



**HAL**  
open science

# Linguistique Interactionnelle: du Corpus à l'Expérimentation

Roxane Bertrand

► **To cite this version:**

Roxane Bertrand. Linguistique Interactionnelle: du Corpus à l'Expérimentation. Linguistique. Aix Marseille Université, 2021. tel-03500277

**HAL Id: tel-03500277**

**<https://hal.science/tel-03500277v1>**

Submitted on 22 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1. Linguistique Interactionnelle: du Corpus à l'Expérimentation

## Volume 1 – Document de synthèse

Dossier présenté pour l'obtention d'une  
**Habilitation à Diriger des Recherches**

Par

**Roxane Bertrand**

Chargée de Recherche Hors Classe au C.N.R.S.  
Aix-Marseille Université  
Laboratoire Parole et Langage – UMR 7309

*Soutenue le 10 novembre 2021 devant:*

**Liesbeth DEGAND**

**Elisabeth DELAIS-ROUSSARIE**

**Simona PEKAREK-DOEHLER**

**Noël NGUYEN**

**Catherine PELACHAUD**

**Philippe BLACHE**

Professeure, Université Catholique de Louvain -Rapporteure

Directrice de Recherche CNRS, Université de Nantes,  
- *Rapporteure*

Professeure, Université de Neuchâtel, - *Rapporteure*

Professeur, Aix-Marseille Université, - *Examineur*

Directrice de recherche CNRS, Sorbonne Université,  
- *Examinatrice*

Directeur de Recherche CNRS, Aix-Marseille Université,  
- *Garant*

## *Quelques mots de remerciements...*

---

*Cette HDR a été longue à venir. Les raisons sont nombreuses mais qu'importe.... Ce qui compte dans de tels moments, c'est le fait de s'arrêter, de se poser un instant, nullement pour s'appesantir sur hier mais plutôt pour savourer où l'on se trouve désormais et surtout apprécier ceux qui sont là. Il s'agit de boulot mais rien ne pourrait advenir sans le reste, tout le reste. Et les occasions ne sont pas si fréquentes pour ne pas profiter de dire à ceux qui comptent combien ils comptent.*

*Mes remerciements vont d'abord au laboratoire Parole et Langage, dont j'ai connu les nombreuses péripéties, de l'Institut de Phonétique d'Aix de mes premières années, au laboratoire Parole et Langage d'aujourd'hui. J'ai eu la chance de participer et de contribuer je l'espère, même modestement, à ses différentes évolutions notamment liées à l'arrivée de nombreux collègues talentueux venant d'horizons si différents. Les directions successives, depuis Bernard Teston, en passant par Philippe Blache, Noël Nguyen et Christine Meunier, Laurent Prévot et Serge Pinto, ont été les moteurs et les meilleurs accompagnateurs de ces évolutions. A chacun et à tous merci.*

*Merci à Liesbeth Degand, Elisabeth Delais-Roussarie, Simona Pekarek-Doehler, Noël Nguyen et Catherine Pelachaux qui m'ont fait l'honneur et le plaisir d'accepter si spontanément de lire mon travail. Je ne pouvais envisager plus beau jury. Merci à Philippe Blache, mon garant. Cela ne pouvait être que toi.*

*Mes remerciements vont aussi aux « femmes du 1<sup>er</sup> étage » Nadera, Stéphanie, Cathy. Merci pour votre soutien, quotidien, depuis tant d'années.*

*Merci aux nombreux collègues qui peuplent les différents étages de Pasteur. Merci Cristel pour nos jolies déambulations prosodiques passées et à venir. A celles que j'espère nombreuses et fructueuses avec Caterina. Aux plaisirs partagés avec Sophie, Amar, Alain, Marion, Anne, Serge, Thierry, Berthille et tous les autres. Et comme tu l'évoquais dans ton HDR, merci Cri, et merci Noël, Marianne, Philippe, Béa et Laurent, Stéphane, Laurent, Yohann pour votre douce et inestimable amitié.*

*Merci aux nombreux étudiants qui font de ce métier, la plus noble et la plus belle des aventures humaines. Un merci tout particulier à mes petites Mathilde, Aurélie, Marie Charlotte, et bienvenue à Auriane...*

*Merci Magalie, ma complice du LIS, pour ton dynamisme et ton enthousiasme.*

*Merci Gaëlle, ma pote, ma copine, pour ta constance et ta fidélité.*

*Merci ma Co, pour tous les moments de bonheur et de rire, parfois de chagrin. Ils sont dans mon cœur.*

*Merci Robert. « Paroles, paroles, paroles »... Les mots sont faibles... Merci pour ta présence, tout simplement.*

*Merci aux Réda, ma famille Laurisienne d'adoption. A Laura, petite blonde lumineuse. A mon Guillaume, fils de cœur.*

*Merci aux Simonin, mes amis de toujours. A nos soirées, nos échappées, et aux prochaines années.*

*Au trio infernal !*

*Aux Bertrand. Papa, Fabio El frangin, Flo, Maxime le magnifique filleul et Axelle la plus jolie des nièces.*

*A Cilou, ma sœur, mon double, mon envers.*

*A ma mère, l'infaillible. Merci pour ta présence, tes encouragements, ta confiance, indéfectibles.*

*A Gino, en dépit de tout. Au lien qui nous unit.*

*Et enfin à mes fils, mes joyaux. J'admire ce que vous devenez jour après jour. J'aime votre singularité et votre détermination. A vous qui me poussez dans mes retranchements, sans concession. Alex, à ton ardeur de vivre. Mathis, à ton art de vivre !*

*A Dostoievski, Balzac, Mankell, Roth, Connelly, Coe, Murakami, parmi tant d'autres. Aux mots que vous faites résonner si fort...*

# Table des matières

Introduction.....	6
Résumé du travail.....	7
Organisation de la synthèse.....	7
Chapitre I. Création et Enrichissement de données interactionnelles multimodales.....	9
Principales contributions.....	9
Préambule.....	10
1. A l'origine de la constitution et de l'enrichissement des données.....	12
1.1 Corpus « Laryngo ».....	13
1.1.1 Contexte.....	13
1.1.2 Création du corpus.....	13
1.2 Méthodologie.....	14
1.2.1 Procédure de détection automatique des événements et d'étiquetage.....	14
1.2.2 Unité d'analyse.....	14
1.2.3 Choix de la séquence analysée.....	14
1.3 Objectif, hypothèse et résultats.....	15
2. Constituer des ressources.....	17
2.1 Le <i>Corpus of Interactional Data (CID)</i> .....	17
2.2 Création du corpus.....	17
2.2.1 Un corpus interactionnel « semi-expérimental ».....	18
2.2.1.1 Protocole expérimental du CID.....	18
2.2.1.2 Motivations.....	19
2.2.1.3 Caractéristiques conversationnelles.....	19
2.2.2 Conclusion de section.....	20
3. Enrichir des ressources : le projet OTIM.....	20
3.1 Transcription.....	21
3.2 Développement d'outils d'aide à l'annotation.....	23
3.3.1 Syllabeur.....	23
3.3.2 Répétitions.....	23
3.2 Annotations discursives et prosodiques.....	24
4. Archivage et mutualisation.....	29
4.1 Contexte.....	29
4.2 Le CID : une ressource pour le français.....	30
Conclusion.....	30
Chapitre II. De la coordination à la convergence.....	32
Principales contributions.....	32
Introduction.....	33
1. Alignement et Compréhension mutuelle du dialogue en psycholinguistique.....	34
1.1 Le "Grounding account".....	35
1.1.1 Processus de Grounding.....	35
1.1.2 Les Conceptual Pacts.....	35
1.2 Le Modèle d'alignement interactif: un processus automatique.....	36
1.2.1 implicite/Explicite Common ground.....	36
1.2.2 Mécanisme de Priming.....	37
1.3 Facteurs stratégiques ou automatiques.....	37
1.3.1 Les marques de promotion du grounding ou de l'alignement : le cas des feedbacks.....	37
1.3.2 Le cas des repair.....	38
1.3.3 Au-delà de l'alignement automatique.....	38
2. Le point de vue de la Linguistique Interactionnelle.....	39
2.1 Formes linguistiques et ressources langagières.....	39

2.2 Le système des Tours de parole.....	40
2.2.1 La composante de construction des tours.....	40
2.2.2 La composante d'allocation des tours .....	40
2.3 Notion de projection et indices .....	41
2.3.1 Indices de projection.....	41
2.3.2 Poids des indices .....	41
2.4 Démarche dans la LI.....	42
3. Collaboration et compréhension mutuelle pour un accomplissement interactionnel réussi.....	43
3.1 Alignement et Affiliation.....	44
3.2 Conversation et activité .....	45
3.3 L'activité de narration.....	45
3.3.1 Le récit : une activité de co-narration.....	46
3.3.2 Attentes et droits dans l'activité asymétrique de narration.....	46
3.3.3 Les marques de collaboration des interlocuteurs .....	47
3.3.3.1 Notion de préférence .....	47
3.3.3.2 Les réponses feedback génériques/spécifiques.....	47
3.3.4 Contexte d'apparition des réponses.....	47
3.3.4.1 Les phases de narration .....	47
3.3.4.2 Résultats.....	48
3.3.5 Co-narration et convergence .....	51
3.3.5.1 Séquence interactionnelle convergente .....	51
3.3.5.2 Les discours rapportés écho.....	52
3.4 Les listes .....	53
3.4.1 Organisation structurelle des listes .....	53
3.4.2 Trajectoires interactionnelles .....	54
3.5 Les hétéro-répétitions .....	56
3.5.1 Aspects formels.....	56
3.5.2 Aspects fonctionnels.....	57
Conclusion .....	59
Chapitre III. Prosodie dans l'interaction .....	63
Principales contributions.....	63
Introduction.....	63
1. Interface Prosodie-énonciation .....	66
1.1 Polyphonie et dialogisme.....	67
1.1.1 Espace interactif et rapports de places.....	67
1.1.2 Discours rapportés directs .....	68
1.2 Rôle de la prosodie .....	69
1.2.1 Corpus .....	69
1.2.2 Paramètres prosodiques .....	69
1.3 Résultats et interprétation.....	70
2. Interface Prosodie-Discours.....	71
2.1 Le groupe ProDiGe : orientation générale .....	71
2.1.1 Cadre théorique .....	72
2.1.2 Résultats principaux.....	72
2.1.3 Diffusion de la recherche .....	75
2.2 Le projet PRO-GRAM: Prosodie dans la Grammaire .....	75
2.3 Les contours intonatifs: valeur pragmatique et réalisation phonétique .....	76
2.3.1 Statut des contours intonatifs.....	76
2.3.2 Résultats principaux .....	78
2.4 Le phrasé prosodique du français.....	81
3. Structuration prosodique et caractéristiques accentuelles du français.....	84

3.1 Objectif du projet.....	84
3.2 Frontières et proéminences.....	84
Conclusion et nouvelles interfaces.....	85
Chapitre IV. Conclusion et Perspectives.....	89
Projet 1: Alignement et affiliation.....	89
Analyse séquentielle des hétéro-répétitions (other-repetitions OR).....	89
Projet 2 : Les feedbacks.....	90
Caractéristiques formelles.....	90
Intégration des aspects informatiques et neuronaux dans l'étude de l'interaction.....	90
Avatar et persuasion.....	92
Impact général sur les agents conversationnels animés.....	93
Projet 3 : Détermination des Unités conversationnelles.....	94
Bibliographie.....	96
Table des figures.....	107
Liste des abréviations et sigles les plus fréquents.....	108
Annexe 1: l'expérience Brain-IHM.....	109
Annexe 2 : le corpus TRUENESS.....	111

## Introduction

---

Si je voulais résumer mon parcours depuis vingt ans, date de mon entrée au CNRS en tant que chargée de recherche à la section 34, j'évoquerais mon intérêt toujours aussi fort aujourd'hui pour ***l'interaction verbale en face-à-face et pour les sujets qui la conduisent et la façonnent tout en étant façonnés par elle***. Comme évoqué dans l'un des chapitres de l'ouvrage collectif intitulé *Le sujet* (Bertrand 2003), l'interaction est la base constitutive du dialogue et le lieu d'accomplissement de la parole, le premier lieu d'apprentissage et de mise en œuvre des premiers mots pour le petit enfant ; le sujet ne se constitue en tant que tel qu'en présence de l'autre et grâce au dialogue qui naît de cette rencontre.

L'interaction verbale est une **situation de parole** sur laquelle j'ai accompli la plupart de mes travaux de recherche. J'ai ainsi travaillé très largement les **corpus** d'interaction, et contribué aux différentes étapes de recueil, d'enrichissement et d'exploitation de données interactionnelles.

J'ai été amenée à m'interroger sur les points suivants : ne pas considérer le sujet producteur de parole comme tout puissant, seul dans sa parole et seul responsable du sens construit et échangé, mais bien comme un sujet multiple au sein duquel « ça » parle (les divers points de vue, la polyphonie, les **discours rapportés**) ; considérer celui ou ceux avec le(s)quel(s) il parle, ses interlocuteurs divers, réels ou potentiels, humains ou virtuels, qui sont à examiner très finement lorsqu'on cherche à comprendre comment le sens **s'élabore et se construit**. J'ai donc d'emblée rejeté la perspective très répandue à la fin des années 90 selon laquelle l'interlocuteur ne serait qu'un simple récepteur passif. Dès 1999, dans ma thèse de doctorat, j'ai notamment envisagé la question des phénomènes de **backchannels** (signaux d'écoute) requalifiés plus tard en **feedbacks** ou encore **listener's responses**, produits par celui qui, plus ou moins ponctuellement, occupe la place de l'interlocuteur.

Dans cette perspective largement inspirée des thèses interactionnistes issues du cadre de la *Conversation Analysis CA* (Sacks et al. 1974) et de *l'Interactional Linguistics IL* (Couper-Kuhlen & Selting 1996), quelques auteurs en psycholinguistique ont proposé à la même époque de s'intéresser au dialogue et à l'interaction. L'interaction est alors définie comme une **construction commune**, dans laquelle tous les participants à l'échange sont conçus comme de réels partenaires élaborant le sens conjointement, ce que le psycholinguiste Clark (1996) formalisera sous le terme de *modèle collaboratif*. Dans le même esprit, Fox Tree (1999), proche de Clark, déclare à propos des feedbacks qu'ils auraient un impact sur le processus non seulement de compréhension et d'écoute mais aussi de production des discours. Une large partie de mes recherches a été guidée par cette affirmation.

Je pourrais par ailleurs évoquer mon autre intérêt d'étude favori qu'est la **prosodie** et son rôle déterminant dans les différents aspects précédemment évoqués. L'émergence de *l'Interactional Linguistics* sous l'impulsion d'auteures telles qu'Elizabeth Couper-Kuhlen (qui m'a honorée de sa présence lors de mon jury de thèse) et Margret Selting, toutes deux spécialistes de prosodie, a été l'un des cadres propices pour illustrer ce rôle. Au sein du courant conversationnaliste de la CA qui s'intéresse plus particulièrement à l'aspect structurel et organisationnel de l'interaction quotidienne, le courant de l'IL fait du langage son objet privilégié d'étude, avec un focus particulier sur la prosodie envisagée comme une **ressource** cruciale pour les interactants. Bien que très attachée à ce courant, ma démarche s'en distingue cependant en raison du fait que je suis issue d'un laboratoire (Parole et Langage, anciennement Institut de Phonétique) très établi dans une autre tradition plus formelle, d'élaboration des modèles (Hirst & Di Cristo 1998). Elle s'inscrit pleinement dans les travaux ancrés phonologie qui s'attachent essentiellement aux aspects formels de la prosodie, bien qu'elle repose sur quelques principes divergents de l'approche métrique autosegmentale AM (depuis la thèse de Pierrehumbert 1980) devenue standard depuis pour de nombreuses langues. Il est à noter que pour le français l'AM est désormais devenu le modèle prégnant (voir les premiers travaux de Jun & Fougeron 2000, 2002, mais aussi l'article plus récent de Delais-Roussarie et al. 2015) qui propose d'adapter l'outil de transcription TOBI, initialement conçu pour l'anglo-américain, au français, à l'instar de nombreuses autres langues.

Ma démarche générale a donc consisté à tenter de relever quelques défis dont celui **d'articuler** une approche associant les aspects **fonctionnels et formels**, et de tenter d'appliquer les **modèles existants** conçus sur de la parole contrôlée, de laboratoire, à des données spontanées en conversation. Des corpus de taille plus importante ont été exploités, nécessitant parfois des méthodes d'analyse plus outillées, issues de la linguistique de corpus notamment. Enfin, un des derniers défis a consisté à tenter de dépasser l'étude de dichotomie relativement classique telle que syntaxe-prosodie et d'élargir l'étude aux liens entre les **différents niveaux de la grammaire**.

## Résumé du travail

Afin de rendre compte de la manière dont les participants à une interaction construisent de manière conjointe et collaborative l'activité dans laquelle ils sont engagés, les différentes ressources disponibles aux locuteurs (syntaxe, prosodie, discours, geste) s'avèrent cruciales pour comprendre ce qui se joue. La méthodologie utilisée consiste à articuler l'approche de *l'Interactional Linguistics* qui consiste en une analyse séquentielle et détaillée des activités conversationnelles et l'approche dite *corpus-based* qui consiste à exploiter et interroger plus systématiquement et quantitativement des corpus de taille conséquente. Cette double approche permet à la fois de mettre à jour de nouveaux phénomènes linguistiques pertinents pour l'interaction conversationnelle à un niveau de granularité très fin tout en quantifiant certains aspects qui nous permettent de rendre compte d'une réelle systématité des échanges. Parmi les ressources linguistiques utilisées par les participants, la prosodie constitue une dimension cruciale.

Nombreuses sont les études qui ont tenté de mettre en évidence les liens entre syntaxe et prosodie, ou bien encore entre discours et geste, mais mon projet vise à dépasser ces dichotomies afin de tenter d'intégrer les différents niveaux dans une approche linguistique large et unifiée, en vue d'aboutir à terme à une réelle théorie unifiée de la langue.

Le goût mais aussi l'exigence à effectuer des détours en vue de me confronter à d'autres questionnements posés par d'autres domaines sont au cœur de ma démarche (aller-retour entre modèles et données, linguistique théorique et linguistique de corpus, prosodie spontanée vs prosodie contrôlée, approche inductive vs hypothético-déductive en prosodie, etc). Ces « détours », bien que très coûteux en termes de temps car ils requièrent de se confronter à ce qui se joue dans les autres domaines, s'avèrent selon moi fondamentaux. En effet, ils permettent d'une part de révéler la complémentarité réelle existant entre les différentes disciplines, en l'occurrence les sciences humaines / sciences de la vie / sciences informatiques. Ils favorisent d'autre part de nouvelles pistes de recherche qui reposent sur des paradigmes et des méthodologies différentes qui permettent de relever de nouveaux défis. Ce projet s'inscrit dans un contexte national et international de plus en plus propice à ce type de recherche, en témoignent les travaux en neurolinguistique et les nouveaux questionnements autour des dialogues et de leurs corrélats neuronaux, regroupant des chercheurs issus de l'analyse des interactions et des expérimentalistes (cf les travaux de l'équipe de S. Levinson, la création de l'ILCB par P. Blache <https://www.ilcb.fr> ou du réseau Cobra par N. Nguyen <https://www.cobra-network.eu/>).

## Organisation de la synthèse

Ce document de synthèse est organisé de la manière suivante : après une introduction, j'ai choisi de résumer ma recherche en trois grands axes. Le chapitre 1 concerne la question des données et de leur enrichissement. Le second chapitre renvoie aux questions de collaboration, alignement et convergence dans l'interaction en face à face. Le troisième est consacré à la prosodie en interaction. Le chapitre 4 présente une conclusion et les perspectives de recherche que j'envisage pour les années à venir.

Dans chacun des trois chapitres centraux, j'insiste sur l'apport des différents domaines ou des différentes traditions qui sont indispensables pour mieux appréhender cet objet d'étude complexe que constitue l'interaction interindividuelle en face à face. Le chapitre 1 présente notamment l'intérêt que représente une approche plus outillée pour rendre compte de phénomènes analysés parfois de manière trop impressionniste ou ne demeurant qu'à l'état d'étude de cas. L'apport d'une partie de la linguistique de corpus est ainsi un atout selon moi pour mieux appréhender certains phénomènes encore méconnus. Ceci ne signifie pas que l'on doive renoncer aux études plus qualitatives (voir chapitre 2 notamment) loin s'en faut. Au contraire, tenter de conjuguer les deux est indispensable et constitue une entreprise en soi. Le chapitre 2 est présenté de manière un peu différente. J'ai choisi de présenter les principes et les présupposés théoriques des deux domaines convoqués, à savoir les modèles en psycholinguistique et le cadre de la linguistique interactionnelle, avant de présenter mes propres travaux. Cette présentation est liée au fait que mes travaux sur la convergence interactionnelle ont pris naissance dans les questionnements issus plutôt de la psycholinguistique tandis que ma démarche s'inscrit dans la linguistique interactionnelle. Le chapitre 3 présente mes travaux en prosodie et son rôle dans l'interaction en lien avec les autres niveaux du système linguistique. Comme dans les autres chapitres, je plaide pour le rapprochement entre les prosodistes/phonologues d'un côté qui élaborent les modèles des systèmes prosodiques des langues et les interactionnistes qui mettent l'accent sur la prise en compte des fonctions et des implications interactionnelles associés aux dispositifs prosodiques mis en jeu dans le discours et l'interaction. Enfin, le chapitre 4 conclut et propose des pistes de recherche pour les années à venir.

# Chapitre I. Création et Enrichissement de données interactionnelles multimodales

---

## Principales contributions

*Dans ce chapitre, je présente une partie de mes travaux qui pose la question de la multimodalité comme enjeu majeur pour l'analyse de la linguistique, ainsi que la question des moyens d'en rendre compte. L'objectif de ces travaux visait à se doter des ressources permettant d'étudier de manière systématique et unifiée les interactions multimodales, du niveau de granularité le plus fin (phonétique) au niveau le plus large (interaction). J'ai participé à la réflexion autour des outils conceptuels et méthodologiques permettant leur exploitation ultérieure, ce qui a nécessité de mieux caractériser chacun des niveaux pour permettre de mieux caractériser leurs relations mutuelles. J'ai ainsi contribué très activement à la mise en œuvre d'une plateforme constituée des ressources brutes et enrichies pour la description du français ainsi qu'au développement de certains outils d'annotation.*

*Ci-après les publications se rapportant à ce chapitre. D'autres les complèteront sur des aspects plus spécifiques présentés au cours du texte.*

## Publications en lien avec le chapitre

---

- Blache, P., **Bertrand**, R., Ferré, G., Pallaud, B., Prévot, L., Rauzy, S. (2017) "The Corpus of Interactional Data: a Large Multimodal Annotated Resource", in *Handbook of Linguistic Annotation* N. Ide & J. Pustejovsky (eds), Springer bookseries Text, Speech, and Language Technology, 1323-1356.
- Bertrand**, R., Ader, M., Blache, P., Ferré, G., Espesser, R. & Rauzy, S. (2009) "Représentation, édition et exploitation de données multimodales: le cas des backchannels du corpus CID", *Cahiers de Linguistique*, Numéro spécial Grands corpus de français parlé, Bruxelles, S., Mondada, L., Simon, A.C, Traverso, V. (eds), 33/2, 183-212. ISSN 0771-6524.
- Blache, P., **Bertrand**, R., Ferré, G. (2009) "Creating and exploiting multimodal annotated corpora: the ToMA project", In Kipp, M. (eds.) *Multimodal Corpora*. Berlin: Springer-Verlag, vol. LNAI 5509, 16, 38-53.
- Bertrand**, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. (2008) "Le CID -Corpus of Interactional Data- Annotation et Exploitation Multimodale de Parole Conversationnelle", *Traitement Automatique des Langues*, 49-3, 105-134
- Bigi, B. & **Bertrand**, R. (2016) "Laughter in French spontaneous conversational dialogs", 10th edition of the *Language Resources and Evaluation Conference (LREC)*, 23-28 May 2016, Portorož (Slovenia).
- Chen, H., Prévot, L., **Bertrand**, R., Priego-Valverde, B., Blache, P. (2012) "Toward a Mandarin-French Corpus of Interactional Data", *SemDial 2012 (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue* (2012 septembre 19-21: Paris, FRANCE).
- Blache, P., **Bertrand**, R., Guardiola, M., Guénot, M.-L., Meunier, C., Nesterenko, I., Pallaud, B., Prévot, L., Priego-Valverde, B., Rauzy, S. (2010) "The OTIM formal annotation model: a preliminary step before annotation scheme", *Proceedings of Language Resource and Evaluation Conference* (2010 mai 19-21: La Valette, MALTA), 3262-3267.
- Blache, P., **Bertrand**, R., Bigi, B., Bruno, E., Cela, E., Espesser, R., Ferré, G., Guardiola, M., Hirst, D., Magro, E.-P., Martin, J.-C., Meunier, C., Morel, M.-A., Murisasco, E., Nesterenko, I., Nocéra, P., Pallaud, B., Prévot, L., Priego-Valverde, B., Seinturier, J., Tan, N., Tellier, M., Rauzy, S. (2010) "Multimodal Annotation of Conversational Data", *Proceedings of Linguistic Annotation Workshop* (2010 juillet 15-16: Uppsala, SWEDEN), 186-191.
- Blache, P., **Bertrand**, R., Ferré, G. (2008) "Creating and exploiting multimodal annotated corpora", *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, (2008: Marrakech, MOROCCO). Marrakech: ELDA, 6 pages. Cederom.
- Bertrand**, R., Ferré, G., Blache, P., Espesser, R., Rauzy, S. (2007) "Backchannels revisited from a multimodal perspective", *Proceedings of Auditory-visual Speech Processing* (2007 août 31-septembre 3: Hilvarenbeek, THE NETHERLANDS), 1-5. Cederom.
- Ferré, G., **Bertrand**, R., Blache, P., Espesser, R., Rauzy, S. (2007) "Intensive Gestures in French and their Multimodal Correlates", *Proceedings of Interspeech* (2007 août 27-31: Antwerp, BELGIUM), 690-693. Cederom.

## Préambule

S'il y a 20 ans, nous, linguistes, étions très peu armés pour créer un corpus (pour diverses raisons, entre autres techniques) et inquiets, voire très craintifs de le voir nous échapper lorsqu'il existait, nous voici désormais dans la posture totalement inverse et tant mieux ! Mais ceci ne s'est pas fait en quelques semaines, et le résultat obtenu, consistant dorénavant à produire du corpus, à l'enrichir et à en obtenir une « gratification », est aujourd'hui un fait. Dans les lignes qui suivent, je voudrais revenir sur cette évolution en rappelant le contexte, les raisons et la manière dont les choses ont changé. Et je voudrais insister sur les personnes (localement) qui ont contribué très activement à ce changement.

Actuellement, la production de données s'effectue de manière très raisonnée, sans être démesurément intensive comme cela a pu être le cas (chacun ne produit pas « son » corpus pour sa « seule » utilisation) mais sans blocage particulier si cela s'avère nécessaire. L'aspect évaluation des « risques » relatif au coût -en termes de temps et d'argent- que peuvent représenter la création puis l'enrichissement d'un corpus, est bien connu. On sait non seulement pourquoi on décide de créer un corpus, mais surtout on sait quoi en faire et comment le faire. Si toutes les questions sont loin d'être résolues, il n'en reste pas moins que les étudiants par exemple sont désormais armés pour « attaquer » de front les données dont ils disposent ou qu'ils souhaitent constituer. Ils peuvent s'appuyer pour cela sur des étapes claires, que nous avons contribué à mettre en place au LPL.

Ma contribution personnelle sur ce point a été effective dès mon entrée au CNRS puisque mon projet comportait l'objectif de création d'une base de données de parole spontanée en interaction. Ce projet permettrait l'analyse simultanée des différents niveaux linguistiques et de leurs relations en vue de mieux comprendre comment le sens s'élabore et s'échange dans nos interactions interindividuelles quotidiennes.

Pour atteindre cet objectif, il s'agissait d'avoir à disposition des données suffisamment importantes et répondant à des critères spécifiques. Mais il s'avérait également indispensable de pouvoir traiter ces données pour les rendre exploitables et répondre à l'objectif de les étudier simultanément, ce qui n'était pas une mince affaire. Longtemps en effet, les chercheurs se sont intéressés à leur domaine d'expertise, la phonétique aux phonéticiens, la syntaxe aux syntacticiens, etc, et chacun d'eux a traité des questions et développé des outils conceptuels et méthodologiques propres à son seul domaine d'expertise.

Une telle entreprise n'a donc été possible que grâce à la collaboration de plusieurs membres du laboratoire qui ont investi ce nouveau champ de recherche avec une énergie et une détermination allant parfois jusqu'à une forme d'abnégation dans la mesure où les étapes extrêmement laborieuses mais nécessaires pour avancer n'étaient que très peu reconnues : quasi impossibilité de publier sur ces questions, peu de valorisation du travail, intérêt plus que limité de la part de ceux qui « faisaient la vraie recherche ». Je souhaite donc ici souligner l'implication de tous ceux qui se sont investis dans ce travail de titan au cours des premières années sans espérer que cela leur rapporterait beaucoup, notamment en termes de carrière académique. Très heureusement, la situation a changé et nous avons à la fois contribué à cet état de fait tout en en bénéficiant, enfin ! Et ceci grâce à d'autres acteurs venant d'horizons tels que le TAL, déjà familiers de ces questions, et qui ont saisi très tôt l'importance de ce virage dans la recherche en linguistique traditionnelle.

Je citerais tout d'abord Philippe Blache, le premier à y avoir cru. Grâce à son expertise scientifique dans divers domaines (de la linguistique au traitement automatique des langues), sa connaissance des enjeux de la recherche, sa capacité à dialoguer, sa curiosité associée à un esprit visionnaire qui sont les principales qualités d'un chercheur mais aussi à son audace et à son esprit entreprenant, il a permis de faire éclore une collaboration entre différents membres du laboratoire (et d'autres laboratoires extérieurs), collaboration formalisée par le projet OTIM qui scelle le début d'une aventure scientifique dont nous mesurons les retombées encore aujourd'hui. Les principaux travaux relatifs aux corpus et à

leur enrichissement trouvent leur origine dans ce projet qui en fédérant des chercheurs d'horizons variés visant tous le même objectif, a véritablement posé les jalons d'une recherche spécifique sur les interactions multimodales au LPL (voir section 3).

Tout cela n'aurait pu se faire non plus sans :

Béatrice Priego-Valverde avec laquelle nous avons créé le CID, sans nous douter un seul instant des retombées qu'il aurait

Robert Espesser, sans lequel rien n'aurait été possible : le travail immense autour de la transcription, de l'édition, des blocs à segmenter, des phases d'overlaps, des étapes de phonétisation et d'alignement du signal audio, entre autres

Gaëlle Ferré, qui a œuvré pour permettre l'analyse des vidéos du corpus à l'époque où cela posait encore de sérieux problèmes et qui nous a offert son expertise incomparable sur la multimodalité

Christine Meunier, phonéticienne, qui n'a pas hésité, alors qu'elle travaillait sur les voyelles isolées, à s'investir dans les interactions spontanées et à risquer de ne plus y retrouver ses voyelles tant aimées

Laurent Prévot, et son expertise en discours et traitement automatique des textes, qui a franchi le pas et nous a rejoint sans la moindre hésitation pour se confronter aux données orales multimodales

Stéphane Rauzy, et son énorme travail autour de l'annotation et des outils (du niveau morphosyntaxique au niveau mimogestuel) qu'il n'a cessé d'améliorer en tenant compte de nos annotations manuelles et qui nous permet de disposer d'annotations particulièrement riches

Berthille Pallaud, et son inépuisable énergie, grâce à laquelle nous disposons d'un corpus de 8 heures de français spontané annoté en disfluences à un niveau de granularité d'une incroyable précision

Brigitte Bigi qui a développé un logiciel et des outils d'interrogation adaptés aux besoins des linguistes

J'ajouterais également ceux et celles qui plus indirectement ou plus ponctuellement ont contribué à l'avancée de ces questions

Cristel Portes avec laquelle nous nous sommes confrontées aux difficultés de l'analyse prosodique en parole spontanée et qui n'a pas boudé son plaisir à annoter de grands pans du CID

Marion Tellier, qui nous a apporté son expertise en mimo-gestuel

Enfin les doctorants/post-doctorants qui dans leurs travaux ou par le biais de campagnes d'annotation, ont permis de constituer un tel corpus : Mathilde Guardiola, Aurélie Goujon, Klim Peskov, Guillaume Roux, Marie-Charlotte Cuartero, Pauline Péri, Océane Granier et tou.te.s les autres...

## 1. A l'origine de la constitution et de l'enrichissement des données

Avant d'entrer dans le détail de ces questions de constitution et d'enrichissement qui seront détaillées en sections 2 et 3 du présent chapitre, je souhaiterais rappeler ma démarche préalable à 2005, en termes d'utilisation de corpus et de méthodologie pour l'analyse de la parole et du langage, démarche qui comporte les prémices à certaines orientations et choix effectués ultérieurement.

Durant mes premières années au CNRS (entre 2000 et 2005), j'ai travaillé sur différents corpus créés « pour l'occasion », ce qui s'avérait courant à cette époque. Mes centres d'intérêt portaient entre autres sur les indices prosodiques en conversation (Bertrand & Espesser 2002 ; Bertrand & Chanet 2005) et le lien entre la voix et le geste (Bertrand et al. 1995 ; Cavé et al. 1996 ; Purson et al. 1998), etc. Il est sans doute important de rappeler que ce type de recherche n'était pas très répandu à cette époque. Il s'agissait donc d'explorer et d'esquisser des pistes permettant de rendre compte de ces phénomènes. Ces derniers s'avéraient selon moi absolument cruciaux pour comprendre la parole et le langage, et ce autrement qu'à travers l'étude de mots ou de phrases isolées, décontextualisées, et totalement dénuées de la dimension interactionnelle inhérente au langage. C'est précisément dans ce contexte qu'a paru l'ouvrage devenu référence depuis en *Linguistique Interactionnelle* (1996), signé des deux auteures majeures que sont Elisabeth Couper-Kuhlen et Margret Selting. Cependant, si cet ouvrage trace la perspective de ce que deviendra le courant de recherche le plus important pour l'analyse de l'interaction dans les deux décennies suivantes (Couper-Kuhlen & Selting 2018), il ne répond pas complètement aux questions qui vont nous intéresser dans ce premier chapitre - à savoir la constitution et l'enrichissement des données orales -. Bien entendu, les tenants de la *Linguistique Interactionnelle* travaillent sur des corpus et développent des outils théoriques et méthodologiques pour les exploiter. L'étape de transcription des données s'avère notamment cruciale pour eux car ils fondent leurs analyses sur les informations issues de cette transcription. Ils y apportent donc un soin tout particulier (voir le système de transcription GAT) et tentent d'intégrer le maximum d'informations dans la mesure où chaque « détail » aux différents niveaux linguistiques peut s'avérer pertinent (Local 2007). Dans notre propre travail, cette étape de transcription, si elle est fondamentale, ne l'est pas pour les mêmes raisons : des choix, et du soin apporté à cette étape, dépendront en effet les annotations ultérieures sur les différents niveaux (cf le schéma du plan de traitement de la parole plus bas) et la robustesse des outils d'annotation automatique utilisés. Concernant la question des corpus, la LI s'inscrit dans la continuité du courant de *l'Analyse Conversationnelle* dont l'un des principes fondateurs est le recours aux données les plus naturelles et écologiques qui soient, et qui ne peuvent être le résultat d'une demande d'un chercheur. De nouveau, notre approche s'est révélée quelque peu différente et nous en donnerons les motivations au cours de ce chapitre.

Le second aspect de mon travail durant le début des années 2000 concerne la perspective d'étude sur la voix et le geste (initié par I. Guaitella), laquelle a donné lieu aux deux premières éditions de la Conférence Oralité et Gestualité (*OraGe*) en 1998 et 2000. De la même manière que la Linguistique Interactionnelle a connu un retentissement énorme au niveau international depuis cette époque, la recherche autour du lien entre les modalités gestuelles et prosodiques de la parole s'est constituée en un domaine à part entière avec des auteurs aussi influents que Swerts, Mc Neill, Kipp, Martin, des conférences telles que *Gesture* et sa revue associée, GESPIN, ISGS, etc.

Pour revenir à mon propre travail durant ces années antérieures à 2005, je souhaiterais évoquer trois points qui selon moi agurent de ce que j'ai effectué ensuite :

- création du corpus « Laryngo »
- mise au point d'un outillage conceptuel et méthodologique pour traiter ces nouvelles données
- exploration des phénomènes de backchannel/feedback

## 1.1 Corpus « Laryngo »

### 1.1.1 Contexte

Les phases de chevauchement de parole (deux ou plus participants parlent simultanément) s'avèrent particulièrement fréquentes dans les interactions verbales spontanées. Elles constituent des moments interactifs cruciaux durant lesquels les participants manifestent leur implication plus ou moins grande dans le discours en cours et où se joue également, en termes de compétition, de négociation ou bien à l'inverse de coopération, l'alternance de parole.

Dans les années 2000, et cela reste en partie vrai aujourd'hui, il existait peu de travaux en prosodie notamment consacrés à ces phases de chevauchement (voir cependant Jefferson 1987 ; Wells & McFarlane 1998 ; Schegloff 2000, et plus récemment Kurtic et al. 2009 ; Heldner & Edlund 2010).

Trois raisons au moins peuvent expliquer ce point :

La première est la rareté des études prosodiques menées sur le dialogue spontané à cette date, à l'exception des conversationnalistes, et notamment le courant *Phonology for conversation* avec des auteurs majeurs comme Local, French, Wells (fin des années 80) et celle de Wells & Macfarlane (1998) pour l'une des plus influentes sur ce thème des chevauchements.

La seconde raison tient au fait que ces phases ont été totalement sous-estimées. En effet, si l'on se reporte à l'article princeps de Sacks, Schegloff et Jefferson (1974) qui décrit le système des tours de parole, on sait que ce dernier est fondé sur des règles précises d'alternance des tours qui tendent précisément à minimiser le nombre de chevauchements de parole. Selon les auteurs, les transitions de tours les plus fréquentes sont les transitions dites « douces » (*smooth-transitions*). Une majorité d'études a donc été consacrée à ces dernières.

