C
researcher	Text	ESLO1: entretien 070_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_070_C
researcher	Text	ESLO1: entretien 069_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_069_C

researcher	Text	ESLO1: entretien 068_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_068_C
researcher	Text	ESLO1: entretien 067_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_067_C
researcher	Text	ESLO1: entretien 066_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_066_C
researcher	Text	ESLO1: entretien 065_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_065_C
researcher	Text	ESLO1: entretien 064_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_064_C
researcher	Text	ESLO1: entretien 063_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_063_C
researcher	Text	ESLO1: entretien 062_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_062_C
researcher	Text	ESLO1: entretien 061_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_061_C
researcher	Text	ESLO1: entretien 142_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_142_C
researcher	Text	ESLO1: entretien 141_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_141_C
researcher	Text	ESLO1: entretien 140_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_140_C
researcher	Text	ESLO1: entretien 139_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_139_C
researcher	Text	ESLO1: entretien 133_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_133_C
researcher	Text	ESLO1: entretien 132_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_132_C
researcher	Text	ESLO1: entretien 131_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_131_C
researcher	Text	ESLO1: entretien 127_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_127_C
researcher	Text	ESLO1: entretien 126_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_126_C
researcher	Text	ESLO1: entretien 125_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_125_C
researcher	Text	ESLO1: entretien 124_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_124_C
researcher	Text	ESLO1: entretien 123_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_123_C
researcher	Text	ESLO1: entretien 122_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_122_C
researcher	Text	ESLO1: entretien 121_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_121_C
researcher	Text	ESLO1: entretien 120_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_120_C
researcher	Text	ESLO1: entretien 119_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_119_C
researcher	Text	ESLO1: entretien 118_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_118_C

researcher	Text	ESLO1: entretien 117_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_117_C
researcher	Text	ESLO1: entretien 115_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_115_C
researcher	Text	ESLO1: entretien 114_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_114_C
researcher	Text	ESLO1: entretien 113_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_113_C
researcher	Text	ESLO1: entretien 112_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_112_C
researcher	Text	ESLO1: entretien 111_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_111_C
researcher	Text	ESLO1: entretien 110_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_110_C
researcher	Text	ESLO1: entretien 109_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_109_C
researcher	Text	ESLO1: entretien 108_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_108_C
researcher	Text	ESLO1: entretien 107_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_107_C
researcher	Text	ESLO1: entretien 106_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_106_C
researcher	Text	ESLO1: entretien 103_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_103_C
researcher	Text	ESLO1: entretien 102_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_102_C
researcher	Text	ESLO1: entretien 101_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_101_C
researcher	Text	ESLO1: entretien 100_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_100_C
researcher	Text	ESLO1: entretien 097_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_097_C
researcher	Text	ESLO1: entretien 096_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_096_C
researcher	Text	ESLO1: entretien 095_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_095_C
researcher	Text	ESLO1: contact avent entretien 217_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_217_C
researcher	Text	ESLO1: contact avent entretien 216_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_216_C
researcher	Text	ESLO1: contact avent entretien 215_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_215_C
researcher	Text	ESLO1: contact avent entretien 214_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_214_C
researcher	Text	ESLO1: contact avent entretien 212_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_212_C
researcher	Text	ESLO1: contact avent entretien 211_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_211_C
researcher	Text	ESLO1: contact avent entretien 208_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_208_C

researcher	Text	ESLO1: contact avent entretien 206_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_206_C
researcher	Text	ESLO1: contact avent entretien 205_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_205_C
researcher	Text	ESLO1: contact avent entretien 204_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_204_C
researcher	Text	ESLO1: contact avent entretien 203_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_203_C
researcher	Text	ESLO1: contact avent entretien 202_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_202_C
researcher	Text	ESLO1: contact avent entretien 201_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_201_C
researcher	Text	ESLO1: entretien 173_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_173_C
researcher	Text	ESLO1: entretien 172_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_172_C
researcher	Text	ESLO1: entretien 170_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_170_C
researcher	Text	ESLO1: entretien 169_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_169_C
researcher	Text	ESLO1: entretien 167_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_167_C
researcher	Text	ESLO1: entretien 166_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_166_C
researcher	Text	ESLO1: entretien 160_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_160_C
researcher	Text	ESLO1: entretien 150_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_150_C
researcher	Text	ESLO1: entretien 149_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_149_C
researcher	Text	ESLO1: entretien 129_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_129_C
researcher	Text	ESLO1: entretien 105_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_105_C
researcher	Text	ESLO1: interaction dans un magasin 290_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_290_C
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 263_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_263_C
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 260_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_260_C
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 258_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_258_C
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 257_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_257_C

researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 254_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_254_C
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 259_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_259_C
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 255_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_255_C
researcher	Text	ESLO1: conversation téléphonique 318_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_318_C
researcher	Text	ESLO1: conversation téléphonique 316_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_316_C
researcher	Text	ESLO1: conversation téléphonique 312_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_312_C
researcher	Text	ESLO1: conversation téléphonique 308_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_308_C
researcher	Text	ESLO1: conversation téléphonique 307_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_307_C
researcher	Text	ESLO1: conversation téléphonique 302_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_302_C
researcher	Text	ESLO1: conversation téléphonique 319_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_319_C
researcher	Text	ESLO1: repas 273_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_273_C
researcher	Text	ESLO1: repas 272_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_272_C
researcher	Text	ESLO1: interview de personnalité 422_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_422_C
researcher	Text	ESLO1: interview de personnalité 421_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_421_C
researcher	Text	ESLO1: interview de personnalité 420_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_420_C
researcher	Text	ESLO1: interview de personnalité 418_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_418_C
researcher	Text	ESLO1: interview de personnalité 417_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_417_C
researcher	Text	ESLO1: interview de personnalité 415_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_415_C
researcher	Text	ESLO1: interview de personnalité 414_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_414_C
researcher	Text	ESLO1: interview de personnalité 409_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_409_C
researcher	Text	ESLO1: interview de personnalité 406_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_406_C
researcher	Text	ESLO1: interview de personnalité 402_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_402_C
researcher	Text	ESLO1: interview de personnalité 401_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_401_C

researcher	Text	ESLO1: contact avent entretien 224_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_224_C
researcher	Text	ESLO1: contact avent entretien 223_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_223_C
researcher	Text	ESLO1: interview de personnalité 419_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_419_C
researcher	Text	ESLO1: interview de personnalité 407_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_407_C
researcher	Text	ESLO1: interaction dans un magasin 630_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_630_C
researcher	Text	ESLO1: interaction dans un magasin 673_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_673_C
researcher	Text	ESLO1: réunion 542_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_542_C
researcher	Text	ESLO1: réunion 519_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_519_C
researcher	Text	ESLO1: interview de personnalité 437_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_437_C
researcher	Text	ESLO1: interview de personnalité 436_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_436_C
researcher	Text	ESLO1: interview de personnalité 431_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_431_C
researcher	Text	ESLO1: interview de personnalité 429_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_429_C
researcher	Text	ESLO1: interview de personnalité 427_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_427_C
researcher	Text	ESLO1: interview de personnalité 426_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_426_C
researcher	Text	ESLO1: interview de personnalité 425_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_425_C
researcher	Text	ESLO1: interview de personnalité 424_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_424_C
researcher	Text	ESLO1: interview de personnalité 423_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_423_C
researcher	Text	ESLO1: interview de personnalité 441_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_441_C
researcher	Text	ESLO1: interview de personnalité 439_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_439_C
researcher	Text	ESLO1: interview de personnalité 438_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_438_C
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 701_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_701_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 702_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_702_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 703_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_703_B

researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 704_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_704_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 707_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_707_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 705_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_705_C
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 709_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_709_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 712_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_712_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 713_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_713_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 717_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_717_C
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 743_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_743_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 751_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_751_C
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 752_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_752_B
researcher	Text	ESLO1: réunion 517_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_517_C
researcher	Text	ESLO1: entretien 002_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_002_B
researcher	Text	ESLO1: entretien 003_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_003_B
researcher	Text	ESLO1: entretien 061_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_061_B
researcher	Text	ESLO1: entretien 004_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_004_B
researcher	Text	ESLO1: entretien 006_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_006_B
researcher	Text	ESLO1: entretien 007_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_007_B
researcher	Text	ESLO1: entretien 008_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_008_B
researcher	Text	ESLO1: entretien 009_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_009_B
researcher	Text	ESLO1: entretien 012_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_012_B

researcher	Text	ESLO1: entretien 013_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_013_B
researcher	Text	ESLO1: entretien 015_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_015_B
researcher	Text	ESLO1: entretien 017_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_017_B
researcher	Text	ESLO1: entretien 018_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_018_B
researcher	Text	ESLO1: entretien 023_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_023_B
researcher	Text	ESLO1: entretien 041_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_041_B
researcher	Text	ESLO1: entretien 042_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_042_B
researcher	Text	ESLO1: entretien 046_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_046_B
researcher	Text	ESLO1: entretien 048_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_048_B
researcher	Text	ESLO1: entretien 052_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_052_B
researcher	Text	ESLO1: entretien 055_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_055_B
researcher	Text	ESLO1: entretien 056_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_056_B
researcher	Text	ESLO1: entretien 059_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_059_B
researcher	Text	ESLO1: entretien 097_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_097_B
researcher	Text	ESLO1: entretien 062_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_062_B
researcher	Text	ESLO1: entretien 064_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_064_B
researcher	Text	ESLO1: entretien 066_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_066_B
researcher	Text	ESLO1: entretien 072_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_072_B
researcher	Text	ESLO1: entretien 073_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_073_B
researcher	Text	ESLO1: entretien 075_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_075_B
researcher	Text	ESLO1: entretien 078_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_078_B
researcher	Text	ESLO1: entretien 080_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_080_B
researcher	Text	ESLO1: entretien 086_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_086_B
researcher	Text	ESLO1: entretien 087_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_087_B
researcher	Text	ESLO1: entretien 090_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_090_B

