

PRÉAMBULES À L'ENSEIGNEMENT DE LA TRANSCRIPTION DE CORPUS

Martine Faraco

A corpus is any collection of speech recordings which is accessible in computer readable form which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by people in other organisations¹.

Gibbon D., Moore R. & Winski R.
(1998 : 79)

Ce travail aborde une réflexion sur l'aspect pédagogique de la transcription de corpus oraux fondée sur une observation contrastive de deux types d'enseignement en sciences du langage : l'un qui s'adresse aux étudiants de licence 1, ouvert aux spécialistes et aux non-spécialistes de cette discipline et l'autre, plus ciblé, s'adressant aux étudiants inscrits en master prenant place dans le parcours pathologie.

Nous montrons comment une expérience d'enseignement de la transcription de corpus pour la parole pathologique a révélé les limites – ou plus exactement les zones perfectibles – de cette transcription, enseignée habituellement. Ces préambules à l'enseignement de la transcription de corpus s'articuleront tout d'abord autour d'une description succincte des publics et objectifs des cours auxquels il est fait référence, ensuite d'une caractérisation des deux formes d'enseignement de transcription de corpus et enfin de quelques propositions de conventions, suivies d'exemples tirés de travaux d'étudiants.

FARACO, Martine (2006), Préambules à l'enseignement de la transcription de corpus, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, vol. 25, p. 97-108.

1. « Un corpus est toute collection d'enregistrements de parole accessible sous une forme lisible sur ordinateur et qui s'accompagne d'une annotation et documentation suffisantes pour permettre la réutilisation des données pour soi ou pour des personnes d'autres organismes. » (t.d.a.).

1. Publics et objectifs du cours

Les publics concernés par cette réflexion préparent l'un ou l'autre des deux diplômes suivants² :

1.1. Licence

La licence – mention sciences du langage – traite des langues et de leurs usages, du langage et de la parole. Cette formation, ouverte et interdisciplinaire pendant les trois premières années, est accessible aux étudiants ayant un baccalauréat ou un DAEU³ Lettres. La plus grande partie des enseignements se fait par groupes de 50 étudiants maximum et des groupes spécifiques accueillent les étudiants d'autres filières, souhaitant découvrir les sciences du langage.

En ce qui concerne le contenu, le semestre 2⁴ se décompose ainsi :

- langage, culture et société
- langage et cerveau
- données et théories
- langue vivante
- discipline d'ouverture

Nous ferons référence ici au cours SCLB03 : Données et Théories, ouvert aux spécialistes et non-spécialistes.

1.2. Master recherche, parcours pathologie

Le parcours pathologie fait partie de la spécialité « recherche » du master en sciences du langage à l'université de Provence. Il s'appuie sur le Laboratoire Parole et Langage (LPL) qui constitue l'équipe d'accueil, en relation avec le laboratoire d'audiophonologie expérimentale et clinique (UPRES-EA 2668, faculté de Médecine de Marseille), et avec le service de neurologie de l'hôpital d'Aix. Ce parcours vise à offrir une formation avancée et une initiation à la recherche dans le domaine des troubles de la voix, de la parole et du langage chez l'adulte et chez l'enfant. Ces questions sont traitées dans une perspective interdisciplinaire par des linguistes et des cliniciens. Le public concerné est constitué d'étudiants en sciences du langage, d'orthophonistes, d'étudiants possédant une qualification de médecin phoniatre, de médecin ORL, ou de médecin neurologue. Le groupe d'étudiants en master sciences du langage, parcours pathologie, a une particularité : il n'est pas homogène. En effet, sont réunis dans ce cours des étudiants issus de la formation

2. Pour des informations plus détaillées, se connecter au site de l'université de Provence <http://www.univ-mrs.fr/> et lire la rubrique « formations », pour l'année 2006-2007.

3. Diplôme d'accès aux études universitaires.

4. Les trois années de licence se divisent en six semestres.

sciences du langage, donc d'une formation assez générale et des orthophonistes déjà en exercice ou sortant de formation. Les uns et les autres ont beaucoup à s'apporter.