La dernière raison pour expliquer la rareté des études sur cette question est d'ordre technique. Que l'on travaille à l'oreille ou que l'on utilise des logiciels de traitement de la parole, la difficulté à démêler les voix demeure une difficulté pour les premiers, une impossibilité pour les seconds. Il est donc indispensable de disposer non seulement d'un enregistrement de qualité mais surtout d'un enregistrement de chacune des voix sur pistes séparées (utilisation par exemple de laryngophones ou de micro-casques), qui rend l'expérimentation plus lourde.

### 1.1.2 Création du corpus

Le corpus « Laryngo » a été constitué en 2000. Ce corpus visait à concilier à la fois des données exploitables acoustiquement et la garantie d'échanges verbaux qui demeurent les plus naturels et conversationnels possibles. Cette nécessité conditionnera pour une part la constitution des prochains corpus et justifiera le recours à des données sans doute moins « écologiques » mais plus exploitables.

Le corpus « Laryngo » est constitué de 6 dialogues. Il a été enregistré au laboratoire Parole et Langage.

Protocole :

- Aucune consigne spécifique n'a été donnée. Les sujets étaient libres de discuter ou pas. Cette situation était ainsi assimilable, selon nous, à n'importe quelle situation quotidienne dans laquelle les individus se trouvent en présence d'une autre personne avec laquelle ils peuvent décider ou non d'entrer en communication.
- Les sujets étaient appareillés (laryngophones). Ils ont également été filmés afin de permettre une potentielle analyse gestuelle.
- Chaque dialogue dure environ 20 minutes. Les dialogues ont été stockés sur CD-ROM puis numérisés sur les stations de travail.

Grâce à ce dispositif expérimental, nous pouvons donc analyser la voix de chacun des locuteurs. En revanche, le laryngophone ne permet pas l'accès au segmental. Voici donc la méthodologie que nous avons déployée ainsi que l'analyse basée sur des indices dits bas-niveau.

## 1.2 Méthodologie

### 1.2.1 Procédure de détection automatique des événements et d'étiquetage

Ce travail a été mené en collaboration avec Robert Espesser, ingénieur au LPL et spécialiste du traitement et de l'analyse des données de parole (Bertrand & Espesser 2001).

Nous avons mis au point une méthode automatique de repérage et d'étiquetage des phénomènes en vue d'éviter la pose manuelle de milliers d'étiquettes nécessaires à l'analyse : le signal de parole a ainsi été découpé en 'groupes phonatoires' (temps de parole borné par deux pauses silencieuses).

### 1.2.2 Unité d'analyse

Similairement à Koiso et collaborateurs (1998), le groupe phonatoire, assimilable à l'*unité interpausale* (*Inter-pausal unit*, IPU) est défini comme un bloc de production de parole entre deux pauses silencieuses. Dans notre travail, les pauses ont été identifiées comme des segments non voisés, égaux ou supérieurs à 200 ms (correspondant à la moyenne de durée des pauses silencieuses brèves). Ce choix des pauses pour délimiter nos unités de parole résoud le problème de l'objectivité et de la robustesse des unités considérées. Contrairement aux unités syntaxiques, aux tours ou encore aux unités prosodiques parfois utilisées (Koiso et al. 1998), et qui nécessitent le recours à des annotations manuelles très coûteuses, les IPU ont les principaux avantages suivants :

- Critère formel d'identification
- Facilité de détection

### 1.2.3 Choix de la séquence analysée

Environ 5000 étiquettes de début et de fin d'IPUs ont été posées. Parmi elles, les chevauchements de parole sont considérés comme deux parties d'IPUs simultanées et ont été automatiquement détectés et indexés (étiquette initiale et finale d'IPUs). Nous comptons environ 2000 étiquettes de début et de fin de chevauchement.

Parmi les nombreux cas de chevauchements, nous avons sélectionné un pattern spécifique, en l'occurrence le plus fréquent dans notre corpus, qui se présente comme suit :

- une « preceding phase » (P), partie de l'IPU précédant le début du chevauchement
- une « overlap phase » (O), phase contenant une partie des deux IPU simultanées des deux participants
- une « continuation phase » (C), partie de l'IPU suivant la phase de chevauchement

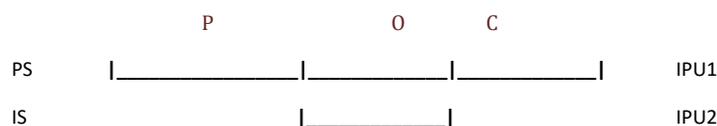


Figure 1a: no speaker change (C=P)

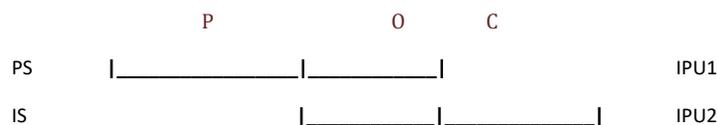


Figure 1b: speaker change (C ≠ P), extraites de Bertrand & Espesser (2001)

### 1.3 Objectif, hypothèse et résultats

Notre travail a consisté à rendre compte quantitativement du phénomène de chevauchement de parole qui était jusqu'alors principalement analysé d'un point de vue qualitatif (voir l'étude très détaillée des différentes catégories de chevauchement établies par Jefferson 1987). Il est à souligner que l'étude de Koiso et al. (1998) qui a fortement inspiré notre travail du point de vue méthodologique, a tout simplement éliminé les cas de chevauchement de parole (pour les raisons évoquées en 1.1.1).

277 séquences POC ont été analysées. 8 variables ont été retenues : durée de P, O, et C en ms, et les valeurs de f0 en 5 points spécifiques de la séquence. Nous avons choisi une fenêtre d'analyse de 130 ms avant, au début et à la fin du chevauchement. Notre hypothèse était que les indices pertinents seraient localisés en ces lieux très spécifiques au sein des séquences de parole examinées<sup>1</sup>.

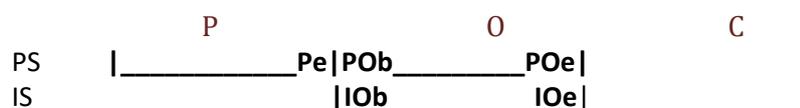


Figure 2: the five selected f0 localization points, extraite de Bertrand & Espesser (2001)

Nous avons modélisé la probabilité du changement de locuteur à l'aide d'une régression logistique, assez novatrice en sciences du langage à cette époque.

Nous avons donc testé le rôle de deux paramètres prosodiques (durée et fréquence fondamentale) dans l'alternance de locuteurs dans la séquence POC. Nous avons vérifié l'hypothèse, inspirée de French et Local (1986), que l'utilisation de ces variables est fortement systématisée. Nos résultats montrent que les paramètres de durée et de f0 significatifs sont la durée de P et C, la moyenne f0 en Pe et IOe. Plus la durée de P est courte, plus la durée de C est longue, plus basse est la moyenne f0 en Pe et plus haute l'est-elle en IOe, plus la probabilité de changement de locuteur est grande. Les autres paramètres ne sont pas significatifs.

Voici quelques éléments d'interprétation de ces résultats. La durée de P courte augmente la probabilité de changement de locuteur. Cela suggère qu'il est plus facile pour le locuteur suivant de prendre le tour quand le premier locuteur est interrompu tôt et n'a pas eu suffisamment de temps pour s'impliquer dans son tour. Concernant l'organisation des tours de parole, ce patron suggère que les deux locuteurs pourraient considérer la présence de la pause avant P comme un indice de place transitionnelle (*Transition-Relevance-Place*, TRP, Sacks et al. 1974), à savoir un lieu potentiel d'alternance de locuteur (voir chapitre 2 pour le modèle des tours). Si tel est le cas, le chevauchement pourrait résulter d'un délai plus long de la part du locuteur 2 à prendre la parole (sachant qu'on estime le temps de réaction à 200 ms, Ward 1996<sup>2</sup>). Le premier quartile de la durée de P associé au changement de locuteur est de 220 ms. Ce type de configuration peut donc illustrer un cas de « blind-spot ». Ce blind-spot appartient à la catégorie de « transitional onset » de Jefferson, défini comme un cas "in which the recipient/next speaker's incoming talk starts just fractionally after the start of further talk by the previous turn-occupant, following a TRP and a pause" (1987: 165-7). Dans notre article, nous précisons ainsi: "If the blind-spot belongs to the class of short P durations, the blind-spot situation would clearly increase the probability of speaker change, durP having a strong effect on the probability of speaker change (odd ratio = 0.52, see table 9)". Jefferson evokes this point when she says "the blind-spot onset can result from the incoming speaker's moving from a recipientship into a speakership mode at the TRP, (i.e after the

<sup>1</sup> Aujourd'hui, nous choisirions sans doute une fenêtre plus grande (au moins de 250 ms correspondant à la durée moyenne d'une syllabe finale accentuée en français, mais ce serait également différent dans la mesure où nous pourrions aussi bénéficier d'une annotation manuelle des tons et des unités prosodiques, voir chap 2)

<sup>2</sup> La pertinence de cette durée d'environ 200 ms dans l'alternance de tours de parole a été notamment confirmée depuis sur différentes langues (Stivers et al. 2009)

*original turn-occupant's previous utterance) and consequently, when s/he starts her/his own turn, not attending to the original turn-occupant's resumption of talk" (cited in Wells & Macfarlane 1998: 277) (Bertrand & Espesser 2002).*

Pour plus de détails, nous renvoyons le lecteur à Bertrand & Espesser (2002) dans lequel nous interprétons chacun de nos résultats à la lumière des catégories établies par Jefferson (1987).

Pour cette synthèse, nous limitons l'exposé ci-après au patron relatif aux backchannels (BC), qui nous occuperont tout au long de notre travail.

La durée de la phase O présente différents patrons. Pour les plus courtes durées, la probabilité de changement est faible. Cette durée courte de O pourrait donc être associée à la présence d'un BC, lequel consiste généralement en une, voire deux syllabes (*mh, d'accord,...*), et dont l'une des fonctions est de signaler l'attention de l'interlocuteur au discours du locuteur 1 tout en lui laissant le tour de parole. Par opposition aux durées de O courtes qui diminuent les probabilités de changement de locuteur, l'allongement des durées de O augmente cette probabilité de changement. Nous avons interprété cette tendance en termes *d'engagement* : étant à l'origine du chevauchement de parole, le locuteur 2 serait particulièrement impliqué dans sa parole et aurait réellement quelque chose à dire. Plus la durée de O est longue, plus son implication est importante. Une durée d'environ 800 ms semble la durée optimale de changement de locuteur. Au delà de 800 ms, cette probabilité de changement décroît. Dans ce cas, il s'agit d'un chevauchement compétitif. Cette zone temporelle de 800 ms pourrait être le seuil au delà duquel le premier locuteur montre sa volonté de préserver à tout prix sa position de locuteur principal, l'interlocuteur devenant alors de plus en plus illégitime.

Pour conclure cette section, nos résultats ont mis en lumière une part des mécanismes basiques à l'œuvre dans le changement de locuteur, et ont permis de caractériser deux types de chevauchements prosodiquement distincts : les chevauchements 'intrusifs' qui impliquent une alternance de locuteur après la phase de chevauchement de parole moyennant une certaine durée de ce dernier, et les chevauchements 'coopératifs' qui eux se particularisent par le fait de ne pas interrompre le premier locuteur. Nous avons fait l'hypothèse que cette dernière catégorie était largement composée des phénomènes de backchannel pour lesquels nous avons mis en lumière un patron prosodique particulier. Ce dernier a été établi sur la base d'une combinaison des paramètres suivants : durée courte de O (Roger 1989), valeurs basses pour mf0I0e, les BC étant souvent produits dans un registre bas (Müller 1996) et valeurs basse pour mf0Pe également, les BC étant préférentiellement précédés par un point (ou une région) bas (Ward 1996), et enfin une durée de P moyenne à longue, les BC apparaissant préférentiellement à des points de complétion (syntaxique notamment) (Ford & Thompson 1996) ou à la fin d'un énoncé ou d'une clause (Koiso et al., 1998). Aucune contrainte n'apparaît sur C.

Dans la suite de cette synthèse, nous reviendrons à plusieurs reprises sur ces backchannels (que nous requalifierons de *feedback* ou *réponses*) qui demeurent l'un de nos objets d'étude privilégiés. En effet, ils reflètent la manière dont les participants se manifestent leur écoute attentive, leur participation, leur compréhension, leur implication plus ou moins active et leur collaboration à créer et échanger du sens, ce qui est au cœur de notre travail (nous reportons le lecteur au chapitre 2).

## Publications

- 
- Bertrand, R. & Espesser, R. (2001)** "About Speech Overlaps: Prosodic Cues Contribution in Predicting a Change of Speaker", In Puppel, S. and Demenko, G. (eds). Proceedings of Prosody 2000, Adam Mickiewicz University, Pologne, 29-35
- Bertrand, R. & Espesser, R. (2002)** "Prosodic Cues of Speaker Changes in Speech Overlap Phenomena in French", Rapport

## 2. Constituer des ressources

Pour revenir au présent chapitre, je souhaite maintenant retracer les étapes principales du travail autour de la constitution et de l'enrichissement des corpus à partir du milieu des années 2000. Ce travail s'inscrit pleinement dans l'esprit qui animait ma recherche dès le début, à savoir étudier les phénomènes langagiers en interaction en tenant compte des différents niveaux linguistiques sans les hiérarchiser, et tenter de mettre en lumière leur poids relatif selon les situations (dans tel cas, le niveau syntaxique peut jouer un rôle prépondérant par rapport aux autres tandis qu'à un autre moment, c'est éventuellement le geste qui prendrait la première place). Mais pour mener à bien ce travail, un travail conséquent, coûteux mais surtout collectif a été nécessaire.

Le projet de constituer des ressources orales enrichies comportant des annotations sur la phonétique, la prosodie, la morphologie, la syntaxe et la pragmatique ainsi que le mimo-gestuel prend très concrètement naissance autour des années 2003-2004. La principale difficulté ne porte pas sur l'annotation de chacun de ces niveaux (il commence à exister pour cela des outils à la fois conceptuels et méthodologiques), mais sur l'annotation de tous les niveaux. Il s'agit donc avant tout de proposer un formalisme répondant à cet objectif. Par ailleurs, de telles ressources doivent pouvoir être manipulées de façon automatique. A la différence des solutions généralement adoptées, l'objectif est d'utiliser un langage permettant non seulement d'effectuer des requêtes sur ces annotations, mais également de les manipuler.

Plus globalement, l'objectif vise donc l'élaboration d'une plate-forme permettant de constituer et d'exploiter des ressources linguistiques comportant plusieurs niveaux d'annotation. Ceci implique de disposer d'un corpus différent du « Laryngo », permettant l'accès aux données segmentales.

Ce projet d'envergure a d'abord été initié dans le cadre de plusieurs « petits » projets, plus ou moins formalisés et/ou financés. Entre 2003 et 2005, nous avons ainsi proposé des normes au niveau de la transcription et développé/adapté des outils d'alignement et de synchronisation des données de parole, grâce à l'obtention d'un projet (PRAX) coordonné par P. Blache et moi-même, financé par l'ILF. Mais le projet dans son ensemble n'a pu se déployer pleinement qu'à travers le programme OTIM qui a permis de regrouper des forces et des moyens (notamment pécuniaires) à la hauteur des objectifs visés. Ce projet sera détaillé au point 3, après la présentation du corpus utilisé.

### 2.1 Le *Corpus of Interactional Data (CID)*

Collaborateurs LPL : P. Blache, R. Espesser, C. Meunier, B. Priego-Valverde, S. Rauzy

Historiquement, le LPL, et avant lui l'Institut de Phonétique d'Aix-en-Provence, a toujours œuvré pour constituer et exploiter des corpus de parole. C'est même le cœur de sa démarche.

Ce qui se joue autour des ressources orales au milieu des années 2000 diffère dans la mesure où plusieurs membres du laboratoire souhaitent désormais travailler ensemble sur un même corpus. Leur motivation première est de mieux comprendre ce qui se joue dans chacun des niveaux linguistiques, dont chacun est expert, et ce en confrontant ses connaissances acquises sur la parole contrôlée de laboratoire à la variabilité de la parole spontanée, encore largement méconnue. Une seconde motivation réside dans la perspective de pouvoir étudier les liens entre les différents niveaux en se dotant des ressources et des outils nécessaires.

Dans les points qui suivent, nous allons décrire le corpus sur lequel ce travail a été mené, en précisant le contexte dans lequel il a été conçu et les raisons pour lesquelles il a pu servir de support à ce projet.

### 2.2 Création du corpus

Le CID est né de la collaboration entre ma collègue B. Priego-Valverde (MCF) et moi-même. Recrutées la même année à Aix, B en analyse des interactions, et moi-même chercheure CNRS sur la prosodie

dans la parole spontanée en interaction, nous avons chacune préalablement travaillé sur des corpus « écologiques » (naturels) durant nos thèses de doctorat respectives. En dépit de leur intérêt indéniable, ces corpus ont cependant montré certaines limites relativement à ce qui nous intéressait : difficulté d'une analyse simultanée des différents niveaux linguistiques en raison d'enregistrements de qualité assez médiocre, impossibilité de prendre en compte la totalité des discours (chevauchements de parole, marqueurs discursifs ou backchannels souvent peu perceptibles en cas d'enregistrement utilisant un seul micro), faiblesse/rareté de certains phénomènes en conversation ne permettant pas de les étudier de manière systématique ou quantitative en dépit de leur rôle, difficulté à comparer des phénomènes extraits de corpus différents présentant des facteurs de variabilité trop importants, etc.

Il s'agissait donc de disposer au départ d'une base de données comportant des *discours rapportés directs* ainsi que des *procédés humoristiques*. Nos corpus précédents nous avaient permis de constater que les discours rapportés par exemple pouvaient être favorisés par la production de récits portant notamment sur les relations professionnelles. De la même manière, obtenir des séquences humoristiques pouvaient être favorisées par des discussions relativement tendues (afin de faire baisser la pression) mais aussi par une certaine familiarité entre les participants<sup>3</sup>.

Le CID a donc été conçu dans cette perspective : un corpus dont la vocation première nous permettrait d'étendre nos résultats à d'autres locuteurs et de tester de nouveaux paramètres (tels que certains indices prosodiques supposant une qualité audio optimale). Les résultats précédents sur les discours rapportés étaient lacunaires, étant extraits d'une conversation enregistrée sur une seule et même piste. Différents paramètres, tels que l'intensité, n'avaient pu être considérés en raison des variations incontrôlées liées au fait que les locuteurs n'étaient pas dans une position stable et fixe par rapport au micro. Grâce au dispositif expérimental déployé pour le CID, nous avons pour ambition de pouvoir contrer ces difficultés et traiter les données obtenues de manière optimale.

## 2.2.1 Un corpus interactionnel « semi-expérimental »

### 2.2.1.1 Protocole expérimental du CID

Les sujets sont enregistrés dans la chambre sourde du laboratoire, assis en biais à une distance d'environ 1m. Ils portent des micro-casques permettant d'enregistrer chacune des voix sur une piste séparée. Ils sont également filmés. Les interactions sont non mixtes. L'âge des participants oscille entre 25 et 45 ans environ. Ils sont pour la moitié originaires de la région PACA, pour l'autre de diverses régions françaises. Enfin, ils sont familiers les uns des autres et familiers des lieux.

Deux séries de dialogues ont été enregistrées, selon deux consignes.

- 1<sup>ère</sup> série (2003) : Nous avons demandé aux sujets d'évoquer le cas d'une ou de relations de travail avec lesquelles ils ont connu des conflits, et d'illustrer en racontant par exemple des scènes précises (d'altercation éventuelle ou autre).

La durée de chaque dialogue est d'une heure environ. La première série du corpus compte 5 interactions.

Ce protocole a effectivement donné lieu à la production de très nombreuses séquences de discours rapportés directs qui ont fait l'objet d'études (Bertrand & Priego-Valverde 2013 ; Guardiola & Bertrand 2013).

---

<sup>3</sup> Bertrand R. (1999). De l'Hétérogénéité de la Parole: Analyse Enonciative de Phénomènes Prosodiques et Kinésiques dans l'Interaction Interindividuelle. *Thèse de Doctorat, Université de Provence*  
Priego-Valverde B. (1999) L'humour dans les interactions conversationnelles : jeux et enjeux. *Thèse de Doctorat, Université de Provence*

- 2<sup>ème</sup> série (2005) : Dans une seconde phase d'enregistrement, nous avons modifié la consigne en vue d'obtenir des séquences humoristiques supplémentaires. La consigne était donc la suivante : racontez-vous des choses ou des situations insolites dans lesquelles vous ou quelqu'un de votre connaissance s'est trouvé impliqué.

3 nouveaux dialogues d'environ 1 heure sont ainsi venus compléter la première série. D'autres n'ont pu être conservés en raison de problèmes techniques (encore fréquents à l'époque, problème de synchronisation audio/vidéo, micro défectueux, etc).



Figure 3 : Dispositif du CID pour deux dialogues

### 2.2.1.2 Motivations

Le choix de constituer un corpus « semi-expérimental » répond à diverses motivations. Par ce terme, nous entendons un corpus enregistré dans des conditions expérimentales tout en s'appuyant sur les résultats issus des analyses conversationnelles pour créer les conditions optimales favorisant l'émergence d'une interaction la plus naturelle possible entre les participants. Les interactions obtenues doivent satisfaire à plusieurs contraintes : il s'agit tout d'abord d'interactions de parole spontanée dans la mesure où il n'y a pas de préparation écrite. Concernant le dispositif lui-même, nous avons veillé aux critères de proximité (distance d'1 m environ), à la position des locuteurs (de biais), et à leur degré de familiarité mutuel (fort en l'occurrence) mais aussi leur familiarité avec le lieu d'enregistrement (ce sont des sujets travaillant sur le site) ainsi qu'avec les méthodes employées (chambre sourde, appareillage avec micro-casques, film). Nous avons également privilégié des interactions non mixtes afin de gommer la variabilité intrinsèque (sexuelle), mais aussi les dialogues au tri/dialogues car les connaissances relatives à ces derniers étaient alors encore relativement limitées.

### 2.2.1.3 Caractéristiques conversationnelles

Durant cette période, ont émergé des corpus tels que la Maptask<sup>4</sup> (Anderson et al. 1991). Le CID se distingue de ce type de corpus par sa proximité plus grande avec une conversation naturelle. Exceptée la consigne que reçoivent les locuteurs avant l'enregistrement, et à laquelle ils peuvent très vite déroger, le CID répond aux critères de la conversation (symétrie entre les locuteurs, rôles et tours de parole non pré-déterminés, pas de réelle tâche à accomplir si ce n'est se raconter quelques anecdotes). Si cette consigne « impose » sans doute une forme de conversation particulière (donnant lieu notamment à de nombreux récits), elle se distingue des interactions en maptask. Par ailleurs, la consigne du CID est destinée à favoriser l'expression de phénomènes apparus et étudiés dans nos corpus conversationnels précédents (discours rapportés, séquences d'humour), ce qui a réussi et qui

<sup>4</sup> Dans ce type de corpus orienté-tâche, deux participants un leader et un follower doivent retrouver un itinéraire sur une carte. Chacun a une carte comportant des informations différentes qui conduisent les participants à négocier le sens.

est en soi très intéressant : cela signifie en effet que l'on dispose de certains leviers pour favoriser l'apparition de certains phénomènes tout en disposant d'une conversation assez naturelle.

### 2.2.2 Conclusion de section

Pour conclure sur l'intérêt, parfois remis en cause par certains, de tels corpus semi-expérimentaux ou semi-spontanés, pour mieux comprendre l'interaction, nous souhaitons souligner deux points : le premier concerne le fait que chaque situation d'interaction est unique et qu'il n'y a pas de hiérarchie à établir entre les différents types. Toute situation de parole, qu'elle soit en interaction ou non (lecture à haute voix par exemple), est digne d'être étudiée car elle contribuera à améliorer nos connaissances sur le langage et la parole. Ce qui est essentiel en revanche, c'est de toujours décrire très précisément les conditions de production d'une situation de parole, qu'elle soit produite en conditions naturelle ou expérimentale, et d'interpréter les résultats obtenus à la lumière de ces conditions.

Par ailleurs, quelles que soient les conversations, les interactionnistes ont montré qu'elles ne pouvaient être étudiées comme un tout homogène. Chacune est constituée d'activités distinctes, liées à des tâches et des objectifs différents, qui conditionnent le rôle des participants. En dépit de cette affirmation, beaucoup reste encore à explorer. C'est l'un des aspects de notre travail (voir la caractérisation de l'activité de narration détaillée au chapitre 2). Comprendre ce qui se joue au sein des interactions requiert donc de tenir compte de ces divers aspects mouvants que sont les buts, rôles, activités, qui créent des attentes et génèrent des réponses particulières au sein de chaque conversation. En ce sens, les *maptasks* constituent un type d'activité spécifique, conditionnée par le but et les rôles discursifs assignés aux participants. Enfin, et pour conclure sur l'intérêt d'étudier des situations de parole variées, plus ou moins écologiques, il est important de noter qu'un fait ou un phénomène observé en milieu écologique peut être rare mais s'avérer absolument crucial pour expliquer un processus particulier. Constituer un corpus semi-expérimental en tentant de favoriser la production d'un tel phénomène peut permettre d'en mieux comprendre les enjeux.

## 3. Enrichir des ressources : le projet OTIM

*OTIM : Outils de Traitement de l'Information Multimodale* (ANR BLANC),  
Philippe Blache (responsable, LPL)  
<http://www2.lpl-aix.fr/~otim/membres.html>  
2009 1/01/2012

Partenaires : LIMSI (Paris), LSIS (UMR CNRS, Toulon), LIA (Université d'Avignon)

Je suis moi-même impliquée à plus de 40% de mon temps de recherche.

Ce projet pose la question de la multimodalité comme enjeu majeur pour l'analyse de la linguistique moderne, mais aussi et surtout la question des moyens d'en rendre compte (à travers les connaissances et les outils existants).

Il a un objectif à la fois linguistique et technologique et se propose d'aborder les points suivants :

- Création d'un format d'encodage multimodal générique
- Spécification d'une chaîne de traitement : définition des étapes, recommandations, outils d'aide à l'annotation
- Création et exploitation d'un outil de requête sur le format créé
- Création et exploitation de nouvelles ressources annotées dans le format spécifié

L'ambition de ce projet est donc de rassembler au sein d'une même plateforme d'une part une base de ressources brutes et enrichies pour la description du français et d'autre part un ensemble d'outils d'annotation et de manipulation ou de requête. Cette plateforme sera expérimentée et validée sur deux types de corpus multimodaux :

- un ensemble de corpus audio et vidéo partiellement annotés que nous possédons (CID)
- des nouveaux corpus annotés sur des domaines variés, permettant d'expérimenter des situations différentes

La plateforme nous permettra d'annoter de façon automatique ou semi-automatique ces données pour enrichir (ou compléter) les corpus sources.

Dans le contexte du programme OTIM, le CID est donc apparu comme un véritable chantier sur lequel chacun a pu faire ses armes. Je voudrais donc et à nouveau souligner le travail des collègues au début même du projet. Avant même en effet de pouvoir développer des outils ou de lancer des campagnes d'annotation de grande envergure, ce sont eux, en tant qu'experts, qui ont ouvert la voie en explorant ce nouveau type de données. L'article dans la revue TAL témoigne de cet effort collectif (Bertrand et al. 2008).

Pour ce qui concerne les différentes étapes du projet, je renvoie le lecteur aux articles ultérieurs relatifs au schéma formel d'annotation multimodale qui a permis d'encoder les informations de la manière la plus homogène et consistante possible et aux annotations concrètes au sein des différents niveaux (cf Blache et al. 2017 et les nombreux papiers publiés sur le sujet par les différents membres du projet).

Dans cette synthèse, je limiterai la présentation à trois points auxquels j'ai contribué de manière active et qui permettent de dérouler quelques étapes cruciales inhérentes à la chaîne de traitement d'OTIM:

- Etape de Transcription
- Développement d'outils d'annotation
- Etape d'Annotations prosodiques et discursives ; campagnes d'annotation, guide d'annotation

### 3.1 Transcription

L'étape préalable à l'enrichissement du CID a consisté en une transcription sur laquelle a été menée une très large réflexion (Bertrand et al. 2008; Blache et al. 2017 entre autres).

Comme mentionné en introduction, j'aimerais rappeler que la transcription est cruciale dans de nombreux travaux. Mais à la différence de la LI et la CA pour lesquelles la transcription doit comporter le maximum d'informations pour permettre l'analyse ultérieure puisque celle-ci s'appuie exclusivement sur elle, l'étape de transcription pour nous est cruciale mais pour d'autres raisons : la transcription représente l'étape initiale à partir de laquelle les autres annotations seront déclinées. L'idée sera ensuite d'ancrer -temporellement en l'occurrence- chacune des annotations afin d'obtenir des données synchronisées qui pourront ainsi être interrogées simultanément.

Dans la figure ci-dessous, la *TOE* renvoie à la *Transcription Orthographique Enrichie*. La motivation initiale de cette transcription est liée au manque de travaux sur la parole spontanée. Nous n'avons pas de certitudes sur son niveau de variabilité, et donc de l'écart à la norme d'un tel jeu de données. Afin de ne pas rater des phénomènes, mais surtout pour optimiser les outils automatiques utilisés ultérieurement, nous avons privilégié cette TOE, qui comporte des éléments tels que les élisions, les prononciations ou liaisons non attendues, etc (voir Bigi et al. 2012 pour plus de détails sur les bénéfices d'une telle TOE sur la phonétisation et l'alignement sur le signal audio). Mais pour permettre également d'utiliser les outils traditionnellement utilisés sur de l'écrit (tagger, chunker pour le niveau morphosyntaxique par exemple), nous avons produit une transcription orthographique standard. Cette double transcription est illustrée dans le schéma ci-dessous. Ce double codage est ancré temporellement (synchronisé) et permettra d'aligner les futures annotations soit sur la ligne « speech » soit sur la ligne NLP (Natural Language Processing) (figure 4).

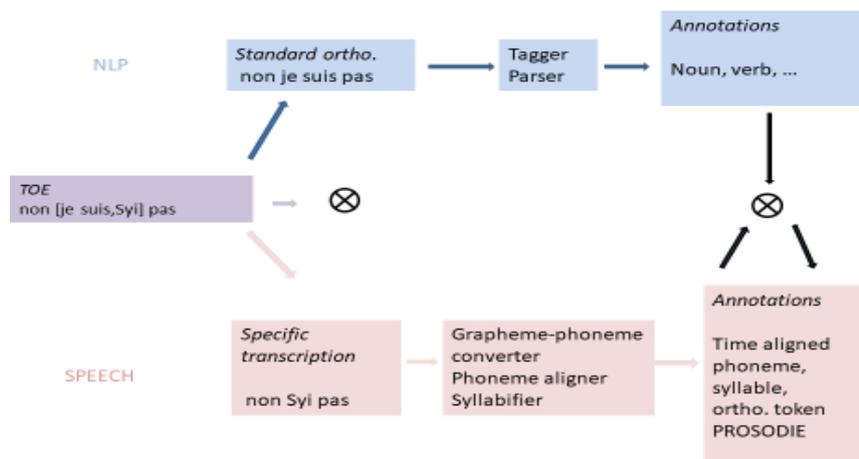


Figure 4 : Plan du traitement automatique

Cette figure 4 permet d'illustrer l'importance de l'étape de transcription qui s'effectue au sein des IPU (segmentation initiale de l'audio). Cette étape est cruciale et les choix de proposer à la fois une transcription ortho et une transcription plus phonétique permet de déployer des outils différents, par exemple les taggers pour les catégories morphosyntaxiques qui se fondent sur la transcription orthographique tandis que les annotations liées à la parole sont fondées sur la transcription phonétique. Grâce à cette double transcription, il est possible d'avoir une correspondance entre les deux « lignes » de traitement (figure 5).

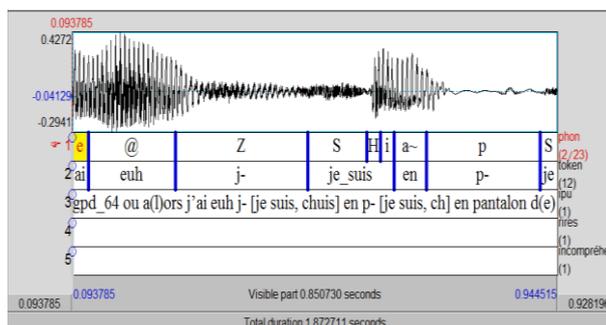


Figure 5: exemple d'annotation sous Praat : signal audio, tire 1 = phonèmes, tire 2 = token, tire 3 = TOE (les deux autres tires 4 et 5 concernent d'autres niveaux d'annotation, rires et incompréhensions). On peut voir l'alignement entre les deux premières tires, et les différentes particularités de prononciation (le mot tronqué j- (je), l'assimilation de /je suis/ en /chui/)

## Publications

Bigi, B., Péri, P., **Bertrand**, R. (2012) "Orthographic Transcription: Which Enrichment is required for Phonetization?" The 8th International Conference on *Language Resources and Evaluation* (2012 mai 23-25: Istanbul, TURKEY), 1756-1763, ISBN 978-2-9517408-7-7.

Bigi, B., Péri, P., **Bertrand**, R. (2012) "Influence de la transcription sur la phonétisation automatique de corpus oraux", *Journées d'Étude sur la Parole*, 449-456, Grenoble (France).

En amont de ces questions plus concrètes, un gros travail a été effectué afin de représenter au mieux les données multimodales selon un même formalisme. Les informations au sein de chaque niveau ont été représentées dans des structures de traits typés (*Type Features Structures*, TFS) qui permettent de décrire les relations hiérarchiques entre les structures, tout en permettant l'homogénéisation des données et leur encodage consistant (pour plus de détails cf. Blache et al. 2017).

## 3.2 Développement d'outils d'aide à l'annotation

Le second point que nous souhaitons évoquer a trait aux outils d'annotation qui ont été créés ou améliorés dans le projet OTIM (cf Blache & Rauzy 2008 entre autres). Pour ma part, j'ai contribué plus particulièrement à deux outils de détection automatique, le syllabeur et l'outil de détection automatique des répétitions lexicales.

### 3.3.1 Syllabeur

Le premier outil concerne l'annotation automatique des syllabes en parole spontanée. Il n'existait que très peu de systèmes automatiques permettant d'extraire les syllabes systématiquement, notamment sur des données conversationnelles. Or, cet outil s'avère particulièrement important car il permet de disposer d'un niveau d'annotation extrêmement précieux pour ceux qui travaillent en phonétique ou en prosodie, la syllabe étant une unité de segmentation et de programmation (Bigi et al. 2010). Il a été conçu grâce à l'annotation manuelle en syllabes d'un sous-corpus du CID effectué par C. Meunier, N. Nesterenko et moi-même. Cette annotation codée par trois expertes a permis d'établir un accord inter-annotateur et de constituer ainsi un gold, à partir duquel ont été extraites les règles que B. Bigi a implémentées dans l'outil. Ce niveau d'annotation est désormais systématiquement proposé dans le cadre de l'enrichissement d'un nouveau corpus (figure 6). Une fois le corpus transcrit, il est ainsi possible de disposer très rapidement des trois niveaux d'alignement sur le signal audio des phonèmes, syllabes et tokens, ce qui permet de mener différentes études telles que leur fréquence d'occurrence ou leurs corrélats acoustiques.

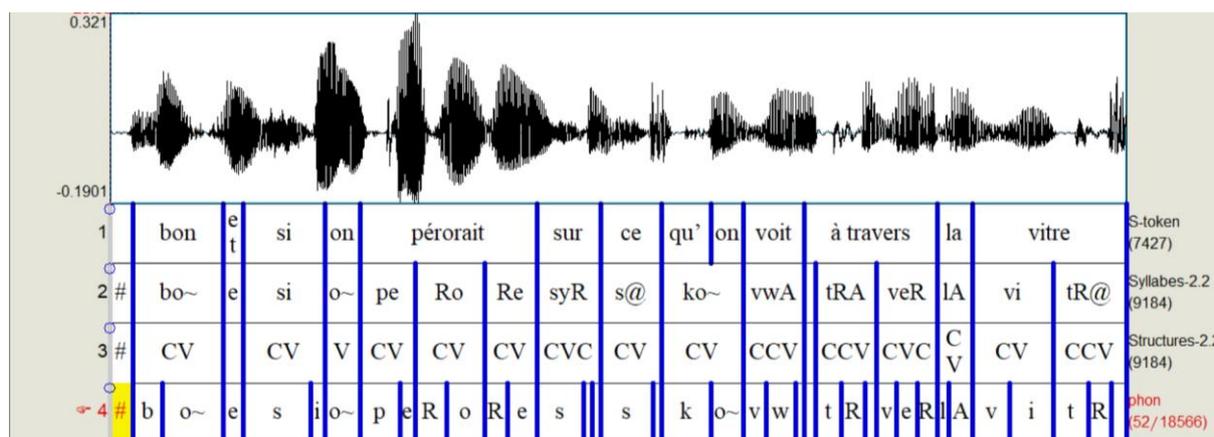


Figure 6 : annotation automatique des syllabes et de leur structure interne (tier 2 et 3), la tier 1 renvoie aux tokens et la tier 4 aux phonemes.

### Publications

- Bigi, B., Meunier, C., Nesterenko, I., **Bertrand**, R. (2010) "Automatic detection of syllable boundaries in spontaneous speech", *Proceedings of Language Resource and Evaluation Conference* (2010 mai 19-21: La Valette, MALTA), 3285-3292. Cederom.
- Bigi, B., Meunier, C., **Bertrand**, R., Nesterenko, I. (2010) "Annotation automatique en syllabes d'un dialogue oral spontané", *Actes XVIIIèmes Journées d'Étude sur la Parole* (2010 mai 24-28: Mons, BELGIUM), 4 pages. Cederom.