researcher	Text	ESLO1: entretien 091_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_091_B
researcher	Text	ESLO1: entretien 092_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_092_A
researcher	Text	ESLO1: entretien 096_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_096_B
researcher	Text	ESLO1: interview de personnalité 419_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_419_B
researcher	Text	ESLO1: entretien 100_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_100_B
researcher	Text	ESLO1: entretien 105_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_105_B
researcher	Text	ESLO1: entretien 109_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_109_B
researcher	Text	ESLO1: entretien 111_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_111_B
researcher	Text	ESLO1: entretien 112_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_112_B
researcher	Text	ESLO1: entretien 113_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_113_B
researcher	Text	ESLO1: entretien 120_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_120_B
researcher	Text	ESLO1: entretien 127_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_127_B
researcher	Text	ESLO1: entretien 140_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_140_B
researcher	Text	ESLO1: entretien 149_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_149_B
researcher	Text	ESLO1: entretien 166_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_166_B
researcher	Text	ESLO1: interview de personnalité 409_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_409_B
researcher	Text	ESLO1: interview de personnalité 414_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_414_B
researcher	Text	ESLO1: interview de personnalité 423_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_423_B
researcher	Text	ESLO1: interview de personnalité 424_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_424_B
researcher	Text	ESLO1: interview de personnalité 438_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_438_B
researcher	Text	ESLO1: repas 272_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_272_B
researcher	Text	ESLO1: entretien 001_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_001_A
researcher	Text	ESLO1: entretien 002_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_002_A
researcher	Text	ESLO1: entretien 003_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_003_A
researcher	Text	ESLO1: entretien 004_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_004_A

researcher	Text	ESLO1: entretien 005_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_005_A
researcher	Text	ESLO1: entretien 006_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_006_A
researcher	Text	ESLO1: entretien 007_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_007_A
researcher	Text	ESLO1: entretien 008_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_008_A
researcher	Text	ESLO1: entretien 009_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_009_A
researcher	Text	ESLO1: entretien 010_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_010_A
researcher	Text	ESLO1: entretien 011_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_011_A
researcher	Text	ESLO1: entretien 012_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_012_A
researcher	Text	ESLO1: entretien 013_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_013_A
researcher	Text	ESLO1: entretien 014_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_014_A
researcher	Text	ESLO1: entretien 015_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_015_A
researcher	Text	ESLO1: entretien 017_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_017_A
researcher	Text	ESLO1: entretien 018_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_018_A
researcher	Text	ESLO1: entretien 019_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_019_A
researcher	Text	ESLO1: entretien 020_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_020_A
researcher	Text	ESLO1: entretien 021_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_021_A
researcher	Text	ESLO1: entretien 022_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_022_A
researcher	Text	ESLO1: entretien 023_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_023_A
researcher	Text	ESLO1: entretien 024_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_024_A
researcher	Text	ESLO1: entretien 025_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_025_A
researcher	Text	ESLO1: entretien 026_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_026_A
researcher	Text	ESLO1: entretien 027_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_027_A
researcher	Text	ESLO1: entretien 028_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_028_A
researcher	Text	ESLO1: entretien 029_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_029_A
researcher	Text	ESLO1: entretien 030_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_030_A

researcher	Text	ESLO1: entretien 041_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_041_A
researcher	Text	ESLO1: entretien 042_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_042_A
researcher	Text	ESLO1: entretien 043_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_043_A
researcher	Text	ESLO1: entretien 044_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_044_A
researcher	Text	ESLO1: entretien 045_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_045_A
researcher	Text	ESLO1: entretien 046_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_046_A
researcher	Text	ESLO1: entretien 047_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_047_A
researcher	Text	ESLO1: entretien 048_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_048_A
researcher	Text	ESLO1: entretien 050_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_050_A
researcher	Text	ESLO1: entretien 051_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_051_A
researcher	Text	ESLO1: entretien 052_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_052_A
researcher	Text	ESLO1: entretien 053_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_053_A
researcher	Text	ESLO1: entretien 054_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_054_A
researcher	Text	ESLO1: entretien 055_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_055_A
researcher	Text	ESLO1: entretien 056_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_056_A
researcher	Text	ESLO1: entretien 057_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_057_A
researcher	Text	ESLO1: entretien 058_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_058_A
researcher	Text	ESLO1: entretien 059_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_059_A
researcher	Text	ESLO1: entretien 060_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_060_A
researcher	Text	ESLO1: entretien 061_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_061_A
researcher	Text	ESLO1: entretien 062_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_062_A
researcher	Text	ESLO1: entretien 063_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_063_A
researcher	Text	ESLO1: entretien 064_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_064_A
researcher	Text	ESLO1: entretien 065_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_065_A
researcher	Text	ESLO1: entretien 066_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_066_A

researcher	Text	ESLO1: entretien 067_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_067_A
researcher	Text	ESLO1: entretien 068_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_068_A
researcher	Text	ESLO1: entretien 069_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_069_A
researcher	Text	ESLO1: entretien 070_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_070_A
researcher	Text	ESLO1: entretien 071_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_071_A
researcher	Text	ESLO1: entretien 072_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_072_A
researcher	Text	ESLO1: entretien 073_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_073_A
researcher	Text	ESLO1: entretien 075_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_075_A
researcher	Text	ESLO1: entretien 076_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_076_A
researcher	Text	ESLO1: entretien 078_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_078_A
researcher	Text	ESLO1: entretien 079_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_079_A
researcher	Text	ESLO1: entretien 080_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_080_A
researcher	Text	ESLO1: entretien 081_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_081_A
researcher	Text	ESLO1: entretien 082_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_082_A
researcher	Text	ESLO1: entretien 083_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_083_A
researcher	Text	ESLO1: entretien 084_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_084_A
researcher	Text	ESLO1: entretien 085_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_085_A
researcher	Text	ESLO1: entretien 086_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_086_A
researcher	Text	ESLO1: entretien 087_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_087_A
researcher	Text	ESLO1: entretien 088_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_088_A
researcher	Text	ESLO1: entretien 089_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_089_A
researcher	Text	ESLO1: entretien 090_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_090_A
researcher	Text	ESLO1: entretien 091_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_091_A
researcher	Text	ESLO1: entretien 092_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_092_B
researcher	Text	ESLO1: entretien 093_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_093_A

researcher	Text	ESLO1: entretien 095_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_095_A
researcher	Text	ESLO1: entretien 096_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_096_A
researcher	Text	ESLO1: entretien 097_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_097_A
researcher	Text	ESLO1: entretien 100_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_100_A
researcher	Text	ESLO1: repas 273_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_273_B
researcher	Text	ESLO1: entretien 122_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_122_B
researcher	Text	ESLO1: entretien 121_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_121_B
researcher	Text	ESLO1: entretien 119_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_119_B
researcher	Text	ESLO1: entretien 118_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_118_B
researcher	Text	ESLO1: entretien 117_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_117_B
researcher	Text	ESLO1: entretien 115_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_115_B
researcher	Text	ESLO1: entretien 114_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_114_B
researcher	Text	ESLO1: entretien 110_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_110_B
researcher	Text	ESLO1: entretien 108_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_108_B
researcher	Text	ESLO1: entretien 107_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_107_B
researcher	Text	ESLO1: entretien 106_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_106_B
researcher	Text	ESLO1: entretien 103_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_103_B
researcher	Text	ESLO1: entretien 102_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_102_B
researcher	Text	ESLO1: entretien 101_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_101_B
researcher	Text	ESLO1: conversation téléphonique 319_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_319_B
researcher	Text	ESLO1: conversation téléphonique 318_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_318_B
researcher	Text	ESLO1: conversation téléphonique 316_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_316_B
researcher	Text	ESLO1: conversation téléphonique 312_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_312_B
researcher	Text	ESLO1: conversation téléphonique 308_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_308_B
researcher	Text	ESLO1: conversation téléphonique 307_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_307_B

researcher	Text	ESLO1: conversation téléphonique 302_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_302_B
researcher	Text	ESLO1: réunion 542_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_542_B
researcher	Text	ESLO1: réunion 519_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_519_B
researcher	Text	ESLO1: réunion 517_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_517_B
researcher	Text	ESLO1: interaction dans un magasin 673_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_673_B
researcher	Text	ESLO1: interaction dans un magasin 630_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_630_B
researcher	Text	ESLO1: interview de personnalité 441_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_441_B
researcher	Text	ESLO1: interview de personnalité 437_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_437_B
researcher	Text	ESLO1: interview de personnalité 431_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_431_B
researcher	Text	ESLO1: interview de personnalité 429_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_429_B
researcher	Text	ESLO1: interview de personnalité 427_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_427_B
researcher	Text	ESLO1: interview de personnalité 426_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_426_B
researcher	Text	ESLO1: interview de personnalité 425_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_425_B
researcher	Text	ESLO1: interview de personnalité 422_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_422_B
researcher	Text	ESLO1: interview de personnalité 421_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_421_B
researcher	Text	ESLO1: interview de personnalité 420_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_420_B
researcher	Text	ESLO1: interview de personnalité 418_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_418_B
researcher	Text	ESLO1: interview de personnalité 417_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_417_B
researcher	Text	ESLO1: interview de personnalité 415_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_415_B
researcher	Text	ESLO1: interview de personnalité 407_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_407_B
researcher	Text	ESLO1: interview de personnalité 406_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_406_B
researcher	Text	ESLO1: interview de personnalité 401_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_401_B
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 263_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_263_B
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 260_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_260_B

researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 257_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_257_B
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 254_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_254_B
researcher	Text	ESLO1: contact avent entretien 224_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_224_B
researcher	Text	ESLO1: contact avent entretien 223_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_223_B
researcher	Text	ESLO1: contact avent entretien 217_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_217_B
researcher	Text	ESLO1: contact avent entretien 216_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_216_B
researcher	Text	ESLO1: contact avent entretien 215_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_215_B
researcher	Text	ESLO1: contact avent entretien 212_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_212_B
researcher	Text	ESLO1: contact avent entretien 211_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_211_B
researcher	Text	ESLO1: contact avent entretien 205_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_205_B
researcher	Text	ESLO1: contact avent entretien 204_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_204_B
researcher	Text	ESLO1: contact avent entretien 203_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_203_B
researcher	Text	ESLO1: contact avent entretien 202_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_202_B
researcher	Text	ESLO1: contact avent entretien 201_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_201_B
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 255_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_255_B
researcher	Text	ESLO1: entretien 173_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_173_B
researcher	Text	ESLO1: entretien 172_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_172_B
researcher	Text	ESLO1: entretien 170_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_170_B
researcher	Text	ESLO1: entretien 169_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_169_B
researcher	Text	ESLO1: entretien 167_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_167_B
researcher	Text	ESLO1: entretien 160_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_160_B
researcher	Text	ESLO1: entretien 150_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_150_B
researcher	Text	ESLO1: entretien 142_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_142_B

researcher	Text	ESLO1: entretien 141_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_141_B
researcher	Text	ESLO1: entretien 139_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_139_B
researcher	Text	ESLO1: entretien 133_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_133_B
researcher	Text	ESLO1: entretien 132_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_132_B
researcher	Text	ESLO1: entretien 131_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_131_B
researcher	Text	ESLO1: entretien 129_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_129_B
researcher	Text	ESLO1: entretien 126_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_126_B
researcher	Text	ESLO1: entretien 125_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_125_B
researcher	Text	ESLO1: entretien 124_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_124_B
researcher	Text	ESLO1: entretien 123_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_123_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 751_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_751_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 717_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_717_B
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 705_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_705_B
researcher	Text	ESLO1: entretien 105_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_105_A
researcher	Text	ESLO1: entretien 103_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_103_A
researcher	Text	ESLO1: entretien 102_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_102_A
researcher	Text	ESLO1: entretien 101_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_101_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 701_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_701_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 702_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_702_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 703_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_703_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 704_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_704_A

researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 705_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_705_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 707_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_707_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 709_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_709_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 712_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_712_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 713_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_713_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 717_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_717_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 743_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_743_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 751_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_751_A
researcher	Text	ESLO1: consultation au centre médico-psycho-pedagogique 752_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_CONSCMPP_752_A
researcher	Text	ESLO1: entretien 106_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_106_A
researcher	Text	ESLO1: entretien 107_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_107_A
researcher	Text	ESLO1: entretien 108_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_108_A
researcher	Text	ESLO1: entretien 109_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_109_A
researcher	Text	ESLO1: entretien 110_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_110_A
researcher	Text	ESLO1: entretien 111_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_111_A
researcher	Text	ESLO1: entretien 112_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_112_A
researcher	Text	ESLO1: entretien 113_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_113_A
researcher	Text	ESLO1: entretien 114_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_114_A
researcher	Text	ESLO1: entretien 115_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_115_A
researcher	Text	ESLO1: entretien 117_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_117_A

researcher	Text	ESLO1: entretien 118_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_118_A
researcher	Text	ESLO1: entretien 119_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_119_A
researcher	Text	ESLO1: entretien 120_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_120_A
researcher	Text	ESLO1: entretien 121_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_121_A
researcher	Text	ESLO1: entretien 122_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_122_A
researcher	Text	ESLO1: entretien 123_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_123_A
researcher	Text	ESLO1: entretien 124_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_124_A
researcher	Text	ESLO1: entretien 125_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_125_A
researcher	Text	ESLO1: entretien 126_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_126_A
researcher	Text	ESLO1: entretien 127_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_127_A
researcher	Text	ESLO1: entretien 129_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_129_A
researcher	Text	ESLO1: entretien 131_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_131_A
researcher	Text	ESLO1: entretien 132_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_132_A
researcher	Text	ESLO1: entretien 133_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_133_A
researcher	Text	ESLO1: entretien 139_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_139_A
researcher	Text	ESLO1: entretien 140_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_140_A
researcher	Text	ESLO1: entretien 141_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_141_A
researcher	Text	ESLO1: entretien 142_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_142_A
researcher	Text	ESLO1: entretien 149_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_149_A
researcher	Text	ESLO1: entretien 150_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_150_A
researcher	Text	ESLO1: entretien 160_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_160_A
researcher	Text	ESLO1: entretien 166_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_166_A
researcher	Text	ESLO1: entretien 167_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_167_A
researcher	Text	ESLO1: entretien 169_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_169_A

researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 255_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_255_A
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 259_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_259_A
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 259_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_259_B
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 264_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_264_A
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 264_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_264_B
researcher	Text	ESLO1: discussion en clôture de la séquence d'entretien 264_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCLO_264_C
researcher	Text	ESLO1: contact avent entretien 201_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_201_A
researcher	Text	ESLO1: contact avent entretien 202_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_202_A
researcher	Text	ESLO1: contact avent entretien 203_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_203_A
researcher	Text	ESLO1: contact avent entretien 204_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_204_A
researcher	Text	ESLO1: contact avent entretien 205_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_205_A
researcher	Text	ESLO1: contact avent entretien 206_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_206_A
researcher	Text	ESLO1: contact avent entretien 206_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_206_B
researcher	Text	ESLO1: contact avent entretien 208_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_208_A
researcher	Text	ESLO1: contact avent entretien 208_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_208_B
researcher	Text	ESLO1: contact avent entretien 211_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_211_A
researcher	Text	ESLO1: contact avent entretien 212_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_212_A
researcher	Text	ESLO1: contact avent entretien 214_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_214_B
researcher	Text	ESLO1: contact avent entretien 215_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_215_A
researcher	Text	ESLO1: contact avent entretien 216_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_216_A
researcher	Text	ESLO1: contact avent entretien 217_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_217_A
researcher	Text	ESLO1: contact avent entretien 223_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_223_A

researcher	Text	ESLO1: contact avent entretien 224_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_224_A
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 254_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_254_A
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 257_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_257_A
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 258_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_258_A
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 258_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_258_B
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 260_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_260_A
researcher	Text	ESLO1: discussion en ouverture de la séquence d'entretien 263_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTOUV_263_A
researcher	Text	ESLO1: interview de personnalité 401_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_401_A
researcher	Text	ESLO1: interview de personnalité 402_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_402_A
researcher	Text	ESLO1: interview de personnalité 406_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_406_A
researcher	Text	ESLO1: interview de personnalité 407_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_407_A
researcher	Text	ESLO1: interview de personnalité 409_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_409_A
researcher	Text	ESLO1: interview de personnalité 414_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_414_A
researcher	Text	ESLO1: interview de personnalité 415_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_415_A
researcher	Text	ESLO1: interview de personnalité 417_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_417_A
researcher	Text	ESLO1: interview de personnalité 418_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_418_A
researcher	Text	ESLO1: interview de personnalité 419_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_419_A
researcher	Text	ESLO1: interview de personnalité 420_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_420_A
researcher	Text	ESLO1: interview de personnalité 421_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_421_A
researcher	Text	ESLO1: interview de personnalité 422_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_422_A
researcher	Text	ESLO1: interview de personnalité 423_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_423_A
researcher	Text	ESLO1: interview de personnalité 424_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_424_A

researcher	Text	ESLO1: interview de personnalité 425_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_425_A
researcher	Text	ESLO1: interview de personnalité 426_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_426_A
researcher	Text	ESLO1: interview de personnalité 427_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_427_A
researcher	Text	ESLO1: interview de personnalité 429_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_429_A
researcher	Text	ESLO1: interview de personnalité 431_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_431_A
researcher	Text	ESLO1: interview de personnalité 437_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_437_A
researcher	Text	ESLO1: interview de personnalité 438_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_438_A
researcher	Text	ESLO1: interview de personnalité 441_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_441_A
researcher	Text	ESLO1: interaction dans un magasin 290_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_290_A
researcher	Text	ESLO1: interaction dans un magasin 630_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_630_A
researcher	Text	ESLO1: interaction dans un magasin 673_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_673_A
researcher	Text	ESLO1: repas 272_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_272_A
researcher	Text	ESLO1: réunion 517_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_517_A
researcher	Text	ESLO1: réunion 519_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_519_A
researcher	Text	ESLO1: réunion 542_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REU_542_A
researcher	Text	ESLO1: conversation téléphonique 302_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_302_A
researcher	Text	ESLO1: conversation téléphonique 307_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_307_A
researcher	Text	ESLO1: conversation téléphonique 308_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_308_A
researcher	Text	ESLO1: conversation téléphonique 312_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_312_A
researcher	Text	ESLO1: conversation téléphonique 316_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_316_A
researcher	Text	ESLO1: conversation téléphonique 318_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_318_A
researcher	Text	ESLO1: conversation téléphonique 319_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_TEL_319_A
researcher	Text	ESLO1: contact avent entretien 214_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENTCONT_214_A
researcher	Text	ESLO1: interview de personnalité 402_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_402_B
researcher	Text	ESLO1: interview de personnalité 436_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_436_A