Pour les modalités pédagogiques, les étudiants de cette filière suivent une série de cours et de séminaires et effectuent, par ailleurs, un stage en milieu hospitalier ; ils sont également intégrés à l'une des équipes de recherche au sein du laboratoire d'accueil. Pour valider cette formation, ils doivent mener à terme un projet de recherche, aboutissant à la rédaction d'un article scientifique. Au semestre 1, ces étudiants ont, d'ores et déjà, été initiés à la transcription de corpus oraux par les formateurs de l'équipe DELIC⁵ et du LPL (*cf.* cours SCLQ02), mais nous intervenons de nouveau au semestre 2, de manière très ponctuelle, dans le cours SCLR11, pathologie 1, s'adressant à ces étudiants spécialistes.

Les publics auxquels nous ferons référence et les objectifs qui sont poursuivis, dans ces deux cours de transcription de corpus, diffèrent donc à deux points de vue. Tout d'abord, le degré de spécialisation varie grandement parmi les publics concernés par la présente réflexion : non-spécialistes et spécialistes dans la discipline générale des sciences du langage d'une part et, d'autre part, étudiants se spécialisant dans l'étude particulière de la parole pathologique. De plus, le but poursuivi par cet enseignement de la transcription de corpus est lui aussi également différent : pour les étudiants de licence, il s'agit d'initiation ou de perfectionnement en transcription de corpus, et, pour ceux de master, de sensibilisation aux symptômes de patients souffrant de troubles de la parole.

2. Transcriptions et conventions

2.1. Données et théories : méthodologie du recueil de données

Cet enseignement comporte deux parties : l'une théorique, présentant les textes fondamentaux en linguistique et phonétique, l'autre pratique, expliquant la méthodologie du recueil de données en sciences du langage. C'est dans ce deuxième volet que prend place notre intervention pédagogique. Après une approche définitionnelle de la dialectologie et de la sociolinguistique, un tour d'horizon, non exhaustif, de travaux effectués dans ces deux perspectives et une présentation des différents types de corpus, il est demandé aux étudiants de rendre un dossier ayant trait au recueil de données. L'informateur doit être un locuteur français d'origine immigrée récente ou étranger de langue maternelle étrangère, et parlant le français. Il s'agit, pour l'étudiant, d'établir (1) une transcription orthographique d'un extrait de trois minutes de conversation entre le locuteur non natif et un locuteur natif (l'enquêteur, par exemple) ; (2) une transcription phonétique de mots

5. Équipe DELIC : Description Linguistique Informatisée sur Corpus (<http://www.up.univ-mrs.fr/delic/>).

étrangers (tirés d'une liste de 100 mots ; (3) une information du corpus : locuteur, lieu... ; (4) une brève description de la langue maternelle, (5) une rapide analyse de la parole du locuteur étranger en français⁶.

Ce enseignement, visant la pratique de la transcription, s'inspire largement de la présentation du Corpus de référence du français parlé de l'équipe DELIC (Description Linguistique Informatisée sur Corpus) de l'université de Provence, s'appuyant lui-même sur les conventions de transcription mises au point par le GARS⁷ (cf. Blanche-Benveniste *et al.*, 1991 : 228-229).

Pour comprendre la philosophie générale de la transcription de corpus oral, il convient de se reporter à la page 11 de la présentation du corpus de référence en français parlé⁸ :

« Ces conventions préconisent une représentation en orthographe standard, sans ponctuation, ni indications prosodiques, signes démarcatifs ou expressifs. Les pauses sont notées, ainsi que les amorces (mots inachevés). On tient à conserver les traces de certaines difficultés de transcription liées à des problèmes d'interprétation. »

Relevons « ni indication prosodique » et ajoutons qu'ont été insérés aux informations données dans le cours de SCLB03, plusieurs marqueurs prosodiques relativement simples :

↑	Intonation montante de question
↓	Intonation descendante d'assertion
↗	Continuité de type énumératif
°	Accent attendu actualisé, placé devant la syllabe accentuée
AAAAaa	Accent emphatique, ex. accent d'insistance à l'initiale
aa _{aaaaaa} aa	Parenthèse basse
aa ^{aaaaaa} aa	Parole qui ne s'adresse pas qu'à l'interlocuteur mais également à des témoins non impliqués directement dans l'interaction. Le locuteur donne à entendre son discours (ou du moins une partie de celui-ci) aux témoins de la conversation.