### 3.3.2 Répétitions

Le second outil concerne la recherche du phénomène des répétitions lexicales produites par l'interlocuteur (dites « other-repetitions » OR, ou répétition hétéro-centrée). Cet outil a été conçu dans un esprit différent du précédent. Il s'agit d'aider au repérage d'un phénomène particulier en conversation, relativement fréquent et important mais qui peut aussi être absent (à la différence des syllabes). Il s'agissait d'aider à définir plus particulièrement ce que l'on considère comme de « réelles » répétitions. L'idée n'était donc pas de repérer systématiquement les mots répétés, quels qu'ils soient, mais d'identifier ceux qui pouvaient être considérés comme des répétitions ostensives de mots ou de

syntagmes de l'interlocuteur. Ainsi, nous n'incluons pas toutes les occurrences répétées d'un item dans la mesure où, par exemple, un mot peut revenir à plusieurs reprises simplement parce qu'il constitue le thème discuté. Nous avons donc dû effectuer un travail préalable d'annotation manuelle sur deux dialogues, afin de déterminer les principaux critères définitoires des OR (fondés principalement sur Tannen 1989; Perrin et al. 2003) en vue de les implémenter ensuite dans l'outil (cf Bigi et al. 2014). Les répétitions sont fréquentes et multifonctionnelles en discours. Les obtenir automatiquement a donc permis non seulement un gain de temps mais il a permis de mieux définir ce que le phénomène recouvre. Ce travail a été mené en lien avec le projet relatif à l'imitation et la convergence dans le dialogue, coordonné par Noël Nguyen (Imitation in Speech, [http://spim.scicog.fr/index\\_francais.htm](http://spim.scicog.fr/index_francais.htm)).

Nous reviendrons plus en détail sur ce projet dans le chapitre 2, notamment sur les raisons qui ont motivé le développement de cet outil d'annotation.

	744.0	744.3	744.5	744.7	745.0	745.2	745.5	745.7	746.0	746.2	746.5	746.7	747.0	747.2	747.5	747.7	748.0	748.2	748.5	748.7	749.0	749.2	749.5
TokensAlign	un	oyer	foyer	foyer	catho	un	foyer	bonne	soeurs				#				ah	ouais					#
OR-Source	S19			S20																			
OR-Repetition													R18						R20		R19	R20	
TokensAlign	#												un	feul	artifice	#			dans	foyer	bonne	oeur	

Figure 7 : illustration de la sortie du système automatique. OR\_Source renvoie à l'occurrence initiale de l'énoncé qui sera répété, OR à l'élément répété. Il s'agit ici de l'énoncé : « un foyer de bonnes sœurs ». On peut remarquer notamment que le terme « foyer » est auto-répété plusieurs fois par le locuteur 1, mais que la source considérée comme la plus pertinente est la S20. Les différentes occurrences sources sont considérées comme des disfluences et la S20 est la source à laquelle se rapportera OR20. Suite à l'identification automatique des occurrences de répétitions, un passage manuel est parfois requis pour confirmer la source pertinente.

## Publications

Bigi B., Bertrand R., Guardiola M. (2014) Automatic detection of other-repetition occurrences: application to French conversational Speech. *Proceedings of The 9th edition of the Language Resources and Evaluation Conference* (2014 mai 26-31: Reykjavik, ICELAND)

Bigi B., Bertrand R., Guardiola M. (2010) Recherche automatique d'hétéro-répétitions dans un dialogue oral spontané. *Actes XVIIIèmes Journées d'Étude sur la Parole* (2010 mai 24-28: Mons, BELGIUM), 4 pages. Cederom

## 3.2 Annotations discursives et prosodiques

Malgré le développement d'outils semi-automatiques d'aide à l'annotation comme présentés ci-dessus, de nombreuses annotations restent largement manuelles. Elles sont ainsi extrêmement coûteuses en temps et en argent. Elles nécessitent une main-d'œuvre importante pour non seulement contribuer à l'enrichissement le plus exhaustif possible du corpus mais également pour permettre de tester leur robustesse (accords inter-annotateurs).

Ces annotations manuelles sont également très coûteuses car elles requièrent la contribution des experts du domaine afin de pouvoir bénéficier d'une annotation de référence et pouvoir la comparer aux annotations plus « naïves » effectuées par plus d'annotateurs. Ma propre contribution sur ce point a été effective aux niveaux prosodique (en lien avec I. Nesterenko et C. Portes) et discursif (en lien avec L. Prévot). Pour chacun des niveaux d'annotation, nous avons élaboré des schémas d'annotation caractérisant les informations devant être encodées ainsi que la manière dont elles doivent l'être.

Les normes d'annotation, fondées sur les standards internationaux existants, ont été adaptées à nos données. Sur le plan discursif notamment, nous avons accordé une part plus grande aux aspects conversationnels qui étaient habituellement très peu pris en compte dans les modèles standards d'annotation discursive. J'ai participé à la campagne d'annotation des unités discursives (Prévot et al. 2010, pour plus de détails voir Prévot et al. 2015) et à celle des fonctions communicatives des feedbacks (cf le projet -CoFEE- Conversational Feedback: multi-dimensional analysis and modelling,

ANR coordonné par Laurent Prévot (<https://cofee.hypotheses.org/>), qui impliquait d’abord la création d’un schéma d’annotation puis d’un guide destiné à l’annotation par des sujets « naïfs ».

Dans le cadre de la thèse de doctorat de M. Guardiola, j’ai également annoté avec cette dernière les *discours rapportés directs* et les *narrations* (nous reviendrons plus en détail sur ces deux annotations qui ont servi de support à nos études présentées dans le chapitre 2). A titre d’illustration, voici le schéma d’annotation des discours rapportés directs, totalement basé sur le schéma d’annotation global proposé dans le projet. Comme nous l’avons déjà évoqué ci-dessus, les annotations au sein de chaque niveau sont représentées dans des *structures de traits typés* (*Type Features Structures*, TFS) qui décrivent les relations hiérarchiques entre les structures mais qui permettent surtout d’homogénéiser les données et rendre leur encodage consistant.

Le discours rapporté direct est représenté par sa « structure » qui renvoie à une structure d’échange (optionnelle) qui révèle si les discours sont rapportés par un ou plusieurs personnage et s’ils sont produits en une seule fois ou s’ils sont discontinus (isolé, échange, discontinu isolé, discontinu échange), « composants » renvoient aux différents composants de la structure (formule introductive, les différentes voix), la « source » renvoie aux voix (soi-même pour l’auto-citation, autres, indéfini) et le « type » renvoie aux fonctions (basées sur la typologie de Vincent & Dubois 1997) tel que reproduction (consiste à présenter le discours rapporté comme ayant été réellement produit dans une autre situation), pseudo-reproduction renvoie au cas où la situation en question n’est pas clairement identifiable, assertion renvoie au fait de rapporter le discours comme un argument d’autorité, l’actualisation est utilisée pour les discours rapportés qui ont été utilisés dans plusieurs situations similaires, enfin l’invention est un cas de discours rapporté qui n’a jamais été produit antérieurement. Nous reportons le lecteur au chapitre 2 sur les discours rapportés directs en écho qui renvoient précisément à ce dernier cas. 590 structures de discours rapportés directs ont été ainsi annotés sur la totalité du CID (figure 8).

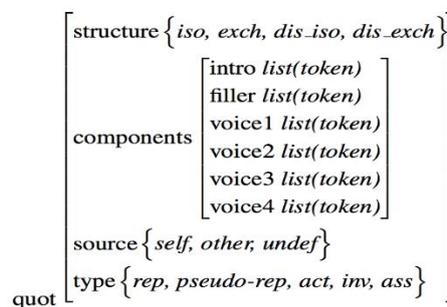


Figure 8 : schéma d’annotation des discours rapportés directs, extrait de Guardiola & Bertrand (2013)

La figure 9 suivante illustre une annotation sur le CID des discours rapportés directs de l’une des locutrices. On observe la tire 1 des IPU dans laquelle est transcrite la parole, en tire 2 les tokens, et dans les tires suivantes les 4 caractéristiques du discours rapporté direct tel que nous les avons décrites dans le schéma d’annotation ci-dessus.

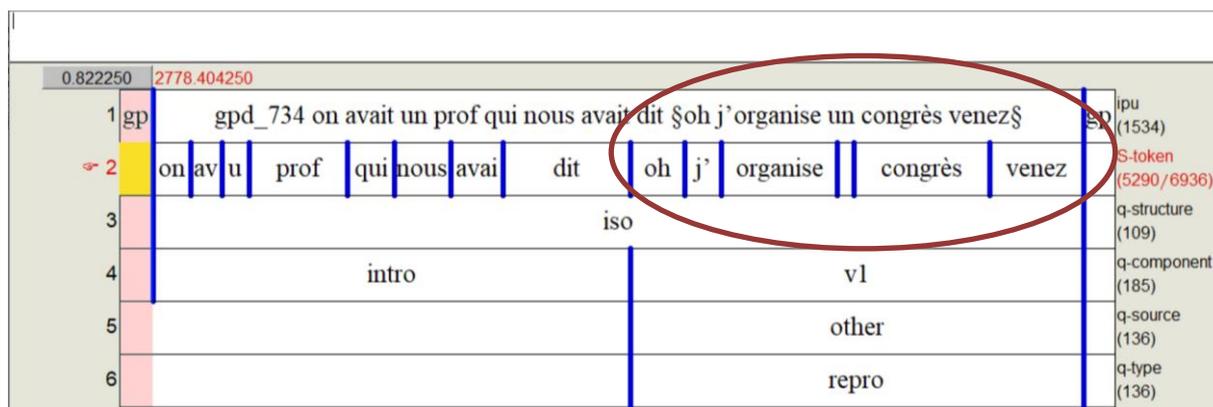


Figure 9 : annotation des discours rapportés directs du CID

Outre les annotations discursives, nous avons également élaboré le schéma d'annotation et annoté les informations du niveau prosodique. Nos annotations ont été également effectuées selon les normes internationales. Inspirées des modèles établis pour le français (Di Cristo 1998 ; Jun & Fougeron 2000), nous avons produit un schéma d'annotation pour la parole spontanée du CID. Dans les termes de l'approche métrique autosegmentale (AM), Jun & Fougeron ont proposé les deux unités suivantes:

- l'AP (*Accentual Phrase*, appelée par ailleurs unité mineure) qui est le domaine de l'accent final et de l'accent initial et dont le patron tonal sous-jacent est /LHLH/ (L = low, H = High). La frontière droite de l'AP est démarquée par la montée finale (LH) et l'allongement de la syllabe finale pleine.
- L'IP (*Intonational Phrase*, ou Unité Intonative/Intonème Majeur) plus large que l'AP. Elle est marquée par un mouvement majeur de  $f_0$  sur la dernière ou les deux dernières syllabes, un allongement final plus important et est souvent suivie d'une pause silencieuse (Hirst & Di Cristo 1984 ; Fougeron & Jun 1998).

Nous avons fait l'hypothèse que les unités établies sur de la parole contrôlée pouvaient, sous réserve d'adaptation à certaines particularités de l'oral, être pertinentes pour des données conversationnelles (Portes & Bertrand 2011). Ces annotations visaient à compléter, voire améliorer nos connaissances encore éparses et peu consensuelles, à l'époque, sur les modèles prosodiques du français. Ainsi, divers auteurs ont proposé d'inclure une troisième unité (*l'intermediate phrase ip*, précisément entre l'AP et l'IP) (Jun & Fougeron 2000, Di Cristo & Hirst 1996, Michelas 2011).

Une campagne d'annotation identique à celle liée aux unités discursives a donc été menée ensuite. Ces campagnes impliquent un nombre important de participants pour permettre l'annotation exhaustive du corpus. Outre l'élaboration d'un guide d'annotation, elles consistent aussi en la recherche, l'entraînement et le suivi des annotateurs, ainsi que l'évaluation des annotations. Ce point est central dans la production de données enrichies. Il est crucial en effet de disposer d'annotations fiables et relativement consensuelles sur de tels corpus qui comportent par ailleurs de nombreux facteurs peu contrôlés. L'estimation de l'accord inter-annotateur (et mesures de kappa) permet ainsi de rendre plus robustes les annotations, en évitant le biais de la variabilité individuelle (Nesterenko et al. 2010), pour l'accord inter-annotateur des unités prosodiques. Nous reviendrons sur ces questions dans le chapitre 3 de cette synthèse.

Les figures suivantes 10a et b et 11a et b illustrent l'annotation du phrasé prosodique à la fois par les deux experts et par les naïfs dans les tiers 3 (IP) et 4 (AP) (1 et 2 renvoyant à la transcription et aux tokens). Le premier chiffre renvoie à l'accord entre les annotateurs experts et le second à celui entre les annotateurs naïfs. IP22 signifie que les deux experts et les deux annotateurs sont d'accord sur cette unité. La figure 10b présente d'autres niveaux d'annotations disponibles pour le CID. Il s'agit des unités discursives (tier 1) du récit et de ses phases formelles (tier 2 et 3), voir chapitre 2 pour plus de détails,

et les informations morphosyntaxiques dans les tiers 6 à 10, issues de Marsatag (Blache & Rauzy 2008). Ces annotations précisent la catégorie morphosyntaxique, le chunk (groupe nominal, groupe prépositionnel, etc) et les scores de fiabilité liés à ces informations (de A excellent à E médiocre).

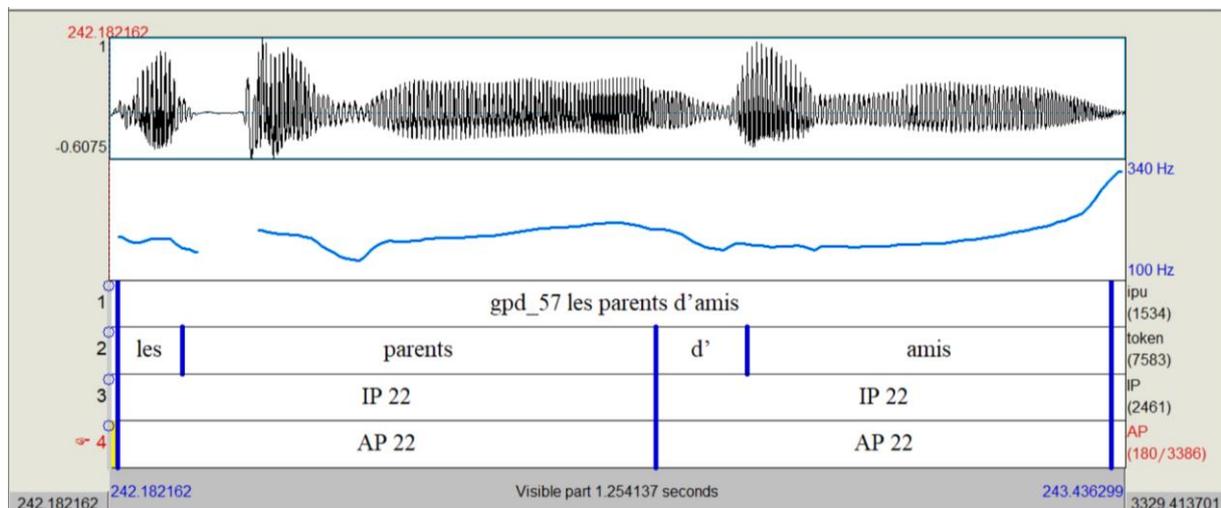


Figure 10a : annotations expertes et naïves du phrasé prosodique

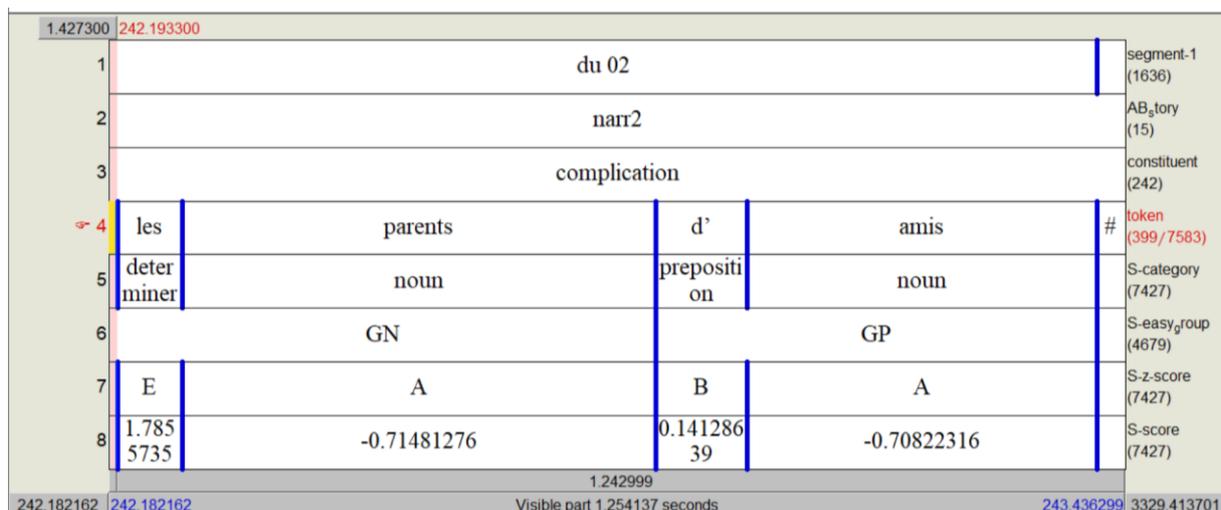


Figure 10b : annotations du niveau morphosyntaxique, des unités discursives et des narrations

Dans la figure 11a, le phrasé prosodique présente plus de désaccords entre les naïfs. On constate que la première IP a été identifiée consensuellement par les experts alors que les naïfs ne l'ont pas retenue (IP20), tandis que pour la seconde IP seul un annotateur naïf a été d'accord avec les deux experts. En revanche on constate que tous les annotateurs sont d'accord sur les AP.

Ces exemples illustrent le fait que l'annotation du phrasé reste encore une tâche délicate notamment sur la parole spontanée. Mais il serait faux de croire que les experts sont toujours d'accord. Il reste encore beaucoup de points à éclaircir notamment sur le lien entre les unités prosodiques et les marqueurs discursifs ou les disfluences, ce que nous discuterons dans le chapitre 3. Cette étape d'annotation s'avère ainsi absolument cruciale pour non seulement mettre en lumière la complexité de la tâche mais surtout pour identifier les difficultés relatives à l'identification des unités en parole continue.

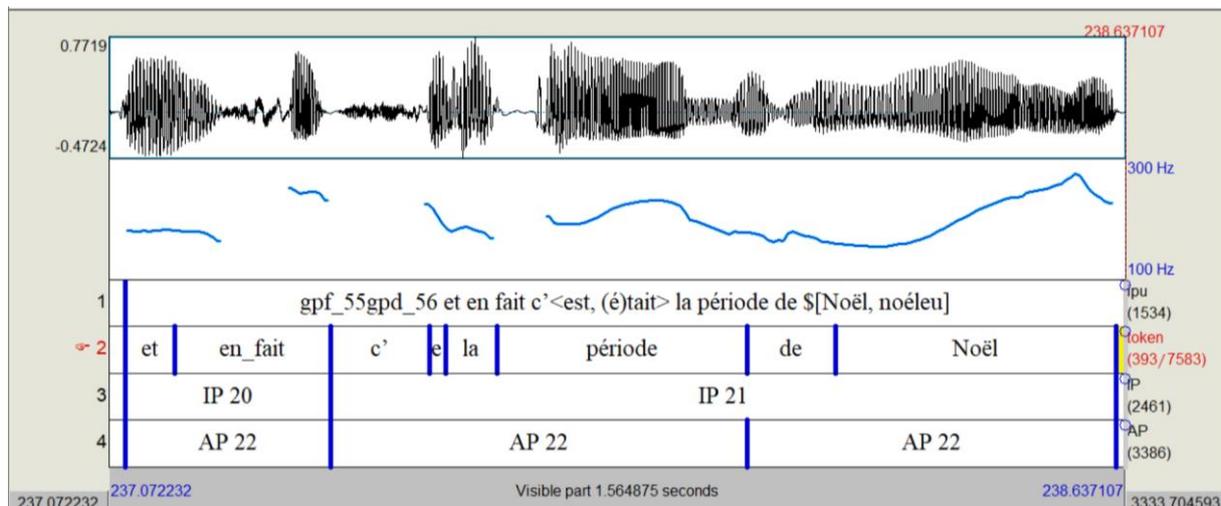


Figure 11a : annotations expertes et naïves du phrasé prosodique

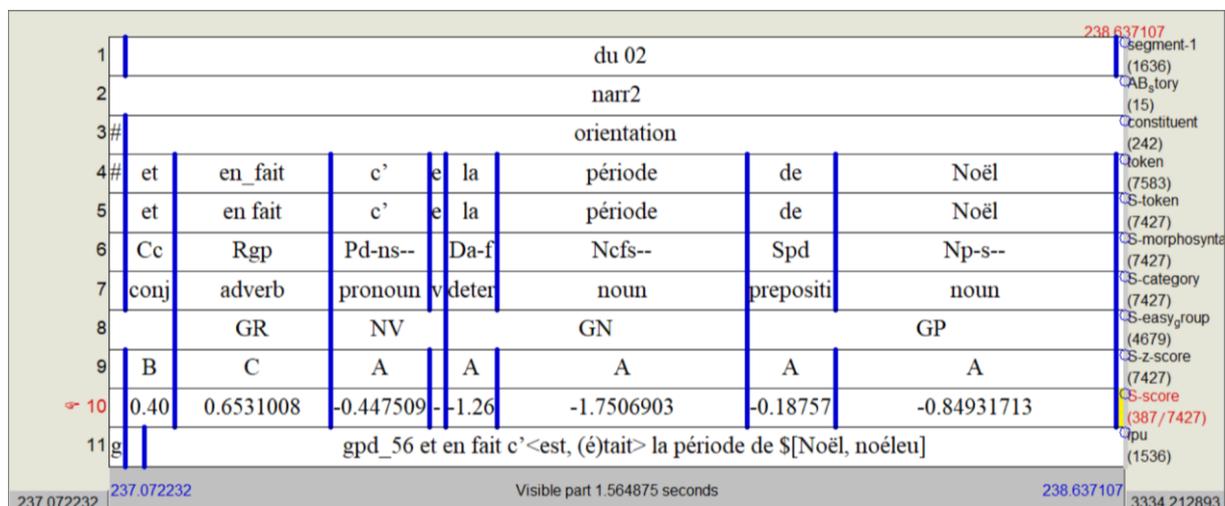


Figure 11b : annotations du niveau morphosyntaxique, du niveau discursif et des narrations

Les campagnes d'annotation prosodique et discursive ont été menées sur environ 3 mois.

Sur le plan prosodique, nous avons par ailleurs effectué une annotation des contours intonatifs sur une partie seulement du corpus avec ma collègue C. Portes (Portes & Bertrand 2006 ; Bertrand & Portes 2007), mais nous n'avons pas lancé de campagne d'annotation plus large auprès de participants naïfs. Cette annotation a été produite pour affiner nos connaissances sur le sens des contours intonatifs que nous avons réinvestis dans un modèle sémantique du dialogue que nous présenterons là encore plus en détail dans le chapitre 3.

## Publications

- Prévot, P., Gorisch, J., **Bertrand**, R. (2016) "A CUP of CoFee: A large Collection of feedback Utterances Provided with communicative function annotations", 10th edition of the *Language Resources and Evaluation Conference (LREC)*, 23-28 May 2016, Portorož (Slovenia).
- Prévot, L., Gorisch, J., **Bertrand**, R., Gorene, E., Bigi, B. (2015) "A SIP of CoFee: A Sample of Interesting Productions of Conversational Feedback", *Proceedings of 16<sup>th</sup> Annual SIGdial Meeting on Discourse and Dialogue*, Prague, Sept 2-4.
- Peshkov, K., Prévot, L., **Bertrand**, R. (2013) "Prosodic phrasing evaluation: measures and tools", *Proceedings of Tools and Resources for Analysing Speech Prosody (TRASP)*, Aix en Provence, août 2013.
- Peshkov, K., Prévot, L., **Bertrand**, R. (2013) "Evaluation of automatic prosodic segmentation", *Proceedings of Prosody-Discourse Interface*, Leuven. September 11-13.
- Nesterenko, I., Rauzy, S., **Bertrand**, R. (2010) "Prosody in a corpus of French spontaneous speech: perception, annotation and prosody ~ syntax interaction", *Proceedings of Speech Prosody* (5. 2010 mai 11-14: Chicago, United States of America), 4 pages. Cederom.

Prévoit, L., **Bertrand**, R., Priego-Valverde, B., Blache, P. (2010) "Discourse Interaction in French Conversation: A Case Study for Interoperable Semantic Annotation", *Proceedings of Workshop on Interoperable Semantic Annotation* (2010 janvier 15-17: Hong-Kong, HONG KONG), 39-48.

Prévoit, L., **Bertrand**, R., Peshkov, K., Rauzy, S., Blache, P. (2015) "Prosody, Discourse and Syntax in French Conversations: Resource creation and evaluation", *Rapport*.

Avant de passer à la dernière section relative à l'archivage et la mutualisation des corpus, nous souhaitons rappeler que les annotations du CID ont également été effectuées sur le versant mimogestuel selon les mêmes principes et le même modèle formel sous-jacent. Les annotations étaient toutes produites manuellement. Nous reportons le lecteur intéressé à Blache et al. (2007) et Tan et al. (2010) pour plus de détails.

Désormais, l'outil d'annotation *Openface* permet d'aider à l'annotation des indices mimogestuels. C'est au cours de sa thèse de doctorat qu'Aurélié Goujon, co-encadrée par ma collègue M. Tellier et moi-même, a inauguré l'utilisation de cet outil pour faciliter la détection des mouvements de sourcils (haussement et froncement) sur trois corpus spécifiques en vue d'affiner leur rôle dans l'élaboration du common ground et de la compréhension mutuelle (Goujon 2019; Bertrand & Goujon 2017). La comparaison entre les sorties de l'outil automatique et les annotations manuelles a montré un score assez médiocre de l'outil, explicité dans la thèse. En revanche, depuis, une autre doctorante M. Amoyal a pu tester l'outil sur l'identification des sourires qui s'est avéré bien plus performant (Rauzy & Amoyal 2020). Dans le cadre du master 2 d'Auriane Boudin, co-encadré par M. Ochs, P. Blache et moi-même, nous avons utilisé *Openface* pour les mouvements de tête. La correction manuelle est en cours afin d'évaluer la performance de l'outil sur ces mouvements.

## 4. Archivage et mutualisation

### 4.1 Contexte

Ce travail aurait été vain sans le souci d'archiver et de mutualiser les données obtenues. Ceci s'est développé au sein d'un mouvement plus général qui a permis de "mettre à l'honneur" la partie jusque là "immergée" du travail sur corpus. Des initiatives importantes ont également vu le jour autour des aspects éthiques relatifs à la constitution des corpus (respect des données personnelles, anonymat des participants, etc) (cf Baude et al. 2006).

La pression pour mutualiser le plus possible les données (étant donné le coût que représente la création ou l'enrichissement de corpus) vise à partager les savoirs-faire, tenter d'homogénéiser les procédures, les rendre transparentes afin d'en permettre l'utilisation ou la reproduction par d'autres chercheurs. A l'image de ce qui se faisait déjà au sein du LDC (*Linguistic Data Consortium*), on peut citer de nombreuses initiatives de soutien à des projets (ILF) ou de création de groupes au cours de cette période en France. Ainsi, au début des années 2010, on a observé par exemple une volonté de mutualisation autour de la transcription. L'idée était de rassembler les différentes versions utilisées et de tenter d'aboutir à un consensus sur les différentes étapes et conventions. Cela a donné lieu à des séances de travail au niveau national. Si l'objectif louable était de partager un système commun pour les conventions de transcription, nous avons tous compris assez rapidement que selon les perspectives des uns et des autres, il était relativement délicat d'aboutir à un système unique. Mais ceci n'a pas été inutile pour la communauté, et la création de consortiums tels que l'IRCOM (*Corpus oraux et Multimodaux*) (pour plus de détails voir <http://ircom.huma-num.fr/site/accueil.php>) auxquels j'ai participé ont été très utiles pour fédérer la communauté autour des ressources orales du français et susciter le dialogue.

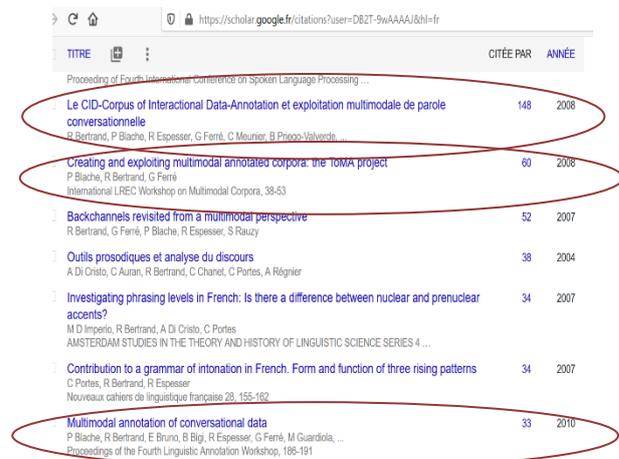
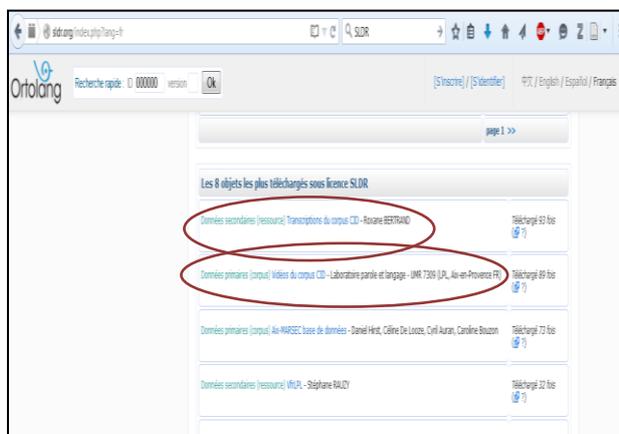
Dans le même sens, j'ai participé également à plusieurs projets financés visant le partage des données orales du français, notamment pour la prosodie (*Rhapsodie : Corpus Prosodique de référence en français parlé*, coordonné par Lacheret 2007-2010).

Enfin, nous avons eu l'occasion de présenter nos travaux dans les conférences et les journaux enfin dédiés à ces questions (tels que la conférence LREC, et sa revue associée LREV), mais aussi dans diverses conférences plus spécialisées telles que *Speech Prosody*, *IcPhS*, *Variations*, *Humor*, *Gesture*, *GespIn*, etc, ou encore lors de séminaires invités, de workshops ou d'écoles d'été.

## 4.2 Le CID : une ressource pour le français

Le CID a été partagé durant les premières années via le CRDO (*Centre de Ressources des Données Orales*) et le SLDR (*Speech & Language Data Repository*) grâce aux différents centres d'archivage et de mutualisation des données orales qui ont été initiés et pour certains pérennisés (Ortolang, [www.ortolang.fr](http://www.ortolang.fr)). Ainsi, les corpus ont pu être exploités pour des recherches très variées par de nombreux chercheurs qui pouvaient selon les cas utiliser tout ou certaines parties des annotations existantes, soit ajouter leurs propres annotations.

Bien que « vieux » d'une quinzaine d'années, le CID représente encore un corpus original dans la mesure où il a été conçu pour en permettre une exploitation large tant par les phonéticiens que les syntacticiens ou encore les gestualistes. En matière d'enrichissement, à savoir les différentes informations ajoutées aux données brutes ou les différentes annotations linguistiques qu'un corpus peut recevoir, il représente une source de richesse très importante, tant au niveau local que plus largement, en témoignent différents indicateurs tels que les nombreux téléchargements via le SLDR (figure ci-dessous), les citations dont il fait l'objet -voir notamment les trois papiers cliqués dans la figure ci-dessous extraite de *Google Scholar*, ou encore son usage dans plusieurs projets ANR auxquels j'ai personnellement participé, parmi lesquels SPIM (NGuyen) Rhapsodie (Lacheret), TYPALOC (Meunier), CoFEE (Prévoit), PhoniACog (Astésano) etc.



## Conclusion

Suite à l'expérience avec le CID qui a permis de montrer tout l'intérêt de disposer de corpus enrichi, documenté et partagé, de nouvelles ressources ont été créées, telles qu'un corpus de 3 dialogues mettant en présence par exemple des sujets aphasiques, ou encore un corpus en mandarin (Chen et al. 2012) sur le modèle précis du CID. L'intérêt de disposer de tels corpus est de permettre la comparaison avec d'autres données. Il est ainsi possible de mener des études cross-linguistiques, ou encore de comparer différents types de parole (genre discursif, parole saine/parole pathologique, etc).

Plus récemment, d'autres corpus sont venus s'ajouter aux corpus de données interactionnelles au LPL (*Maptask*, Gorish & Prévot 2014 ; *corpus DVD*, Prévot et al. 2015 ; *Cheese*, Priego-Valverde et al. 2020 ou *PACO*, Amoyal et al. 2020 entre autres). Ces corpus ont bénéficié de l'expérience acquise sur le CID et ont ainsi reçu diverses améliorations sur différents aspects (conditions d'enregistrements, qualité des films, amélioration des problèmes techniques liés à la synchronisation des données, (pré)-traitement des données).

Nous souhaitons citer également le développement par B. Bigi du logiciel *Spass* (<http://www.sppas.org/>), dont l'un des intérêts est de faciliter les étapes et les procédures successives élaborées au sein d'OTIM en particulier, auxquelles s'ajoutent de nouvelles potentialités pour interroger les données.

Ce travail sur la constitution et l'enrichissement des données interactionnelles et multimodales est un apport sur le plan à la fois des ressources mais également des aspects théoriques, conceptuels et méthodologiques. La prise en compte des phénomènes interactionnels qui incombait traditionnellement à la *Linguistique Interactionnelle* s'effectue désormais également via une analyse plus outillée, issue de la Linguistique de Corpus, et modifie profondément l'approche que l'on avait aux différents niveaux. En effet, cette approche a un impact également au sein de chaque niveau considéré : en prosodie par exemple, lorsqu'on se contentait d'une transcription effectuée par un seul expert, il s'avère désormais difficile de ne pas passer par un contre-codage impliquant donc le recours à plusieurs annotateurs, garantissant ainsi une plus grande fiabilité des données considérées.

L'analyse de plus en plus outillée et automatisée des conversations pourrait sembler contradictoire avec les principes de l'analyse conversationnelle ou de la linguistique interactionnelle. Or, nous avons la certitude que cette approche permettra de révéler de nouveaux faits et de comprendre les phénomènes d'une autre manière. Sur le plan méthodologique, ceci se traduit par deux grands types de méthodologies, l'une dite horizontale (linéaire, syntagmatique, plus classique en linguistique) et l'autre verticale (linguistique de corpus) qui selon nous ne sont nullement mutuellement exclusives. Dans les deux prochains chapitres, nous illustrerons l'intérêt de cette double approche (via nos propres travaux notamment), approche qui s'inscrit dans un mouvement très largement international.

Ce travail en corpus trouve également un écho particulièrement positif dans les approches psycholinguistiques centrées sur les dialogues. De plus en plus de projets sont conçus sur la base de données et de résultats issus de corpus d'interactions à partir desquels les chercheurs vont tenter de construire des dispositifs expérimentaux extrêmement contrôlés sur un phénomène en particulier mais « écologiquement pertinent ». Dans ce cadre apparaissent également de nouvelles techniques d'investigation (très sophistiquées en neuro-linguistiques par exemple, ou des mesures physiologiques sur la respiration, etc) qui permettent parfois de répondre à des questionnements (quelquefois anciens), auxquels il était impossible de répondre en l'état des techniques ou des données disponibles. Ces études en corpus permettent également d'améliorer les interactions humain-machine, en favorisant la comparaison entre dialogue humain-humain et humain-machine.

L'émergence d'un institut tel que le BLRI puis l'ILCB (<https://www.ilcb.fr/>) que l'on doit à P. Blache ou des réseaux tels que COBRA (<https://www.cobra-network.eu/>), que l'on doit à N. Nguyen, illustre cette orientation de la recherche actuelle visant à mettre en commun les connaissances issues des sciences linguistiques et sociales sur les interactions, celles liées aux aspects cognitifs et cérébraux et enfin celles relatives à l'informatique et à l'intelligence artificielle.

## Chapitre II. De la coordination à la convergence

---

### Principales contributions

*Dans ce chapitre, je présente un second axe de mes travaux relatifs à la « convergence interactionnelle ». Tant dans les modèles psycholinguistiques (Clark 1996, Pickering & Garrod 2004) que dans le cadre de la Linguistique Interactionnelle (Couper-Kuhlen & Selting 2018), le postulat selon lequel les partenaires à l'échange n'ont de cesse de collaborer pour se comprendre et s'aligner est commun. La différence essentielle réside dans l'objet d'étude qui pour les premiers concerne les processus cognitifs à l'œuvre dans cet alignement tandis que les seconds s'intéressent aux marques langagières explicites de la collaboration. Nos études s'inscrivent dans ce dernier. Nous examinons les marques explicites de la collaboration que sont les réponses feedbacks des interlocuteurs dans l'activité conversationnelle de narration. Nous empruntons à la Linguistique Interactionnelle ses concepts « d'alignement » et « d'affiliation » ainsi que les principes de « l'analyse séquentielle » qui lui sont propres, tout en combinant à cette analyse une approche plus quantitative inspirée de la linguistique de corpus. La convergence interactionnelle est ainsi définie comme un phénomène ponctuel, local, dans lequel les rôles des participants peuvent parfois s'inverser. Nous montrons l'intérêt d'une étude séquentielle, examinant tour par tour ce qui se déroule en termes de progressivité de l'interaction et de choix de trajectoire interactionnelle pour aboutir ou non à une telle convergence. Cette étude séquentielle, si elle ne permet pas de trancher entre des modèles plaidant plutôt pour un processus d'alignement « négocié » versus « automatique », permet cependant d'affiner nos connaissances. Elle s'avère en outre cruciale pour contribuer à construire de nouveaux dispositifs expérimentaux fondés sur ces nouveaux résultats grâce auxquels cette question pourra potentiellement être tranchée.*

*Ci-après les publications majeures se rapportant à ce chapitre. Nous compléterons par d'autres au fil du chapitre*

### Publications en lien avec le chapitre

---

- Bertrand, R. & Espesser, R.** (2017) "Storyteller and listener's behavior in French conversational storytelling", *Journal of Pragmatics* 111, 33-53.
- Bertrand, R., Priego-Valverde, B.** (2017) "Listing practice in French conversation. From collaborative achievement to interactional convergence", *Discours* (en ligne).
- Bertrand, R. & Goujon, A.** (2017) "(Dis)aligning for improving mutual understanding in talk-in-interaction", *Revue Française de Linguistique Appliquée (RFLA)*, vol XXII-2, 53-70.
- Guardiola, M. & **Bertrand, R.** (2013) "Interactional convergence in conversational storytelling : when reported speech is a cue of alignment and affiliation", *Frontiers in Cognitive Science*, 30 pages.
- Bertrand, R., Ferré, G., Guardiola, M.** (2013) "French face-to-face interaction: Repetition as a multimodal resource", In N. Campbell & M. Rojc (eds) *Verbal Synchrony in Human-Machine Interaction*. USA: Science Publishers, 30 pages.