researcher	Text	ESLO1: interview de personnalité 436_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_436_B
researcher	Text	ESLO1: interview de personnalité 439_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_439_A
researcher	Text	ESLO1: interview de personnalité 439_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_INTPERS_439_B
researcher	Text	ESLO1: interaction dans un magasin 290_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_MAG_290_B
researcher	Text	ESLO1: repas 273_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_REPAS_273_A
researcher	Text	ESLO1: entretien 173_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_173_A
researcher	Text	ESLO1: entretien 172_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_172_A
researcher	Text	ESLO1: entretien 170_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_170_A
researcher	Text	ESLO1: entretien 098_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_098_C
researcher	Text	ESLO1: entretien 098_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_098_B
researcher	Text	ESLO1: entretien 098_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO1_ENT_098_A
researcher	Sound	ESLO2: entretien 1005	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1005
researcher	Text	ESLO2: Entretien 1005	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1005_C
researcher	Sound	ESLO2: entretien 1004	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1004
researcher	Sound	ESLO2: entretien 1009	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1009
researcher	Sound	ESLO2: entretien 1001	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1001
researcher	Text	ESLO2: entretien 1009_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1009_C
researcher	Text	ESLO2: entretien 1004_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1004_C
researcher	Sound	ESLO2: entretien 1014	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1014
researcher	Text	ESLO2: entretien 1001_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1001_C
researcher	Text	ESLO2: entretien 1014_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1014_C
researcher	Sound	ESLO2: entretien 1016	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1016
researcher	Sound	ESLO2: entretien 1024	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1024
researcher	Sound	ESLO2: entretien 1025	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1025
researcher	Text	ESLO2: entretien 1016_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1016_C

researcher	Sound	ESLO2: entretien 1039	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1039
researcher	Text	ESLO2: entretien 1025_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1025_C
researcher	Text	ESLO2: entretien 1024_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1024_C
researcher	Sound	ESLO2: entretien 1052	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1052
researcher	Sound	ESLO2: entretien 1055	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1055
researcher	Sound	ESLO2: entretien 1056	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1056
researcher	Sound	ESLO2: entretien 1083	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1083
researcher	Sound	ESLO2: entretien 1272	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1272
researcher	Text	ESLO2: entretien 1039_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1039_C
researcher	Sound	ESLO2: demande d'itinéraire 1072	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1072
researcher	Sound	ESLO2: demande d'itinéraire 1073	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1073
researcher	Sound	ESLO2: demande d'itinéraire 1086	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1086
researcher	Sound	ESLO2: demande d'itinéraire 1087	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1087
researcher	Sound	ESLO2: demande d'itinéraire 1088	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1088
researcher	Sound	ESLO2: demande d'itinéraire 1089	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1089
researcher	Sound	ESLO2: demande d'itinéraire 1090	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1090
researcher	Sound	ESLO2: demande d'itinéraire 1091	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1091
researcher	Sound	ESLO2: demande d'itinéraire 1092	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1092
researcher	Sound	ESLO2: demande d'itinéraire 1093	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1093
researcher	Sound	ESLO2: demande d'itinéraire 1094	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1094
researcher	Sound	ESLO2: demande d'itinéraire 1095	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1095
researcher	Sound	ESLO2: demande d'itinéraire 1096	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1096
researcher	Sound	ESLO2: demande d'itinéraire 1097	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1097
researcher	Sound	ESLO2: demande d'itinéraire 1098	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1098
researcher	Sound	ESLO2: demande d'itinéraire 1099	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1099

researcher	Sound	ESLO2: demande d'itinéraire 1100	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1100
researcher	Sound	ESLO2: demande d'itinéraire 1101	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1101
researcher	Sound	ESLO2: demande d'itinéraire 1102	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1102
researcher	Sound	ESLO2: demande d'itinéraire 1103	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1103
researcher	Sound	ESLO2: demande d'itinéraire 1104	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1104
researcher	Sound	ESLO2: demande d'itinéraire 1105	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1105
researcher	Sound	ESLO2: demande d'itinéraire 1106	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1106
researcher	Sound	ESLO2: demande d'itinéraire 1107	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1107
researcher	Sound	ESLO2: demande d'itinéraire 1108	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1108
researcher	Sound	ESLO2: demande d'itinéraire 1109	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1109
researcher	Sound	ESLO2: demande d'itinéraire 1110	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1110
researcher	Sound	ESLO2: demande d'itinéraire 1111	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1111
researcher	Sound	ESLO2: demande d'itinéraire 1112	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1112
researcher	Sound	ESLO2: demande d'itinéraire 1113	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1113
researcher	Sound	ESLO2: demande d'itinéraire 1114	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1114
researcher	Sound	ESLO2: demande d'itinéraire 1115	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1115
researcher	Sound	ESLO2: demande d'itinéraire 1116	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1116
researcher	Sound	ESLO2: demande d'itinéraire 1117	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1117
researcher	Sound	ESLO2: demande d'itinéraire 1119	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1119
researcher	Sound	ESLO2: demande d'itinéraire 1120	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1120
researcher	Sound	ESLO2: demande d'itinéraire 1121	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1121
researcher	Sound	ESLO2: demande d'itinéraire 1122	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1122
researcher	Sound	ESLO2: demande d'itinéraire 1123	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1123
researcher	Sound	ESLO2: demande d'itinéraire 1124	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1124
researcher	Text	ESLO2: entretien 1052_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1052_C

researcher	Text	ESLO2: entretien 1083_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1083_C
researcher	Text	ESLO2: entretien 1056_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1056_C
researcher	Text	ESLO2: entretien 1055_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1055_C
researcher	Text	ESLO2: entretien 1272_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1272_C
researcher	Sound	ESLO2: demande d'itinéraire 1118	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1118
researcher	Sound	ESLO2: demande d'itinéraire 1125	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1125
researcher	Sound	ESLO2: demande d'itinéraire 1126	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1126
researcher	Sound	ESLO2: demande d'itinéraire 1127	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1127
researcher	Sound	ESLO2: demande d'itinéraire 1128	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1128
researcher	Sound	ESLO2: demande d'itinéraire 1129	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1129
researcher	Sound	ESLO2: demande d'itinéraire 1130	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1130
researcher	Sound	ESLO2: demande d'itinéraire 1131	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1131
researcher	Sound	ESLO2: demande d'itinéraire 1132	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1132
researcher	Sound	ESLO2: demande d'itinéraire 1133	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1133
researcher	Sound	ESLO2: demande d'itinéraire 1134	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1134
researcher	Sound	ESLO2: demande d'itinéraire 1135	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1135
researcher	Sound	ESLO2: demande d'itinéraire 1136	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1136
researcher	Sound	ESLO2: demande d'itinéraire 1137	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1137
researcher	Sound	ESLO2: demande d'itinéraire 1138	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1138
researcher	Sound	ESLO2: demande d'itinéraire 1139	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1139
researcher	Sound	ESLO2: demande d'itinéraire 1140	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1140
researcher	Sound	ESLO2: demande d'itinéraire 1141	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1141
researcher	Sound	ESLO2: demande d'itinéraire 1142	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1142
researcher	Sound	ESLO2: demande d'itinéraire 1143	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1143
researcher	Sound	ESLO2: demande d'itinéraire 1144	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1144

researcher	Sound	ESLO2: demande d'itinéraire 1145	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1145
researcher	Sound	ESLO2: demande d'itinéraire 1146	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1146
researcher	Sound	ESLO2: demande d'itinéraire 1147	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1147
researcher	Sound	ESLO2: demande d'itinéraire 1148	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1148
researcher	Sound	ESLO2: demande d'itinéraire 1149	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1149
researcher	Sound	ESLO2: demande d'itinéraire 1150	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1150
researcher	Sound	ESLO2: demande d'itinéraire 1151	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1151
researcher	Sound	ESLO2: demande d'itinéraire 1152	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1152
researcher	Sound	ESLO2: demande d'itinéraire 1153	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1153
researcher	Sound	ESLO2: demande d'itinéraire 1154	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1154
researcher	Sound	ESLO2: demande d'itinéraire 1155	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1155
researcher	Sound	ESLO2: demande d'itinéraire 1156	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1156
researcher	Sound	ESLO2: demande d'itinéraire 1157	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1157
researcher	Sound	ESLO2: demande d'itinéraire 1158	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1158
researcher	Sound	ESLO2: demande d'itinéraire 1159	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1159
researcher	Sound	ESLO2: demande d'itinéraire 1160	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1160
researcher	Sound	ESLO2: demande d'itinéraire 1161	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1161
researcher	Sound	ESLO2: demande d'itinéraire 1162	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1162
researcher	Sound	ESLO2: demande d'itinéraire 1163	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1163
researcher	Sound	ESLO2: demande d'itinéraire 1164	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1164
researcher	Sound	ESLO2: demande d'itinéraire 1165	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1165
researcher	Sound	ESLO2: demande d'itinéraire 1166	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1166
researcher	Sound	ESLO2: demande d'itinéraire 1167	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1167
researcher	Sound	ESLO2: demande d'itinéraire 1168	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1168
researcher	Sound	ESLO2: demande d'itinéraire 1169	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1169