En ce qui concerne le résumé des symboles de transcription, correspondant à une révision des conventions du GARS par l'équipe DELIC⁹, on s'arrêtera, dans le paragraphe intitulé « notations phonétiques », sur le passage suivant :

6. Ce cours a été conçu en collaboration avec Yohann Meynadier, enseignant-chercheur à l'université de Provence et membre du LPL, la partie expérimentale a été révisée à la lumière de la linguistique de terrain par Alice Vittrant, également enseignant-chercheur à l'université de Provence.

7. Groupe Aixois de Recherche en Syntaxe.

8. <http://www.up.univ-mrs.fr/veronis/pdf/2004-presentation-crfp.pdf>.

9. Voir site <http://www.up.univ-mrs.fr/delic/>, rubrique « corpus français parlé » <http://www.up.univ-mrs.fr/delic/corpus/index.html>, puis onglet « transcription ».

« Pour les textes “ordinaires”, la transcription phonétique est indiquée entre crochets au sein d’une balise “pron” placée entre accolades :

Dans le corps du texte : je me suis cherché une {pron = [yna]} place d’apprenti

Pour les textes “particuliers” (jeunes enfants, pathologies, étrangers), des conventions spécifiques sont établies, selon les besoins, pour placer des notations phonétiques dans le texte. »

Or, en ce qui concerne la partie « recueil » du cours SCLB03, les étudiants doivent retranscrire en phonétique les mots (ou groupes de mots) prononcé(s) par un locuteur non natif s’exprimant en français, s’écartant probablement de la prononciation attendue. Bien entendu, comme pour une transcription phonétique fine, on peut faire appel aux traditionnels diacritiques. On en trouve la définition suivante sur l’un des sites consacrés à l’API¹⁰ :

« Un diacritique est un élément ajouté à une lettre d’un alphabet pour en modifier la valeur. Cet élément peut être souscrit (en indice), surscrit (en chef, en exposant) à cette lettre, à sa droite ou encore à sa gauche. [...]

˜ nasalisation — ã = a nasalisé dans fanfan [...]

◦ surdité — ʄ = l assourdi

· sonorité — ʒ = z [...]

De temps à autre, il faut avouer que les diacritiques semblent relativement insuffisants et que les étudiants se trouvent confrontés à de telles difficultés de transcription qu’il leur a été suggéré d’utiliser une transcription à la fois plus harmonisée et plus simple. Cette transcription est d’autant plus appréciée que certains d’entre eux ne maîtrisent pas bien la transcription phonétique et ont beaucoup de problèmes pour s’appuyer sur des critères de classement phonétique des différents sons sur la base d’indices articulatoires (tels que palatalisation, antériorisation, postériorisation, labialisation, pour n’en citer que quelques-uns). Une des méthodes adoptées par ces étudiants de licence 1 est la mise en exposant du phonème vers lequel tend la prononciation de locuteurs non natifs : ex. [rⁱ] pour « riz » prononcé par un japonophone, pour qui la différence [r] / [ʀ] n’existe pas dans la langue maternelle ou encore [ʃ³ak³ʒ³ur] pour “chaque jour” prononcé par un germanophone qui n’identifie pas bien ces deux fricatives post-alvéolaires, puisque le phonème voisé /ʒ/ n’existe pas en allemand. Notons encore que lorsque ce phonème initial a vraiment été remplacé par un autre, est adoptée la notation pleine du phonème prononcé (ex. [pɔnœr] pour le mot « bonheur »).

10. <http://alis.isoc.org/langues/api.fr.htm#diacritiques>.

2.2. Pathologie 1 : transcription de parole pathologique

Il serait bon de se reporter, pour prendre la mesure de ce qu'est la parole pathologique, aux travaux de Teston (2000, 2001) qui portent sur deux types de troubles : la dysarthrie et la dysphonie. Pour la dysarthrie, trouble d'origine centrale, ce sont les activités motrices qui sont généralement atteintes. Quelques-uns des symptômes sont les suivants : difficultés d'articulation ; débit altéré, rythme saccadique ; altération de la voix : grinçante, assourdie. La dysphonie, quant à elle, correspond à une atteinte organique ou fonctionnelle du larynx. Au niveau des symptômes, ce trouble réalise un enrrouement durable, variant dans son intensité et sans tendance à la guérison spontanée. La dysphonie correspond à une modification de la voix portant sur son timbre et son intensité¹¹ (pour le détail des différents types de dysphonies et de leurs symptômes, cf. Teston, 2001).