## Introduction

*How people coordinate is one of the fundamental issues of language use (Clark 1996).*

Une conversation est une production collaborative dont la réussite implique une coordination entre les participants. L'usage du terme coordination renvoie au fait que le dialogue est conçu comme une activité conjointe impliquant la contribution de tous les interlocuteurs. Mais ce terme de coordination ne recouvre pas toujours le même sens. Comme le souligne Pickering et Garrod (2004), on trouve au moins deux conceptions dans la littérature, ce qui a pu conduire à une certaine confusion :

*According to one notion (Clark 1985), interlocutors are coordinated in a successful dialogue just as participants in any successful joint activity are coordinated (e.g., ballroom dancers, lumberjacks using a two-handed saw). According to the other notion, coordination occurs when interlocutors share the same representation at some level (Branigan et al. 2000; Garrod & Anderson 1987). To remove this confusion, we refer to the first notion as coordination and the second as alignment. Specifically, alignment occurs at a particular level when interlocutors have the same representation at that level. Dialogue is a coordinated behavior (just like ballroom dancing). However, the linguistic representations that underlie coordinated dialogue come to be aligned (2004: 4).*

Suite à ces auteurs, nous adoptons cette première distinction qui permet de clarifier les niveaux considérés. Cependant, la notion d'*alignement* telle qu'elle est définie ici renvoie au domaine de la psycholinguistique dont sont précisément issus les auteurs cités. Ainsi, lorsqu'ils évoquent l'alignement, ils cherchent à mettre en lumière les processus cognitifs à l'œuvre pour le réaliser.

En linguistique, et plus particulièrement dans l'approche de la Linguistique Interactionnelle -LI- (Couper-Kuhlen & Selting 1996 ; Couper-Kuhlen & Selting, 2018), l'intérêt porte moins sur les processus que sur les marques et les ressources langagières déployées dans l'organisation structurelle de la conversation. Si le terme d'alignement a parfois été employé dans les études sur les interactions pour rendre compte de la manière dont les participants à une conversation tendaient à s'accorder pour coopérer, s'ajuster et avancer ensemble dans une même direction, le terme n'a été réellement conceptualisé qu'avec Stivers (2008). Selon l'auteure, l'*alignement* renvoie dès lors au niveau structurel de la coopération de l'interaction, terme qu'elle couple à celui d'*affiliation* qui renvoie au niveau affectif de la coopération.

Cette mise au point terminologique s'avère fondamentale pour comprendre les enjeux des études présentées dans la suite de ce document. Car si l'essentiel des travaux dont je souhaite faire état ici s'inscrit dans le cadre de l'approche interactionnelle, ils ont vocation, bien que modestement, à apporter d'éventuels éléments de discussion relatifs au processus d'alignement lui-même.

De SPIM -*Imitation in Speech*- à CoBra -*Conversational Brains*-  
Noël Nguyen (responsable)

Le travail sur la *convergence interactionnelle* a pris naissance dans le projet d'une ANR portée par Noël Nguyen : SPIM (Imitation in Speech): From sensori-motor integration to the dynamics of conversational interaction. ANR BLANC; 1/01/2009-31/12/2011 [http://spim.scicog.fr/index\\_francais.htm](http://spim.scicog.fr/index_francais.htm)

Partenaires : Grenoble Images Parole Signal Automatique (GIPSA), CNRS : UMR 5216, St Martin D'hères (France) Interactions, Corpus, Apprentissages et Représentations (ICAR), CNRS : UMR 5191, École normale supérieure Lettres et Sciences humaines, Lyon (France)

Ce projet est à l'origine de mon intérêt pour les phénomènes dits de convergence et une occasion unique de confronter des perspectives de recherche sur la conversation qui n'aurait pu se faire sans l'esprit éclairé et visionnaire de Noël, que je remercie pour m'avoir offert cette opportunité.

Le projet partait du postulat selon lequel le comportement des interlocuteurs les uns envers les autres au cours d'une conversation, peut évoluer de deux manières différentes : il peut devenir plus semblable à celui de l'autre (ce qui est décrit sous le terme de convergence, ou d'imitation) ou plus dissemblable. Les effets de convergence, coeur de ce projet, ont été décrits comme étant systématiques et récurrents, et comme pouvant se manifester sous différentes formes tant gestuelles que vocales (via la posture par exemple, mais aussi les mouvements de tête ou les expressions faciales, ou encore l'intensité vocale, la courbe mélodique ou le débit de parole). Ces phénomènes facilitent les échanges conversationnels en contribuant aux savoirs partagés, au common ground des interlocuteurs. Ces phénomènes de convergence ont le même effet que les mécanismes d'alignement et permettent d'améliorer la compréhension commune du discours échangé (Garrod & Pickering, 2004).

Notre propre contribution, inscrite dans le courant de la LI, a cherché à nuancer cette vision « tout automatique », sans cependant permettre encore de trancher la question. Mais la diversité des membres de ce projet en a fait un lieu d'échange tout-à-fait fructueux et a permis de faire dialoguer les deux approches dont nous avons fait état au cours de ce chapitre.

Une dizaine d'années après la fin de ce projet, la question de l'alignement est au centre de la recherche non seulement au LPL mais également au niveau international. La période se révèle propice à une prise en compte conjuguée des connaissances sur l'interaction en face-à-face, afin d'en améliorer à la fois sa compréhension et sa modélisation. C'est dans ce contexte qu'a été créé le réseau COBRA *Conversational Brains* (2020-2024), sous l'impulsion de N. Nguyen, Marie Curie Innovative training Networks (European Program) <http://cobra-network.eu/>. Ce réseau, auquel j'ai l'honneur d'appartenir, rassemble des chercheurs de disciplines et d'universités différentes en Europe. Il a vocation à former une génération de chercheurs capables de modéliser les mécanismes linguistiques, cognitifs et cérébraux déployés par les locuteurs humains en conversation, à la fois dans des conversations humain-humain et humain-machine.

## 1. Alignement et Compréhension mutuelle du dialogue en psycholinguistique

En psycholinguistique, les travaux précurseurs sur le dialogue (années 90) ont d'abord cherché à montrer que son étude s'avérait absolument cruciale pour comprendre les processus de compréhension à l'œuvre dans le langage. Jusqu'alors, les études portaient pour l'essentiel sur les monologues. Or ceux-ci ne seraient qu'une réalisation très particulière de la mise en œuvre du langage qui se manifesterait en première instance dans et par le dialogue. Parmi les auteurs les plus influents, on trouve les travaux fondateurs de Garrod & Anderson (1987), Clark & Brennan (1991), Brennan & Clark (1996), Garrod & Pickering (2004), Pickering et Garrod (2004)<sup>5</sup>, entre autres. Premier point fondamental pour le domaine, il s'avère tout à fait possible de proposer des dispositifs expérimentaux pour l'étude du dialogue et d'en contrôler les paramètres. Autre point important, les auteurs soutiennent l'idée que le dialogue ne serait pas aussi complexe qu'imaginé (Garrod & Pickering 2004) sous réserve de le considérer comme une activité conjointe au même titre que d'autres telles que le jeu ou la danse (Clark 1996). Dans cette activité conjointe, les interlocuteurs chercheraient principalement à *s'aligner* (Garrod & Pickering 2004, 2007 ; Pickering & Garrod 2004). Ainsi, comme le

---

<sup>5</sup> Je serai amenée à citer fréquemment les travaux de Pickering & Garrod. Je me permettrais donc parfois de les évoquer sous cette forme : P & G ou G & P

soulignent les auteurs, si dans le monologue les interlocuteurs cherchent à encoder et/ou décoder des messages, en dialogue ils cherchent à aligner leurs états mentaux :

*One aspect of joint action that is important concerns what we call 'alignment'. To come to a common understanding, interlocutors need to align their situation models, which are multi-dimensional representations containing information about space, time, causality, intentionality and currently relevant individuals. The success of conversations depends considerably on the extent to which the interlocutors represent the same elements within their situation models (e.g. they should refer to the same individual when using the same name) ... (Garrod & Pickering 2004 : 8)*

Globalement, il y aurait consensus sur le fait que non seulement le dialogue implique un alignement entre les états mentaux mais aussi que cet alignement conditionne la réussite de l'interaction (Garrod & Pickering 2007). La différence résiderait plutôt dans la manière dont cet alignement se réalise, ce dernier pouvant être le résultat de facteurs stratégiques (intentionnels, délibérés, voir le point 1.1) ou bien de facteurs automatiques (voir le point 1.3).

Dans la section suivante, nous exposons quelques éléments relatifs aux deux modèles les plus célèbres sur ces questions afin de comprendre où se jouent les différences. Pour ce faire, nous nous inspirons des écrits de P & G (2004, 2007) qui en ont produit une explication très claire.

## 1.1 Le "Grounding account"

Comprendre les questions relatives à la compréhension et à l'alignement implique de faire un détour par la notion de *common ground* tel que Clark & Brennan (1991) l'ont défini à la suite de quelques auteurs tels que Stalnaker (1978) entre autres.

### 1.1.1 Processus de Grounding

#### **GROUNDING**

**It takes two people working together to play a duet, shake hands, play chess, waltz, teach, or make love. To succeed, the two of them have to coordinate both the content and process of what they are doing. Alan and Barbara, on the piano, must come to play the same Mozart duet. This is coordination of content. They must also synchronize their entrances and exits, coordinate how loudly to play forte and pianissimo, and otherwise adjust to each other's tempo and dynamics. This is coordination of process. They cannot even begin to coordinate on content without assuming a vast amount of shared information or common ground—that is, mutual knowledge, mutual beliefs, and mutual assumptions (Clark & Carlson, 1982; Clark & Marshall, 1981; Lewis, 1969; Schelling, 1960). And to coordinate on process, they need to update their common ground moment by moment. All collective actions are built on common ground and its accumulation.**

Dans cet extrait, Clark & Brennan (1991: 222) parlent du processus de communication pour renvoyer à l'activité de discuter ensemble. Les auteurs précisent que lorsqu'on parle à quelqu'un, on fait plus que planifier et émettre des énoncés : on se coordonne à différents niveaux et cette coordination n'est possible qu'à condition de partager des informations, connues sous le terme de *common ground*. Ce common ground renvoie aux connaissances et aux croyances mutuelles des participants. Se coordonner passe alors par l'élaboration conjointe de ce common ground et par sa mise à jour constante, à chaque moment de l'interaction : le *processus de grounding*. Techniquement parlant, le grounding est donc le processus collectif par lequel les participants essaient d'atteindre des croyances mutuelles.

### 1.1.2 Les Conceptual Pacts

En 1996, Brennan & Clark proposent la notion de *pactes conceptuels* grâce auxquels les participants vont s'accorder. Cela peut concerner l'usage d'un terme spécifique et sa signification. Car trouver le

bon référent ne suffit pas, il faut partager la même conceptualisation. Ces pactes sont établis sur la base de quelques éléments clés tels que la fréquence d'usage ou le partenaire avec lequel on converse. Il serait hors de propos de détailler ces différents éléments permettant de faire pacte, et nous reportons le lecteur qui souhaite en savoir davantage à cet article.

Ce que nous retenons, c'est que l'usage et le choix d'un terme est ponctué par un accord, souvent tacite, entre les participants. Le pacte conceptuel ainsi établi est donc le résultat d'un travail de coordination qui se déroule comme suit: si le loc A choisit d'utiliser un terme spécifique et que le locuteur B utilise aussi ce terme, cela signifie qu'il est tacitement d'accord avec la référence associée à l'objet en question. Dans ce cas, A et B assument que B a accepté ce terme, et que les deux le connaissent. Par contraste, si B s'interroge sur l'expression de A, alors le pacte ne saurait être formé. Dans ce cas, cet aspect d'alignement est le résultat d'un processus de négociation (grounding) qui est typique du dialogue et implique l'inférence (extrait de Garrod & Pickering 2007, ma traduction).

Tant qu'ils sont en désaccord, les participants doivent négocier jusqu'à trouver le point d'accord qui leur permettra d'aboutir à ce fameux pacte sur la base duquel la conversation pourra se poursuivre. Ce point relatif aux désaccords ou aux incompréhensions, plus globalement aux troubles ponctuels qui surviennent fréquemment en conversation, est crucial car il est l'une des clés pour comprendre les divergences entre les deux modèles évoqués. Nous y reviendrons dans la section suivante.

## 1.2 Le Modèle d'alignement interactif: un processus automatique

Le modèle d'alignement interactif repose sur la même idée que celle à l'origine du modèle de Clark :

*It is much better to understand dialogue as a joint activity, like ballroom dancing or using a two-handed saw (Clark 1996), and to assume that alignment follows from the inherently interactive process (Garrod & Pickering 2007 : 443)*

Le dialogue est une activité conjointe et c'est de ce processus intrinsèquement interactif que découle l'alignement. Les interlocuteurs cherchent à s'aligner non seulement aux niveaux linguistiques mais également au niveau de leurs états mentaux. Mais les auteurs soulignent que l'alignement complet pour les participants au dialogue n'existe pas et qu'il subsiste inmanquablement des ratés et des incompréhensions.

*Full alignment (in which interlocutors have identical mental states) may never occur, but interlocutors attempt to align just as isolated speakers and listeners attempt to encode and decode (Garrod & Pickering 2007 : 443)*

Une fois établi ce point, les auteurs se donnent pour mission essentielle de caractériser les manières dont les interlocuteurs essaient d'atteindre cet état, et de comprendre notamment comment sont traitées ces incompréhensions.

### 1.2.1 implicite/Explicite Common ground

*Just like other complex cognitive processes, alignment in dialogue involves both automatic and strategic components (Garrod & Pickering 2007: 443)*

Si les pactes conceptuels sont négociés explicitement entre les participants pour s'aligner et renvoient donc à une conception largement fondée sur des processus stratégiques, P & G considèrent que l'alignement relève pour une large part de processus automatiques. Cette affirmation repose sur l'existence de deux niveaux de common ground, qu'ils définissent comme un *common ground implicite* ~ *explicite*:

*As a conversation proceeds, interactive alignment predicts that interlocutors build up a body of aligned representations, which we call the 'implicit common ground'. When this is sufficiently extensive, interlocutors do not have to infer each others' state of mind. What this means, crucially, is that people routinely have no need to construct separate representations for themselves and for their interlocutors, or to reason with such representations (2004 :10)*

### 1.2.2 Mécanisme de Priming

Pour P & G, il n'est pas nécessaire de prendre la perspective de l'autre (c'est-à-dire de modéliser ses propres représentations et celles de l'autre) pour converser. L'alignement entre les différents niveaux linguistiques favorisent l'alignement des modèles de situation, par un mécanisme de priming qui permet de faire que si le locuteur produit une forme particulière (au niveau syntaxique par exemple), l'autre va avoir davantage tendance à produire dans son énoncé suivant une forme similaire.

Ces processus de priming opèrent à tous les niveaux :

*global alignment of models seems to result from « local » alignment at the level of the linguistic representations being used ... Via a priming mechanism, whereby encountering an utterance that activates a particular representation makes it more likely that the person will subsequently produce an utterance that uses this representation (2004 : 5)*

Grâce à ce phénomène de priming, les conversations seraient alors moins coûteuses et largement « facilitées » (Garrod & Pickering 2004) pour les interactants puisque cet alignement automatique à différents niveaux ne nécessiterait pas de raisonnement complexe, excepté dans certains cas de réelles incompréhensions que nous abordons dans la section suivante.

### 1.3 Facteurs stratégiques ou automatiques

Le modèle interactif de Pickering & Garrod dépendrait donc d'un alignement des modèles de situation qui ne nécessiterait pas de négociation explicite, ni d'une modélisation et d'une mise à niveau dynamique de chacun des aspects des états mentaux des interlocuteurs, mais d'un processus largement inconscient « d'alignement interactif » (2004 : 9, ma traduction).

Ce paragraphe est une traduction personnelle des écrits de 2004 et 2007. Les auteurs opposent leur vision automatique de l'alignement à celle plus stratégique et négociée de Clark selon laquelle les interlocuteurs construiraient et interprèteraient les contributions de leurs partenaires sur la base d'une modélisation extensive des états mentaux et décisions de leurs partenaires sur la base des contributions qui leur sont les plus utiles. Dans cette approche plus stratégique des pactes conceptuels, les locuteurs construisent leurs énoncés largement pour le bénéfice de leurs interlocuteurs. La plupart des aspects de leur parole sont donc intentionnels. A l'opposé, pour P & G, les interlocuteurs sont largement égocentriques mais opèrent dans un environnement de traitement visant à promouvoir l'alignement. En ce sens, l'alignement interactif promeut un common ground implicite, par le biais notamment de l'imitation, la résonance, l'attention conjointe à travers la co-présence physique. Le traitement du dialogue balance ainsi entre une dépendance sur le common ground implicite et sur le common ground complet. Tant que l'alignement est automatique (cas d'imitation par exemple), s'appuyer sur le common ground implicite est relativement économique. Au contraire, prendre la perspective de l'autre relèverait d'un processus plus stratégique et beaucoup plus coûteux qui ne peut être que rare, et soumis à des circonstances particulières qui vont bien au-delà d'un simple cas d'échec (failure) dans le discours qui, selon les auteurs, ne nécessitent pas non plus, toujours, de recourir aux états mentaux de l'interlocuteur. Nous revenons sur ce point dans la section 1.3.2.

#### 1.3.1 Les marques de promotion du grounding ou de l'alignement : le cas des feedbacks

Pour Clark & Schaefer (1987), la mise à niveau du CG requiert non seulement une phase de présentation mais aussi une phase d'acceptation. Pour qu'une contribution soit mise à jour et/ou acceptée dans le CG, il faut qu'elle soit comprise par l'interlocuteur, lequel doit fournir des preuves positives de cette compréhension (vs négatives en cas d'échec) (voir Clark & Brennan 1991 ; Clark 1996). Parmi les marques positives -comme l'initiation d'un nouveau tour pertinent de la part de l'interlocuteur-, les feedbacks d'acquiescement notamment ont un rôle majeur.

Chez P & G (2007), au contraire, le modèle d'alignement automatique ne requiert pas de preuve positive d'acceptation (p. 448). Loin de nier l'importance des feedbacks, et citant en ce sens les études

de Schober et Clark (1989) ou Bavelas et al. (2000) qui attestent que la compréhension (dans une tâche de description ou de narration respectivement) serait fortement améliorée lorsque les interlocuteurs ont pu produire librement leurs feedbacks, les auteurs considèrent que les feedbacks sont de réelles marques de promotion de l'alignement. Mais ces feedbacks servent à promouvoir le common ground implicite et non le common ground complet (à savoir le modèle explicite de ce qui est disponible conjointement par les interlocuteurs). Ces feedbacks positifs apparaissent quand les interlocuteurs sont alignés, sans que cela soit nécessaire pour l'interlocuteur de représenter le fait qu'ils sont bien alignés. Ils attestent simplement de la croyance selon laquelle l'interlocuteur a fait une mise à jour cohérente de son état mental, et promeuvent donc un alignement automatique.

### *1.3.2 Le cas des repair*

Les auteurs vont plus loin et reconnaissent qu'en dialogue, il est fréquent que surviennent des troubles affectant la compréhension. Selon eux, les interlocuteurs vont alors faire usage de feedbacks spécifiques (*comment ? pardon ?*, etc) ou encore d'énoncés tels que des demandes de clarification (répétition d'une partie de l'énoncé par exemple). Mais de nouveau les auteurs affirment que les participants, bien que non alignés dans ces cas précis, n'auraient nullement besoin de représenter qu'ils ne sont pas alignés. Ces cas illustreraient simplement le fait que l'interlocuteur croit qu'il n'a pas fait une mise à jour cohérente de son état mental. A ce point précis, le locuteur peut donc chercher à clarifier son énoncé en utilisant un repair. Le premier des locuteurs qui essaie de réparer peut employer une simple stratégie pour être plus explicite sur la partie de l'énoncé que l'interlocuteur interroge. Mais bien que ceci requière certaines décisions concernant le fait que la recherche réfère au dernier constituant ou à l'énoncé total par exemple, il n'y a nul besoin encore de modéliser ce que l'interlocuteur croit ne pas savoir. Il suffit d'un feedback correctif tel que remplacer certains mots du locuteur par l'interlocuteur quand il réalise que certains aspects de l'énoncé sont incompatibles avec sa connaissance background, puis le locuteur qui intègre cette correction, et ce sans modélisation nécessaire (p 448). Ces cas « d'échec » ou de « trouble » en discours ne requièrent donc pas le recours aux états mentaux de l'interlocuteur. Loin de devoir faire des inférences « profondes » impliquant un raisonnement complexe et conscient sur le problème considéré, les locuteurs se contentent de réparer. Les auteurs parlent d'un processus de réparation interactionnel qui ne se situe qu'à un niveau « superficiel » du common ground implicite.

### *1.3.3 Au-delà de l'alignement automatique*

P & G ne réfutent pas totalement l'existence de moments où l'alignement automatique ne serait plus suffisant. Mais le seul cas nécessitant une stratégie de négociation est ce qui requiert la recherche de ce que « pense » ou « veut signifier » l'autre, et qui pose explicitement la question suivante : est-ce qu'on attribue bien à ce mot, cette phrase, la même représentation ? Mais ces cas sont loin d'être majoritaires, excepté lorsque l'on veut cacher quelque chose ou tromper quelqu'un.

En vue d'étayer leur hypothèse, P & G évoquent des travaux sur des enfants relativement jeunes (< 7 ans) qui auraient du mal à contourner le processus d'alignement automatique lorsqu'il y a un problème dans la description des cartes qui leur ont été distribuées. Selon les auteurs, il faut en effet être capable de raisonner pour proposer un nouveau schéma, ce qui est le fruit d'une réelle stratégie consciente. Mais ces stratégies, très coûteuses en termes de traitement, peuvent ainsi être parfois au-delà des capacités de certains usagers du langage (facteurs liés à l'âge ou à certaines pathologies entre autres).

Enfin, l'usage de ces stratégies, certes coûteux, peut être particulièrement pertinent lorsqu'on cherche précisément à ne pas s'aligner. L'étude de Danet (1980) fréquemment citée, fait état de ces deux avocats adverses qui refusent d'utiliser le même terme pour une même référence afin de montrer qu'ils ne partagent pas la même conceptualisation. Il s'agit en l'occurrence des termes *fœtus* versus *unborn child*, chacun de ces termes étant choisis par chacun des deux avocats, afin de ne pas donner l'impression à l'autre d'accepter une certaine conceptualisation. On peut imaginer que ce type de stratégie est particulièrement pertinent dans le cadre de séquences argumentatives au sein de débats

par exemple. Les répétitions lexicales peuvent être parfois utilisées de manière similaire pour nier ce qui vient d'être dit par le locuteur, et qui semble attester d'un refus de s'aligner sur le terme pour éviter d'abonder dans le sens de la conceptualisation qu'il revêt ou véhicule.

## 2. Le point de vue de la Linguistique Interactionnelle

Les conversations ont été également le focus du courant de l'*Interactional Linguistics* (IL) (Couper-Kuhlen & Selting 1996 ; 2018). Celle-ci s'inscrit dans le prolongement de la *Conversation Analysis* (CA) dont les auteurs sont issus de divers domaines tels que l'anthropologie, la sociologie ou la psychologie (Sidnell & Stivers 2013). L'approche théorique et méthodologique qui caractérise à la fois la CA et l'IL contraste avec la linguistique du XXe siècle largement fondée sur la langue écrite en ce sens que :

*... some of the most fundamental features of natural language are shaped in accordance with their home environment in co-present interaction, as adaptations to it, or as part of its very warp and weft* (Schegloff 1996: 54; cité par IL Couper-Kuhlen & Selting 2018: 3)

Cette approche considère ainsi que le langage doit être analysé dans son cadre naturel, à savoir dans une interaction co-présente puisque ses caractéristiques les plus fondamentales seraient façonnées en fonction de cette interaction et de leur adaptation à cette dernière.

Le principe de l'IL consiste à appliquer les mêmes méthodes empiriques que celles de la CA dont le but est de révéler la structure et l'organisation de l'interaction quotidienne. L'IL plus particulièrement explore la structure et l'organisation du langage, tel qu'il est utilisé dans l'interaction sociale, afin de contribuer à l'émergence d'une nouvelle théorie du langage. Si la perspective de la Linguistique Interactionnelle poursuit donc l'objectif de la CA en ce qui concerne l'analyse de la structure et de l'organisation de l'interaction sociale, elle considère donc bien *le langage comme son objet privilégié*.

### 2.1 Formes linguistiques et ressources langagières

L'approche de l'IL fait des formes linguistiques déployées par les locuteurs dans les rencontres ordinaires et institutionnelles le focus de son attention. Mais plutôt que d'étudier les formes grammaticales dans le cadre du système linguistique, elle les étudie en tant que *ressources qui façonnent les pratiques utilisées pour concevoir et donner un sens aux tours de parole* (Sidnell & Stivers 2013 : 476).

*La pensée linguistique interactionnelle consiste en la conceptualisation de la structure linguistique vue comme une ressource pour accomplir des actions dans l'interaction sociale et les rendre interprétables* (Selting 1996, 2017, traduction personnelle)

Les études en IL portent sur la manière dont les participants usent des différentes ressources langagières pour accomplir des actions (au sein de séquences), obéir et faire face aux contraintes imposées par les situations de communication diverses dans lesquelles ils se retrouvent quotidiennement, et par la mise en œuvre conjointe, collaborative et coordonnée de ces ressources.

L'objectif de l'IL consiste donc à établir une *grammaire de l'organisation* de la conversation. Il s'agit donc d'étudier comment le langage est déployé  systématiquement  et  méthodiquement  dans les interactions sociales, afin de rendre compte des aspects structurels de la conversation notamment ceux relatifs aux tours de parole (point suivant).

Au sein de la LI, le courant très actif de la *grammaire en interaction* concerne précisément l'émergence de la grammaire, qui est le produit d'un double mouvement de l'interaction et du langage :

*En insistant sur la dimension actionnelle, temporelle et profondément indexicale du langage, l'interrogation majeure de la linguistique interactionnelle peut se formuler ainsi : comment les participants à l'interaction exploitent-ils les ressources grammaticales dans la parole en interaction et, à l'inverse, comment l'organisation interactive contribue t-elle à son tour à*

*façonner les ressources linguistiques mobilisées ? La LI préconise un travail systématique sur les données empiriques analysées dans leur contexte d'occurrence (Pekarek Doehler 2005), remettant en cause les principes d'une grammaire décontextualisée et insistant sur la nécessité d'une modélisation qui tienne compte des activités conversationnelles (De Stefani et Horlacher 2017 : 16)*

Bien que cruciale, nous ne détaillerons pas cette question liée à l'émergence de la grammaire, mais nous reportons le lecteur aux publications sur ce thème (Ochs et al. 1996 ; Auer 2005 ; Pekarek-Doehler, De Stefani & Horlacher 2015, etc), notamment au récent ouvrage de Maschler et al. (2020).

Dans la section suivante, nous présentons succinctement les aspects structurels de la conversation relatifs à l'organisation des tours de parole, qui ont occupé les interactionnistes depuis les origines.

## 2.2 Le système des Tours de parole

Dès l'article princeps (Sacks et al 1974), les conversationnalistes ont mis en lumière certaines règles d'organisation structurelle sous-jacentes à l'interaction et qui permettent sa réussite. Le modèle des tours de parole a ainsi permis de montrer que les conversations, loin d'être aléatoires et hasardeuses, nécessitent le respect de certaines règles et principes de construction et d'alternance de la parole.

Ce modèle de l'organisation des tours de parole cherche à résoudre la question de la manière dont s'opèrent les transitions de parole entre les locuteurs, lesquelles sont majoritairement « douces » (smooth-transitions) : dans la plupart des cas, un tour est produit, un locuteur parle durant un certain temps, puis les tours se succèdent avec un minimum de pauses et de chevauchements.

Cette alternance de parole fluide repose sur l'existence de règles et de principes fondés sur un mécanisme qui s'appuie lui-même sur deux composantes : la composante de construction des tours (construction des unités) et la composante d'allocation des tours (régulation et négociation pour l'alternance).

### 2.2.1 La composante de construction des tours

Le tour de parole est construit à partir d'unité de construction des tours que l'on appelle le *Turn-Constructional Unit* = TCU. Ces TCU ont des formes variées et peuvent être une construction phrastique, une unité lexicale (1 seul mot, « salut », « quoi », etc). Les auteurs préfèrent cette notion à celle de sentence.

Lorsque ces unités de construction de tours sont considérées comme « achevées », elles donnent lieu à des points de complétion possible rendant la transition de tour pertinente, bien que non obligatoire :

*[...] on their (TCUs) possible completion, transition to a next speaker becomes relevant (although not necessarily accomplished) (Schegloff, 1996: 55)*

Ces points de complétion possible sont appelés des *Transition Relevance Place (TRP)*. Entre autres, le design syntaxique du tour en cours s'avère une ressource majeure (et la plus étudiée) pour évaluer si le locuteur est encore légitime à conserver le tour et à prédire où le locuteur atteindra un point de complétion possible rendant ce changement de locuteur potentiel.

### 2.2.2 La composante d'allocation des tours

Il existe des techniques d'allocation des tours qui se distribuent ainsi : la sélection du locuteur suivant par le locuteur courant ou l'auto-sélection par le locuteur du prochain tour.

Les règles d'allocation de tour sont énoncées ci-dessous :

- Règle 1. A la place transitionnelle d'un TCU initial

(a) dans un tour (sélection d'un locuteur par le loc courant), celui qui a été sélectionné a le devoir de prendre le prochain tour. Aucun autre n'a ce droit ou cette obligation, et le changement s'opère en ce lieu

(b) si le tour n'implique pas la technique de sélection d'un prochain locuteur, alors peut s'opérer une auto-sélection des prochains locuteurs. Le premier qui commence est celui qui acquiert les droits au tour, et le changement s'opère en ce lieu

(c) si le tour n'implique pas la technique de sélection d'un prochain locuteur, alors le locuteur courant peut (mais sans y être obligé), continuer jusqu'à ce qu'un autre s'auto-sélectionne et gagne ainsi le droit au tour

- Règle 2. Si, à la *TRP* initiale d'un *TCU* initial, ni 1a ni 1b n'ont opéré, et suivant la disposition de 1c, le locuteur courant a continué, alors la série a – c s'applique à la prochaine *TRP*, et ce de manière récursive à chaque prochaine *TRP*, jusqu'à ce que le transfert s'effectue.

Ces règles simples mais opérationnelles rendent compte d'une variété de pratiques en conversations. Ces pratiques normatives sont donc utilisées pour accomplir de manière ordonnée les changements de tours, et se fondent sur la connaissance tacite des participants. Elles sont par ailleurs considérées comme des descriptions de pratiques dans lesquelles sont engagés les locuteurs et grâce auxquelles ils montrent leur orientation envers un moment particulier, localisé, de parole.

### 2.3 Notion de projection et indices

Sur la seule base des deux composantes énoncées ci-dessus, les locuteurs sont donc capables d'alterner leur tour de la manière la plus harmonieuse qui soit. Ceci est rendu possible par la capacité des locuteurs à prédire/anticiper la fin d'un tour, ainsi qu'à entamer leur prochain tour après la fin du précédent, dans un empan temporel de 200 ms environ, empan relativement similaire dans de nombreuses langues (Stivers et al 2009). Cette capacité est liée à ce que l'on appelle la *projection* (« projectability »), propriété qui permet d'identifier à quel moment le tour sera complet afin de prédire et d'anticiper sur le suivant, et donner lieu à une place possible de transition.

#### 2.3.1 Indices de projection

De nombreuses informations (lexicales, syntaxiques, prosodiques, sémantiques, discursives, gestuelles) sont susceptibles de jouer un rôle dans la projection de la fin d'un tour : les *indices de projection*. Dès 1974, Sacks et al. évoquaient le rôle des indices lexico-syntaxiques dans cette prédiction. D'autres travaux tels que ceux de Duncan (1972) ont aussi montré la présence de certains indices en fin vs medial tour. Les « turn yielding cues » en anglais US notamment seraient réalisés par des courbes intonatives spécifiques, des indices gestuels ou encore des expressions stéréotypées. Un autre travail très célèbre, mené par Ford et Thompson (1996), consiste en l'étude systématique de l'alternance de parole en corpus. Les auteurs montrent qu'un tour est considéré comme complet lorsque les indices syntaxique, prosodique et pragmatique coexistent.

#### 2.3.2 Poids des indices

Au cours de la dernière décennie, des travaux en psycholinguistique ont émergé. Grâce au développement de nouvelles techniques d'investigation, les auteurs peuvent dès lors mesurer directement le rôle et le poids des indices considérés (activité cérébrale par ex EEG), ce qui n'était pas possible en corpus. En effet, montrer qu'un changement de tour est associé préférentiellement à la présence d'un facteur ou d'un autre ne permet pas de trancher sur le poids réel de ce facteur, contrairement aux études menées plus récemment par De Ruiter et al. (2006) ou Bögels et Torreira (2015). Une controverse désormais célèbre oppose ces travaux. Si pour les premiers les indices lexico-syntaxiques sont nécessaires et suffisants pour marquer la complétude d'un tour, les seconds accordent un rôle majeur aux indices prosodiques. Nous n'entrerons pas dans le détail de cette

controverse très intéressante pour les tenants de chacun des niveaux examinés (cf les articles en question) mais hors de propos ici.

Nous tenions simplement à souligner l'intérêt d'un tel rapprochement entre les interactionnistes et les psycholinguistes notamment, que de plus en plus d'auteurs appellent de leurs vœux (Meyer et al. 2018).

## 2.4 Démarche dans la LI

Ce qui précède illustre combien la conversation se déroule selon des principes stricts bien que ni appris ni explicites. Dans la conversation, très tôt apparentée à un ballet ou un orchestre (voir les écrits de l'école de Palo Alto, Goffman, etc), chacun joue sa partie (de sa place et dans son rôle), tout en s'ajustant et en s'adaptant à/aux autre(s). Ce sont les règles, les principes sous-jacents ainsi que les mécanismes mis en œuvre pour comprendre comment la conversation se déroule et s'accomplit avec succès que les interactionnistes tentent de caractériser.

Deux types de démarche peuvent être observés. Nous reproduisons quasi tels-quels ici les propos de De Stefani & Horlacher (2017) relatifs à ces deux types de démarche :

- soit on focalise sur une forme particulière qui sera ensuite explorée dans les séquences qui la caractérisent et les actions qu'elle contribue à effectuer. On trouve dans ce premier cas de nombreux travaux en syntaxe, lesquels ont contribué à alimenter plusieurs études empiriques sur la modélisation de la grammaire en interaction (Hakulinen & Selting 2005 ; Szczepek Reed & Raymond 2013), mais aussi sur la grammaticalisation (Couper-Kuhlen 2011).

Egalement, plusieurs études ont décrit le « formatage des tours et leur déploiement temporel, en traitant de questions de projection, à savoir *ce qui rend les locuteurs aptes à reconnaître la trajectoire d'un tour en cours et ses achèvements potentiels* » (Corminboeuf et Horlacher 2016). Ceci renvoie également aux différents travaux sur les tours collaboratifs, les incréments, les tours inachevés entre autres. Enfin, soulignons que cette démarche est très liée à l'articulation syntaxe-prosodie qui demeure un fort enjeu au sein de la LI.

- soit on focalise sur une action particulière (évaluation, réparation, désaccord, etc) et on étudie les ressources à travers lesquelles elle est réalisée par les participants (avec la difficulté qu'une action est moins identifiable qu'une construction)..., on articule la description des ressources utilisées (prosodie, syntaxe, lexique, multimodalité) avec l'action et que les locuteurs accomplissent dans un environnement séquentiel spécifique (requête, ordres, instructions...)  
...

Ces études ont aussi permis de réfléchir à la notion de « *format (a recurrent linguistic practice or formulation for accomplishing a particular action* » (Fox & Henemann 2016, 500) ou encore à celle de « *pattern (a recurrent interactional practice which has not become sedimented as a grammatical format but is instead a pragmatic routine* (Couper Kuhlen & Thompson 2008, 445). »

Dans mon propre travail, j'ai expérimenté les deux démarches. Par exemple, j'ai examiné les feedbacks en tant que forme particulière apparaissant dans des séquences spécifiques. Nous avons ainsi pu dégager un patron spécifique de feedback (backchannel) réalisé dans des séquences particulières de chevauchement (voir chapitre 1), ou étudier certaines actions qu'ils permettent de réaliser. Sur ce point, nous renvoyons aux travaux suivants sur les feedbacks spécifiques (Discours Rapportés écho, les Other-Repetition et les listes) pour lesquels nous avons également recouru à ces deux démarches. Une différence reste mon approche parfois plus outillée, empruntée à la linguistique de corpus, complémentaire à la démarche de la LI, et propice au développement d'une *Linguistique Interactionnelle Expérimentale*.

### 3. Collaboration et compréhension mutuelle pour un accomplissement interactionnel réussi

Cette section 3 présente ma propre contribution à ces questions. Ma démarche s'inspire des questionnements inhérents aux deux domaines présentés précédemment.

La réussite de l'interaction requiert la compréhension mutuelle. Du point de vue psycholinguistique évoqué précédemment, *l'alignement* faciliterait la compréhension. Du point de vue linguistique, pour les analystes de la conversation, la compréhension n'est pas considérée comme :

*a cognitive, private, individual phenomenon but a collaborative one involving situated, contingent, embodied, and intersubjective dimensions* (Mondada 2011, 542).

La compréhension mutuelle est donc un accomplissement collaboratif réalisé par le déploiement des différentes ressources disponibles aux participants grâce auxquelles ils analysent ce qui est en train de se produire afin d'anticiper la suite via des mécanismes de projection.