researcher	Sound	ESLO2: demande d'itinéraire 1170	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1170
researcher	Sound	ESLO2: demande d'itinéraire 1171	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1171
researcher	Sound	ESLO2: demande d'itinéraire 1172	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1172
researcher	Sound	ESLO2: demande d'itinéraire 1173	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1173
researcher	Sound	ESLO2: demande d'itinéraire 1174	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1174
researcher	Sound	ESLO2: demande d'itinéraire 1175	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1175
researcher	Text	ESLO2: demande d'itinéraire 1119_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1119_C
researcher	Text	ESLO2: demande d'itinéraire 1117_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1117_C
researcher	Text	ESLO2: demande d'itinéraire 1116_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1116_C
researcher	Text	ESLO2: demande d'itinéraire 1115_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1115_C
researcher	Text	ESLO2: demande d'itinéraire 1113_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1113_C
researcher	Text	ESLO2: demande d'itinéraire 1112_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1112_C
researcher	Text	ESLO2: demande d'itinéraire 1111_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1111_C
researcher	Text	ESLO2: demande d'itinéraire 1110_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1110_C
researcher	Text	ESLO2: demande d'itinéraire 1109_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1109_C
researcher	Text	ESLO2: demande d'itinéraire 1108_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1108_C
researcher	Text	ESLO2: demande d'itinéraire 1107_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1107_C
researcher	Text	ESLO2: demande d'itinéraire 1106_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1106_C
researcher	Text	ESLO2: demande d'itinéraire 1105_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1105_C
researcher	Text	ESLO2: demande d'itinéraire 1103_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1103_C
researcher	Text	ESLO2: demande d'itinéraire 1102_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1102_C
researcher	Text	ESLO2: demande d'itinéraire 1101_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1101_C
researcher	Text	ESLO2: demande d'itinéraire 1100_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1100_C
researcher	Text	ESLO2: demande d'itinéraire 1099_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1099_C
researcher	Text	ESLO2: demande d'itinéraire 1098_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1098_C

researcher	Text	ESLO2: demande d'itinéraire 1097_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1097_C
researcher	Text	ESLO2: demande d'itinéraire 1096_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1096_C
researcher	Text	ESLO2: demande d'itinéraire 1095_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1095_C
researcher	Text	ESLO2: demande d'itinéraire 1094_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1094_C
researcher	Text	ESLO2: demande d'itinéraire 1093_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1093_C
researcher	Text	ESLO2: demande d'itinéraire 1092_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1092_C
researcher	Text	ESLO2: demande d'itinéraire 1091_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1091_C
researcher	Text	ESLO2: demande d'itinéraire 1090_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1090_C
researcher	Text	ESLO2: demande d'itinéraire 1089_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1089_C
researcher	Text	ESLO2: demande d'itinéraire 1088_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1088_C
researcher	Text	ESLO2: demande d'itinéraire 1087_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1087_C
researcher	Text	ESLO2: demande d'itinéraire 1086_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1086_C
researcher	Text	ESLO2: demande d'itinéraire 1073_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1073_C
researcher	Text	ESLO2: demande d'itinéraire 1072_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1072_C
researcher	Text	ESLO2: demande d'itinéraire 1124_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1124_C
researcher	Text	ESLO2: demande d'itinéraire 1123_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1123_C
researcher	Text	ESLO2: demande d'itinéraire 1122_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1122_C
researcher	Text	ESLO2: demande d'itinéraire 1121_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1121_C
researcher	Text	ESLO2: demande d'itinéraire 1120_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1120_C
researcher	Sound	ESLO2: discours 1237	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1237
researcher	Sound	ESLO2: discours 1238	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1238
researcher	Sound	ESLO2: discours 1239	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1239
researcher	Sound	ESLO2: entretien de 14-25 ans 1232	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1232
researcher	Sound	ESLO2: entretien de 14-25 ans 1234	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1234
researcher	Text	ESLO2: demande d'itinéraire 1175_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1175_C

researcher	Text	ESLO2: demande d'itinéraire 1174_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1174_C
researcher	Text	ESLO2: demande d'itinéraire 1173_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1173_C
researcher	Text	ESLO2: demande d'itinéraire 1172_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1172_C
researcher	Text	ESLO2: demande d'itinéraire 1171_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1171_C
researcher	Text	ESLO2: demande d'itinéraire 1170_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1170_C
researcher	Text	ESLO2: demande d'itinéraire 1169_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1169_C
researcher	Text	ESLO2: demande d'itinéraire 1168_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1168_C
researcher	Text	ESLO2: demande d'itinéraire 1167_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1167_C
researcher	Text	ESLO2: demande d'itinéraire 1166_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1166_C
researcher	Text	ESLO2: demande d'itinéraire 1165_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1165_C
researcher	Text	ESLO2: demande d'itinéraire 1164_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1164_C
researcher	Text	ESLO2: demande d'itinéraire 1163_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1163_C
researcher	Text	ESLO2: demande d'itinéraire 1162_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1162_C
researcher	Text	ESLO2: demande d'itinéraire 1161_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1161_C
researcher	Text	ESLO2: demande d'itinéraire 1160_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1160_C
researcher	Text	ESLO2: demande d'itinéraire 1159_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1159_C
researcher	Text	ESLO2: demande d'itinéraire 1158_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1158_C
researcher	Text	ESLO2: demande d'itinéraire 1157_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1157_C
researcher	Text	ESLO2: demande d'itinéraire 1156_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1156_C
researcher	Text	ESLO2: demande d'itinéraire 1155_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1155_C
researcher	Text	ESLO2: demande d'itinéraire 1154_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1154_C
researcher	Text	ESLO2: demande d'itinéraire 1153_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1153_C
researcher	Text	ESLO2: demande d'itinéraire 1152_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1152_C
researcher	Text	ESLO2: demande d'itinéraire 1151_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1151_C
researcher	Text	ESLO2: demande d'itinéraire 1150_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1150_C

researcher	Text	ESLO2: demande d'itinéraire 1149_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1149_C
researcher	Text	ESLO2: demande d'itinéraire 1148_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1148_C
researcher	Text	ESLO2: demande d'itinéraire 1147_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1147_C
researcher	Text	ESLO2: demande d'itinéraire 1146_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1146_C
researcher	Text	ESLO2: demande d'itinéraire 1145_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1145_C
researcher	Text	ESLO2: demande d'itinéraire 1144_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1144_C
researcher	Text	ESLO2: demande d'itinéraire 1143_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1143_C
researcher	Text	ESLO2: demande d'itinéraire 1142_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1142_C
researcher	Text	ESLO2: demande d'itinéraire 1141_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1141_C
researcher	Text	ESLO2: demande d'itinéraire 1140_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1140_C
researcher	Text	ESLO2: demande d'itinéraire 1139_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1139_C
researcher	Text	ESLO2: demande d'itinéraire 1138_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1138_C
researcher	Text	ESLO2: demande d'itinéraire 1137_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1137_C
researcher	Text	ESLO2: demande d'itinéraire 1136_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1136_C
researcher	Text	ESLO2: demande d'itinéraire 1135_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1135_C
researcher	Text	ESLO2: demande d'itinéraire 1134_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1134_C
researcher	Text	ESLO2: demande d'itinéraire 1133_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1133_C
researcher	Text	ESLO2: demande d'itinéraire 1132_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1132_C
researcher	Text	ESLO2: demande d'itinéraire 1131_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1131_C
researcher	Text	ESLO2: demande d'itinéraire 1130_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1130_C
researcher	Text	ESLO2: demande d'itinéraire 1129_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1129_C
researcher	Text	ESLO2: demande d'itinéraire 1128_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1128_C
researcher	Text	ESLO2: demande d'itinéraire 1127_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1127_C
researcher	Text	ESLO2: demande d'itinéraire 1126_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1126_C
researcher	Text	ESLO2: demande d'itinéraire 1125_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1125_C

researcher	Sound	ESLO2: entretien de 14-25 ans 1228	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1228
researcher	Sound	ESLO2: entretien de 14-25 ans 1229	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1229
researcher	Sound	ESLO2: entretien de 14-25 ans 1230	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1230
researcher	Sound	ESLO2: entretien de 14-25 ans 1231	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1231
researcher	Sound	ESLO2: entretien de 14-25 ans 1233	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1233
researcher	Sound	ESLO2: entretien de 14-25 ans 1235	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1235
researcher	Sound	ESLO2: entretien de 14-25 ans 1236	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1236
researcher	Text	ESLO2: demande d'itinéraire 1118_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1118_C
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1176	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1176
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1177	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1177
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1178	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1178
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1179	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1179
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1181	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1181
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1182	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1182
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1183	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1183
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1184	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1184
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1185	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1185
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1186	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1186
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1187	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1187
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1188	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1188
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1189	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1189
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1190	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1190
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1191	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1191
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1192	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1192
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1193	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1193

researcher	Sound	ESLO2: interaction à la sortie du cinéma 1194	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1194
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1195	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1195
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1196	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1196
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1199	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1199
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1200	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1200
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1201	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1201
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1202	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1202
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1203	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1203
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1204	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1204
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1205	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1205
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1206	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1206
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1207	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1207
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1208	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1208
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1209	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1209
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1210	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1210
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1211	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1211
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1212	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1212
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1213	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1213
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1214	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1214
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1215	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1215
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1216	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1216
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1217	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1217
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1218	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1218
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1219	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1219
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1220	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1220

researcher	Text	ESLO2: entretien de 14-25 ans 1228_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1228_C
researcher	Text	ESLO2: entretien de 14-25 ans 1232_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1232_C
researcher	Text	ESLO2: entretien de 14-25 ans 1231_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1231_C
researcher	Text	ESLO2: entretien de 14-25 ans 1230_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1230_C
researcher	Text	ESLO2: entretien de 14-25 ans 1229_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1229_C
researcher	Text	ESLO2: entretien de 14-25 ans 1236_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1236_C
researcher	Text	ESLO2: entretien de 14-25 ans 1235_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1235_C
researcher	Text	ESLO2: entretien de 14-25 ans 1234_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1234_C
researcher	Text	ESLO2: entretien de 14-25 ans 1233_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1233_C
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1198	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1198
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1197	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1197
researcher	Sound	ESLO2: interaction à la sortie du cinéma 1180	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1180
researcher	Text	ESLO2: interaction à la sortie du cinéma 1191_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1191_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1190_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1190_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1189_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1189_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1188_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1188_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1187_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1187_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1186_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1186_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1185_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1185_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1184_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1184_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1183_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1183_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1182_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1182_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1181_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1181_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1179_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1179_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1178_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1178_C