Cet aperçu, même très rapide, de ces deux troubles de la parole, appelle une remarque : quand on compare les phénomènes à noter en parole pathologique à ceux que l'on note habituellement, on se rend aisément compte que, dans le premier cas, il s'agit d'annoter des particularités renvoyant davantage à la parole : intensité, débit, F0, rythme... qu'au langage lui-même.

2.2.1. Transcription de lecture pathologique et enseignement

L'expérience acquise dans l'enseignement de la transcription de corpus de natifs et non-natifs s'est avérée très insuffisante pour affronter les différentes transcriptions de parole pathologique.

Le travail effectué lors des quelques séances proposées aux étudiants de master a consisté à établir, à partir d'enregistrements anonymes de patients effectués à l'hôpital de la Timone de Marseille, des grilles d'analyse de parole lue. Le corpus n'était pas du tout documenté, nous n'avions surtout aucune indication sur le type de diagnostic fait à l'hôpital. Plusieurs exercices orthophoniques étaient effectués lors de l'examen, mais c'est le test de lecture qui a été choisi car il était possible de partir ainsi d'un texte-source identique pour tous les sujets. Voici le passage de *la Chèvre de monsieur Seguin*, tiré des *Lettres de mon moulin* d'Alphonse Daudet, que le patient lit habituellement :

Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon : un beau matin, elles cassaient leur corde, s'en allaient dans la montagne, et là-haut le loup les mangeait. Ni les caresses de leur maître, ni la peur du loup, rien ne les retenait. C'était, paraît-il, des chèvres indépendantes, voulant à tout prix le grand air et la liberté.

Le travail en classe a été le suivant : chacun des intervenants a lu à haute voix le texte et un texte commun de référence a été établi, après discussion.

11. <http://www.med.univ-rennes1.fr/cerf/edicerf/strat/SR049.html>, consulté le 27 septembre 2006.

Le texte commun de référence est le suivant :

monsieur seguin n'avait jamais eu de bonheur avec ses °chèvres↓ +++ il les perdait
TOUTES de la même fa°çon↓ +++ un beau ma°tin↗ ++ elles cassaient leur
°corde↗ ++ s'en allaient dans la mon°tagne↗ ++ et là-haut le loup les man°geait↓
+++ **NI** les caresses de leur °maî::tre↗ +++ **NI** la peur du °loup↗ +++ **RIEN** ne
les rete°nait↓ ++ c'é°tait↗ ++ paraî°t-il ++ des chèvres indépen°dantes↗ +++ voulant
A TOUT PRIX↗ ++ le grand air et la liber°té↓ ////

Pour le travail personnel demandé aux étudiants, il s'agissait, à partir de la transcription commune, de choisir deux enregistrements de lecture de ce texte par deux patients différents. Il leur fallait noter tout ce qui semblait pertinent pour établir le profil pathologique de chaque patient : phonologie, prosodie (intonation, accents, débit, allongements,...), problèmes d'intensité et/ou hauteur de la voix. Enfin, pour chacun des patients, devait être rédigée une brève conclusion dans laquelle les problèmes qui leur étaient propres devaient être recensés.

2.2.2. Quelques propositions de conventions

Ces propositions vont être de deux ordres : annotations prosodiques et annotations renvoyant à des caractères physiques de la voix.

Pour l'annotation prosodique de la lecture à haute voix du patient, il suffirait de prendre un texte vierge de toute annotation et de noter de la même façon les pauses, les continuatifs et terminatifs ainsi que les accents attendus ou non. Il ne resterait alors qu'à superposer la lecture annotée du patient et la lecture-type pour en déceler les différences.

Toujours est-il qu'il faudrait penser à certains phénomènes qui pourraient se produire en lecture pathologique :

- | | |
|--|---|
| • pauses intra-mots | la mon # tagne |
| • silences ou arrêts | monsieur seguin ////////////// |
| • accents inattendus | c' était para ît-il |
| • syllabation | mon-sieur-se-guin-n'a-vait-ja-mais-
eu-de-bon-heur |
| • rythme accéléré ou tachyphémie ¹² | <aaaaaaaaa> |
| • rythme ralenti | >aaaaaaaaa< |

12. Trouble de la parole qui consiste à accélérer de façon importante le débit de celle-ci, en prononçant les phrases trop rapidement, de façon saccadée et d'une voix de plus en plus faible.