*Le fait que la conversation soit à la fois un déploiement dynamique et son aboutissement constitués au fil de leur réalisation interactionnelle par les participants – dans des termes ethnométhodologiques : un accomplissement (achievement) interactionnel – peut être décrit en observant des phénomènes hétérogènes. Parmi eux, Sacks a identifié dans ses cours des années soixante-dix, ce qu'il a appelé des énoncés collaboratifs, des énoncés qui étaient construits par deux locuteurs, où l'un poursuivait ce que le premier avait commencé. Ces phénomènes, qui ont ensuite déclenché un intérêt considérable, permettent d'interroger les processus par lesquels les locuteurs analysent ce qui est en train de se dérouler et rendent immédiatement disponible et exploitable le résultat de cette analyse pour compléter le pattern identifié (Sacks 1992, I : 654). Ils permettent de montrer que non seulement la conversation, mais aussi la réalisation des formes syntaxiques sont des accomplissements interactionnels (Ono & Thompson, 1996). Ces remarques se fondent sur la description de divers processus de projection par lesquels les interlocuteurs sont en mesure d'anticiper la fin de l'unité de construction du tour émergeant du travail réflexif d'énonciation (Mondada, 1999 : 15).*

De la même manière que la compréhension n'est pas interrogée de la même manière en psycholinguistique et en LI, la notion d'alignement revêt également un sens différent. Tout d'abord, notons que le terme est assez peu utilisé en LI alors qu'il s'est généralisé en psycholinguistique.

*When psycholinguistics focuses on cognitive mechanisms involved in the understanding process (Pickering & Garrod (2004) among others), conversational analysis are more often concerned not with understanding in itself but rather in the basic methods through which understanding is established and maintained (Koschmann 2011 : 436, cité dans Bertrand & Goujon 2017 : 54).*

Dans Bertrand & Goujon (2017), et Goujon (2019), thèse que j'ai co-encadrée avec M. Tellier, nous avons ainsi analysé comment les locuteurs se rendent mutuellement reconnaissables l'état de leurs connaissances partagées. Totalement inspirées par les objectifs et les méthodes issues de la LI, à travers une analyse séquentielle, nous avons observé dans plusieurs types de corpus comment, tour par tour, de manière incrémentale, les participants construisent pas à pas l'interaction, en se donnant, via leurs mouvements de sourcils, des marques explicites de compréhension/incompréhension et de (dés)alignement.

Les deux domaines apportent des éléments distincts mais complémentaires qui permettent de mener une réflexion à la fois sur les processus et les mécanismes de la compréhension. Ce point est essentiel pour les perspectives de recherche récentes qui s'ouvrent et qui cherchent à articuler les approches linguistiques, psycholinguistiques et informatiques (cf le réseau Cobra).

Me situant plutôt du côté de la linguistique interactionnelle, je n'ai cependant pas les moyens me permettant de caractériser les processus cognitifs sous-jacents à l'alignement, mais avoir simplement à l'esprit que les différentes pratiques examinées pourraient être parfois le résultat d'un alignement plutôt automatique ou plutôt négocié, permet de s'interroger un peu différemment et d'infléchir éventuellement, parfois, nos interprétations.

De la même manière que certaines questions pourraient être résolues par les méthodes de la psycholinguistique, cette dernière peut également s'enrichir de tels résultats en corpus et favoriser le développement de protocoles expérimentaux de plus en plus sophistiqués tenant compte des différents facteurs influençant l'émergence de telle ou telle forme en discours. Nous reportons par exemple aux travaux sur la dislocation droite (De Stefani & Horlacher 2017), ou la forme « je sais pas » chez Pekarek Doehler (2016, 2019) notamment. Les auteurs montrent que si ces éléments sont pour une part standardisés, ils sont cependant flexibles et mouvants car soumis aux contingences interactionnelles que leurs analyses séquentielles permettent de mettre en lumière.

Dans les sections suivantes, je vais présenter quelques éléments issus notamment de mes propres travaux centrés davantage sur les « réponses » des interlocuteurs (occurrence les plus fréquentes dans le tour 2). Ces travaux visent à mieux comprendre comment tous les participants collaborent pour accomplir de manière optimale leur interaction. Nos résultats sont fondés sur des analyses de corpus systématiques et quantitatives ainsi que sur des analyses séquentielles. L'ensemble des analyses porte sur la manière dont les participants construisent leurs tours et élaborent le sens au fur et à mesure du déroulement du discours, en tenant compte de facteurs aussi cruciaux que l'activité en cours, les rôles discursifs des participants et les attentes générées par ces derniers.

### Publications

---

**Bertrand, R. & Goujon, A. (2017)** " (Dis)aligning for improving mutual understanding in talk-in-interaction", *Revue Française de Linguistique Appliquée (RFLA)*, vol XXII-2, 53-70.

Goujon, A., **Bertrand, R.**, Tellier, M. (2017) "Eyebrow Movements: What Occurs in Sequences of Misunderstanding in a Maptask Dialogue", *IGesto International Conference on Gesture and Multimodality*, 02-03/02/17, University of Porto, Portugal.

Goujon, A., **Bertrand, R.**, Tellier, M (2015) "Eyebrows in French talk-in-interaction? " *Proceedings of Gesture and Speech in Interaction (GESPIN)*, Nantes, September 2-4.

## 3.1 Alignement et Affiliation

Je commencerais par présenter les concepts *d'alignement* et *d'affiliation* qui sont au coeur de mon travail depuis plusieurs années.

Comme précédemment mentionné, peu de travaux en LI utilisent le terme d'alignement, et encore moins à la manière des psycholinguistes pour lesquels il renvoie aux processus cognitifs. On trouve dans quelques études conversationnelles le terme d'alignement mais sans qu'il n'en soit donné une définition claire. Le terme ne sera réellement conceptualisé qu'avec le travail de Stivers (2008):

*[...] we conceptualize alignment as the structural level of cooperation and affiliation as the affective level of cooperation (Stivers, 2008). Thus, aligning responses cooperate by facilitating the proposed activity or sequence; accepting the presuppositions and terms of the proposed action or activity; and matching the formal design preference of the turn. By contrast, affiliative responses cooperate at the level of action and affective stance. Thus, affiliative responses are maximally pro-social when they match the prior speaker's evaluative stance, display empathy and/or cooperate with the preference of the prior action. (Stivers et al. 2011: 21)*

L'alignement renvoie donc à un niveau de coopération structurel en facilitant l'activité ou la séquence proposée, en acceptant les présuppositions et les conditions de l'action ou de l'activité proposée [...] (Stivers et al. 2011, 21).

Cette définition requiert donc une connaissance de l'activité déployée, ce qui nous conforte dans l'idée qu'il est impératif de caractériser les différentes activités au sein d'une conversation pour rendre compte d'une forme d'alignement quelconque.

### 3.2 Conversation et activité

Une conversation est souvent composée de plusieurs « moments » eux-mêmes distincts les uns des autres en termes de but poursuivi ou de rôle discursif occupé par les participants. Ces moments peuvent être caractérisés comme des activités spécifiques telles que la narration, l'argumentation, la recherche d'un itinéraire, la confiance, la description, etc.

La prise en compte de l'activité déployée dans l'interaction en face-à-face est déterminante pour comprendre comment les participants construisent leurs échanges en collaborant et en s'ajustant sans cesse. Il est donc crucial que l'activité dans laquelle ils s'engagent soit « rendue mutuellement reconnaissable ». Il existe des indices d'identification (lexicaux, prosodiques, syntaxiques, etc.) permettant de rendre les activités saillantes. Ceci permet donc non seulement aux uns et aux autres de savoir ce qu'ils « sont en train de faire » mais surtout d'adapter leur comportement aux attentes générées par le déploiement de cette activité. Les modalités et les moyens de cette adaptation mis en œuvre au niveau langagier constituent l'enjeu de nos études. Caractériser très précisément le type d'activité dans lequel les interactants sont engagés permet de mieux comprendre pourquoi tel énoncé est plus ou moins adéquat et adapté à celui qui le précède. Ceci passe par la mise en évidence de trajectoires interactionnelles spécifiques qui rendent compte de ce qui se joue au cœur des conversations, ce qui constitue le cœur de ma démarche.

L'entrée dans une activité particulière requiert l'approbation de tous les protagonistes. Un locuteur désireux de se confier par exemple ne pourra le faire sans avoir préalablement pris les précautions qui s'imposent en s'assurant notamment que le contexte et les circonstances sont favorables, et que son interlocuteur est bien d'accord pour s'engager dans cette activité et dans ce qu'elle implique. Cela peut concerner une prise de pouvoir à travers une prise de parole intensive par un seul des locuteurs (exemple du récit). Par ailleurs, s'engager dans une activité particulière ne peut pas se faire de manière hasardeuse : la confiance notamment implique une relation faite de complicité et de connivence entre les participants (Rabatel 2005).

Enfin, si toute interaction globale vise un objectif particulier, les activités qui la composent peuvent viser elles-mêmes des objectifs plus spécifiques, subordonnés à l'objectif plus général. Dès lors, il faut tenir compte de ces doubles objectifs pour rendre compte des modalités d'ajustement des interlocuteurs. Produire un récit peut permettre d'illustrer un propos ou d'enrichir une argumentation. Ces objectifs seront atteints via des actions spécifiques qui auront des incidences sur le niveau structurel de l'organisation des tours de parole (notamment dans des activités dissymétriques telles que le récit ou la confiance), ou encore au niveau plus global (telles que les phases formelles des récits, voir le modèle de Labov et Waletzky 1966).

### 3.3 L'activité de narration

Parmi différentes activités conversationnelles, la narration, à laquelle j'ai consacré plusieurs études, s'avère être l'une des plus fréquentes en conversation.

Depuis les travaux pionniers de Sacks (1992, Lecture 2) dans le cadre de la CA, plusieurs auteurs ont montré que raconter une histoire est une activité qui peut apparaître dans différentes situations interactionnelles, parmi lesquelles les conversations familiales et ordinaires. Norrick (2007 : 127) va jusqu'à dire que "storytelling is a fundamental mode of everyday conversation" qui remplit des fonctions cruciales incluant "sharing personal news, entertaining listeners, revealing attitudes, constructing identity, inviting counter-disclosure, and so on".

### 3.3.1 *Le récit : une activité de co-narration*

Plus formellement, les auteurs définissent la *storytelling* comme une activité partagée résultant d'un transfert d'information du narrateur à l'auditeur. Cependant, en dépit de sa nature asymétrique, les narrations affichent une co-construction impliquant l'ensemble des participants. Cette notion de co-construction a été beaucoup explorée en CA (Sacks et al. 1974 ; Jefferson 1978 ; Goodwin 1979, 1981, 1984 ; Schegloff, 1982, 1997; Norrick, 2007, 2010, 2012, etc). Par co-construction, nous entendons un « accomplissement interactionnel collaboratif » tel que défini initialement par Schegloff (1992) (voir aussi Rühleman & Gries 2015 pour une revue).

Comment est accomplie cette co-construction et plus spécifiquement comment l'activité de l'auditeur au cours de la narration contribue pleinement à la narration et à son bon déroulement est l'un des thèmes centraux de ma recherche.

L'étude de Bavelas et collègues (2000) est le premier travail expérimental visant à montrer que les réponses requises (par l'interlocuteur) pour faire de l'activité de narration une réussite, sont les réponses dites appropriées, à savoir les réponses attendues en fonction de l'activité en cours.

Pour ma part, ce n'est pas tant l'analyse de la réussite de l'histoire en tant que telle qui m'intéresse mais plutôt l'analyse de celui qui, lors de la narration, est en position d'écoute. Plus particulièrement, nous avons tenté de mettre en évidence le rôle déterminant qu'il joue dans la construction et l'aboutissement du récit, via ses réponses feedbacks principalement, afin de montrer leur rôle de promotion de l'alignement. La question a donc été de définir ce que peut être une réponse appropriée dans le cadre d'un récit, en fonction des objectifs et des rôles discursifs qui lui sont attachés.

### 3.3.2 *Attentes et droits dans l'activité asymétrique de narration*

L'activité de raconter est envisagée comme une activité asymétrique impliquant des rôles discursifs différents de narrateur et narrataire/destinataire. Il n'en demeure pas moins que les deux participants participent activement et travaillent de manière conjointe à construire l'histoire (Jefferson 1978 ; Goodwin 1984 ; Norrick, 2010,2012 ; Bavelas et al, 2000 ; Guardiola & Bertrand 2013 ; Bertrand & Espesser 2017).

Afin de comprendre comment les participants construisent ensemble le récit, il est important de caractériser les attentes qui sont liées à l'activité dans laquelle ils sont engagés. Dans le cadre d'une activité de récit, les participants ont tout d'abord à s'assurer que leur interlocuteur est prêt à entendre leur récit, et qu'il accepte ainsi les règles de fonctionnement qui vont accompagner cette activité. Ainsi, devenir destinataire d'un récit implique d'accepter de laisser pour un temps parfois relativement long la parole au locuteur. Plusieurs unités de tours de parole sont donc nécessaires pour accomplir un récit, ce qui justifie le terme de « large project » proposé par Selting (2000) pour le caractériser. D'autre part, le locuteur doit s'assurer de la « racontabilité » (« tellability ») (Norrick 2007) de son récit. Celui-ci est-il pertinent? Est-il adéquat à ce qui est attendu ? Si le locuteur a planifié un récit doté d'une valence particulière, est-il en mesure de transmettre correctement la valence en question ? Au-delà du contenu, il est important également de déterminer avec pertinence la localisation, en temps et lieu, d'un récit. Comme déjà évoqué dans la section précédente, produire un récit est dépendant de l'objectif de l'interaction en cours et des relations qu'entretiennent les participants. Selon le degré de familiarité de ces derniers et leur histoire conversationnelle, il s'agira d'être pertinent et de présenter alors des événements inattendus ou nouveaux selon les cas. Selon Norrick, cette question de « tellabilité » est cruciale dans la mesure où elle va influencer très largement les frontières du récit : jusqu'où peut-on ou doit-on aller selon les destinataires ? L'ensemble de ces contraintes doit donc être vu et géré par l'ensemble des participants, en tant que construction conjointe. Tant le locuteur que le destinataire doivent donc adopter des comportements appropriés, qui satisfont à certaines contraintes. Cette vision s'oppose à celle du destinataire muet ou invisible ou à celle du locuteur « en attente » qui a prévalu longtemps dans de nombreux modèles. A la différence de ces visions de la

conversation, il est désormais clair que l'auditeur en tant que « story-recipient » est activement impliqué dans le récit et on attend de lui qu'il fournisse les réponses attendues (Jefferson, 1978; Norrick, 2008).

### 3.3.3 Les marques de collaboration des interlocuteurs

#### 3.3.3.1 Notion de préférence

La compréhension mutuelle requiert de fournir des réponses appropriées qui favoriseront le déroulement de l'interaction jusqu'à son terme de la manière la plus réussie possible. Parmi ces réponses, différentes alternatives sont possibles parmi lesquelles certaines sont plus ou moins préférées/non préférées. L'alternative préférée est celle qui correspond aux attentes et aux droits/devoirs des participants. Cette notion de préférence a été largement explorée au niveau séquentiel, particulièrement dans les séquences de paires adjacentes. Par exemple, à une question ou à une offre, les réponses préférées qui constitueront la paire adjacente vont être une réponse ou une acceptation respectivement (les paires adjacentes respectant les principes de préférence et de pertinence conditionnelle : *an answer is conditionally relevant after a question* Schegloff & Sacks 1973). La compréhension et la production de réponse préférée requiert donc de pouvoir prédire, anticiper et les participants doivent sans cesse s'appuyer sur le discours en cours et les indices de projection fournis afin de satisfaire aux attentes en produisant le plus de réponses préférées. Comme nous le montrerons, les réponses préférées ne se limitent pas aux séquences de paires adjacentes. Pour rendre compte de la manière dont peut s'établir ou se rompre la compréhension mutuelle au cours d'une conversation, nous avons exploré les items de feedbacks en tant que réponses susceptibles d'être les réponses préférées par excellence au sein d'une activité asymétrique telle que la narration.

#### 3.3.3.2 Les réponses feedback génériques/spécifiques

Dès 1970, Yngve a distingué un canal parallèle et subordonné par lequel le participant qui ne parle pas produit des messages brefs (tels que mh, oui, etc.) vers le locuteur principal sans l'interrompre. Selon l'auteur, ces messages en « backchannel » jouent un rôle crucial dans la régulation des dialogues. Similairement, des auteurs tels que Bunt (1994) ont caractérisé les énoncés produits par les interlocuteurs comme des items de « feedback », tandis que Bavelas et collègues (2000) parlent de « réponses » produites par celui qui, durant un récit, écoute. Les auteurs montrent que ces réponses peuvent être simples ou complexes, les premières pouvant apparaître dans n'importe quel récit en raison de leur nature *générique*, tandis que les secondes, davantage liées au contenu des récits, sont dites *spécifiques*. Ces réponses génériques / spécifiques sont à mettre en lien avec les *continuers / assessments* de Schegloff (1982). Les réponses génériques renvoient le plus souvent aux items tels que *mh, oui, d'accord*, ou encore les items mimo-gestuels (*hochements de tête*) qui expriment l'attention, l'intérêt et la compréhension du discours en cours. En tant que marques explicites du processus d'écoute et de compréhension des discours, ces réponses permettent au narrateur de traquer des incompréhensions potentielles, et ce quel que soit leur lieu d'apparition dans le récit. Quant aux réponses spécifiques, elles peuvent inclure des *énoncés verbaux relativement brefs*, des *gestes de mains*, des *grimaces*, ou encore *un ton particulier de voix* (Tomlinson & Fox Tree 2011). Ces réponses spécifiques sont fréquemment utilisées pour exprimer des commentaires sur le discours produit. Quelle que soit la modalité dans laquelle elles sont produites, les réponses spécifiques ont une fonction d'évaluation relative au contenu sémantique des récits qui en font donc des réponses plus dépendantes du récit considéré.

### 3.3.4 Contexte d'apparition des réponses

#### 3.3.4.1 Les phases de narration

Les réponses ont été étudiés non seulement dans leurs caractéristiques fonctionnelles mais également dans l'environnement spécifique dans lequel elles apparaissent (Ward 1996 ; Ward & Tsukahara 2000 ; Koiso et al 1998 ; Bertrand et al 2007 ; Gravano & Hirschberg 2010 ; Benus et al. 2011) notamment en

lien avec l'organisation des tours de parole (voir Tolins & Fox Tree 2014 pour une revue). Plus spécifiquement en lien avec notre sujet, il a été montré que ces réponses apparaissent à différents endroits dans la narration. Les réponses génériques apparaissent préférentiellement dans les phases initiales des récits lorsque le common ground est en cours d'élaboration tandis que les réponses spécifiques apparaissent préférentiellement une fois que le common ground est suffisant (Goodwin 1984 ; Bavelas et al. 2000 ; Blondal 2005 ; Stivers 2008 ; Guardiola & Bertrand 2013).

La plupart de ces travaux présentent des résultats de nature qualitative, à l'exception de Bavelas et al. (2000) qui ont établi leur résultat sur la base d'une seule mesure pour chacune des deux moitiés de l'histoire. Dans Bertrand & Espesser (2017), nous avons eu pour ambition de confirmer les résultats de cette étude sur les récits conversationnels du CID qui par nature étaient beaucoup moins contrôlés et aléatoires. Nous avons tenté de caractériser la manière dont les réponses feedback génériques~spécifiques évoluaient au cours du temps. Nous avons interprété l'évolution temporelle des différents feedbacks en lien avec les phases formelles du modèle de Labov et Waletzky (1966) lesquelles, suite à Norrick (207 : 129), peuvent être décrites à travers les questions qu'elles posent:

- *Abstract* : répond à la question « what is about ? »
- *Orientation*; répond à la question “who, what, when, where”?
- *Complication*: consiste en des clauses narratives ordonnées séquentiellement
- *Evaluation* : répond à la question « so what »
- *Resolution* : répond à la question « what finally happened » ?
- *Coda* : repousse toute autre question sur ce qui s'est passé ou pourquoi cela importait

Ces phases sont liées aux différentes étapes relatives au traitement de l'information ainsi qu'à l'élaboration du common ground : en tant que narrateur, le locuteur principal doit rendre disponibles à son interlocuteur plusieurs éléments d'information selon un ordre séquentiel relativement précis.

La phase d'orientation présente les éléments temporels et spatiaux ainsi que les personnages impliqués dans le récit. La phase de complication est relative aux actions ou aux événements qui vont conduire à l'apex, c'est-à-dire au point culminant de l'histoire. Après l'apex, l'évaluation est une phase consistant essentiellement en des commentaires sur le récit. Ces trois phases d'orientation, complication (incluant l'apex) et évaluation peuvent être suffisantes pour rendre l'histoire satisfaisante. Même si toutes les phases ne sont pas requises pour chaque récit, les attentes en termes structurels, sont très hautes. Le respect des différentes étapes permet à l'interlocuteur d'être guidé dans le récit et de fournir ainsi des réponses appropriées. Quelques études relatives à l'activité d'écoute avaient déjà évoqué le fait que les actions des interlocuteurs variaient relativement à la phase d'orientation et la fin du récit, ou bien qu'elles apparaissaient avant ou après l'apex (Goodwin 1984). Norrick (2007) affirmait quant à lui que les phases d'évaluation seraient plus propices à la production de séquences de co-telling (« co-narration ») lesquelles impliqueraient des réponses spécifiques.

### 3.3.4.2 Résultats

Nos résultats (Bertrand & Espesser 2017) sur 147 récits produits par 16 locuteurs en français conversationnel révèlent l'existence de phases de co-narration qui se traduisent par une évolution temporelle des réponses produites par les interlocuteurs. Une tendance significative montre à la fois une augmentation du taux des tours de parole (cf figure 12) et de leur durée, au fur et à mesure que le récit progresse. Ceci montre que les listeners non seulement tendent à parler plus souvent mais aussi plus longtemps au cours du récit. Ces tours plus longs et plus fréquents sont l'occasion pour les interlocuteurs de fournir des réponses spécifiques, qui apparaissent principalement à partir de la moitié du récit, c'est-à-dire proche de l'apex. A l'intérieur des tours de parole, la distinction entre les trois catégories retenues (mot de contenu, mot outil et interjections) permet de retrouver la distinction générique/spécifique puisque l'évolution des mots de contenu augmentent pour les listeners qui ont non seulement des tours plus longs et plus fréquents mais aussi plus riches en termes

sémantique (figure 13). A l'inverse, les interjections, assimilables pour l'essentiel aux réponses génériques, apparaissent de manière significative avant l'apex et marquent ainsi les étapes plus formelles de l'élaboration du common ground. Elles ont tendance à baisser à partir de l'apex. Concernant les narrateurs, les tours de parole restent constants. La richesse des tours de parole reste importante mais on peut observer une augmentation de leurs interjections à partir de l'apex. Ces résultats soutiennent de manière assez convaincante la proposition de Bavelas et al. (2000) selon laquelle le listener devient, fréquemment mais brièvement, co-narrateur, en utilisant des tours qui sont plus fréquents, plus longs et plus riches sémantiquement (voir figure 14). Enfin, les rires montrent également une évolution intéressante puisqu'un pic apparaît autour de 0.58 (la durée des récits étant normalisée entre 0 et 1) pour les listeners, à savoir proche de l'apex. Etant donné le type de données considérées, ce pic apparaît comme le fait que le récit fonctionnerait parfaitement puisqu'à un récit insolite, une réponse spécifique, qui pourrait être aussi la réponse préférée, peut-être la production d'un rire. Le rire fournit donc une preuve de la tellability, et du succès du récit, tels que l'évoque Norrick (2007). Nous pouvons également constater que le locuteur rit plutôt après l'interlocuteur, ce qui est là aussi plutôt attendu : il pourrait paraître saugrenu de rire de sa propre histoire alors qu'il est davantage approprié de rire avec l'autre, ce qui traduit également une forme de co-action. L'ensemble de ces résultats tend à confirmer que l'élaboration d'une activité, aussi asymétrique que la narration, est cependant le résultat d'une élaboration conjointe, de la part de chacun des participants. Nous reportons le lecteur à l'article original dans lequel nous avons également présenté des résultats propres à chaque locuteur.

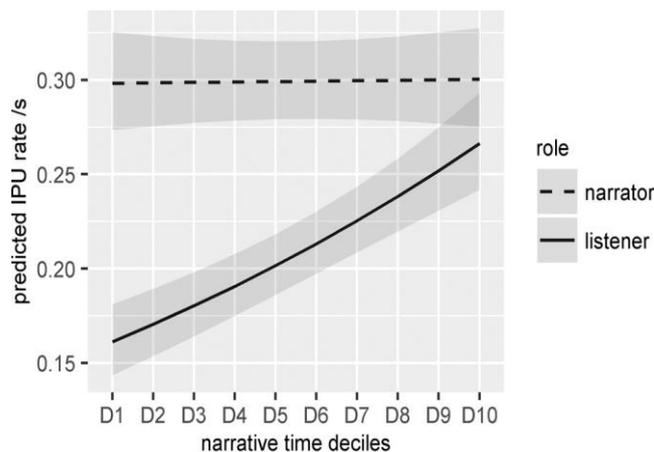


Figure 12 : évolution temporelle des IPU en fonction du rôle discursif (narrateur/interlocuteur), extrait de Bertrand & Espesser (2017)

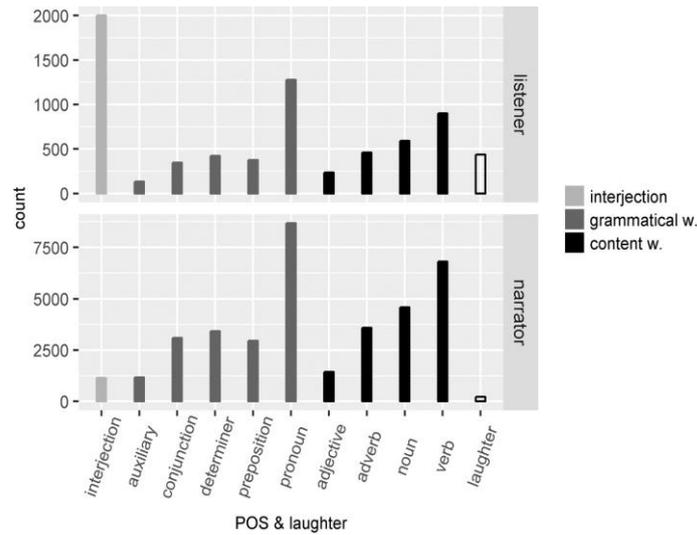


Figure 13 : distribution des catégories morphosyntaxiques (POS) et des rires en fonction du rôle discursif (narrateur/interlocuteur), extrait de Bertrand & Espesser 2017

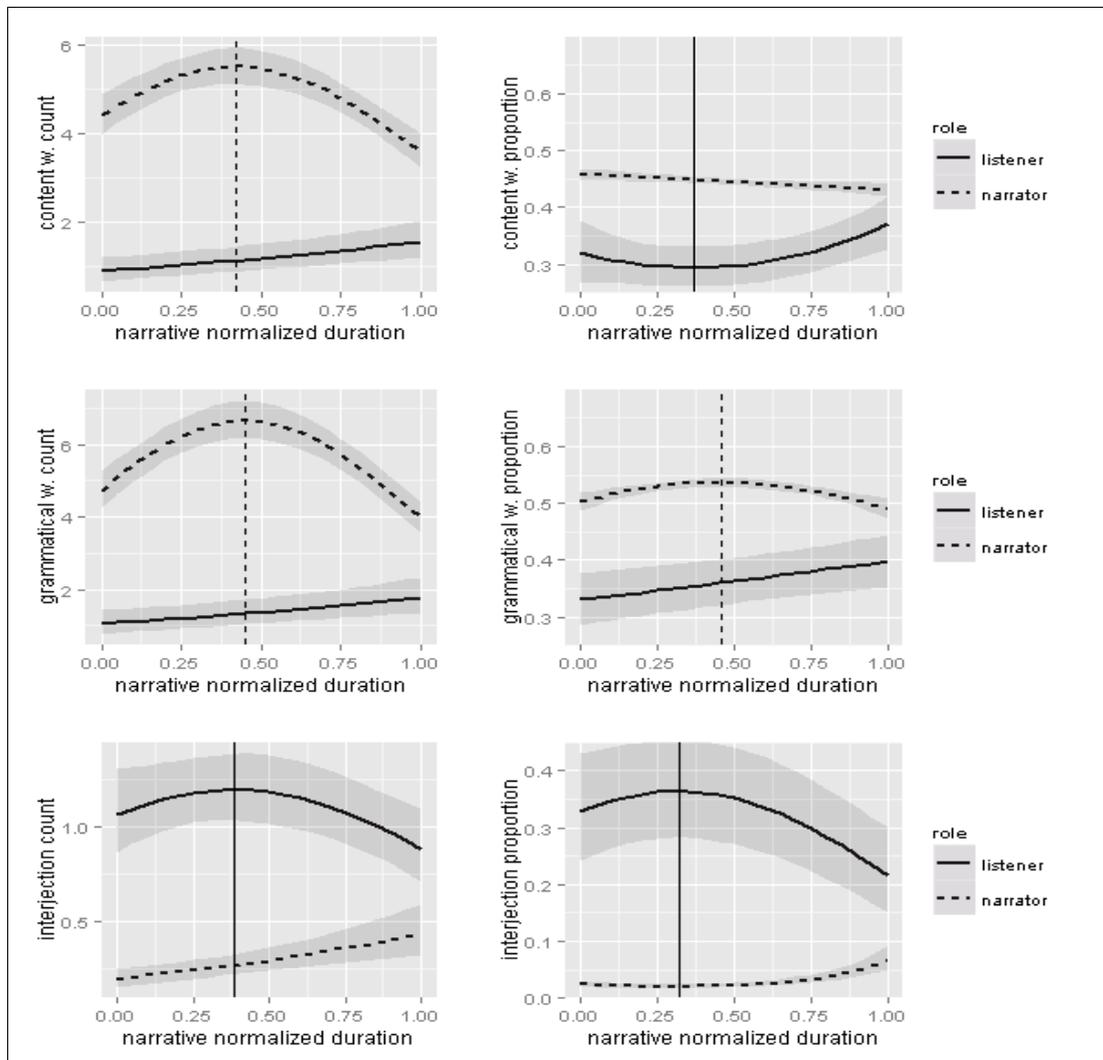


Figure 14 : évolution temporelle des trois catégories retenues (mot de contenu, mot outil et interjections dont feedbacks) pour caractériser la production des narrateurs et de leurs interlocuteurs, extrait de Bertrand & Espesser 2017

Grâce à une procédure quantitative inédite sur des corpus de conversations peu contrôlées, notre contribution a donc permis de confirmer certains résultats relatifs à la construction conjointe des récits par les deux participants. Nous avons pu montrer que l'évolution temporelle des feedbacks était étroitement liée aux phases formelles du modèle Labovien. Ainsi, les réponses génériques apparaissent préférentiellement avant l'apex (donc plutôt dédiées aux phases d'orientation et de complication) tandis que les réponses spécifiques, supposant un common ground suffisamment important, apparaissent après. Le caractère approprié, préféré, des réponses des auditeurs en récit est donc non seulement dépendant de leur nature (générique~spécifique) mais aussi de leur localisation dans la conversation. En l'occurrence, une réponse spécifique ne peut survenir à n'importe quel lieu du récit. Par ailleurs, ce travail donne une vision en parallèle et simultanée de l'évolution de la parole tant chez l'auditeur (via ses feedbacks) que chez le narrateur. Ceci a permis de confirmer que le récit est une réelle construction collaborative de la part de chacun, tout en démontrant une grande systématisme dans cette collaboration malgré la forte variabilité inhérente aux récits observés. Enfin, cette étude constitue une contribution originale à l'étude de l'organisation structurale de l'interaction, thème largement sous étudié en CA selon Robinson (2013).

### Publications

- 
- Bertrand, R. & Espesser, R.** (2017) "Storyteller and listener's behavior in French conversational storytelling", *Journal of Pragmatics* 111, 33-53.
- Guardiola, M., **Bertrand, R.**, Espesser, R., Rauzy, S. (2012) "Listener's responses during storytelling in French Conversation", *Proceedings of Interdisciplinary Workshop on Feedback Behaviors in Dialog* (2012 sept 7-8: Stevenson, UNITED STATES OF AMERICA), 6 pages.

### 3.3.5 Co-narration et convergence

La question qui se pose alors à nous est de savoir comment cette co-narration s'articule à la notion de convergence en interaction.

La convergence est définie comme le rapprochement d'un comportement qui devient de plus en plus similaire au cours du temps. L'examen des résultats précédents peut conduire en partie à une telle interprétation. Cependant, étant donné les différentes activités qui ont cours durant une conversation, lesquelles mettent en œuvre des rôles discursifs variés, il est délicat de déterminer automatiquement et a priori une seule mesure de convergence qui serait basée sur la similarité.

Les résultats de Bertrand & Espesser (2017) tendent à montrer que des lieux spécifiques favoriseraient certaines réponses typiques de l'interlocuteur qui pourraient être interprétées comme convergentes. Déterminer si ces réponses reflètent une réelle convergence a été l'objet d'une étude menée dans le cadre de la thèse de Mathilde Guardiola (2009), que j'ai co-encadrée avec P. Blache.

#### 3.3.5.1 Séquence interactionnelle convergente

Les réponses génériques produites préférentiellement entre le début et l'apex du récit servent plusieurs fonctions qui permettent au narrataire notamment d'accepter de devenir auditeur le temps du récit, puis elles donnent des indices de l'élaboration du common ground. Ces réponses génériques fournissent ainsi des preuves de la compréhension de l'auditeur tout en révélant une forme de « préférence pour la progressivité » (Stivers & Robinson 2006 ; Bertrand & Goujon 2017). Sacks (1992), avait parlé de ces « understanding positions », c'est-à-dire les environnements séquentiels spécifiques au sein de la parole pour « faire compréhension » (doing comprehension). Pour l'auteur, par exemple, la complétion d'une histoire en serait un cas : « *where recipients can exhibit their possible understanding, for example, by using a proverbial expression* » (op.cit., II, 426). Similairement, et grâce au concept d'affiliation de Stivers (2008) relatif au point de vue et de l'expression du stance (mood, emotion, etc), nous avons montré que les réponses spécifiques apparaissent préférentiellement à la fin des récits, quand les interlocuteurs ont suffisamment d'information pour non seulement

comprendre mais surtout produire une réponse plus évaluative. Ces understanding positions ont donc bien à voir avec la localisation et le timing des feedbacks.

Plus particulièrement, dans Guardiola & Bertrand (2013), nous avons interrogé la notion de convergence interactionnelle à travers les deux concepts d’alignement et d’affiliation. Nous avons formulé l’hypothèse que la convergence interactionnelle ne pouvait se réduire à une simple coopération ou un simple alignement établi via la production des réponses génériques. Nous avons montré que la convergence interactionnelle ne pouvait se révéler que sous certaines conditions et en certains endroits, précisément en lieu et place des réponses spécifiques. Nous avons également montré que toutes les réponses spécifiques ne donnaient toutefois pas lieu à une séquence convergente.

### 3.3.5.2 Les discours rapportés écho

Dans Guardiola et Bertrand (2013), nous avons mis en lumière une réponse spécifique que nous avons appelé *discours rapporté écho (DR-écho)*. Une analyse séquentielle a été effectuée sur une centaine de DR-écho du CID.

Les DR échos révèlent un fort degré d’alignement et d’affiliation, dans la mesure où ils s’ajustent parfaitement à l’activité en cours et au « stance » du locuteur. Par ailleurs, ils apparaissent majoritairement en fin de récit, dans la phase d’évaluation. Ils se caractérisent en outre par une sorte de renversement des rôles, l’interlocuteur pouvant devenir, ponctuellement, locuteur principal.

Grâce à ce travail, nous avons montré également que la convergence dans l’interaction ne peut se réduire à la « similarité ». Bien que ces DR échos soient toujours produits dans l’environnement immédiat d’autres discours rapportés (produits par le locuteur), donc « similaires » au discours de l’autre, ces réponses pourtant alignées et affiliées ne donnent pas toujours lieu à des séquences convergentes. Plusieurs études ont montré que les énoncés se forment de manière réflexive en montrant une adéquation permanente à l’énoncé précédent (Hakulinen & Selting, 2005). Les conversations sont ainsi faites d’ajustements mutuels locaux et constamment négociés par les interactants, ce qui est au cœur de l’organisation interactionnelle (Lerner, 1996, Goodwin, 1979, 1981). Au-delà de la réponse alignée et affiliée que constituent les DR-écho, nous avons également analysé si et comment le locuteur s’ajustait à ce DR-écho lui-même, via la trajectoire interactionnelle empruntée à sa suite par le narrateur. Seule une ratification de sa part peut donner lieu à une séquence conversationnelle convergente telle que nous l’avons définie.

L’exemple 1 suivant illustre une complétion collaborative dans laquelle AB produit une première partie de tour qui projette une suite. Cet espace de complétion potentiel est utilisé par l’interlocuteur pour anticiper sur ce qui suit et proposer un énoncé (au nom de AB, en DR-écho) que le locuteur AB peut ou non accepter. En l’occurrence elle l’accepte puisqu’elle le répète explicitement comme argument dans son propre discours.

Cet extrait illustre un cas typique de complétion d’une structure disloquée: prises ensemble, les parties - produites respectivement par le narrateur et l’interlocuteur – forment un tout cohérent, comme nous pouvons le voir grâce à l’anaphore (le pronom “les” correspondant à l’antécédent “ces bonnes soeurs”) et le temps verbal (présent simple).

#### Exemple 1 (extrait de Guardiola et Bertrand 2013)

AB_203	et puis on avait dit Son va faire un truc\$ etc elle avait <u>dit oui</u>
CM_336	<u>ah ouais</u>
AB_203b	oui euh faites un truc de toute façon ces bonnes sœurs euh (0.305)
CM_337	ouais (0.523)
→ CM_338	<b><u>faut les bouger</u></b> @
AB_204	<u>et patin et couffin</u> etc faut les bouger etc puis
CM_339	<u>mh</u>

AB\_204b façon tout le monde va croire que ce sont les carabins parce qu'à  
chaque fois qu'il y a des conneries qui sont faites  
 CM\_340 @  
 CM\_341 ouais + ouais ouais  
 AB\_205 c'est la faute des carabins etc  
 CM\_342 ouais

(voir les Conventions de transcriptions dans Bertrand et al 2008<sup>6</sup>)

AB relate une scène de sa jeunesse où en compagnie de plusieurs copains, ils se sont amusés à tracasser une institution de soeurs, notamment en lançant des feux d'artifice de manière impromptue, en pleine nuit, etc.

Parler de séquence convergente (= hautement interactionnellement accomplie) suppose la prise en compte des 3 contributions suivantes :

- LOC A: contribution 1 (C1)
- LOC B: contribution 2 (C2) (réponse spécifique alignée et affiliée)
- LOC A: contribution 3 (C3) (ratification de c2)

Nous nous sommes particulièrement intéressées à ce troisième élément (présent ou absent) pour mieux comprendre comment certaines réponses a priori alignées et affiliées pouvaient cependant ne pas donner lieu à une séquence convergente. Nous en donnons un aperçu dans le point suivant relatif à l'étude des listes.