researcher	Text	ESLO2: interaction à la sortie du cinéma 1177_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1177_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1176_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1176_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1192_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1192_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1193_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1193_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1207_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1207_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1206_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1206_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1205_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1205_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1204_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1204_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1203_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1203_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1202_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1202_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1201_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1201_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1200_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1200_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1199_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1199_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1198_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1198_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1197_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1197_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1196_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1196_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1195_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1195_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1194_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1194_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1180_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1180_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1220_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1220_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1219_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1219_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1218_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1218_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1217_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1217_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1216_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1216_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1215_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1215_C

researcher	Text	ESLO2: interaction à la sortie du cinéma 1214_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1214_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1213_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1213_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1212_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1212_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1211_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1211_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1210_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1210_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1209_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1209_C
researcher	Text	ESLO2: interaction à la sortie du cinéma 1208_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1208_C
researcher	Sound	ESLO2: conférence 1074	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1074
researcher	Sound	ESLO2: conférence 1240	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1240
researcher	Sound	ESLO2: conférence 1241	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1241
researcher	Sound	ESLO2: conférence 1242	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1242
researcher	Sound	ESLO2: conférence 1243	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1243
researcher	Sound	ESLO2: conférence 1244	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1244
researcher	Text	ESLO2: discours 1239_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1239_C
researcher	Text	ESLO2: discours 1239_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1239_B
researcher	Text	ESLO2: discours 1239_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1239_A
researcher	Text	ESLO2: discours 1238_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1238_C
researcher	Text	ESLO2: discours 1238_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1238_B
researcher	Text	ESLO2: discours 1238_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1238_A
researcher	Text	ESLO2: discours 1237_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1237_C
researcher	Text	ESLO2: discours 1237_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1237_B
researcher	Text	ESLO2: discours 1237_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_DISC_1237_A
researcher	Text	ESLO2: demande d'itinéraire 1121_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1121_A
researcher	Text	ESLO2: demande d'itinéraire 1120_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1120_B
researcher	Text	ESLO2: demande d'itinéraire 1120_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1120_A

researcher	Text	ESLO2: demande d'itinéraire 1119_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1119_B
researcher	Text	ESLO2: demande d'itinéraire 1119_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1119_A
researcher	Text	ESLO2: demande d'itinéraire 1118_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1118_B
researcher	Text	ESLO2: demande d'itinéraire 1118_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1118_A
researcher	Text	ESLO2: demande d'itinéraire 1117_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1117_B
researcher	Text	ESLO2: demande d'itinéraire 1117_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1117_A
researcher	Text	ESLO2: demande d'itinéraire 1116_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1116_B
researcher	Text	ESLO2: demande d'itinéraire 1116_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1116_A
researcher	Text	ESLO2: demande d'itinéraire 1115_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1115_B
researcher	Text	ESLO2: demande d'itinéraire 1115_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1115_A
researcher	Text	ESLO2: demande d'itinéraire 1114_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1114_B
researcher	Text	ESLO2: demande d'itinéraire 1114_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1114_A
researcher	Text	ESLO2: demande d'itinéraire 1113_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1113_B
researcher	Text	ESLO2: demande d'itinéraire 1113_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1113_A
researcher	Text	ESLO2: demande d'itinéraire 1112_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1112_B
researcher	Text	ESLO2: demande d'itinéraire 1112_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1112_A
researcher	Text	ESLO2: demande d'itinéraire 1111_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1111_B
researcher	Text	ESLO2: demande d'itinéraire 1111_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1111_A
researcher	Text	ESLO2: demande d'itinéraire 1110_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1110_B
researcher	Text	ESLO2: demande d'itinéraire 1110_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1110_A
researcher	Text	ESLO2: demande d'itinéraire 1109_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1109_B
researcher	Text	ESLO2: demande d'itinéraire 1109_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1109_A
researcher	Text	ESLO2: demande d'itinéraire 1108_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1108_B
researcher	Text	ESLO2: demande d'itinéraire 1108_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1108_A
researcher	Text	ESLO2: demande d'itinéraire 1107_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1107_B

researcher	Text	ESLO2: demande d'itinéraire 1107_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1107_A
researcher	Text	ESLO2: demande d'itinéraire 1106_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1106_B
researcher	Text	ESLO2: demande d'itinéraire 1106_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1106_A
researcher	Text	ESLO2: demande d'itinéraire 1105_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1105_B
researcher	Text	ESLO2: demande d'itinéraire 1105_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1105_A
researcher	Text	ESLO2: demande d'itinéraire 1104_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1104_B
researcher	Text	ESLO2: demande d'itinéraire 1104_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1104_A
researcher	Text	ESLO2: demande d'itinéraire 1103_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1103_B
researcher	Text	ESLO2: demande d'itinéraire 1103_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1103_A
researcher	Text	ESLO2: demande d'itinéraire 1102_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1102_B
researcher	Text	ESLO2: demande d'itinéraire 1102_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1102_A
researcher	Text	ESLO2: demande d'itinéraire 1101_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1101_B
researcher	Text	ESLO2: demande d'itinéraire 1101_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1101_A
researcher	Text	ESLO2: demande d'itinéraire 1100_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1100_B
researcher	Text	ESLO2: demande d'itinéraire 1100_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1100_A
researcher	Text	ESLO2: demande d'itinéraire 1099_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1099_B
researcher	Text	ESLO2: demande d'itinéraire 1099_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1099_A
researcher	Text	ESLO2: demande d'itinéraire 1098_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1098_B
researcher	Text	ESLO2: demande d'itinéraire 1098_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1098_A
researcher	Text	ESLO2: demande d'itinéraire 1097_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1097_B
researcher	Text	ESLO2: demande d'itinéraire 1097_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1097_A
researcher	Text	ESLO2: demande d'itinéraire 1096_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1096_B
researcher	Text	ESLO2: demande d'itinéraire 1096_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1096_A
researcher	Text	ESLO2: demande d'itinéraire 1095_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1095_B
researcher	Text	ESLO2: demande d'itinéraire 1095_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1095_A

researcher	Text	ESLO2: demande d'itinéraire 1094_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1094_B
researcher	Text	ESLO2: demande d'itinéraire 1094_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1094_A
researcher	Text	ESLO2: demande d'itinéraire 1093_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1093_B
researcher	Text	ESLO2: demande d'itinéraire 1093_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1093_A
researcher	Text	ESLO2: demande d'itinéraire 1092_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1092_B
researcher	Text	ESLO2: demande d'itinéraire 1092_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1092_A
researcher	Text	ESLO2: demande d'itinéraire 1091_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1091_B
researcher	Text	ESLO2: demande d'itinéraire 1091_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1091_A
researcher	Text	ESLO2: demande d'itinéraire 1090_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1090_B
researcher	Text	ESLO2: demande d'itinéraire 1090_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1090_A
researcher	Text	ESLO2: demande d'itinéraire 1089_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1089_B
researcher	Text	ESLO2: demande d'itinéraire 1089_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1089_A
researcher	Text	ESLO2: demande d'itinéraire 1088_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1088_B
researcher	Text	ESLO2: demande d'itinéraire 1088_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1088_A
researcher	Text	ESLO2: demande d'itinéraire 1087_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1087_B
researcher	Text	ESLO2: demande d'itinéraire 1087_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1087_A
researcher	Text	ESLO2: demande d'itinéraire 1086_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1086_B
researcher	Text	ESLO2: demande d'itinéraire 1086_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1086_A
researcher	Text	ESLO2: demande d'itinéraire 1073_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1073_B
researcher	Text	ESLO2: demande d'itinéraire 1073_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1073_A
researcher	Text	ESLO2: demande d'itinéraire 1072_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1072_B
researcher	Text	ESLO2: demande d'itinéraire 1072_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1072_A
researcher	Text	ESLO2: entretien de 14-25 ans 1236_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1236_B
researcher	Text	ESLO2: entretien de 14-25 ans 1236_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1236_A
researcher	Text	ESLO2: entretien de 14-25 ans 1235_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1235_B

researcher	Text	ESLO2: entretien de 14-25 ans 1235_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1235_A
researcher	Text	ESLO2: entretien de 14-25 ans 1234_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1234_B
researcher	Text	ESLO2: entretien de 14-25 ans 1234_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1234_A
researcher	Text	ESLO2: entretien de 14-25 ans 1233_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1233_B
researcher	Text	ESLO2: entretien de 14-25 ans 1233_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1233_A
researcher	Text	ESLO2: entretien de 14-25 ans 1232_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1232_B
researcher	Text	ESLO2: entretien de 14-25 ans 1232_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1232_A
researcher	Text	ESLO2: entretien de 14-25 ans 1231_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1231_B
researcher	Text	ESLO2: entretien de 14-25 ans 1231_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1231_A
researcher	Text	ESLO2: entretien de 14-25 ans 1230_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1230_B
researcher	Text	ESLO2: entretien de 14-25 ans 1230_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1230_A
researcher	Text	ESLO2: entretien de 14-25 ans 1229_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1229_B
researcher	Text	ESLO2: entretien de 14-25 ans 1229_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1229_A
researcher	Text	ESLO2: entretien de 14-25 ans 1228_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1228_B
researcher	Text	ESLO2: entretien de 14-25 ans 1228_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENTJEUN_1228_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1220_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1220_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1220_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1220_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1219_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1219_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1219_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1219_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1218_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1218_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1218_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1218_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1217_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1217_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1217_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1217_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1216_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1216_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1216_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1216_A