La tachyphémie est souvent associée à la palilalie (désordre de la parole en rapport avec un trouble de l'intelligence et consistant à répéter involontairement et spontanément, deux ou plusieurs fois de suite, la même phrase ou le même mot).

Pour l'annotation de la parole, il convient ici de citer, à titre indicatif, les diacritiques proposés pour transcrire la parole perturbée ou pathologique qui donnent la possibilité de visualiser dans la transcription quelques particularités.

034B	◌̥	DIACRITIQUE HOMOTHÉTIQUE EN CHEF • API : nasillement
034C	◌̥̥	DIACRITIQUE PRESQUE ÉGAL À EN CHEF • API : friction vélopharyngale
034D	◌̥̥̥	DIACRITIQUE FLÈCHE BILATÉRALE SOUSCRITE • API : écartement labial
034E	◌̥̥̥̥	DIACRITIQUE FLÈCHE VERS LE HAUT SOUSCRITE • API : articulation sifflée

Figure 1
*Diacritiques de l'API pour la parole perturbée ou pathologique*¹³

Pour trouver des propositions de notation plus complètes, il convient de se reporter à l'extension de l'Alphabet Phonétique International (*the Extended IPA*) qui a été conçu pour transcrire le discours pathologique (voir Ball *et al.* (1995) et Duckworth *et al.* (1990)). Ces lettres et diacritiques de l'API étendu ont l'avantage de prendre en compte, entre autres, des phénomènes liés au courant d'air phonatoire (par exemple ingressif ou égressif), à la qualité de la voix ou encore à la force articulatoire. Quelques annotations prosodiques sont également prises en compte.

Pourtant ces conventions semblent quelque peu complexes ou encore insuffisamment harmonisées, et, par suite, au fil des travaux des étudiants¹⁴, nous ferons quelques propositions simplifiées de notation. Mis à part les problèmes d'ordre phonologique, pourraient être notés les phénomènes suivants :

- variations anormales d'intensité <⚡ aaaa ⚡>
- hausse d'intensité <aaaa>
- baisse d'intensité <aaa>
- baisse continue d'intensité aaaaaaa
- raucité <R aaaa R>

13. Pour des précisions ou compléments se reporter à :
<<http://www.unicode.org/fr/charts/symbols.html#CombiningDiacriticalMarks>> et ouvrir le fichier PDF « diacritiques ».

14. Nous tenons à remercier les étudiantes de master SCL, parcours pathologie de l'année 2005-2006, pour leur travaux de transcription et suggestions.

- voix grinçante¹⁵ <G aaaa G>
- nasillement <N aaaa N>
- tremblements aaaaaaaaaaaa
- approximations d'articulation aaaaaaaaaa
- parasitages se produisant seuls (respiration) (déglutition)
- parasitages pendant la phonation aaa (respiration) aaaaaaaa

Il va sans dire que les conventions de transcription ne sont absolument pas exhaustives, qu'elles sont perfectibles et qu'elles doivent être mises à l'épreuve pour tester leur lisibilité et leur efficacité. Ce travail pourrait, à long terme, permettre, non pas de diagnostiquer tel ou tel trouble, puisque le bilan orthophonique¹⁶ joue déjà ce rôle, mais de sensibiliser les étudiants à la transcription de la parole pathologique à laquelle les orthophonistes s'affrontent quotidiennement.

2.2.3. Exemples de transcriptions de lecture

Voilà ce que note une étudiante à l'écoute de la lecture d'un des patients :

monsieur se°guin↗ + (inspiration) n'avait jamais eu de bon°heur↗ + avec ses
 chè°vres↘ +++ (inspiration) il les perdait °TOUTES↗ + (inspiration) de la même
 fa°çon↘ + un beau ma°tin↗ + (inspiration et déglutition) elles cassaient leur c- leur
 °corde↗ (inspiration et déglutition) s'en allaient dans la mon°tagne↗ et là-°haut↗
 (inspiration) le loup les man°geait↘ +++ (inspiration et déglutition) **NI** les caresses
 de leur °maître↗ +(inspiration) **NI** la peur du loup↗ + (inspiration) **RIEN** ne les
 rete°nait↘ ////