## Publications

Guardiola, M. & Bertrand, R. (2013) "Interactional convergence in conversational storytelling : when reported speech is a cue of alignment and affiliation", *Frontiers in Cognitive Science*, 30 pages.

Guardiola, M. & Bertrand, R. (2012) "Discursive convergence in conversation", *International Symposium on Imitation and Convergence in Speech (ISICS)* (1 : 2012 Septembre 3-5 : Aix en Provence).

## 3.4 Les listes

Plutôt que d'activité, on a parlé de « pratique » conversationnelle pour référer aux listes (Bertrand & Priego-Valverde 2017). Similairement aux activités, les pratiques variées qui ont cours en conversation impliquent des rôles discursifs particuliers qui ont une incidence sur l'organisation structurale globale de l'interaction. Avant d'aborder ce point, faisons un point sur leur organisation structurale.

### 3.4.1 Organisation structurale des listes

La liste est une pratique qui requiert des unités multi-tours produites par un locuteur. Cependant, elle comporte des caractéristiques qui en font un candidat idéal pour être accomplie collectivement par les divers participants. En proposant une liste, le locuteur projette explicitement une suite tout en permettant à son interlocuteur d'anticiper cette suite jusqu'à la produire éventuellement lui-même. D'un point de vue prosodique, on sait que cette pratique comporte un aspect quelque peu figé (Selting 2007), stylisé (Portes, Bertrand & Espesser 2007), qui accroît ce potentiel de projection (cf chapitre 3).

En ce sens, les listes nous sont apparues comme des ressources interactionnelles très pertinentes pour les participants qui peuvent en faire usage pour s'aligner, s'affilier et donner potentiellement lieu à une séquence convergente, sous réserve d'être acceptées et ratifiées par le locuteur 1.

<sup>6</sup> Pour les exemples présentés ici: les initiales renvoient à chacun des locuteurs. Les nombres après les initiales renvoient aux IPUS: présence d'une pause silencieuse de plus de 200 ms entre chaque IPU. Les + renvoient à des pauses internes aux IPUS (< 200 ms). Le nombre entre parenthèse correspond à la durée de la pause silencieuse. Le soulignement renvoie aux énoncés chevauchés. @ renvoie à un rire.

Selting (2007) décrit la liste comme une structure composée de trois éléments :

- the *projection component* PC (projecting more to come): projette le tour multiple qui va être produit
- the *list items* : éléments font partie d'un nombre fermé ou ouvert d'éléments de la liste
- the *post-detailing component (PDC)* : complète la structure de liste tout en reliant la liste au sujet ou à l'activité en cours

Notre étude portait sur les conditions d'émergence d'une séquence interactionnelle convergente. Nous nous sommes donc concentrées exclusivement sur les listes construites à deux. Toutefois, la structure de Selting exposée précédemment s'applique également à ces cas de figure. En voici une illustration:

#### Exemple 2 (extrait du CID)

```
YM gpd_273   je suis allé dormir un peu dans m enfin bon par rapport à
              • elle c'est sûr que bon @ mais bon elle elle est occupée
AG gpd_243   ouais ouais non mais *
              -> elle a toi tu es là tu @
AG gpd_244   @ elle a un but ouais
YM gpd_274   ouais elle a euh elle fait quelque chose quoi toi tu es
              => là tu @
AG gpd_245   ouais
```

#### Conventions

- indique le PC
- > indique la liste d'items
- => indique le PDC

Cet extrait relate un épisode relatif à la nuit durant laquelle la compagne de YM a donné naissance à leur enfant. YM raconte que cela a duré très longtemps, ce qui explique le début de l'extrait dans lequel il explique qu'il est allé dormir un peu. Mais il se reprend immédiatement pour minimiser sa propre fatigue par rapport à celle de sa compagne, et propose un énoncé humoristique en guise d'initiateur de liste.

Le premier énoncé (en 273) est donc le composant de projection de Selting, qui porte un contour intonatif de liste typique mis en évidence dans Portes, Bertrand & Espesser (2007). Ce contour en particulier rend reconnaissable à AG le fait qu'YM est en train d'initier une liste. Simultanément, et grâce à ce fort potentiel de projection, AG peut alors saisir l'opportunité de montrer sa collaboration, en proposant une suite à cette liste (list items) qui se présente dans un format syntaxique et prosodique similaire (AG 244). En 274, YM acquiesce (*ouais*) avant de conclure par un troisième item ponctué par un *quoi* conclusif. Relativement à la structure décrite, on constate ici qu'un seul item dans la liste d'items peut être suffisant. On observe par ailleurs que des éléments peuvent être insérés entre les trois composants de la structure sans que cela n'entrave réellement l'effet de liste.

### 3.4.2 Trajectoires interactionnelles

Les analyses séquentielles au sein d'une activité puis l'étude plus systématique des trajectoires interactionnelles qui suivent les réponses sur lesquelles nous portons notre focus, sont l'un des moyens de montrer les stratégies discursives et pragmatiques dans lesquelles sont engagés les participants à une interaction, et de mieux comprendre 'l'envers' des conversations. L'examen de la trajectoire déployée après l'occurrence d'un item de liste produit par l'interlocuteur nous permet de mettre en lumière les conditions d'apparition d'une séquence convergente. Ce travail fait tout naturellement suite à celui relatif aux DR-écho dans lequel nous avons seulement esquissé l'importance de ce troisième élément dans la séquence examinée.

Les items produits par l'interlocuteur de la liste sont considérés comme des réponses spécifiques, similairement aux DR-écho (les deux pouvant d'ailleurs être mêlés parfois). Ces items de liste ont en

effet la particularité d’être alignés et affiliés puisqu’ils s’intègrent parfaitement à l’activité courante (en l’occurrence la liste) mais sont également totalement en adéquation avec le “stance” de l’initiateur de la liste. Dans l’exemple précédent, on remarque que le premier item de liste est produit dans une tonalité humoristique. AG, en collaborant à la structure de liste initiée par YM, produit un item qui s’inscrit dans la même tonalité humoristique (affiliation).

Grâce à l’examen de plusieurs extraits, nous avons montré que la liste construite collaborativement par les deux participants, pouvait donner lieu à des séquences interactionnellement réussies en termes de convergence. Dans ces cas, le locuteur initiateur de la liste accepte, reprend à son compte, évalue, entre autres, l’item proposé par son interlocuteur: ces listes construites à deux sont alors de réels moments de convergence, tels que nous les avons définis dans Guardiola & Bertrand (2013).

Mais ce qui nous intéresse plus particulièrement ici, ce sont les cas moins canoniques que nous avons pu mettre en évidence. Ces cas de listes construites collaborativement, dans lesquelles, malgré des items présentant un alignement et une affiliation certains, ne donnent pas lieu à l’occurrence d’une séquence hautement convergente. Ces cas réfèrent à des extraits dans lesquels l’initiateur de la liste ne ratifie pas l’item de l’interlocuteur. L’analyse séquentielle tenant compte de l’activité, de l’objectif et plus globalement des enjeux de production de cette liste en un lieu précis, permet d’en expliquer les raisons.

L’exemple suivant illustre ce cas.

**Exemple 3** (*Tu fais que tchatcher*) (extrait de l’article Bertrand & Priego-Valverde 2017: 24)

1	MB_316	donc ça c’était je sais plus y a deux ans ou autre
2	MB_317	y a une nana qui est arrivée dans l’école
3	MB_318	et ce que je supportais plus c’était les
4	● MB_319	c’est qu’en fait comme ça se passe à midi
5	MB_320	tout le monde c- d- commence à arriver avec de la bouffe
6	->	et des bouteilles et tu bois tu bouffes et tout et en
7	=> ●	fait de réunion tu fais rien du tout tu fais que tu fais que bouffer
8	-> AC_403	ouais tu fais que dalle tu fais tu tchatches quoi *
9	-> MB_321	dire des conneries tu avances pas
10	=> MB_322	et euh je disais moi si je veux bouffer avec des gens je choisis
11		j’ai je mange avec qui je veux avec mes copains mais pas
12		avec euh tronche machin euh sous prétexte de
13	AC_404	ah ouais

Dans cet exemple, MB relate un conflit professionnel qui l’a opposée à une collègue à propos des déjeuners qui pour certains sont l’occasion de partager un bon moment tandis qu’ils ne sont pas du tout envisagés ainsi par MB, qui explique qu’il n’est pas question pour elle de considérer la pause déjeuner comme un moment de détente durant lequel elle devrait en plus « subir » ses collègues.

Ce qu’il faut souligner, c’est que la liste qu’elle initie est elle-même intégrée dans une activité plus large de narration. Grâce à cette liste, MB présente les personnages et les circonstances de son récit. Il s’agit donc de la phase d’orientation du récit. Plusieurs items de liste se succèdent (ligne 6) puis un autre en 7 sur lequel AC va produire sa réponse sous la forme d’un item de liste de même nature. Cet item semble tout à fait adéquat, aligné et affilié à ce qui précède. Or, MB poursuit son propre énoncé en ignorant tout simplement l’item proposé par AC. Nous avons interprété ce fait de la manière suivante : MB est engagée dans la phase d’orientation de son récit. Dans les lignes 10-12 elle va finalement exprimer sa propre conception d’un repas pris entre collègues de bureau, disant qu’elle estime avoir le droit de choisir avec qui déjeuner. Ainsi, le but des listes qu’elle a produit précédemment ne visait pas à illustrer quelque chose qu’elle aurait déjà dit précédemment par exemple, et qui pourrait dans ce cas être co-élaboré avec son interlocutrice, mais sa liste est produite pour présenter les personnages et les circonstances dans lesquelles les événements de son récit vont prendre place. Elle prépare ainsi le « ground » qui lui permettra d’établir sa propre opinion et de la

justifier par avance. A cet égard, AC ne peut intervenir avant que cela ne soit achevé. A la différence de nombreuses réponses examinées dans nos travaux qui exhibent une nature réellement proactive qui fait progresser le discours, MB, en faisant si peu de cas de la contribution de AC, la rétrograde au statut de contribution parasite. On constate qu'en 321, MB exprime quelque chose de totalement similaire à la contribution de AB, ce qui confirme que ce qui se joue n'est pas lié au contenu mais bien à son rôle discursif et au timing qui n'est décidément pas le bon pour qu'AC intervienne.

Nous reportons le lecteur à l'article original dans lequel nous examinons d'autres exemples illustrant l'importance du timing dans lequel sera produite la réponse de l'interlocuteur, et notamment lorsque l'activité plus large dans laquelle s'intègre la liste implique d'autres objectifs, eux-mêmes se superposant aux objectifs de celle-ci. Si un item de liste arrive prématurément alors que son initiateur n'a pas pu réellement exprimer ce qu'il voulait exprimer (tel que l'enjeu de son récit), les réponses spécifiques sont alors jugées comme illégitimes aux yeux du locuteur principal, eu égard aux règles de participation et de prise de parole. Nos résultats corroborent ceux de Stivers (2008 : 36) selon lesquels une réponse spécifique arrivant prématurément - le récit étant jugé incomplet par son producteur - renvoie à un problème d'interprétation de ce qui s'est réellement passé et devient alors *a posteriori* une réponse désalignée.

L'examen des trajectoires interactionnelles qui se déploient après une réponse considérée comme alignée et affiliée est donc un point important à étudier sur d'autres activités/pratiques conversationnelles, afin de comprendre quels sont les éléments qui jouent un rôle dans la réussite de la progressivité de l'interaction, à savoir son bon déroulement jusqu'à son aboutissement réussi. La prise en compte systématique de cet élément de ratification/absence de ratification par le locuteur après les réponses examinées a été peu analysé. Il semble cependant crucial pour expliquer la promotion des séquences non seulement collaboratives (construites par les différents participants) mais surtout convergentes telles que nous les avons définies dans nos travaux.

### Publication

---

Bertrand, R & Priego-Valverde, B. (2017) " Listing practice in French conversation. From collaborative achievement to interactional convergence", *Discours* (en ligne).

## 3.5 Les hétéro-répétitions

### 3.5.1 Aspects formels

Quand on parle de convergence et d'imitation (SPIM), on peut comprendre aisément la raison pour laquelle les hétéro-répétitions (*other-repetitions*, OR, déjà évoquées au chapitre 1) ont retenu notre attention. Les OR, répétitions de ce que notre interlocuteur a dit, sont une illustration explicite d'une forme de similarité lexicale, au cœur du projet SPIM dans lequel s'inscrivent les travaux pré-cités. En tant que telles, on peut imaginer que ces OR seront des marques de promotion d'un alignement entre les individus. Mais au-delà de cet alignement de forme, nous nous sommes de nouveau interrogées sur leur potentiel à véhiculer une forme de convergence plus profonde, relative à celle que nous avons tenté de mettre en lumière dans nos études sur les DR-écho et les listes. Avant de pouvoir répondre à cette question, nous avons dû passer par différentes étapes, dont l'identification et la mise au point d'un outil automatique pour repérer les OR, exposé dans le chapitre 1 (Bigi et al. 2014). Le développement de cet outil a permis un gain de temps important en termes d'annotation, mais cet outil a également impliqué de mieux définir formellement le phénomène considéré. Enfin, une fois les énoncés repérés et identifiés, une évaluation manuelle (par M. Guardiola et moi-même) a été effectuée pour ne conserver que les occurrences consensuelles.

La collection ainsi obtenue (environ 400 répétitions) s'avère une ressource considérable.

Les OR peuvent être strictes, splittée en 2, 3 par ex (2 mots ou 3 mots répétés mais pas dans le même ordre), réduites, modifiées par rapport à la source. On peut également avoir le nombre de POS qui les composent, la distance entre la source et l'OR.

On observe des OR de longueur variable : elles peuvent être composées d'un seul item ou aller jusqu'à 7 ou 8. La plus longue est composée de 11 items : AB raconte une histoire concernant une vieille dame de sa famille qu'on avait retrouvée le matin toute frigorifiée (mais vivante !) dans son lit, l'explication en étant le fait que la fenêtre était ouverte :

#### Exemple 4 (OR composée de 11 items similaires) (extrait du CID)

CM gpd\_1209 et y a de la neige qui était rentrée dans la chambre et euh  
AB gpd\_938 et y a de la neige qui était rentrée dans la fenêtre oui un peu

### Publications

- 
- Guardiola, M. & **Bertrand**, R. (2015) "Other-repetition as a display of alignment and/or affiliation", 14th International Pragmatics Conference, Antwerp, Belgium, 26-31 July 2015.
- Bigi, B., **Bertrand**, R., Guardiola, M. (2014) "Automatic detection of other-repetition occurrences: application to French conversational Speech", *Proceedings of The 9th edition of the Language Resources and Evaluation Conference* (2014 mai 26-31: Reykjavik, ICELAND).
- Bertrand**, R., Ferré, G., Guardiola, M. (2013) "French face-to-face interaction: Repetition as a multimodal resource", In N. Campbell & M. Rojc (eds) *Verbal Synchrony in Human-Machine Interaction*, USA: Science Publishers, 30 pages.

### 3.5.2 Aspects fonctionnels

Le travail a consisté ensuite à annoter les fonctions des OR. Nous nous sommes fondées sur deux typologies (Perrin 2003 et Tannen 2007).

Selon Tannen (1987) les répétitions réfèrent à 4 niveaux de discours : production, compréhension, connexion, et interaction, qui fournissent une fonction supérieure (de surplomb) dans l'établissement de la cohérence et de l'implication personnelle. Nous ne détaillerons pas ces 4 points, mais nous pouvons souligner le fait que Tannen considère que le niveau production permet au locuteur de produire le langage d'une manière plus efficace et à moindre coût (plus de quantité et plus fluide). La répétition est ainsi vue comme une ressource qui favorise la production de parole (ample et en chevauchement coopératif). L'auteure, dès 1987, affirme que la relative « automaticité » de production des répétitions facilite la production du langage en conversation. Ceci plaide pour la thèse de P & G considérant que de nombreux phénomènes langagiers seraient produits selon un processus automatique. Une analyse plus détaillée des OR en corpus pourrait apporter des éléments en faveur ou pas de cette thèse (voir chapitre 4).

Au niveau interaction, voici ce que Tannen déclare à propos des OR :

*Interaction. The functions of repetition discussed under the headings of production, comprehension, and connection all refer to the creation of meaning in conversation –what might be called, following Bateson 1972, the 'message level' of talk. But repetition also functions on an interactional level –accomplishing social goals, or simply managing the business of conversation. Some functions observed in transcripts ... include the following : getting or keeping the floor, showing listenership, providing backchannel responses, stalling, gearing up to answer or speak, humor and play, savoring and showing appreciation of a good line or a good joke, persuasive effect (what Koch 1983 calls 'presentation of proof'), linking one speaker's ideas to another's, ratifying another's contribution (including another's ratification), and including in an interaction a person who did not hear a previous utterance. In other words, repetition not only ties parts of discourse to other parts, but ties participants to the discourse and to each other, linking individual speakers in a conversation (1987: 583)*

Les fonctions retenues sont empruntées aux deux typologies de Tannen et Perrin (2003). Pour ce dernier, il s'agit de 4 fonctions correspondant aux fonctions des feedbacks :

- *Taking into account* (TIA): fonction de simple prise en compte de la parole d'autrui : il s'agit de répéter les mots de l'autre locuteur pour lui montrer que l'on « accuse réception » de ses propos
- *Confirmation request*: demande de confirmation
- *Positive reply* : souvent après une question, reprise pour montrer son accord
- *Negative reply*: plutôt pour montrer son désaccord

A ces fonctions, nous avons ajouté quelques fonctions extraites de Tannen qui n'a pas réellement proposé de typologie mais plutôt quelques usages observés :

- *Ratifying listening participation* : répétition de l'énoncé de l'interlocuteur par le locuteur principal qui le réintègre dans son discours
- *Bounding episode* : fonction de structuration, sorte de borne dans les discours (clôture)
- *Savoring* : répéter pour montrer son plaisir des mots, de l'énoncé

Et enfin nous avons ajouté la fonction de *Pining* qui est utilisé pour détourner, souvent en l'exagérant, un mot, un énoncé, etc (Guardiola 2014).

L'analyse d'une centaine de cas a montré que les deux fonctions de "ratifying listenership" et "positive reply" semblent étroitement liées à la notion de convergence.

Au stade actuel, nous avons achevé l'annotation des fonctions communicatives des OR sur l'ensemble du CID. Notre objectif est de couvrir l'ensemble des usages du phénomène pour ensuite procéder à l'analyse de chaque fonction et la mettre en relation avec la notion de convergence.

Nous reportons le lecteur au chapitre 4 dans lequel nous détaillerons davantage le travail que nous projetons désormais de mener sur les OR. Ce travail s'inscrit dans la continuité des travaux présentés ici. Il s'agira donc d'étudier si les répétitions sont propices à la production de séquences convergentes. Une analyse séquentielle sera menée en tenant compte de leurs différentes fonctions, et nous les étudierons en lien avec les principaux concepts utilisés dans ce chapitre, à savoir l'alignement, l'affiliation, les activités et les trajectoires interactionnelles. Pour l'heure, nous présentons deux exemples issus de nos répétitions et qui donnent une piste de ce que nous souhaitons faire systématiquement sur les 400 occurrences de répétitions disponibles.

Dans ces deux exemples, les répétitions assument une fonction de pining, c'est-à-dire que l'interlocuteur/trice reprend un ou des mots du locuteur pour les détourner. En l'occurrence, ces deux répétitions apparaissent au début de chaque récit. En ce lieu, il est généralement attendu une approbation sous forme d'un feedback générique de l'interlocuteur qui prend acte que le locuteur va se lancer dans un récit et le valide ainsi dans son rôle. Dans l'esprit de ce que nous avons présenté au cours de ce chapitre, ces répétitions apparaissent donc inappropriées en ce lieu du discours : non alignées et non affiliées. Nous allons en expliquer brièvement les raisons et montrer les implications interactionnelles différentes de l'une et de l'autre sur la relation interlocutive et sur la progressivité de l'interaction.

#### Exemple 5 (extrait du dialogue ML-IM du CID)

ML gpd\_13 si j'ai j'ai fait une fois non deux f- deux étés j'ai fait la secrétaire

IM gpd\_17 oh j'adore cette expression \$j'ai fait la secrétaire ça c'est vraiment ML gpd\_14 j'ai fait la secrétaire le midi hein @

IM gpd\_18 oui donc tu faisais la secrétaire c'est-à-dire @

ML gpd\_15 et c'était absolument atroce

ML gpd\_16 non mais en plus tu sais c'est compliqué pa(r)ce que t- c'était euh un boulot [que, que] mon père m'avait trouvé

IM gpd\_19 mh

ML gpd\_17 donc c'était un boulot euh dans une euh

ML gpd\_18 une entreprise qui s'occupait d(e) la réparation des containers

ML gpd\_19 et j'étais euh je remplaçais la secrétaire euh pendant un mois  
euh pendant ses vacances quoi + ouais  
IM gpd\_20 l'été quoi + ouais

**Exemple 6 (extrait du dialogue EB-SR du CID)**

EB gpd\_62 ouais ça c'était ap ob assez particulier c'était euh c'était quand  
je bossais en Belgique euh  
EB gpd\_63 et euh  
SR gpd\_49 ah oui c'est vrai tu as bossé en Belgique je euh retire tout ce  
que j'ai dit euh @  
EB gpd\_63a @  
EB gpd\_64 @ j'ai une expérience internationale  
SR gpd\_50 @ euh international  
EB gpd\_64a @  
SR gpd\_51 @ même en Belgique ils nous l'achètent @  
EB gpd\_65 une fois  
EB gpd\_65a @  
EB gpd\_66 et euh  
EB gpd\_67 et c'était euh c'est pourquoi on s'était pris la tête

Comme rappelé, ces répétitions en début de récit ne sont pas les réponses attendues qui serviraient l'alignement en ce début de récit. Au contraire, elles désalignent explicitement en pointant sur un mot (secrétaire/Belgique), pour le détourner de son usage initial et pour ouvrir une séquence latérale. Dans le premier exemple, il s'agit d'une séquence de moquerie destinée aux méridionaux et à travers eux à la locutrice principale dont la face n'est pas vraiment préservée. En répétant l'expression cible et en la commentant, elle semble mettre un peu de distance et ne pas prendre trop ombrage de l'agression subie pendant que l'interlocutrice elle-même n'insiste pas davantage. On constate dans la difficulté pour ML de reprendre le tour que cette séquence latérale a été relativement délicate à amorcer. Au contraire, dans le second exemple, on constate que SR rebondit sur le terme pour dévier vers une trajectoire humoristique, que EB accepte pleinement en s'associant à lui pour ajouter à la dérision (voir les rires mutuels). Après un désalignement créé par la répétition de SR, il y a donc un réaligement des deux protagonistes pour converger vers une séquence latérale humoristique. On observe que le retour à la trajectoire du récit pour EB au gp\_66 est là aussi un peu compliqué mais il s'agit de retrouver le « fil ».

Ces deux exemples illustrent combien des phénomènes de même nature, de surcroît sensés refléter un certain mimétisme, une certaine similitude, dont les fonctions sont semblables (de pining en l'occurrence, -fonction qui pourrait favoriser davantage de divergence ?-), et qui sont localisées au même lieu au sein de l'activité courante, peuvent avoir des implications interactionnelles différentes et renvoyer à des formes de convergence/divergence en termes de relation interlocutive, qui nécessitent d'être encore largement explorées pour mieux en mettre à jour les ressorts.

## Conclusion

Les études en psycholinguistique rendent compte de l'alignement en conversation comme d'un processus cognitif complexe tandis que les interactionnistes s'intéressent plutôt aux marques langagières de l'alignement afin de comprendre l'organisation interne de la conversation et la manière dont on fait usage du langage pour organiser et structurer l'interaction. L'étude des marques alimente toutefois les questionnements relatifs à l'étude du processus, puisque les psycholinguistes les utilisent pour confirmer/infirmer leur théorie.

Savoir si les marques et les procédés observés en conversation sont le fruit d'un simple priming (répétition de l'énoncé qui précède, de la forme syntaxique ou l'usage d'un item lexical particulier) ou s'il s'agit d'un processus plus conscient, stratégique et négocié, est une question à laquelle il s'avère délicat de répondre dans le cadre de la LI, mais qui peut l'être également et encore en psycholinguistique. Dans Brennan et Clark (1996), le fait d'utiliser un terme générique (shoe) après avoir dû utiliser un terme plus spécialisé, est expliqué comme le résultat d'un accord tacite par les

auteurs, tandis que Pickering & Garrod (2007 : 145) l'expliquent en termes d'imitation. Même s'il reste encore à faire pour trancher la question, elle me paraît intéressante à garder à l'esprit lorsqu'on tente de comprendre ce qui se déroule en interaction. Notre étude future sur les répétitions, candidats dont on pourrait penser qu'ils seraient plutôt produits de manière automatique, nous apportera sans doute un éclairage intéressant.

Si l'on s'en tient aux études précédentes, nous avons montré comment les différents participants, et plus particulièrement les « listeners » dans le cadre des narrations, collaborent pour accomplir de la manière la plus réussie possible leur interaction. Nous avons vu que les listeners, lorsque le CG est suffisamment important, peuvent prendre la perspective de l'autre, via les marques/procédés mis en lumière. On peut donc penser que ceci est rendu possible par le fait que leurs états mentaux sont alignés : les marques langagières examinées (réponses spécifiques telles que les DR-écho, les items de liste), se présentent donc comme de réelles ressources pour afficher un alignement fort à la fois aux différents niveaux linguistiques (similarités des procédés, imitation via les répétitions lexicales, etc) et au niveau des représentations.

Nous avons montré également que la convergence était un phénomène très ponctuel et très localisé. Des analyses détaillées, telles que les analyses séquentielles de la LI, semblent encore indispensables pour déterminer avec précision si les participants convergent ou non. Selon comment les interlocuteurs orientent leur discours dans un sens ou dans l'autre, selon la trajectoire interactionnelle pour laquelle ils optent, selon qu'elle soit similaire à celle envisagée par l'autre participant, et selon que ce dernier la ratifie ou pas, et de quelle manière, résulte une séquence interactionnelle plus ou moins convergente. Cette convergence repose en partie sur les niveaux d'alignement et d'affiliation qui s'établissent, mais nous avons vu que ces derniers étaient nécessaires mais non suffisants pour donner lieu à une séquence interactionnelle convergente.

Ce qui semble ressortir, de prime abord, c'est que la convergence telle que nous la concevons, semble donc plutôt se rattacher à un processus stratégique, au sens de Clark. Cependant, on pourrait aussi imaginer que le détournement d'un mot, d'une formule pour transiter vers de l'humour, puisse se produire de manière routinisée. Imaginons des personnes très familières les unes des autres, très habituées à fonctionner ensemble, pourrait-on exclure que des procédés ou des pratiques apparemment « stratégiques » soient peu à peu le résultat d'une forme de routinisation, en lien possible avec la nature automatique du processus d'alignement proposé par P & G ?

*There is something of a tension between the two previous sections, with the former emphasizing decisions to mold one's contributions so that they accord with beliefs about one's interlocutor, and the latter emphasizing automatic processes of imitation. We have already suggested one response: that some apparently strategic decisions may in fact reflect automatic resonance processes (Horton and Gerrig 2005)*

Si l'on reprend les propos de P & G à propos des « repair », rappelons que les auteurs considèrent que les cas d'échec ou de trouble en discours ne requièrent pas le recours aux états mentaux de l'interlocuteur. Le processus de réparation suite à une incompréhension par exemple ne requiert pas de faire des inférences « profondes » qui impliqueraient un raisonnement complexe et conscient sur le problème considéré. Pour P & G le processus de réparation interactionnel ne se situe donc qu'à un niveau « superficiel » du common ground implicite. Si l'on suit cette voie, est-il inimaginable que les différentes réponses feedback étudiées (DR-écho, item de liste, OR) soient produites à ce même niveau superficiel (eu égard à la similarité des procédés considérés) ? Grâce aux indices forts permettant à l'interlocuteur de reconnaître ce qui se déroule, celui-ci « n'aurait qu'à » calquer sa production sur l'énoncé pertinent (DR précédent, composant projetant une liste, source pour un élément répété), ce qui donnerait lieu à une réponse préférée (le plus fréquemment) qui, à condition qu'elle soit ratifiée par le locuteur, illustrerait une séquence convergente (alignement des états mentaux). Ces items de liste ou ces DR-échocs seraient ainsi les promoteurs du common ground implicite. L'absence de ratification par le locuteur (plus rare, à la manière des réponses non préférées), serait une éventuelle

preuve de non prise en compte du common ground complet, créant ainsi un désalignement plus profond, nécessitant le recours explicite aux représentations mutuelles, afin de réaligner les deux modèles de situation. Étant donné la rareté de tels cas en corpus, et la difficulté à rendre compte de procédés dans lesquels il n'y a pas de réelles marques explicites (absence de ratification !), il s'avère délicat de conclure sur le caractère automatique ou plus stratégique de cet alignement. On souhaiterait pouvoir étudier ce qui se passe au niveau des processus cognitifs quand l'interlocuteur se trouve confronté à un tel cas. On pourrait faire l'hypothèse qu'il réévalue alors immédiatement le modèle de situation de son partenaire pour ne plus se retrouver dans une telle situation, qui pourrait mettre à mal son image, celle de son partenaire, et nuire à leur interaction, voire leur histoire conversationnelle. Hélas, si tel est le cas, rien ne nous permet de conclure à une telle interprétation même si au cours de l'interaction il ne se retrouve plus dans un tel cas. Une réflexion relative aux dispositifs expérimentaux permettant de tester ce point nous semble importante à mener. Elle nous permettra de mieux apprécier la nature automatique, stratégique ou hybride (à l'image du modèle de Bard et al. 2000) du processus d'alignement.

Ce que nous avons souhaité montrer enfin à travers ce chapitre, c'est qu'une conversation plus spontanée est soumise à une mouvance des rôles, des activités et des objectifs qui s'y déploient. Diverses activités, parmi lesquelles la narration que j'ai étudiée particulièrement, mais d'autres telles que l'explication, l'argumentation, la consultation d'itinéraire, etc, comportent des objectifs divers et peuvent être initiés par l'un ou l'autre des participants, lesquels occupent ainsi des rôles différents, entraînant des différences au niveau de l'organisation structurale de l'interaction. L'analyse de l'accomplissement interactionnel d'une conversation vers sa réussite, ne peut donc s'affranchir d'une analyse plus précise de ces facteurs de variation. L'analyse séquentielle propre à la Linguistique Interactionnelle est particulièrement fructueuse et encore très largement requise en ce sens pour l'analyse des nombreuses activités largement sous-décrites à ce jour. Ces analyses permettront de comprendre les interactions dans toute leur complexité, et de mettre en lumière notamment des lieux particuliers qui vont favoriser une activité collaborative davantage différenciée (collaboration = co-construction = convergence ?) telle que nous cherchons à en rendre compte dans un travail en cours (Bertrand, Nguyen & Priego-Valverde).

Les conversations sont donc un matériau extrêmement précieux qui nécessite d'être encore très largement exploré. Particulièrement, les feedbacks jouent un rôle crucial dans le déroulement de l'interaction. En tant que marques explicites de l'élaboration et de la mise à niveau du common ground, mais aussi grâce à leur nature proactive (Tolins & Fox Tree 2014, Bertrand & Priego-Valverde 2017), les feedbacks selon leur nature, leur localisation, le timing dans lequel ils sont produits, méritent d'être encore largement étudiés étant donné les informations précieuses qu'ils fournissent. Au-delà d'un simple accord ou d'une simple adhésion à un mot, à une idée, ils se révèlent de véritables ressources langagières pour garder le contrôle quand on le souhaite, garder le cap quand on ne veut pas être entraîné sur une piste ou sur une autre, refuser une trajectoire particulière, etc. Sous une apparente collaboration se cache ainsi parfois une réelle volonté de se distancer, de s'ignorer, de se contredire, de conserver une image positive de soi, etc, et c'est toute cette mécanique complexe qu'il faut mettre à jour pour comprendre les raisons pour lesquelles une conversation sera, ou pas, pleinement réussie.

Un autre point sur lequel nous voudrions conclure ce chapitre concerne la question de la similarité, son lien à la convergence et la question que cela pose pour le type de mesures à privilégier. Dans de nombreuses études en phonétique ou en syntaxe, la convergence a été estimée par un indice de rapprochement des comportements des deux locuteurs. La plupart des travaux concernés ont été menés sur des corpus très contrôlés et très homogènes. Il s'agit généralement d'une interaction entre deux locuteurs autour d'une tâche, dans lequel on ne trouve donc qu'une seule activité bien définie (jeu, récit) dans laquelle chaque participant a un rôle pré-établi qui ne varie pas au cours de l'interaction. En lien avec le paragraphe précédent, comment mesurer la convergence par une simple mesure de similarité quand l'un parle et l'autre quasiment pas, et ce alternativement, dans le cas de

récits par exemple ? Il est crucial de montrer que l'éventuel rapprochement entre les locuteurs ne dépend peut-être pas d'une certaine similarité dans leur production, tel que cela a pu être montré dans le cas d'un accent qui influencerait sur l'autre. Mais il s'agit plutôt de la production adéquate d'éléments, en bonne place et dans le bon timing. Sur la base de la littérature existante et de nos résultats, nous savons que les locuteurs occupant des rôles discursifs différents, vont se trouver confrontés à des attentes et des droits différents. Tenir compte de ces paramètres nous semble primordial, y compris pour l'étude des phénomènes de similarité ou d'imitation qui étudiés ainsi, pourraient donner lieu à de nouvelles interprétations (lien entre similarité et (dés)alignement).

D'un point de vue méthodologique, le point précédent plaide pour la recherche d'une mesure qui soit adaptée à cette variation inhérente à la conversation (voir Fuscone 2020 pour une tentative en ce sens). Les mesures de convergence appliquée sur une seule fenêtre d'analyse par exemple ne semblent pas adaptées, du moins ne rendent pas compte de la même chose que ce qui se passe réellement sur le décours d'une interaction.

Notre proposition n'est pas d'exclure une mesure (globale) pour ne conserver qu'une mesure plus locale. Peut-être les deux mesures sont-elles pertinentes : ainsi, nous considérons qu'il y a une forme de convergence globale qui pourrait se définir de manière relativement large (qui se mesure notamment par le choix d'utiliser les mêmes mots, les mêmes formules, pour finalement aligner les modèles de situation qui permettront de se comprendre, voir Pickering & Garrod ou Clark, dont les différences résident notamment dans les mécanismes de convergence qui seraient plus ou moins automatiques) et une convergence plus ponctuelle, plus *fine-grained level*, qui ne pourrait être que locale. Il n'est pas exclu d'imaginer que les deux puissent informer respectivement sur des aspects relativement différents : la mesure d'une forme de convergence globale sur plusieurs interactions entre les mêmes individus, auxquels on aurait proposé des scénari différents ayant une incidence sur le niveau de *common ground* ainsi que sur les différents rôles occupés par l'un et l'autre, aurait-elle une incidence sur la présence plus ou moins fréquente de moments de convergence plus locaux?

## Chapitre III. Prosodie dans l'interaction

---

### Principales contributions

*Dans ce chapitre, je présente un dernier axe de mes travaux relatifs aux aspects prosodiques dans l'interaction conversationnelle. Ces travaux, menés sur des données spontanées du français, visent à rendre compte des fonctions et des implications interactionnelles de divers procédés et dispositifs prosodiques. Nous présenterons plus particulièrement les travaux relatifs aux contours intonatifs et ceux dévolus au phrasé prosodique dans la parole en interaction. Dans le même temps, les modèles liés aux aspects formels, plus ancrés historiquement dans les approches prosodiques, sont appliqués à la parole en interaction. Ce travail a pour objectif de vérifier la permanence des unités établies sur de la parole contrôlée de laboratoire. Si tel est le cas, si les unités existantes résistent à la variabilité des données de parole en interaction, quelles sont les adaptations nécessaires pour les rendre les plus pertinents possibles ? Nous exposons quelques propositions relatives notamment au traitement des disfluences, élément incontournable lorsqu'on aborde la parole spontanée.*

*Ci-après les publications majeures se rapportant à ce chapitre. D'autres seront présentées au cours du chapitre.*

#### Publications en lien avec le chapitre

---

- Astésano, C. & **Bertrand**, R. (2016) "Accentuation et niveaux de constituance en français: enjeux phonologiques et psycholinguistiques", *Langue Française* 191, 11-30.
- Bertrand**, R. & Portes, C. (2012) "Pour une approche phonologique de la prosodie dans l'interaction", *Langue française*, vol. 2012/3, no. 175, 19-36.
- Bertrand**, R., Portes, C. & Sabio F. (2008) "Distribution syntaxique, discursive et interactionnelle des contours intonatifs du français dans un corpus de conversation", *Travaux Neuchâtelois de Linguistique*, 47, 59-77. ISSN 1010-1705.
- Di Cristo, A. ; Auran, C. ; **Bertrand**, R. ; Chanet, C. ; Portes, C., Regnier, A. (2004) "Outils prosodiques et analyse du discours", in A.C. Simon, A. Auchlin et A. Grobet (eds), *Cahiers de Linguistique de Louvain* 30/1-3, Louvain-la-neuve : Peeters, vol. 28, p. 27-84.
- D'Imperio, M., **Bertrand**, R., Di Cristo, A., Portes, C. (2007) "Investigating phrasing levels in French: Is there a difference between nuclear and prenuclear accents? " In J. Camacho, V. Deprez, N. Flores, L. Sanchez, *Selected Papers from the 36th Linguistic Symposium on Romance Languages (LSRL)*. New Brunswick: John Benjamins Publishing Company, 97-110.
- Portes C. & **Bertrand** R. (2011). Permanence et variation des unités prosodiques dans le discours et l'interaction. *Journal of French Language Studies*, 97-110. Internet : <<http://hal.archives-ouvertes.fr/hal-00576851>>
- Portes, C., **Bertrand**, R., Espesser, R. (2007) "Contribution to a grammar of intonation in French. Form and function of three rising patterns", *Nouveaux Cahiers de Linguistique Française*, 2, 155-162.