researcher	Text	ESLO2: interaction à la sortie du cinéma 1215_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1215_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1215_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1215_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1214_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1214_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1214_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1214_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1213_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1213_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1213_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1213_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1212_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1212_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1212_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1212_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1211_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1211_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1211_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1211_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1210_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1210_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1210_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1210_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1209_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1209_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1209_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1209_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1208_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1208_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1208_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1208_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1207_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1207_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1207_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1207_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1206_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1206_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1206_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1206_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1205_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1205_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1205_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1205_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1204_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1204_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1204_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1204_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1203_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1203_B

researcher	Text	ESLO2: interaction à la sortie du cinéma 1203_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1203_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1202_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1202_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1202_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1202_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1201_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1201_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1201_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1201_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1200_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1200_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1200_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1200_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1199_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1199_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1199_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1199_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1198_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1198_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1198_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1198_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1197_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1197_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1197_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1197_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1196_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1196_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1196_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1196_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1195_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1195_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1195_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1195_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1194_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1194_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1194_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1194_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1193_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1193_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1193_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1193_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1192_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1192_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1192_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1192_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1191_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1191_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1191_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1191_A

researcher	Text	ESLO2: interaction à la sortie du cinéma 1190_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1190_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1190_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1190_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1189_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1189_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1189_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1189_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1188_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1188_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1188_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1188_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1187_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1187_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1187_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1187_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1186_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1186_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1186_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1186_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1185_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1185_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1185_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1185_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1184_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1184_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1184_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1184_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1183_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1183_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1183_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1183_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1182_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1182_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1182_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1182_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1181_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1181_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1181_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1181_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1180_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1180_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1180_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1180_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1179_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1179_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1179_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1179_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1178_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1178_B

researcher	Text	ESLO2: interaction à la sortie du cinéma 1178_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1178_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1177_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1177_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1177_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1177_A
researcher	Text	ESLO2: interaction à la sortie du cinéma 1176_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1176_B
researcher	Text	ESLO2: interaction à la sortie du cinéma 1176_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CINE_1176_A
researcher	Text	ESLO2: demande d'itinéraire 1129_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1129_B
researcher	Text	ESLO2: demande d'itinéraire 1129_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1129_A
researcher	Text	ESLO2: demande d'itinéraire 1128_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1128_B
researcher	Text	ESLO2: demande d'itinéraire 1128_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1128_A
researcher	Text	ESLO2: demande d'itinéraire 1127_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1127_B
researcher	Text	ESLO2: demande d'itinéraire 1127_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1127_A
researcher	Text	ESLO2: demande d'itinéraire 1126_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1126_B
researcher	Text	ESLO2: demande d'itinéraire 1126_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1126_A
researcher	Text	ESLO2: demande d'itinéraire 1125_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1125_B
researcher	Text	ESLO2: demande d'itinéraire 1125_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1125_A
researcher	Text	ESLO2: demande d'itinéraire 1124_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1124_B
researcher	Text	ESLO2: demande d'itinéraire 1124_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1124_A
researcher	Text	ESLO2: demande d'itinéraire 1123_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1123_B
researcher	Text	ESLO2: demande d'itinéraire 1123_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1123_A
researcher	Text	ESLO2: demande d'itinéraire 1122_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1122_B
researcher	Text	ESLO2: demande d'itinéraire 1122_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1122_A
researcher	Text	ESLO2: demande d'itinéraire 1121_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1121_B
researcher	Text	ESLO2: demande d'itinéraire 1175_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1175_B
researcher	Text	ESLO2: demande d'itinéraire 1175_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1175_A
researcher	Text	ESLO2: demande d'itinéraire 1174_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1174_B

researcher	Text	ESLO2: demande d'itinéraire 1174_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1174_A
researcher	Text	ESLO2: demande d'itinéraire 1173_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1173_B
researcher	Text	ESLO2: demande d'itinéraire 1173_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1173_A
researcher	Text	ESLO2: demande d'itinéraire 1172_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1172_B
researcher	Text	ESLO2: demande d'itinéraire 1172_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1172_A
researcher	Text	ESLO2: demande d'itinéraire 1171_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1171_B
researcher	Text	ESLO2: demande d'itinéraire 1171_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1171_A
researcher	Text	ESLO2: demande d'itinéraire 1170_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1170_B
researcher	Text	ESLO2: demande d'itinéraire 1170_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1170_A
researcher	Text	ESLO2: demande d'itinéraire 1169_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1169_B
researcher	Text	ESLO2: demande d'itinéraire 1169_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1169_A
researcher	Text	ESLO2: demande d'itinéraire 1168_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1168_B
researcher	Text	ESLO2: demande d'itinéraire 1168_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1168_A
researcher	Text	ESLO2: demande d'itinéraire 1167_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1167_B
researcher	Text	ESLO2: demande d'itinéraire 1167_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1167_A
researcher	Text	ESLO2: demande d'itinéraire 1166_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1166_B
researcher	Text	ESLO2: demande d'itinéraire 1166_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1166_A
researcher	Text	ESLO2: demande d'itinéraire 1165_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1165_B
researcher	Text	ESLO2: demande d'itinéraire 1165_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1165_A
researcher	Text	ESLO2: demande d'itinéraire 1164_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1164_B
researcher	Text	ESLO2: demande d'itinéraire 1164_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1164_A
researcher	Text	ESLO2: demande d'itinéraire 1163_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1163_B
researcher	Text	ESLO2: demande d'itinéraire 1163_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1163_A
researcher	Text	ESLO2: demande d'itinéraire 1162_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1162_B
researcher	Text	ESLO2: demande d'itinéraire 1162_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1162_A

researcher	Text	ESLO2: demande d'itinéraire 1161_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1161_B
researcher	Text	ESLO2: demande d'itinéraire 1161_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1161_A
researcher	Text	ESLO2: demande d'itinéraire 1160_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1160_B
researcher	Text	ESLO2: demande d'itinéraire 1160_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1160_A
researcher	Text	ESLO2: demande d'itinéraire 1159_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1159_B
researcher	Text	ESLO2: demande d'itinéraire 1159_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1159_A
researcher	Text	ESLO2: demande d'itinéraire 1158_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1158_B
researcher	Text	ESLO2: demande d'itinéraire 1158_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1158_A
researcher	Text	ESLO2: demande d'itinéraire 1157_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1157_B
researcher	Text	ESLO2: demande d'itinéraire 1157_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1157_A
researcher	Text	ESLO2: demande d'itinéraire 1156_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1156_B
researcher	Text	ESLO2: demande d'itinéraire 1156_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1156_A
researcher	Text	ESLO2: demande d'itinéraire 1155_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1155_B
researcher	Text	ESLO2: demande d'itinéraire 1155_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1155_A
researcher	Text	ESLO2: demande d'itinéraire 1154_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1154_B
researcher	Text	ESLO2: demande d'itinéraire 1154_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1154_A
researcher	Text	ESLO2: demande d'itinéraire 1153_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1153_B
researcher	Text	ESLO2: demande d'itinéraire 1153_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1153_A
researcher	Text	ESLO2: demande d'itinéraire 1152_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1152_B
researcher	Text	ESLO2: demande d'itinéraire 1152_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1152_A
researcher	Text	ESLO2: demande d'itinéraire 1151_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1151_B
researcher	Text	ESLO2: demande d'itinéraire 1151_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1151_A
researcher	Text	ESLO2: demande d'itinéraire 1150_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1150_B
researcher	Text	ESLO2: demande d'itinéraire 1150_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1150_A
researcher	Text	ESLO2: demande d'itinéraire 1149_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1149_B

researcher	Text	ESLO2: demande d'itinéraire 1149_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1149_A
researcher	Text	ESLO2: demande d'itinéraire 1148_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1148_B
researcher	Text	ESLO2: demande d'itinéraire 1148_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1148_A
researcher	Text	ESLO2: demande d'itinéraire 1147_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1147_B
researcher	Text	ESLO2: demande d'itinéraire 1147_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1147_A
researcher	Text	ESLO2: demande d'itinéraire 1146_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1146_B
researcher	Text	ESLO2: demande d'itinéraire 1146_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1146_A
researcher	Text	ESLO2: demande d'itinéraire 1145_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1145_B
researcher	Text	ESLO2: demande d'itinéraire 1145_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1145_A
researcher	Text	ESLO2: demande d'itinéraire 1144_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1144_B
researcher	Text	ESLO2: demande d'itinéraire 1144_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1144_A
researcher	Text	ESLO2: demande d'itinéraire 1143_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1143_B
researcher	Text	ESLO2: demande d'itinéraire 1143_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1143_A
researcher	Text	ESLO2: demande d'itinéraire 1142_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1142_B
researcher	Text	ESLO2: demande d'itinéraire 1142_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1142_A
researcher	Text	ESLO2: demande d'itinéraire 1141_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1141_B
researcher	Text	ESLO2: demande d'itinéraire 1141_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1141_A
researcher	Text	ESLO2: demande d'itinéraire 1140_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1140_B
researcher	Text	ESLO2: demande d'itinéraire 1140_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1140_A
researcher	Text	ESLO2: demande d'itinéraire 1139_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1139_B
researcher	Text	ESLO2: demande d'itinéraire 1139_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1139_A
researcher	Text	ESLO2: demande d'itinéraire 1138_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1138_B
researcher	Text	ESLO2: demande d'itinéraire 1138_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1138_A
researcher	Text	ESLO2: demande d'itinéraire 1137_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1137_B
researcher	Text	ESLO2: demande d'itinéraire 1137_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1137_A