Son commentaire est le suivant :

*« Il paraît évident, à l'écoute de ce corpus, que monsieur A. ne présente aucune dysphonie (trouble de la voix). En effet, la voix de celui-ci ne montre aucun caractère de raucité, de souffle, d'altération globale du timbre, de caractère hypo- ou hypertonique. Ce patient ne souffre pas non plus de troubles arthritiques. La totalité de ses phonèmes sont en effet bien formés.
 Par contre, on note dans cette production, de nombreuses pauses qui ne correspondent ni à la ponctuation, ni aux pauses du texte de référence. Il ne semble pas que ces pauses soient liées à une quelconque recherche lexicale ou programmation de la parole, mais plutôt à des phases atypiques d'inspiration et/ou de déglutition. Notons que ces troubles peuvent se produire soit pendant un silence soit pendant la phonation : (inspiration) RIEN. Comme la prise de souffle est insuffisante, ceci semble obliger monsieur A. à découper son énoncé en groupes de souffle plus courts et plus*

15. Pour la raucité et la voix grinçante, on ajoutera une note en bas de page si phénomène constant.

16. <http://www.orthophonistes.fr/Theme.php?NumTheme=20&Article=5>

fréquents que ceux qui sont attendus. Cette structuration amène le patient, semble-t-il, à multiplier les intonèmes continuatifs pour lier les différents éléments de son discours. »

Voilà la transcription qui est faite pour un autre patient :

monsieur [segɛ], n'avait <jamais eu de> <N [bʰɔ̃nœr] N>↓ + avec ses chèvres↓
+++ (respiration) il les perdait +++ [tʷus] **TOUTES**↗ de la même fa^oçon↓ +++
(respiration) [a] beau ma^otin + elles cassaient leur [°ko^ard]↓ +++ >s'en allaient
dans la mon^otagne< et là-haut + (respiration) le loup les [mã^aʒe]↓ +++
(respiration) **NI** les [sarɛs] de leur [mɛt]↗ +++ **NI** la peur du loup, {là}17 **RIEN**
ne les rete^onait↓ +++ >c'était [paetɪl] des chèvres indépen^odantes <voulant> à tout
prix↗ +> le [gã] air et la <N [lib^oɛrte] N>↓ +++ (souffle) ////

Cette lecture est commentée ainsi :

*« Le sujet présente des écarts par rapport à la prononciation attendue. Il faut noter, par exemple le dévoisement de la bilabiale occlusive [b] --> [bʰ] (ex. [bʰɔ̃nœr] ou encore [lib^oɛrte]). Par ailleurs, la difficulté que le sujet éprouve dans la prononciation du “r” post-consonantique ou post-vocalique, le conduit à une stratégie d'évitement (ex. [ã] pour “grand”. Un nasillement se produit également sur certains mots (ex. <N [bʰɔ̃nœr] N>). Sans doute, le patient connaît-il une baisse d'amplitude des mouvements articulatoires conduisant à une dégradation de l'articulation ? Au plan prosodique, il faut noter des accélérations et des ralentissements assez fréquents. Les pauses sont généralement longues et, parfois non attendues, comme par exemple dans « il les perdait +++ [tʷus] **TOUTES**↗ » et peut-être à cause d'une anticipation de la difficulté à prononcer cette occlusive en début de mot. »*

Enfin, un dernier exemple :

<monsieur seguin n'avait jamais eu de bonheur avec ses °chèvres↗ + il les perdait
toutes de la même fa^oçon↓ + un beau ma^otin↗ + elles cassaient leur °corde↗ +
s'en allaient dans la mon^otagne↗ + et là-haut le loup les man^ogeait↓ ++ ni les
caresses de leur °maître↗ + ni la peur du °loup↗ + rien ne les rete^onait↓ + c'était

17. Nous notons, entre accolades, les mots rajoutés.

paraît-il des chèvres indépen°dantes↗ +voulant à tout prix le grand air et la
liber°té↓ ////>

Voici ce qu'en dit l'étudiante :

« Dans ce texte lu, on ne relève aucune déformation phonologique mais on ressent une certaine monotonie à l'écoute de ce texte. Ce dernier est lu très vite, comme en témoignent les pauses relativement courtes. Aucune variation de type parenthèse basse, par exemple sur "paraît-il", aucun accent emphatique. Dans sa tonalité d'ensemble, la voix de madame S. est éraillée. »