### Introduction

Mon intérêt pour la prosodie depuis le début de ma recherche concerne son rôle majeur dans le dialogue en face à face, à travers ses diverses fonctions de segmentation de la parole, de structuration et d'organisation des énoncés et des tours de parole, de hiérarchisation et de mise en relief, ou encore d'expression des attitudes et des émotions. L'interaction interindividuelle est le lieu par excellence où chacune de ces fonctions peut s'instancier, de manière plus ou moins conjuguée avec les autres.

Le titre de ce chapitre renvoie au fait que je m'inscris dans une perspective visant à étudier de manière totalement intégrée les deux aspects. Depuis deux décennies, on assiste à un engouement pour l'étude de ces deux domaines qui longtemps ont été étudiés plutôt en parallèle, comme en témoignent Swerts et Hirschberg (1998) dans le numéro spécial intitulé *Prosody AND Interaction*. Quand je parle de prosodie DANS l'interaction, je souhaite donc exprimer le fait que j'ai toujours considéré la prosodie comme un élément crucial et inhérent aux interactions en face à face, qu'il s'avérait impensable de ne

pas prendre en compte. Mais au début des années 2000, le manque d'intérêt de la part de nombreux prosodistes pour l'usage et le sens dans l'interaction, laisse ce champ béant. Il serait cependant faux de laisser penser qu'il n'y a pas de travaux sur les aspects prosodiques de la parole en interaction. Dès le début des années 80 en effet, on recense de nombreuses études qui s'inscrivent dans le courant de la CA, sous l'axe *Phonology for conversation* (Local & French 1983 ; French & Local 1986, Local & Walker 2005, Local et al. 2005 parmi de nombreux autres). Mais ces travaux présentent une approche radicalement opposée -en termes de présupposés théoriques et de méthodologies- aux tenants de la première mouvance qui eux prônent une approche phonologique de la prosodie (voir Pierrehumbert 1980 ; Ladd 1996, 2008 ; Hirst & Di Cristo 1998, parmi d'autres). Pour plus de clarté, je parlerai des conversationnalistes versus les prosodistes pour référer aux uns et aux autres. Si les prosodistes se concentrent donc sur les aspects formels de la prosodie afin de mettre en évidence les primitives phonologiques des langues, les conversationnalistes considèrent la prosodie comme une « ressource » au même titre que d'autres (syntaxe, geste) permettant de comprendre comment l'interaction -au cœur de leur questionnement- fonctionne. Cette approche considère que le moindre *détail* (prosodique, phonétique) peut être pertinent pour l'interaction. Ce postulat selon lequel le moindre détail pourrait avoir une implication interactionnelle telle que délimiter les tours de parole, céder son tour, produire un chevauchement de parole pour interrompre son interlocuteur, etc, est particulièrement intéressante en soi car elle implique la prise en compte de tous les indices, quels qu'ils soient, et sans les hiérarchiser : tout peut faire sens ! Les auteurs vont ainsi s'intéresser à la prosodie de phénomènes très originaux (cf « oh », Local 2007) peu ou jamais décrits jusque-là car inhérents à la conversation spontanée, et les examiner sous toutes les facettes. Outre les aspects prosodiques, la pertinence interactionnelle de paramètres phonétiques est également envisagée. Ogden (2001, 2004) étudie ainsi la qualité vocale (-voix craquée, voix soufflée, voix chuchotée) pour marquer par exemple l'approche imminente de la fin d'un tour de parole. Enfin, dans cette perspective, il est impensable de travailler sur des données non écologiques, et d'appliquer des modèles préalablement établis sur de la parole de laboratoire. Les phénomènes sont donc étudiés d'une manière complètement déconnectée des modèles phonologiques qui se développent pour la prosodie à la même époque. C'est d'ailleurs le regret d'un manque de cadre théorique partagé qui nuirait selon les tenants de l'approche phonologique à l'étude de la prosodie en interaction :

*Conversational analysts (...) have had difficulty dealing with prosodic phenomena due to the lack of a reliable and widely shared framework for annotating suprasegmental information. Without such a framework, prosodic information could neither be captured nor communicated successfully to others, nor could the analysis of prosodic function be made reliably (Swerts et Hirschberg 1998 : 229)*

Il nous paraît cependant important de rappeler qu'un modèle non seulement établi mais surtout consensuel n'est pas entièrement disponible à l'époque, notamment pour certaines langues telles que le français, ce qui ne facilite donc pas l'adhésion au modèle en question. Mais pour les tenants de la *Phonology for Interaction*, il s'agit surtout de déployer une méthodologie intégrant à la fois les fonctions communicatives des détails prosodiques et phonétiques et leur lien à l'organisation interactionnelle. Un item tel que « so » ou « oh » (Local 2007) qui présentent des caractéristiques prosodiques ou phonétiques différentes, assument une fonction différente et ont des implications particulières sur l'organisation interactionnelle qui les conduit à définir une vision spécifique de la phonologie :

*We argue that if phonology is to be truly concerned with function and linguistic contrast we need to induce those functions and domains of contrast from a thorough-going phonetic and sequential analysis of talk-in-interaction (Local & Walker 2005 : 2)*

Face à l'approche conversationnaliste, l'approche phonologique de la prosodie fait donc référence plus particulièrement aux différents modèles (quelles que soient leurs divergences internes qu'il serait hors de propos d'aborder ici), qui considèrent que la prosodie ne peut être caractérisée sans faire référence à ce niveau phonologique abstrait et aux liens étroits qu'entretiennent les entités abstraites et les

paramètres phonétiques. De manière classique, il s'agit donc d'établir un inventaire limité de catégories distinctes (à partir des régularités identifiées dans le continuum acoustique) qui pourront être mises en correspondance avec la description physique des paramètres :

*At a minimum, a complete phonological description includes (a) a level of description in which the sounds of an utterance are characterised in relatively small number of categorically distinct entities –phonemes, features, or the like- and (b) a mapping between such a description and a physical description of the utterance in terms of continuously varying parameters (Ladd 2008: 11)*

Les divergences théoriques entre les deux approches se doublent donc de divergences très fortes sur les questionnements et la méthodologie mise en œuvre dans chacune d'elles. Je n'entrerai pas dans le détail des modèles en prosodie. Mais à partir des années 80, avec la publication de la thèse de Pierrehumbert (1980) à l'origine du développement de l'approche métrique auto-segmentale (AM), devenue le standard depuis, le fossé se creuse encore. L'objectif de l'AM consiste à établir un modèle phonologique de la prosodie, d'abord pour la langue anglaise, qui se généralisera ensuite à un grand nombre de langues. L'idée est de disposer d'un arsenal formel et conceptuel partagé permettant d'avoir un vocabulaire commun favorisant les comparaisons interlangues. Si le français a longtemps résisté pour des raisons qui seraient là encore hors de propos de détailler ici, plusieurs modèles du français dont celui élaboré à Aix, s'inscrivent dans ce type de démarche. Nous reportons le lecteur aux articles de Di Cristo & Hirst 2002, Di Cristo 2004, pour une revue exhaustive des modèles du français ainsi qu'à la thèse de Welby (2003) qui après avoir exposé les différents modèles pour le français, apporte des arguments en faveur de l'approche AM. Les travaux de Jun et Fougeron (2000, 2002) sont les premiers menés sur le français dans cette mouvance. L'arrivée au LPL au début des années 2000 de chercheurs ancrés dans l'AM comme M. D'Imperio, P. Welby, puis plus récemment C. Petrone, A. Michelas, J. German ainsi que la proposition récente de nombreux auteurs en prosodie du français au niveau national, ont conforté le modèle AM comme un standard (cf Delais-Roussarie et al. 2015 pour le développement d'un « Tobi French »).

Mon propre travail s'est inscrit dans ce contexte relativement complexe puisque j'ai effectué ma thèse de doctorat (1999) au LPL (anciennement *Institut de Phonétique*) qui est historiquement établi dans une tradition d'élaboration des modèles (ancrée phonologie), tout en cherchant à rendre compte des aspects prosodiques en conversation. Or, l'Institut de Phonétique à l'époque est encore relativement peu versé vers les aspects de l'usage. Cependant, durant les années qui vont suivre, des initiatives très novatrices telles que la création d'un groupe sur les liens entre prosodie et discours avec A. Di Cristo, vont émerger. L'un des points particulièrement importants pour aborder cette problématique est la spécification des niveaux d'analyse. Pour le français, Di Cristo représente l'un des auteurs les plus attachés à décrire très précisément ces questions et regrette que la spécification des niveaux d'analyse ne soit pas toujours suffisamment explicite. En raison de la nature complexe de la prosodie qui peut être analysée des points de vue tant acoustique, que perceptif ou phonologique, il est en effet crucial de savoir à quel niveau on se situe et de caractériser ces derniers très précisément (voir Di Cristo 2013, 2016, et plus particulièrement le premier dans lequel il fait un point très précis sur ces questions).

La démarche dans laquelle je m'inscris consiste en l'étude de la prosodie et des liens qu'elle entretient avec un ou plusieurs autres domaines. L'objectif est de les analyser sans jamais les hiérarchiser et d'être très explicite en ce qui concerne précisément les niveaux d'analyse évoqués ci-dessus, quels qu'ils soient. Ma position, intermédiaire entre les deux propositions, a toujours été teintée de ce souci de spécification hérité de Di Cristo. Mon cheminement reflète également cette position intermédiaire dans la mesure où je n'ai pas eu de position absolument tranchée entre les deux, mais j'ai plutôt tenté d'avancer en tenant compte de l'une et de l'autre, notamment en essayant d'être toujours très explicite sur les niveaux d'analyse et de représentation auxquels je me situais (que je m'intéresse aux aspects acoustiques ou aux aspects plutôt phonologiques). Dans ce contexte, le lien avec d'autres niveaux de l'analyse linguistique, excepté la syntaxe qui constitue l'un des domaines les plus envisagés

en lien avec la prosodie, pouvait sembler prématuré. Ajouter en effet à toute cette complexité celle de la parole spontanée, dont on connaissait si peu les ressorts, était un vrai défi. Mais malgré les nombreux écueils rencontrés, j'ai tenté de mener des études, modestes, mais qui ont permis selon moi d'ouvrir des perspectives intéressantes tout en contribuant à l'essor des études prosodiques sur les dimensions discursive et interactionnelle des conversations au LPL.

Bien que cela se soit souvent révélé difficile, je ne pouvais me résoudre à étudier la prosodie comme un élément secondaire, « pertinent à ses heures », dans le cadre et le contexte du LPL dont la réputation sur le plan prosodique n'était plus à faire. Simultanément, les modèles prosodiques existants pour le français ont été revisités et adaptés à un ordre nouveau international sur lequel certains ont pu avoir des réticences et des doutes, mais qui permet toutefois de combler le manque de travaux sur le français au niveau international et de pouvoir dialoguer désormais plus facilement avec la communauté (essor du modèle AM). Pour ma part, je ne suis donc ni du côté de ceux qui « font » les modèles ni du côté de ceux qui les réfutent. Je pense qu'une voie entre les deux est possible mais qu'elle requiert d'être à l'écoute des deux plans pour en permettre éventuellement la rencontre. C'est ce dont témoignent certains de mes travaux présentés ci-après, mais également des travaux très récents visant à réconcilier ces approches qui se sont si longtemps ignorées.

Une dernière difficulté que je voudrais évoquer pour l'analyse de la prosodie dans la parole spontanée en interaction concerne le point déjà largement évoqué dans le chapitre 1, à savoir le manque de données<sup>7</sup>. Au cours des deux dernières décennies, nous avons déjà évoqué l'intérêt grandissant pour des corpus de données orales spontanées en conversation qui s'est manifesté à tous les niveaux de l'analyse linguistique, la prosodie n'y échappant pas. Le mouvement de constitution et de mutualisation des données documenté dans le chapitre 1 de cette synthèse s'étend ainsi aux prosodistes eux-mêmes qui se sont emparés des corpus disponibles pour tenter de valider leur théorie, (Anderson et al. 1991 pour *The HCRC map task*, Pitt et al 2005 pour *The Buckeye Corpus*, Dubois et al. 2000-2005 pour le *Santa Barbara Corpus of Spoken American*, etc). C'est le sens de ma participation au projet ANR Rhapsodie, coordonné par Anne Lacheret ([http://ircom.humanum.fr/site/description\\_projet.php?projet=rhapsodie](http://ircom.humanum.fr/site/description_projet.php?projet=rhapsodie)) qui consistait précisément à collecter et rassembler plusieurs extraits de parole spontanée afin de constituer une base de données prosodiques de référence pour le français. Nous avons été sollicités notamment dans ce projet pour tout le travail de réflexion et d'élaboration méthodologique déployé autour du CID. Cet engouement pour les données spontanées s'est également traduit par le développement de diverses bases de données pour le français. Nous renvoyons également le lecteur intéressé à la base de données VALIBEL et ses récentes évolutions (dont le corpus LOCAS pour l'étude de la prosodie et de la syntaxe) (Bolly et al. 2016 ; Degand & Simon 2009) ou encore au projet ORFEO (Benzitoun et al. 2016).

Dans les sections suivantes, je présenterai quelques étapes de mon cheminement prosodique dans le contexte décrit dans les lignes précédentes.

## 1. Interface Prosodie-énonciation

J'ai choisi de présenter ce volet interface prosodie-énonciation à part de celui qui concerne l'interface prosodie-discours. Ceci ne signifie pas que l'énonciation serait vue comme extérieure au discours, mais il s'agit plutôt d'une différence de démarche à la fois conceptuelle et méthodologique entre l'étude que je développerai ci-après, qui vise à montrer le rôle énonciatif de la prosodie, et les études ultérieures rassemblées dans l'interface prosodie-discours. Dans l'étude présentée ci-après, nous avons travaillé sur le niveau phonético-acoustique de la parole, contrairement aux travaux ultérieurs

---

<sup>7</sup> D'un côté les corpus écologiques des interactionnistes qui ne présentaient pas toujours une qualité audio suffisante ou la possibilité d'étudier les liens entre les différents niveaux, de l'autre les corpus des prosodistes comportant des phrases isolées, décontextualisées et produites par un locuteur unique

dans lesquels nous tenterons de contribuer à l'analyse des différentes interfaces, et parmi elles l'interface phonétique-phonologie.

### 1.1 Polyphonie et dialogisme

Autour des années 2000, dans la poursuite de mes travaux de thèse de doctorat, je me suis intéressée aux théories énonciatives nouvellement en vogue pour appréhender l'interaction interindividuelle.

Cette orientation théorique trouve son origine chez des auteurs extérieurs à la linguistique, pour lesquels le sujet ne se constitue en tant que tel que dans le dialogue qu'il noue avec l'autre. Cette hypothèse contribue à remettre en cause l'unicité du sujet parlant qui a longtemps prédominé dans le champ des sciences humaines. Dès lors qu'on conteste l'idée de 'l'unicité du sujet de conscience', on peut avancer l'idée selon laquelle les productions langagières du sujet parlant soient soumises à d'autres influences. Le sujet n'est plus conçu alors comme étant l'unique dépositaire de sa parole (Bertrand 2003).

A la différence de l'analyse conversationnelle plus centrée sur les aspects structuraux et organisationnels de la parole en interaction, la théorie énonciative vise à étudier davantage le sujet parlant en tant que tel (Bakhtine, Authier-Revuz, entre autres). L'approche énonciative est basée sur les aspects polyphoniques des discours. L'idée principale réside dans le fait que les sujets sont traversés de voix, hypothèse selon laquelle les sujets convoquent dans leur discours une pluralité, une hétérogénéité de voix qu'ils font dialoguer pour construire du sens mais aussi pour se construire en tant que sujet parlant. Notre objectif vise ainsi à valider la théorie énonciative selon laquelle un locuteur produit non seulement un discours qui est traversé par plusieurs « voix » (au sens énonciatif du terme), c'est à dire des "points de vue", mais que ces différentes voix se mêlent et se combinent à celles des différents participants à l'interaction pour produire du sens.

#### 1.1.1 Espace interactif et rapports de places

Suite à Vion (1995), l'interaction est un espace interactif dans lequel les participants construisent le sens et s'adaptent les uns aux autres à partir de rapports de places qui définissent à la fois leur relation sociale et leur relation interlocutive (figure 15).

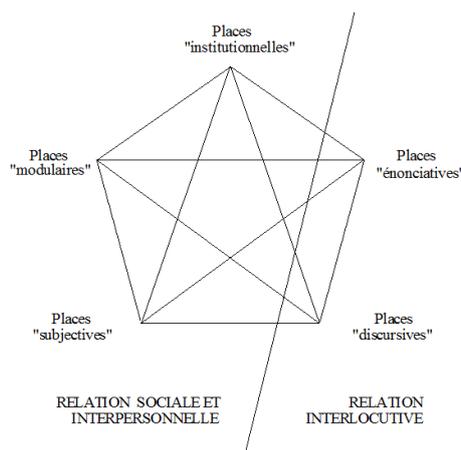


Figure 15 : Schéma des places, extrait de (Vion 1995), dans Bertrand & Espesser (1998)

Dans ce schéma des places, l'auteur rend compte des divers positionnements adoptés par les interactants au cours de leurs échanges. L'intérêt de ce schéma est de montrer que les places ne se limitent pas aux seuls aspects sociaux : les places institutionnelles (statutaire ou professionnelle) définissent les rôles *a priori*, tels que ceux de médecin / patient par exemple. Mais l'interaction dépend également d'autres rapports de places qui vont s'instaurer au fur et à mesure. Les places modulaires

renvoient aux sous-types qui seront mobilisés dans l'interaction plus large : dans la consultation médicale, on peut trouver d'autres séquences que celles consacrées au diagnostic (prise de nouvelle d'un tiers par exemple). Les places subjectives renvoient aux images que les participants transmettent d'eux-mêmes notamment via la manière dont ils se nomment ou sont nommés. On cite fréquemment le cas des débats avec l'exemple célèbre de Chirac et Mitterrand dans lequel ce dernier nie le rapport de place symétrique que tente d'imposer le premier en l'appelant « Monsieur le premier ministre » face à lui qui est président de la République. Mais comprendre l'interaction requiert de tenir compte aussi des rapports de places établis dans la relation interlocutive. Les places discursives et énonciatives jouent ainsi un rôle fondamental. Les places discursives renvoient notamment au type d'activité dans lequel sont engagés les participants à l'interaction (argumentation, narration, etc). C'est celles auxquelles nous avons ultérieurement consacré des études (voir chapitre 2).

Dans cette section, nous rapportons le travail sur les places énonciatives qui sont le niveau le plus fin de positionnement des individus. Elles concernent le mode d'implication des sujets dans leur parole. Le postulat selon lequel le locuteur ne serait pas l'unique dépositaire de sa parole est au cœur de la problématique. Cette parole est traversée de voix que le locuteur fait dialoguer entre elles. A partir de ces places, les locuteurs vont « mettre en scène » leur discours à travers divers énonciateurs invoqués. Plusieurs « mises en scène énonciatives » sont possibles.

### 1.1.2 Discours rapportés directs

Nos travaux portent sur l'étude d'un procédé langagier particulier permettant de mettre en évidence cette diversité de voix au sein de la parole : les discours rapportés directs (DRD). A ce stade, rappelons les premiers corpus sur lesquels nous avons porté notre intérêt et les motivations à créer le CID par la suite. Il s'agissait de favoriser entre autres un maximum de DRD. Ces derniers, étudiés durant ma thèse de doctorat (Bertrand 1999) et qui ont fait l'objet d'articles par ailleurs (Bertrand & Espesser 1998, 2002), constituent donc le phénomène énonciatif par excellence pour tester cette hypothèse énonciative de diversité de voix au sein d'un même locuteur. Un discours rapporté direct est en effet l'une des illustrations les plus explicites de cette présence d'autres voix dans la voix du locuteur :

Le DR direct consiste en la retransmission différée qu'un locuteur (...) L fait des paroles antérieurement prononcées par un locuteur (...) L' (L et L' peuvent représenter une seule et même personne dans deux situations différentes) (De Gaulmyn 1992 : 22)

En termes de mise en scène énonciative, le DRD peut relever de la *dualité énonciative* (lorsque le sujet construit deux positions énonciatives différentes, deux points de vue). Il s'agit alors d'un discours *auto-rapporté* qui peut concerner ses propres paroles antérieures et ses paroles présentes. Le DRD peut également relever du *parallélisme* ou de l'*opposition énonciative* dans lesquels le sujet convoque sa propre parole et celle d'autres énonciateurs avec lesquels il est par exemple en accord ou pas respectivement<sup>8</sup>.

Selon nous le discours du/des locuteur(s), et au-delà, l'interaction elle-même, se construisent et s'expliquent par la prise en considération des niveaux à la fois dialogaux (l'échange hic et nunc) et dialogiques (ces diverses voix au sein du même locuteur).

L'étude a porté sur une centaine de discours rapportés directs produits par une locutrice dans une conversation (d'une heure environ) à trois participantes. Nous avons validé l'hypothèse selon laquelle un changement de voix (au sens énonciatif) correspondait à un changement de voix (au sens prosodique).

---

<sup>8</sup> A l'époque, le terme n'avait pas encore été défini par Stivers (2008) mais nous aurions pu parler à ce propos d'affiliation (cf chap 2).

## 1.2 Rôle de la prosodie

Dans ce contexte nous avons cherché à étudier la prosodie en tant que composante essentielle de l'interaction au même titre que la composante discursive ou énonciative. L'une de nos interrogations portait sur les fonctions sociales et interactionnelles de la prosodie, largement négligées à cette époque (voir introduction). Dans ce cadre, nous nous sommes interrogées sur le rôle de la prosodie en tant que marqueur polyphonique, et nous avons cherché à déterminer son rôle dans la manifestation de l'hétérogénéité énonciative inhérente aux discours. Les indices prosodiques ont été envisagés comme des *indices de contextualisation* au sens d'Auer ou de Couper Kuhlen. A cette période, les seules marques linguistiques associées aux DRD sont syntaxiques. Authier-Revuz (1982) parle "d'hétérogénéité montrée" pour rendre compte de ces *formes (...) qui offrent des représentations en discours du discours d'autrui* (1982: 114). Notre but est donc de mettre en lumière le rôle de la prosodie dans cette hétérogénéité montrée à travers quelques paramètres spécifiques.

### 1.2.1 Corpus

Le travail a été mené sur un corpus naturel de conversation enregistré entre trois personnes familières (Traverso 1998) lors d'une réunion familiale. L'enregistrement est effectué à l'insu de deux des trois participantes par la troisième, moi-même. Ce type de recueil de données s'avérait relativement fréquent à l'époque, sous réserve d'informer les participants immédiatement après et de leur demander l'autorisation d'exploiter les données ainsi constituées. D'un point de vue éthique, ceci était moins discutable dans la mesure où les corpus étaient la plupart du temps créés pour l'occasion, et nullement destinés à être partagés.

La conversation obtenue voit l'une des protagonistes particulièrement engagée dans cette rencontre: elle occupe une place discursive de narratrice principale, dont le thème principal est sa relation tumultueuse avec son patron. Elle produit environ 250 discours rapportés directs durant 1 heure. Les deux autres participantes sont principalement en position de destinataires, écoutant, évaluant et relançant fréquemment la narratrice. Grâce à ses DRD, la locutrice met en scène divers énonciateurs et s'implique très fortement pour narrer ses "exploits" professionnels pour lesquels elle convoque de nombreuses voix qui lui permettent à la fois de se construire une image positive d'elle-même tout en offrant une image très dévalorisée de celui avec lequel elle est en conflit (Bertrand & Espesser 2002).

### 1.2.2 Paramètres prosodiques

Les paramètres prosodiques que nous avons étudiés sont le débit, l'intensité et la  $f_0$ . Nous avons déployé une méthodologie relativement similaire à celle décrite dans le chap 1 pour étudier les chevauchements de parole, en nous concentrant sur le niveau phonéto-acoustique. Nous ne disposons pas à l'époque des différents niveaux d'annotation ni d'un modèle phonologique suffisamment établi et consensuel pour envisager les choses autrement.

Différentes étapes ont donc été élaborées. Pour étudier le changement de voix, nous avons retenu des paires comportant la fin d'un discours direct et le début d'un DRD. Ces paires ont été extraites automatiquement en fonction d'un critère de durée : chaque élément doit être compris entre 1 et 4 secondes, avec un minimum de 400 ms de voisement. Afin d'avoir des éléments symétriques en durée, l'élément le plus long est borné à la longueur du plus court. La contrainte de 400 ms permet d'éviter des cas limites (cas typique de présence de pauses silencieuses longues). Par ailleurs, nous avons dû recourir à une segmentation en pseudo-syllabes (ne disposant pas de cette annotation très coûteuse en temps): une pseudo syllable est donc définie comme une zone voisée bornée par ses propres frontières (de voisement) soit par une durée maximale de 140 ms.

Pour le détail entre les pseudo-syllabes et les syllabes réelles nous renvoyons le lecteur à l'article original (Bertrand & Espesser 1998). Les résultats présentés ci-après sont issus de cet article et du suivant (Bertrand & Espesser 2002).

Les paires segmentées retenues doivent contenir au moins 5 PS de part et d'autre. Nous obtenons ainsi 94 paires de DD-DRD.

L'analyse porte sur des différences éventuelles entre les deux populations (DD-DRD) en fonction des trois paramètres prosodiques retenus. Le débit est calculé sur les syllabes réelles tandis que l'intensité et la  $f_0$  sont calculées sur la moyenne de chaque PS en DD et DRD.

### 1.3 Résultats et interprétation

Les résultats sont les suivants : le débit et l'intensité en DD-DRD ne diffèrent pas de manière significative. A l'inverse, la  $f_0$  présente des différences significatives : soit un mode constant à 220 Hz, nous observons l'existence d'une dissymétrie à droite largement plus marquée pour les valeurs élevées de la  $f_0$ . D'autres tests nous permettent de confirmer une différence particulièrement nette au dessus de 270 Hz.

Pour affiner les résultats sur la  $f_0$ , nous avons cherché un autre indice relatif à l'évolution séquentielle de la  $f_0$ . Nous avons recherché une dépendance entre les  $f_0$  de 2 PS consécutives. Deux analyses de régression linéaire nous montrent que les  $R^2$  des modèles respectifs diffèrent significativement: la variabilité des DRD est mieux expliquée par une droite de régression que celle en DD.

Si le débit et l'intensité ne semblent pas des marqueurs pertinents pour le changement de voix entre DD et DRD, la  $f_0$  semble donc avoir un impact important. Nous observons en effet un élargissement de la dynamique (gamme) tonale de la locutrice vers les valeurs élevées de  $F_0$  en DRD. Nous l'interprétons comme la présence d'accents mélodiques marqués par des pics accentuels vers le haut ou bien une étendue tonale plus large en DRD. Le DRD est interprété comme étant un énoncé véhiculant un plus fort engagement, une plus grande implication de la locutrice qui le produit. Ceci se traduirait par une forme d'*emphatisation* (Selting 1994) marquée par des valeurs élevées, ou encore comme un effet de *crystallisation* selon les termes de De Gaulmyn (1992). Outre ce décrochage mélodique significatif dans le registre haut de la locutrice pour marquer le début de ses DRD, nous constatons également la présence fréquente de pauses silencieuses qui démarquent les deux discours, comme pour révéler la présence explicite des deux voix dans la parole de la locutrice (figure 16).

Dans le même sens, lorsque la locutrice produit deux DRD successifs référant à deux voix différentes, nous constatons une différence prosodique tranchée entre les deux voix. Celle qui rapporte ses propres paroles (auto-citation) ou les personnages qui transmettent plutôt une image positive d'elle versus les personnages qui véhiculent une image plus négative.

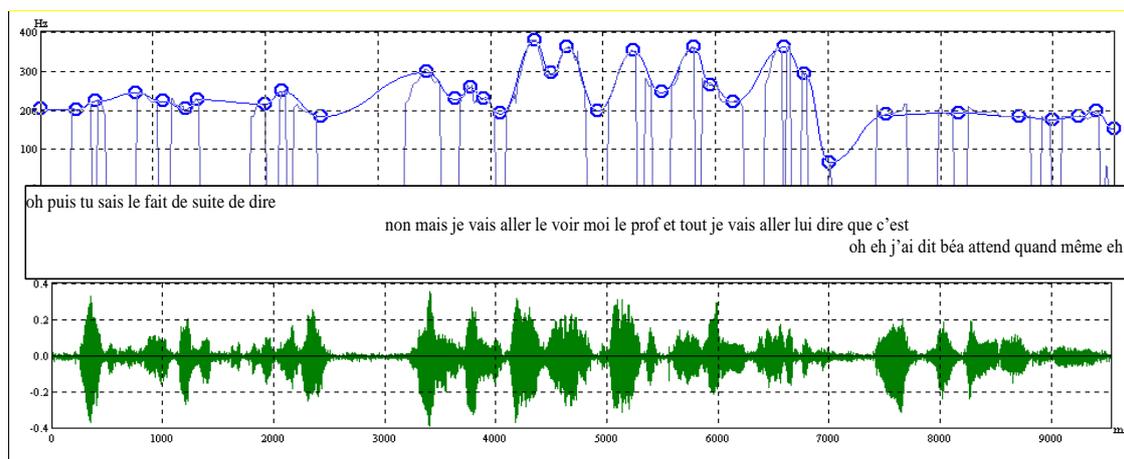


Figure 16: la locutrice produit un premier DRD montrant un élargissement mélodique net et global sur l'ensemble du DRD associé à une image négative du personnage cité tandis que le second DRD illustre un aplatissement de la gamme tonale associé à une image positive. Chacun des DRD est précédé d'une pause silencieuse relativement longue (720 ms et 330 ms respectivement), extrait de Bertrand 1999.

Nous avons montré par ailleurs une meilleure prédictibilité des PS en DRD qui serait le reflet d'une plus grande cohérence des PS en DRD. Ceci peut s'expliquer par les blocs examinés. Pour les DRD, le fait qu'ils marquent le début d'une séquence dans laquelle la locutrice se met en scène et s'implique fortement, pourrait se traduire par des mouvements plus continus pouvant être associés à moins d'hésitations que sur les DD qui sont les fins de blocs juste avant les DRD. Nous faisons l'hypothèse que ces fins de blocs (fin d'énoncés), pourraient être précédés d'éléments plus aléatoires comme des introductions, des préparations qui seraient moins bien planifiées, et donc moins "stylisées" que le début de ses DRD. Les lignes mélodiques continues en DRD pourraient refléter des schémas intonatifs spécifiques contribuant à créer un effet de citation qui participe de l'effet de cristallisation évoqué plus haut. A l'inverse, durant certains DD, la locutrice semble se désimpliquer pour mieux s'impliquer ensuite dans ses DRD. On observerait donc des moments successifs forts et faibles, dépendants de l'implication de la locutrice dans son discours, donnant lieu à une certaine forme de rythmisation énonciative.

### Publications

**Bertrand, R. & Espesser, R. (1998)** "Prosodie et discours rapporté: La mise en scène des voix", In Verschuereen J. (ed.) *Pragmatics in 1998: Selected papers from the 6th International Pragmatics Conference*. Antwerp, Belgium: Int. Pragmatics Association, vol.2, 45-56.

**Bertrand, R. & Espesser, R. (2002)** "Voice diversity in conversation: a case study", In B. Bel; I. Marlien (eds.) *Proceedings of Speech Prosody 2002*. Aix-en-Provence: Laboratoire Parole et Langage, 171-174.

## 2. Interface Prosodie-Discours

Pour évoquer ce pan de ma recherche, j'évoquerai les deux projets principaux auxquels j'ai contribué et qui partageait le même constat : le domaine de la prosodie, en ce début des années 2000, suscite un intérêt de plus en plus vif dans autant de disciplines connexes que sont la phonologie, la syntaxe du français parlé, la sémantique, la pragmatique, la psycholinguistique, la neurolinguistique, le traitement automatique des langues, sans oublier l'étude des pathologies du langage. Il se trouve cependant que la majorité des études réalisées se réfère à des modèles conçus pour l'interprétation de la phrase ou de l'énoncé. En conséquence, l'un des objectifs est de vérifier dans quelles limites les prédictions de ces modèles peuvent également rendre compte de la parole continue, spontanée, en interaction.

Une question centrale est donc de savoir quelles sont les primitives et les constructions prosodiques nécessaires pour rendre compte de l'organisation prosodique de ce type de production.

Un autre point concerne le fait que les relations examinées restent cantonnées à la sphère des relations entre micro et macro-syntaxe. L'analyse des interfaces plus complexes relative aux faits discursifs ou interactionnels et aux faits de prosodie reste programmatique dans la plupart des travaux. Enfin, les résultats disponibles en psycholinguistique concernant les stratégies cognitives des locuteurs-auditeurs proviennent d'études très souvent effectuées dans des conditions particulières. Comme nous l'avons déjà évoqué pour d'autres niveaux (chap 2), les études en compréhension portent généralement sur des énoncés isolés et non sur des discours. En production, les études portent sur des monologues ou sur des dialogues construits (contenus imposés aux participants). La problématique du discours spontané comme objet d'étude demeure encore largement négligée. Or, comme le propose Fodor dès 2002, la prosodie est une composante incontournable des recherches sur le langage oral, qu'il s'agira désormais de mettre au coeur de l'analyse en mettant en oeuvre de nouveaux paradigmes.

### 2.1 Le groupe ProDiGe : orientation générale

Collaborateurs : A. Di Cristo (PR) (responsable); C. Auran (MCF), R. Bertrand (CR), C. Chanet (MCF), A. Colas (IE), C. Portes (MCF), A. Régnier (Doctorant), M. Vion (PR).

Le groupe de recherche ProDiGe rassemble ainsi plusieurs chercheurs issus de la prosodie, l'analyse des discours et de la psycholinguistique. La perspective de recherche proposée vise à articuler les connaissances acquises dans les divers domaines afin de contribuer à mettre en évidence les stratégies

cognitives qui sous-tendent la production et la compréhension de la parole spontanée, ainsi que leurs déterminants multiples. Ces questionnements restent largement d'actualité.

Plus particulièrement, nos travaux ont concerné l'étude des relations entre les phénomènes prosodiques (l'accentuation, le rythme, l'intonation, les phénomènes temporels tels que les pauses, le tempo, etc.) et l'organisation des différentes formes de discours oraux. Ces relations ont été appréhendées comme une *interface complexe*. L'hypothèse de travail procédait de l'idée que cette interface est constituée par un *réseau de relations non-bijectives* entre des ressources langagières et para-langagières (lexicales, syntaxiques, prosodiques, mimo-gestuelles) et des fonctions linguistiques et communicatives hétérogènes (grammaticales, pragmatiques et discursives).

Ce groupe de recherche interdisciplinaire, créé en 2001, a été très actif durant plusieurs années. Il a donné lieu à l'édition du premier Symposium International sur l'interface Prosodie-Discours que nous avons organisé à Aix-en-Provence en 2005, édition suivie de 2007 à Genève (Suisse), 2009 à Paris (France), 2011 à Salford (Grande Bretagne) et 2013 à Louvain (Belgique).

### 2.1.1 Cadre théorique

L'objectif principal du groupe visait à dégager les stratégies cognitives à l'œuvre dans la production et dans l'interprétation des discours. Dans cette perspective, A. Di Cristo propose le cadre de la *Grammaire Ecologique* dans lequel le discours est conçu comme un écosystème, « *c'est-à-dire un mode d'expression qui s'adapte en permanence au milieu dans lequel il se déploie, en fonction de la versatilité des situations, de l'évolution de la mémoire discursive, de l'état des connaissances mutuellement partagées et non partagées, des pressions exercées par les forces interactionnelles, de la subjectivité et de l'intersubjectivité des interlocuteurs* » (Di Cristo et al. 2004: 33).

Par nature, le discours est un objet pluridimensionnel et d'une grande complexité. Il doit donc être interprété en ne se fondant pas uniquement sur les données verbales mais en intégrant également les aspects prosodiques. L'hypothèse globale sous-tendant cette démarche réside dans le fait que les données verbales et prosodiques interagissent en permanence pour construire les significations véhiculées dans le discours. Soulignons le point crucial, à rapprocher de la démarche de la LI (chapitre 2), que la grammaire est ainsi vue comme en partie émergente: une grammaire à l'état fini dont les formes et les fonctions seraient pré-définies ne peut être opératoire, du moins suffisante. Une autre question concerne notamment la spécification des interactions entre prosodie et discours, thème cher à A. Di Cristo et dont nous avons déjà fait état en introduction. Enfin, la dernière question concerne le choix des dispositifs qui permettraient de représenter ces interactions, en vue de les intégrer dans un schéma d'annotation global (tel que développé en parallèle dans les projets évoqués au chapitre 1). Ce point fait l'objet du paragraphe suivant.

### 2.1.2 Résultats principaux

Un apport de ce groupe a été le développement d'une méthodologie permettant de mettre en relation deux types de représentations. Les premières concernaient les représentations plurilinéaires de la structure du discours, informées par des modèles d'analyse du discours aboutis (Grosz & Sidner, 1986; Mann & Thompson, 1988; Roulet et al., 2001). Les secondes se rapportaient à des représentations plurilinéaires de la composante prosodique, ces dernières ayant la particularité d'intégrer à la fois des aspects métrique, tonal et temporel (Di Cristo et al. 2003; 2004).

En ce sens, nous avons développé un outil d'analyse favorisant la mise en correspondance des représentations du module discursif et du module prosodique qui se présente sous la forme d'une grille plurilinéaire à n. « tires » (tiers). L'élaboration de cette grille a nécessité deux démarches complémentaires mais néanmoins indépendantes : le choix des facteurs qui doivent faire partie de la représentation plurilinéaire de la composante « organisation discursive » et le choix des marques qui doivent figurer dans la représentation plurilinéaire de la composante prosodique. En ce qui concerne

la composante prosodique, le cadre d'analyse proposé s'est référé à deux types de marques : syntagmatiques et orthogonales qui permettent de procéder à une annotation et au codage de la prosodie en termes de symboles catégoriels (validés expérimentalement).