researcher	Text	ESLO2: demande d'itinéraire 1136_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1136_B
researcher	Text	ESLO2: demande d'itinéraire 1136_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1136_A
researcher	Text	ESLO2: demande d'itinéraire 1135_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1135_B
researcher	Text	ESLO2: demande d'itinéraire 1135_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1135_A
researcher	Text	ESLO2: demande d'itinéraire 1134_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1134_B
researcher	Text	ESLO2: demande d'itinéraire 1134_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1134_A
researcher	Text	ESLO2: demande d'itinéraire 1133_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1133_B
researcher	Text	ESLO2: demande d'itinéraire 1133_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1133_A
researcher	Text	ESLO2: demande d'itinéraire 1132_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1132_B
researcher	Text	ESLO2: demande d'itinéraire 1132_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1132_A
researcher	Text	ESLO2: demande d'itinéraire 1131_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1131_B
researcher	Text	ESLO2: demande d'itinéraire 1131_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1131_A
researcher	Text	ESLO2: demande d'itinéraire 1130_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1130_B
researcher	Text	ESLO2: demande d'itinéraire 1130_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ITI_1130_A
researcher	Text	ESLO2: entretien 1025_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1025_B
researcher	Text	ESLO2: entretien 1025_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1025_A
researcher	Text	ESLO2: entretien 1024_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1024_B
researcher	Text	ESLO2: entretien 1024_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1024_A
researcher	Text	ESLO2: entretien 1016_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1016_B
researcher	Text	ESLO2: entretien 1016_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1016_A
researcher	Text	ESLO2: entretien 1014_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1014_B
researcher	Text	ESLO2: entretien 1014_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1014_A
researcher	Text	ESLO2: entretien 1009_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1009_B
researcher	Text	ESLO2: entretien 1009_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1009_A
researcher	Text	ESLO2: entretien 1005_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1005_B

researcher	Text	ESLO2: entretien 1005_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1005_A
researcher	Text	ESLO2: entretien 1039_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1039_B
researcher	Text	ESLO2: entretien 1039_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1039_A
researcher	Sound	ESLO2: repas 1253	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1253
researcher	Sound	ESLO2: repas 1254	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1254
researcher	Sound	ESLO2: repas 1255	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1255
researcher	Sound	ESLO2: repas 1256	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1256
researcher	Sound	ESLO2: repas 1257	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1257
researcher	Sound	ESLO2: repas 1258	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1258
researcher	Sound	ESLO2: repas 1259	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1259
researcher	Sound	ESLO2: repas 1260	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1260
researcher	Sound	ESLO2: repas 1261	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1261
researcher	Sound	ESLO2: repas 1262	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1262
researcher	Sound	ESLO2: repas 1263	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1263
researcher	Sound	ESLO2: repas 1265	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1265
researcher	Sound	ESLO2: repas 1266	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1266
researcher	Sound	ESLO2: repas 1267	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1267
researcher	Sound	ESLO2: repas 1269	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1269
researcher	Sound	ESLO2: repas 1270	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1270
researcher	Sound	ESLO2: repas 1271	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1271
researcher	Text	ESLO2: entretien 1052_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1052_B
researcher	Text	ESLO2: entretien 1052_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1052_A
researcher	Text	ESLO2: entretien 1083_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1083_B
researcher	Text	ESLO2: entretien 1083_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1083_A
researcher	Text	ESLO2: entretien 1056_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1056_B

researcher	Text	ESLO2: entretien 1056_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1056_A
researcher	Text	ESLO2: entretien 1055_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1055_B
researcher	Text	ESLO2: entretien 1055_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1055_A
researcher	Text	ESLO2: entretien 1272_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1272_B
researcher	Text	ESLO2: entretien 1272_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1272_A
researcher	Sound	ESLO2: repas 1247	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1247
researcher	Sound	ESLO2: repas 1264	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1264
researcher	Sound	ESLO2: repas 1268	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1268
researcher	Text	ESLO2: repas 1271_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1271_C
researcher	Text	ESLO2: repas 1271_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1271_B
researcher	Text	ESLO2: repas 1271_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1271_A
researcher	Text	ESLO2: repas 1270_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1270_C
researcher	Text	ESLO2: repas 1270_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1270_B
researcher	Text	ESLO2: repas 1270_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1270_A
researcher	Text	ESLO2: repas 1269_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1269_C
researcher	Text	ESLO2: repas 1269_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1269_B
researcher	Text	ESLO2: repas 1269_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1269_A
researcher	Text	ESLO2: repas 1268_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1268_C
researcher	Text	ESLO2: repas 1268_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1268_B
researcher	Text	ESLO2: repas 1268_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1268_A
researcher	Text	ESLO2: repas 1267_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1267_C
researcher	Text	ESLO2: repas 1267_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1267_B
researcher	Text	ESLO2: repas 1267_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1267_A
researcher	Text	ESLO2: repas 1266_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1266_C
researcher	Text	ESLO2: repas 1266_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1266_B

researcher	Text	ESLO2: repas 1266_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1266_A
researcher	Text	ESLO2: repas 1265_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1265_C
researcher	Text	ESLO2: repas 1265_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1265_B
researcher	Text	ESLO2: repas 1265_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1265_A
researcher	Text	ESLO2: repas 1264_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1264_C
researcher	Text	ESLO2: repas 1264_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1264_B
researcher	Text	ESLO2: repas 1264_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1264_A
researcher	Text	ESLO2: repas 1263_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1263_C
researcher	Text	ESLO2: repas 1263_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1263_B
researcher	Text	ESLO2: repas 1263_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1263_A
researcher	Text	ESLO2: repas 1262_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1262_C
researcher	Text	ESLO2: repas 1262_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1262_B
researcher	Text	ESLO2: repas 1262_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1262_A
researcher	Text	ESLO2: repas 1261_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1261_C
researcher	Text	ESLO2: repas 1261_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1261_B
researcher	Text	ESLO2: repas 1261_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1261_A
researcher	Text	ESLO2: repas 1260_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1260_C
researcher	Text	ESLO2: repas 1260_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1260_B
researcher	Text	ESLO2: repas 1260_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1260_A
researcher	Text	ESLO2: repas 1259_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1259_C
researcher	Text	ESLO2: repas 1259_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1259_B
researcher	Text	ESLO2: repas 1259_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1259_A
researcher	Text	ESLO2: repas 1258_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1258_C
researcher	Text	ESLO2: repas 1258_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1258_B
researcher	Text	ESLO2: repas 1258_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1258_A

researcher	Text	ESLO2: repas 1257_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1257_C
researcher	Text	ESLO2: repas 1257_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1257_B
researcher	Text	ESLO2: repas 1257_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1257_A
researcher	Text	ESLO2: repas 1256_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1256_C
researcher	Text	ESLO2: repas 1256_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1256_B
researcher	Text	ESLO2: repas 1256_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1256_A
researcher	Text	ESLO2: repas 1255_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1255_C
researcher	Text	ESLO2: repas 1255_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1255_B
researcher	Text	ESLO2: repas 1255_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1255_A
researcher	Text	ESLO2: repas 1254_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1254_C
researcher	Text	ESLO2: repas 1254_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1254_B
researcher	Text	ESLO2: repas 1254_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1254_A
researcher	Text	ESLO2: repas 1253_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1253_C
researcher	Text	ESLO2: repas 1253_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1253_B
researcher	Text	ESLO2: repas 1253_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1253_A
researcher	Text	ESLO2: repas 1247_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1247_C
researcher	Text	ESLO2: repas 1247_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1247_B
researcher	Text	ESLO2: repas 1247_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_REPAS_1247_A
researcher	Text	ESLO2: conférence 1074_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1074_C
researcher	Text	ESLO2: conférence 1074_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1074_B
researcher	Text	ESLO2: conférence 1074_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1074_A
researcher	Text	ESLO2: conférence 1244_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1244_C
researcher	Text	ESLO2: conférence 1244_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1244_B
researcher	Text	ESLO2: conférence 1244_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1244_A
researcher	Text	ESLO2: conférence 1243_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1243_C

researcher	Text	ESLO2: conférence 1243_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1243_B
researcher	Text	ESLO2: conférence 1243_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1243_A
researcher	Text	ESLO2: conférence 1242_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1242_C
researcher	Text	ESLO2: conférence 1242_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1242_B
researcher	Text	ESLO2: conférence 1242_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1242_A
researcher	Text	ESLO2: conférence 1241_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1241_C
researcher	Text	ESLO2: conférence 1241_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1241_B
researcher	Text	ESLO2: conférence 1241_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1241_A
researcher	Text	ESLO2: conférence 1240_C	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1240_C
researcher	Text	ESLO2: conférence 1240_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1240_B
researcher	Text	ESLO2: conférence 1240_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_CONF_1240_A
researcher	Text	ESLO2: entretien 1001_B	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1001_B
researcher	Text	ESLO2: entretien 1001_A	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-ESLO2_ENT_1001_A

Travaux :

***Corpus de la parole* dans Cocoon**

[fr] Corpus de la parole

Baude, Olivier (compiler)

(archivage: 2010-06-26T11:05:33+02:00; mise à disposition: 2010-06-26; dernière modification de la notice: 2014-07-15)

Editeur(s):	Délégation générale à la langue française et aux langues de France Typologie et universaux linguistiques Institut de linguistique Française
Description(s):	[fr] Le programme Corpus de la parole du ministère de la culture et de la communication a pour but de valoriser le patrimoine linguistique de la France. Il donne accès en ligne à des fonds sonores transcrits et numérisés, en français et dans différentes langues parlées sur le territoire national, en métropole et outremer. Ces langues sont considérées comme "Langues de France"..
Type(s):	Collection
Sujet(s):	Le français et les langues de France
Droits:	Freely available for non-commercial use
Pour citer la ressource:	http://purl.org/poi/crdo.vjf.cnrs.fr/crdo-COLLECTION_LANGUESDEFRANCE ou ark:/87895/1.5-124203