A ce stade du travail, il est demandé aux étudiantes de rester très prudentes, puisque les enregistrements sont dépourvus de toute information sur les patients enregistrés et sur le diagnostic porté sur leur maladie. Il s'agit, pour le moment, essentiellement d'une découverte pratique de ce que peut être la transcription de parole pathologique. Cette méthode pourrait permettre de suivre, de façon longitudinale, l'évolution de la lecture pour un patient en rééducation orthophonique. Il faudrait alors garder un même texte pour comparaison, mais aussi penser à contrôler les phénomènes d'habituatation à la tâche en lui proposant, à chaque séance où serait pratiqué un recueil, un texte nouveau. Si la recherche se poursuit, dans ce domaine, il serait bon de recueillir un certain nombre d'enregistrements correspondant à un type de trouble, par exemple la maladie de Parkinson ou la sclérose latérale amyotrophique (SLA), et de trouver les grandes caractéristiques symptomatiques et leur notation¹⁸. Ainsi, seraient mises au point des conventions de transcription non seulement communes mais aussi spécifiques à ces deux types d'affection.

3. Conclusion

Lorsqu'on parle de transcription de corpus oraux, on s'inquiète très légitimement du fait que certaines spécificités de l'oral ne soient pas retransmises. Or, cette contribution a montré que la transcription manuelle de corpus peut se trouver prise en défaut, même pour des types particuliers d'informateurs. Afin d'illustrer ce constat, les exemples de corpus de non-natifs et de corpus de patients atteints de troubles de la parole ont été pris.

Nous souhaitons que les prochaines recherches s'aventurent plus loin sur ce terrain pour que soient mises au point des conventions permettant de rendre compte certes de la phonétique, phonologie, mais aussi de la prosodie et des caractéristiques de la voix¹⁹ au sens physique du terme.

18. Actuellement des étudiantes, Lila Chaouch et Christine Bellot, viennent d'entreprendre ce travail dans le cadre d'un master 2.

19. Bien entendu, nous rejoignons ici Bernard Teston (2000, 2001) pour souligner que cette notation contient une part de subjectivité et que des analyses objectives s'imposent.

4. Bibliographie

- BALL, M.J. ; ESLING, J.H. ; DICKSON, C. (1995). The VoQS system for the transcription of voice quality, *Journal of the International Phonetic Alphabet*, 25 (2), p. 71-80.
- BLANCHE-BENVENISTE, C. ; BILGER, M. ; ROUGET, M ; VAN DEN EYNDE, K. (1991). *Le français parlé*, Paris : CNRS éditions.
- DUCKWORTH, M. ; ALLEN, G. ; HARDCASTLE, W. ; BALL, M.J. (1990). Extensions to the International Phonetic Alphabet for the transcription of atypical speech, *Clinical Linguistics and Phonetics*, 4, p. 273-280.
- GIBBON, D. ; MOORE, R. ; WINSKI, R. (1998). *Handbook of Standards and Resources for Spoken Language Systems*, vol. 1. *Spoken Language System and Corpus Design*, XV. vol. 2, *Spoken Language Reference Materials*, XVI, Berlin : Mouton de Gruyter.
- TESTON B. (2000). L'évaluation objective des dysfonctionnements de la voix et de la parole. 1^e partie : les dysarthries, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 19, p. 115-154.
- TESTON B. (2001). L'évaluation objective des dysfonctionnements de la voix et de la parole. 2^e partie : les dysphonies, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 20, p. 169-232.

Liste des sites consultés, avec la date de consultation :

- <<http://www.med.univ-rennes1.fr/cerf/edicerf/strat/SR049.html>>, (27/09/2006).
- <<http://www.med.univ-rennes1.fr/iidris/cache/fr/18/1860>> (27/09/2006).
- <<http://www.orthophonistes.fr/Theme.php?NumTheme=20&Article=5>> (16/10/2006) pour le bilan orthophonique.
- <<http://www.unicode.org/fr/charts>>, (6/11/2006).
- <<http://www.up.univ-mrs.fr>>, (4/11/2006).
- <<http://www.up.univ-mrs.fr/delic>>, (5/11/2006).
- <<http://www.up.univ-mrs.fr/delic/corpus/index.html>> (5/11/2006).