Une illustration de la grille est présentée ci-après (figure 17). On y observe des informations provenant de différentes sources (notations auditives, représentations analytiques et codages automatiques et semi automatiques). Très brièvement nous passons en revue les différents niveaux des informations présentées:

- **Osc.** renvoie au signal de parole
- **Momel** renvoie à la courbe de f0 modélisée sous la forme de l'interpolation d'une séquence de points cibles
- **UI:** Informations relatives au découpage en unités intonatives
- **ortho.** Renvoie à la transcription orthographique du texte
- **INTSINT** présente le codage catégoriel des points-cibles de Momel
- **ReSp** (Register Span) présente les variations de la gamme tonale (codage des variations de l'intervalle de hauteur entre les cibles hautes et basses dans une séquence donnée)
- **ReLe** (Register Level) correspond au niveau global du registre pour un domaine donné et à celle des réajustements de ce paramètre
- **Brea** (Breathing) correspond aux prises de souffle, et sont alignées avec les pauses auxquelles elles sont associées
- **Pauses**, notées # pour une pause courte, et ## pour une pause longue
- **Length.** renvoie aux variations de durées syllabiques
- **Tempo** renvoie à l'organisation temporelle du débit
- **Proe.** concerne les proéminences perçues
- **Met.** renvoie à la quantification des niveaux de proéminence selon une échelle de 1 à 3
- **F.** renvoie aux fonctions prosodiques

Nous n'entrerons pas dans le détail des autres niveaux qui ne sont pas aussi détaillés que celui de la prosodie. Le lecteur pourra se reporter à l'article original (Di Cristo et al. 2003).



### 2.1.3 Diffusion de la recherche

Outre les aspects méthodologiques, plusieurs recherches publiées ont porté sur des aspects particuliers de l'interface prosodie/discours comme la configuration prosodique des marqueurs discursifs ou des particules de discours (« enfin » Bertrand & Chanet 2005), exposée plus loin, mais aussi plus largement au niveau du groupe, le rôle de la prosodie dans la détermination des unités minimales du discours, la planification des Unités Intonatives dans la narration, la prosodie et l'anaphore, le rôle de la prosodie dans l'écologie du discours, la reconnaissance de la question en français, l'analyse de la variabilité interactionnelle, les contraintes du phrasage prosodique, etc.

Les recherches au sein du groupe ont donné lieu à plusieurs publications collectives ou individuelles et ont fait l'objet de nombreuses présentations dans des congrès internationaux (Congrès international de Pragmatique, Toronto, 2003, Congrès international « Prosody and Pragmatics » de Preston (UK, 2003), Congrès International des Sciences Phonétiques (Barcelona, 2003). Le nombre des publications pour la période 2003/2006 s'élève à 22.

#### Publications

- Bertrand**, R. & Chanet, C. (2005) "Fonctions pragmatiques et prosodie de *enfin* en français spontané", *Revue de Sémantique et Pragmatique*, 17, 41-68.
- Di Cristo, A., Auran, C., **Bertrand** R., Chanet, C., Portes, C., Regnier, A. (2004) "Outils prosodiques et analyse du discours", in A.C. Simon, A. Auchlin et A. Grobet (eds), *Cahiers de Linguistique de Louvain* 30/1-3, Louvain-la-neuve : Peeters, 28, 27-84.
- Di Cristo, A., Auran, C., **Bertrand**, R., Chanet, C., Portes, C. (2003) "An integrative approach to the relations of prosody to discourse : towards a multilinear representation of an interface network", in A. Mettouchi & G. Ferré (eds), *Interfaces Prosodiques*, Nantes, 29-34.
- Auran, C., **Bertrand** R., Chanet, C., Colas, A., Di Cristo, A., Portes, C., Régnier, A., Vion, M. (2006) *Proceedings of Discourse-Prosody Interface Symposium - IDP05* (Eds), Cederom.

En guise de conclusion, je souhaiterais insister sur l'importance de la création de ce groupe et sur le caractère novateur de la perspective déployée. Il a également été l'occasion de travailler avec des collègues d'autres disciplines autour d'un objet commun, ce qui a largement été fondateur dans la suite de ma carrière. Ce projet a également été un moteur pour avancer sur la question des liens entre les niveaux dans l'interaction, qui m'occupent depuis le début, quels que soient les niveaux envisagés. Enfin, ce travail a également été extrêmement formateur et m'a conduit à toujours prêter attention aux niveaux de représentation et d'analyse auxquels on se réfère et à les spécifier, thème cher à A. Di Cristo. Seule une telle spécification permet de savoir de quoi l'on parle, ce que l'on met en lien, et peut contribuer à une meilleure compréhension du langage et de la parole en interaction dans toute sa complexité, en permettant un dialogue clair avec la/les communauté(s) de recherche intéressée(s).

### 2.2 Le projet PRO-GRAM: Prosodie dans la Grammaire

*PRO-GRAM. Prosodie dans la grammaire. Interfaces syntaxique, sémantique et pragmatique de la prosodie dans la grammaire du français*

J.M. Marandin (responsable, LLF) <http://pro-gram.linguist.jussieu.fr/>

(2006-2010)

A peu près à la même époque, nous avons eu l'occasion de participer à un projet financé par l'ANR, porté par J. M. Marandin (2006-2010).

Partenaires : Laboratoire de Linguistique Formelle (LLF), Paris ; DELIC (Description Linguistique sur Corpus), Aix-en-Provence; Atelier de Recherches sur la Parole (ARP), Paris

Similairement à la perspective de *ProDIGe*, ce projet d'envergure nationale proposait de mettre en lien différents niveaux d'analyse, complémentaire du précédent, mais davantage centré sur les aspects théoriques.

Ce projet, inscrit dans le double mouvement qui traversait la recherche en linguistique contemporaine, concernait :

- la prise en compte de plus en plus fréquente des caractéristiques prosodiques des phénomènes syntaxiques, sémantiques et pragmatiques
- l'intérêt croissant pour des phénomènes réputés périphériques (l'ellipse, l'incidente) et qui, pourtant, s'observent avec une grande fréquence dans le discours (oral ou écrit).

Les deux mouvements sont à l'évidence corrélés puisque ces phénomènes traités comme périphériques mettent crucialement en jeu l'organisation prosodique. Suite à ce constat, on ne peut plus négliger l'organisation prosodique des énoncés de la même façon, et on cesse enfin de considérer la pragmatique comme un simple « supplément d'âme ».

Ce projet s'est également inscrit dans une conjoncture marquée en France par le peu d'interactions intellectuelles ou institutionnelles entre différentes communautés de linguistes : les linguistes à visée modélisatrice et les linguistes à visée descriptive, les syntacticiens, sémanticiens et pragmaticiens auxquels s'ajoutent les phonologues.

C'est dans ce double contexte que ce projet visait à favoriser le développement d'interfaces entre modèles explicites de la prosodie d'une part, de la syntaxe, de la sémantique et de la pragmatique d'autre part, au sein de grammaires formalisées qui aient la plus large couverture empirique possible.

Les objectifs du projet étaient donc de :

- fournir des généralisations détaillées d'aspects prosodiques de constructions qui puissent fournir la base empirique d'une modélisation grammaticale de la dimension prosodique ;
- combler des lacunes dans la connaissance grammaticale du français ;
- de manière générale, créer une synergie entre la communauté des syntacticiens, sémanticiens et pragmaticiens formels et celle des phonologues travaillant en prosodie.

Il est à noter que cette nouvelle perspective s'inscrit également dans le contexte décrit dans le chapitre 1 de cette synthèse. Pour mener à bien ce type d'études, il s'avérait fondamental de disposer de données orales multi-niveaux permettant de mettre en lumière les différentes interfaces. Ce projet PRO-GRAM, coordonné par un sémanticien féru de prosodie, a constitué une réelle avancée et permis le développement d'une vision du sens intégrant les autres niveaux de la grammaire, dont la prosodie.

### *2.3 Les contours intonatifs: valeur pragmatique et réalisation phonétique*

Collaboration : C. Portes

C'est dans ce contexte de recherche novateur que nous avons initié avec ma collègue C. Portes une collaboration très étroite, qui a été l'occasion d'un dialogue particulièrement riche et enthousiasmant. Nous avons cherché à mieux caractériser les contours intonatifs du français en termes formel et fonctionnel, en les réinvestissant dans une approche du dialogue et de l'interaction. En effet, les constructions prosodiques et en particulier les contours intonatifs tirent leur signification notamment de la relation qui s'instaure entre le locuteur et ses interlocuteurs, en dialogue (Hirschberg & Ward 1995 ; Marandin et al. 2004).

#### *2.3.1 Statut des contours intonatifs*

Dans la perspective que nous adoptons, inspirée des travaux menés dans l'ANR Program, les contours ne sont plus conçus comme des morphèmes mais plutôt comme des *constructions* (Marandin 2004) : ils ne sont vus ni comme des mouvements holistiques ni comme des primitives. Ils conjuguent un mouvement tonal linguistiquement pertinent, définissable formellement par une succession de cibles tonales qui s'associent de manière spécifique avec le niveau segmental, et ils sont fonctionnellement contrastifs. La particularité de ce modèle réside dans son ancrage fortement sémantique, qui le

distingue des modèles de la phonologie métrique autosegmentale dont il se réclame pourtant. En effet, pour ces derniers, les contours contrastifs sont identifiés sur la base de leurs propriétés formelles (composition tonale) dans lesquels il n'est pas fait allusion au sens de l'intonation (Post 2000 ; Jun et Fougeron 2000). Il serait injuste de ne pas citer cependant l'une des rares études ayant abordé la question du sens au sein de cette mouvance dans un article princeps (Pierrehumbert & Hirschberg 1990) et auxquelles les études plus récentes font invariablement référence. Dans cet article, les auteurs proposent alors de concilier une approche du sens et de la phonologie prosodique. Selon elles, les caractéristiques intonatives de phrasé, le placement de l'accent, la gamme tonale sont d'importantes sources d'information pour rendre compte des structures attentionnelles et intentionnelles du discours :

*we propose that a speaker chooses a particular tune to convey a particular relationship between an utterance, currently perceived beliefs of a hearer or hearers, and anticipated contributions of subsequent utterances* (Pierrehumbert & Hirschberg 1990:271).

Nous sommes séduites par le modèle de Marandin pour la double raison liée à son inscription dans un cadre tel que celui de la phonologie métrique autosegmentale tout en assumant son ancrage sémantique : pour les auteurs en effet, la signification est cruciale pour définir les contrastes intonatifs, toute construction linguistique ayant pour vocation de transmettre du sens. Une difficulté subsiste cependant pour le français à l'époque, et non des moindres : l'inventaire phonologique des primitives n'est pas encore établi de manière consensuelle, et l'outil de transcription TOBI associé à l'approche AM pour l'anglais américain n'est donc pas applicable à nos données. Malgré cette difficulté, nous faisons l'hypothèse forte avec C. Portes que le français peut être décrit selon les mêmes principes, moyennant un travail en ce sens. Nous estimons également que l'analyse de la parole spontanée pourra bénéficier de cet inventaire phonologique, lorsque nous en disposerons, en nous permettant de mieux caractériser les variations inhérentes à la parole spontanée.

Si dans la littérature discursive, 2 mouvements seulement sont pertinents (montant non final et descendant final), la littérature prosodique présente plusieurs contours montants, formellement distincts, et qui remplissent des fonctions différentes. Le modèle très influent de Delattre (1966) sur les contours de base du français a été une source importante pour notre travail bien que nous n'accordions pas le même statut que lui aux contours intonatifs. Chez Delattre, mais aussi chez Rossi (1981), les éléments du système intonatif sont considérés comme des morphèmes intonatifs. Chaque contour, appelé aussi intonème, est défini comme un mouvement de f0 contrastif en relation avec une fonction spécifique (contour de finalité marqué par un mouvement descendant et qui signale la fin de l'énoncé) (figure 18).

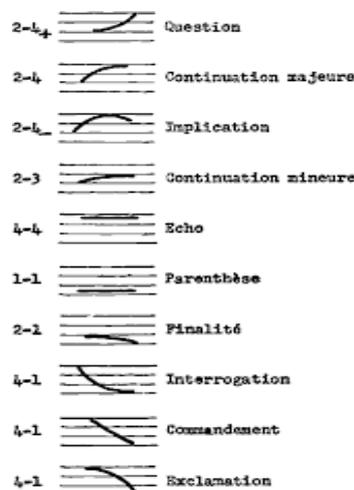


Figure 18 : Les 10 intonations de base du français, extrait de Delattre (1966)

Le modèle de Marandin (figure 19) présente en revanche un inventaire des contours terminaux, conçus comme nous l'avons mentionné précédemment comme conjuguant un mouvement tonal linguistiquement pertinent, définissable formellement par une succession de cibles tonales qui s'associent de manière spécifique avec le niveau segmental, et fonctionnellement contrastifs.

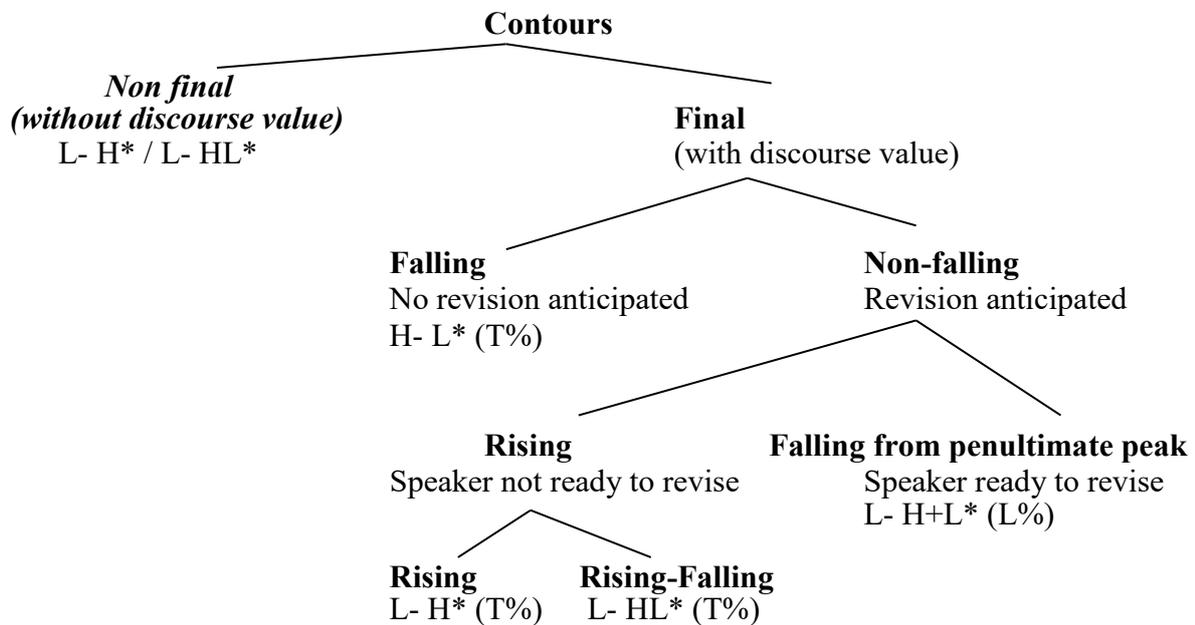


Figure 19: The semantics of French intonation according to Marandin et al. (2004). Italics signals the absence of discourse value of non final F0 movements, i.e. Delattre's continuation contour, extrait de Portes & Bertrand (2006)

Notre travail a porté entre autres sur le contour montant de continuation majeure de Delattre, absent notamment du modèle de Marandin et al. (2004). En effet, selon les auteurs, ce contour n'entrerait pas dans le paradigme des contours terminaux, ce qui a été précisément l'objet d'une de nos propositions.

### 2.3.2 Résultats principaux

A travers trois études sur le CID, nous avons tenté de distinguer formellement et fonctionnellement les différents contours montants. Préalablement, nous avons montré que l'inventaire phonologique des contours intonatifs du français devait réintroduire la différence entre contours montants (continuatifs) majeur et mineur, tant à cause de différences formelles (D'Imperio et al. 2006) qu'en raison de différences fonctionnelles (Portes & Bertrand 2006; Portes et al. 2007; Bertrand et al. 2008) (figure 20).

Selon Marandin et al (2004) mais également pour Delais Roussarie (2005), les contours intonatifs non terminaux n'auraient pas de valeur discursive et ne seraient finalement que des mouvements de f0 par défaut. Si nous avons adhéré à l'idée que les contours montants mineurs n'avaient pas de fonction discursive particulière, nous avons cherché à montrer que les contours montants de continuation devaient être réintroduits dans le modèle.

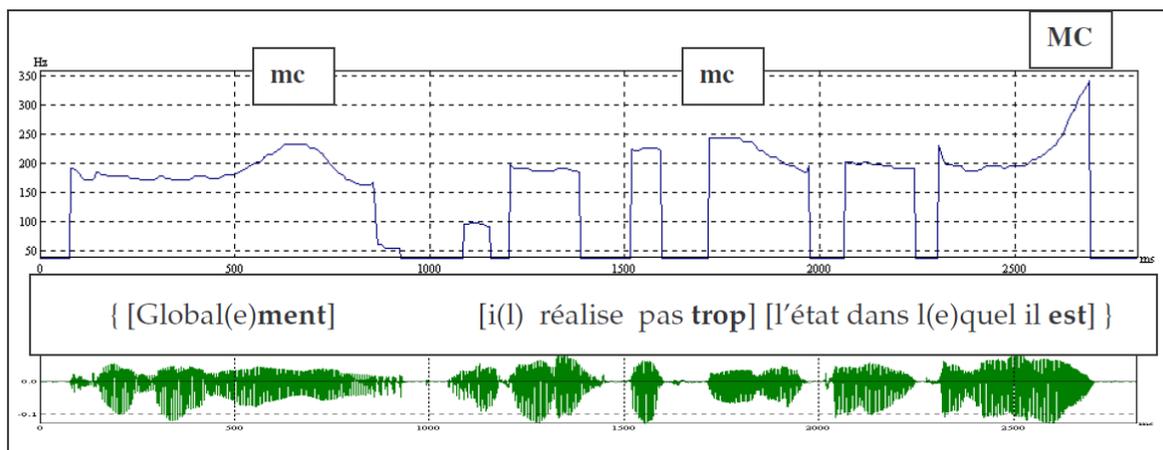


Figure 20 : exemple du CID présentant un contour continuatif majeur (MC ou LH%) composé de deux contours mineurs (mc ou LH\*), extrait de Portes & Bertrand 2006

Les études menées sur les conversations du CID ont permis de montrer que ces contours intonatifs montants de continuation jouaient un rôle en termes discursifs et interactionnels (Portes & Bertrand 2005). D'abord ils regroupent les unités marquées par des contours mineurs en une unité plus large. Ils semblent se placer à l'intersection des contours mineurs et des contours majeurs. En fait, et suite à l'idée proposée par Ginzburg cité par Engdahl (2005), nous avons proposé de les considérer au même titre que les contours terminaux dans la mesure où ils jouent notamment un rôle identique au contour terminal descendant dans l'élaboration du common ground. Le locuteur produisant un contour descendant ou un contour montant continuatif inscrit en effet l'énoncé dans les connaissances partagées des interlocuteurs, à condition cependant qu'il ne soit pas interrompu ou qu'il soit suivi d'un simple backchannel (feedback). En ce sens, nos résultats ont confirmé le rôle des contours montants de continuation dans l'organisation des tours de parole (TCU finaux et non finaux). Ils sont également particulièrement propices à la production de feedbacks. A priori à l'intersection des contours mineurs et des contours majeurs terminaux, les contours montants continuatifs sont selon nous tout à fait légitimes à être promus dans ce modèle dialogique de l'intonation car comme les contours terminaux, ils signalent le type de réception que le locuteur anticipe pour son énoncé : en l'occurrence, le contenu est inscrit dans le contenu des connaissances partagées, le locuteur n'anticipant aucune révision à la condition que l'interlocuteur n'intervienne pas. En outre, le contour de continuation est utilisé pour informer l'interlocuteur que le contenu de l'énoncé suivant est à interpréter en lien avec l'énoncé courant (fonction de liage chez Chafe 1988, Matsumoto 2003) que nous pouvons également mettre en lien avec la notion de projection (chapitre 2).

L'examen des différents contours intonatifs a été également mené dans une étude concernant leur distribution syntaxique, discursive et interactionnelle (Bertrand et al. 2008). En raison du sens qu'ils véhiculent, ces différents contours ne peuvent pas apparaître de manière hasardeuse. Outre leur apparition en certains points syntaxique et discursif spécifiques, ils apparaissent également en certains lieux interactionnels particuliers. Nous avons montré l'importance de leur localisation au sein des phases qui constituent le récit qui était l'activité principale considérée (voir chap 2). A titre d'illustration, nous avons montré que les phases d'orientation, dans lesquelles le narrateur présente des informations (personnages, contexte) qui vont s'inscrire dans les connaissances partagées par les participants, sans lesquelles le récit ne peut se construire de manière optimale, comportent préférentiellement des contours montants de continuation. Les contours montants descendants ou les contours d'implication apparaissent plutôt dans les phases d'évaluation.

Une autre étude nous a enfin permis de montrer que le cadre théorique proposé par Marandin et al. (2004) permettait de rendre compte à la fois de la parenté et de la distinction entre contour montant continuatif majeur et contour montant d'énumération, le second dérivant du premier par stylisation

(Portes et al. 2007). Le contour de liste se caractérise notamment par une pente moins abrupte que le contour de continuation.

Du point de vue des implications interactionnelles, ce contour de liste a permis de réaffirmer le rôle de la prosodie comme indice privilégié de projection. Nous avons vu au chapitre 2 l'importance de cette notion de projection pour favoriser la progressivité de l'interaction et en permettre un accomplissement interactionnel optimal. Pour que l'interaction progresse, il est en effet crucial que les interlocuteurs identifient, reconnaissent et interprètent les actions au sein des diverses activités conversationnelles. Des indices de nature diverse ont été mis en évidence pour annoncer par exemple lorsqu'un locuteur souhaite achever ou céder son tour, initier un récit, ou qu'il a besoin de réparer en cas d'incompréhension (voir par exemple les mouvements de sourcils pour projeter une telle séquence de réparation dans Goujon, Bertrand & Tellier 2017; Goujon 2019). Dans Portes et al. (2007), nous avons montré que les contours intonatifs de liste signalent à la fois l'intention communicative de garder son tour mais aussi le type d'énoncé qui va suivre, en l'occurrence un item de liste, ainsi que les attentes que cela génère au niveau des réponses de l'interlocuteur. Ces contours intonatifs de liste sont de tels indices de projection qu'ils sont propices à la création d'une liste que l'on appelle liste virtuelle, caractérisée par un seul item marqué prosodiquement et identifié en tant que tel par l'interlocuteur. Ces contours de liste sont également propices à la production collaborative d'une séquence de liste, l'interlocuteur pouvant collaborer avec l'initiateur de la liste et produire lui-même l'item de liste suivant pour donner lieu potentiellement à des séquences de convergence interactionnelle (Bertrand & Priego-Valverde 2017, chapitre 2). Dans l'exemple 7, l'interlocuteur produit le dernier item de la liste avec un contour terminal descendant puisqu'il clôture la liste totale.

**Exemple 7**, extrait de Bertrand & Priego-Valverde (2017)

```

IM_726 @ je cours à l'école je lui dis
IM_727 vous savez que ça fait cinquante ans qu'on laisse les enfants écrire
avec la main qu'ils veulent hein
ML_690 @
IM_728 => oui oui mais vous comprenez il aura une écriture horrible je ne peux
=> pas tolérer euh une chose pareille euh il faut absolument qu'il
=> s'entraîne de la main droite euh
IM_729 => il a pris un mauvais pli euh vous l'avez laissé faire enfin à
ML_691 => et moi je vais le redresser quoi
IM_730 han
  
```

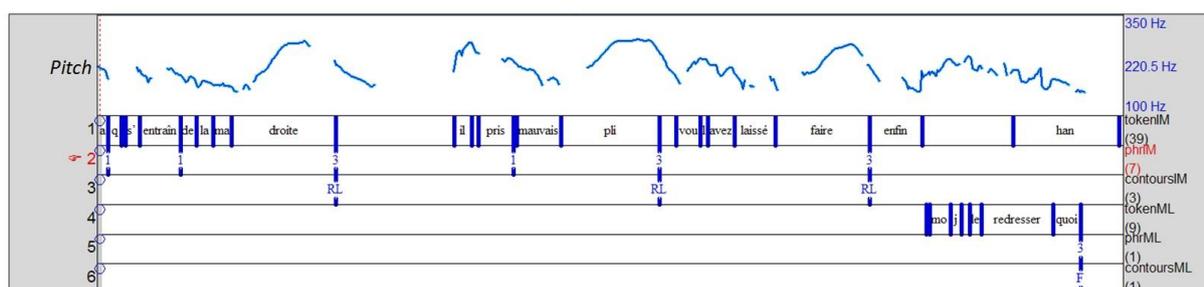


Figure 21 : liste comportant plusieurs items terminés par un contour de liste typique et le dernier qui clôture la séquence énumérative est produite par l'interlocuteur avec un contour intonatif descendant final adapté, extrait de Bertrand & Priego-Valverde (2017)

Pour le lecteur intéressé, nous renvoyons à d'autres travaux conduits dans cette perspective visant à mieux caractériser les contours intonatifs du français (Portes et al 2014 ; Portes & Beyssade 2015).

**Publications**

**Bertrand, R., Portes, C. & Sabio, F. (2008)** "Distribution syntaxique, discursive et interactionnelle des contours intonatifs du français dans un corpus de conversation", *Travaux Neuchâtelois de Linguistique*, 47, 59-77. ISSN 1010-1705.

- Portes, C., **Bertrand**, R., Espesser, R. (2007) "Contribution to a grammar of intonation in French. Form and function of three rising patterns", *Nouveaux Cahiers de Linguistique Française*, 2, 155-162.
- D'Imperio, M., Bertrand, R., Di Cristo, A., Portes, C. (2006) "The phonology and phonetics of prenuclear and nuclear accents in French", *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics* (2006 août 28-30: Athens, Greece), 121-124.
- D'Imperio, M., **Bertrand**, R., Di Cristo, A., Portes, C. (2007) "Investigating phrasing levels in French : Is there a difference between nuclear and prenuclear accents? " In J. Camacho, V. Deprez, N. Flores, L. Sanchez, *Selected Papers from the 36th Linguistic Symposium on Romance Languages* (LSRL). New Brunswick: John Benjamins Publishing Company, 97-110.
- Portes, C. & **Bertrand**, R. (2006) "Some cues about the interactional value of the «continuation» contour in French", *Actes Discours et Prosodie comme Interface Complexe* (IDP) (2005 octobre 8-9: Aix-en-Provence, France). Cederom (14 pages).

## 2.4 Le phrasé prosodique du français

Le second aspect sur lequel nous avons travaillé avec Cristel a concerné le phrasé prosodique.

Notre ambition était de montrer que les unités phonologiques traditionnelles, élaborées sur des corpus très contraints, pouvaient être opératoires dans la parole spontanée. Ceci était possible sous réserve de quelques adaptations. Celles-ci concernaient notamment la question des disfluences. Les disfluences sont en effet l'un des écueils auxquels on peut rapidement se heurter lorsque l'on s'attache à caractériser les unités prosodiques établies dans les modèles traditionnels, sur de la parole contrôlée. En effet, il s'avère que certains paramètres phonétiques tels que l'allongement syllabique par exemple sont des éléments non seulement cruciaux pour la détermination des frontières prosodiques mais qu'ils sont également une marque potentielle de disfluence. Ces réalisations communes peuvent ainsi opacifier le phrasé et rendre son identification très délicate. L'une de nos propositions a été de clairement distinguer les deux niveaux. Ceci fait écho à la problématique de la spécification des niveaux. Une autre de nos propositions a été de considérer que les unités intonatives majeures, de la même manière que les unités syntaxiques par exemple, pouvaient être non seulement interrompues mais aussi inachevées. Dans ce cas, cela pouvait laisser le loisir également à l'interlocuteur de les compléter, les unités intonatives pouvant donc être également construites à deux.

Le travail très conséquent que nous avons mené en termes d'annotation sur le CID en particulier, (chapitre 1) associé à une réflexion très intense menée sur les modèles phonologiques nous a permis d'affiner notre approche théorique de la prosodie dans l'interaction. Ce positionnement théorique a été présenté dans JFLS (Portes & Bertrand 2011) et pour l'essentiel repris dans le numéro spécial destiné à dresser un panorama des nouvelles perspectives de recherche en analyse conversationnelle pour le français (Bertrand & Portes 2012). Nous reportons le lecteur qui souhaiterait plus de détails et des exemples en corpus à ces deux publications.

### Publications

- Bertrand**, R. & Portes, C. (2012) "Pour une approche phonologique de la prosodie dans l'interaction", *Langue française*, vol. 2012/3, no. 175, 19-36.
- Portes, C. & **Bertrand** R. (2011) "Permanence et variation des unités prosodiques dans le discours et l'interaction", *Journal of French Language Studies*, 97-110. <http://hal.archives-ouvertes.fr/hal-00576851>.

Avant de conclure cette section, soulignons que notre proposition relative aux unités intonatives qui pourraient être interrompues, est un point que nous souhaiterions poursuivre théoriquement et empiriquement avec ma collègue Cristel. Cette perspective paraît envisageable aujourd'hui en raison des travaux que j'ai également effectués sur les disfluences elles-mêmes avec Berthille Pallaud, experte de ces questions. En effet, depuis quelques années, le CID a été totalement annoté sur le plan des disfluences par Berthille elle-même. Suivant le modèle de Schriberg, les disfluences ont été annotées en trois phases: le *reparandum*, le *break*, et le *reparans*. Nous n'entrerons pas dans le détail de l'annotation effectuée, nous reportons aux différents articles consacrés à ces questions (Pallaud & Bertrand 2020; Pallaud et al. 2019). Mais il nous semble important de souligner cependant que nous avons abouti notamment à une dichotomie opposant les disfluences/interruptions réelles aux

disfluences/interruptions dites suspensives. Ces dernières sont liées à la présence d'éléments tels que des pauses silencieuses, remplies, des marqueurs discursifs ou encore des incidentes, qui occupent le break, mais après lesquelles il n'y a pas nécessité d'une réparation.

Ces interruptions suspensives pourraient pour certaines être traitées comme de « simples » unités prosodiques interrompues telles que nous l'avons proposé avec Cristel, à la différence de nombreux auteurs pour lesquels la présence d'une pause silencieuse induit nécessairement une frontière d'unité intonative. Or, si l'on observe la figure suivante (figure 22), dans laquelle une pause silencieuse de 250 ms sépare le verbe de son complément, notre proposition consisterait à considérer ici une seule unité intonative, malgré l'interruption par la pause silencieuse, ceci étant renforcé par la continuité mélodique observée avant et après la pause. La question soulevée dans Portes & Bertrand (2011) selon laquelle on aurait tantôt à faire à des pauses pragmatiques tantôt à des pauses liées aux difficultés d'encodage, renvoie en partie à la même question et fera l'objet d'une étude ultérieure systématique.

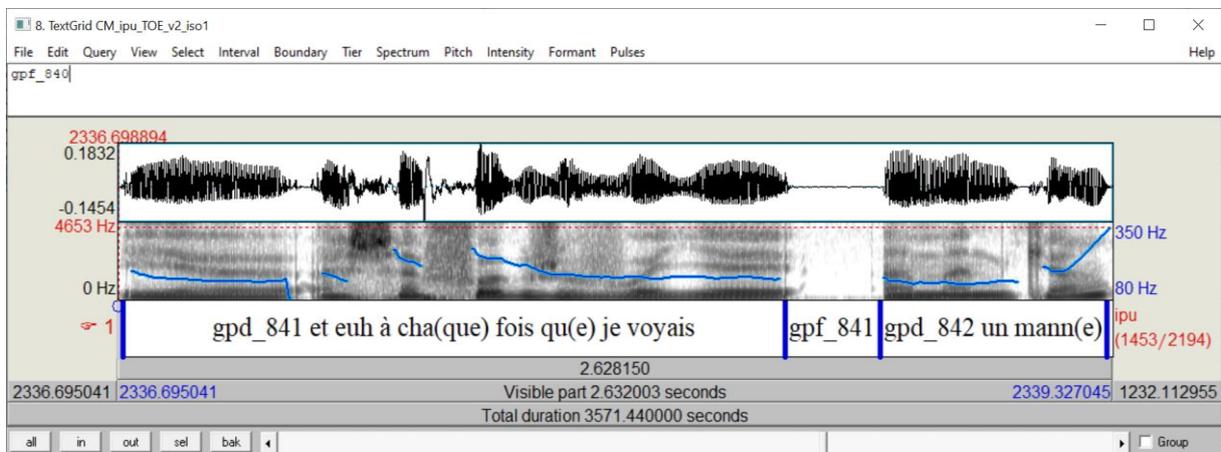


Figure 22 : pause silencieuse interrompant une seule unité intonative potentielle, extrait du CID

L'exemple suivant (figure 23), extrait du corpus CHEESE, présente de la même manière une pause silencieuse très longue entre le syntagme nominal sujet "et le train" et le syntagme verbal qui le suit. Si l'on opte préférentiellement ici pour considérer que l'on a bien deux unités intonatives, dont la première se termine par un contour intonatif montant, nous avons d'ores et déjà pu observer des configurations intonatives relativement différentes (entre autres des AP descendantes) qu'il s'agit maintenant d'étudier en détail. En outre, les implications interactionnelles relatives aux différents types de dispositif seront examinées.

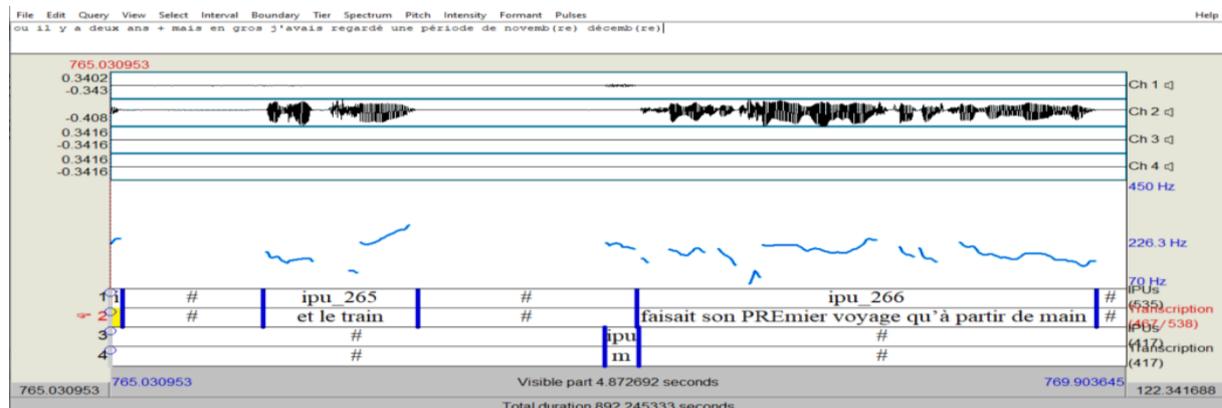


Figure 23: très longue pause silencieuse séparant le syntagme nominal sujet du syntagme verbal suivant, extrait du corpus CHEESE

Dans le cas suivant enfin (figure 24), se pose le même problème quant à la détermination d'une unité intégrant les items de disfluences ou pas. Le découpage en unités intonatives en 7 et 8, par deux annotateurs distincts, a été effectué avant que ne soient disponibles les annotations de disfluences. On remarque qu'en 7 l'annotateur a intégré les items de disfluences dans l'IP alors qu'en 8 l'annotateur a noté un ?? pour rendre compte d'une difficulté.

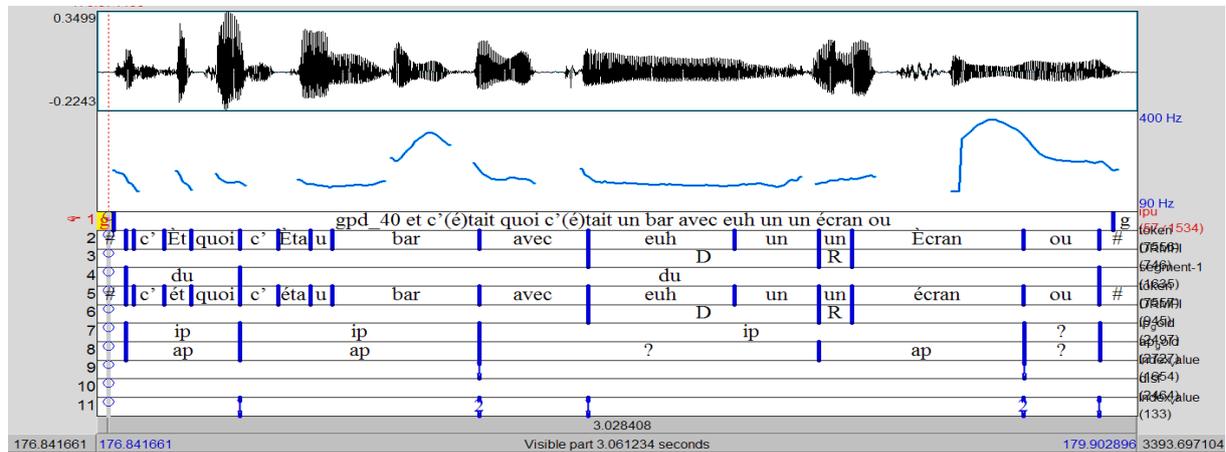


Figure 24 : unités intonatives et disfluences ; les annotations sur les tires 3 et 5 correspondent à D (disfluence) et R (réparation), extrait du CID

## Publications

- Pallaud, B. & Bertrand, R. (2020) "Auto-interruptions et disfluences à l'oral. C'était euh tu vois complètement loufoque comme si ouais euh comme situation. Dans de recherches sur les pauses présentes dans la parole et le discours", in Hirsch, F., Didirková, I. & Dodane, C. (Eds.). *Manuel de Pausologie*. Recueil. Paris: L'Harmattan, coll. Langue et Parole, 21-49.
- Pallaud, B., Bertrand, R., Blache, P., Prévot, L. & Rauzy, S. (2019) "Suspensive and disfluent self interruptions in French language interactions", In L., Degand, G., Gilquin, L., Meurant & A. C., Simon (eds.) *Fluency and Disfluency across Languages and Language Varieties*. Corpora and Language in Use -Proceedings 4, Louvain-la-Neuve: Presses universitaires de Louvain, 109-13.

Pour conclure sur cet aspect, nous souhaitons rappeler que le travail effectué ultérieurement, lors de la campagne d'annotation intensive sur les unités prosodiques exposé au chapitre 1, n'a pu se déployer que grâce à ce travail théorique mais aussi très concret (en termes d'annotation) que nous avons conduit avec ma collègue C. Portes.

L'amélioration de nos connaissances sur les unités de phrasé est absolument cruciale dans la mesure où elles sont nécessaires pour développer des études sur divers aspects prosodiques, mais aussi pour permettre l'étude des liens entre les différentes unités (syntaxique, discursive, etc). Les travaux relatifs aux unités conversationnelles se fondent pour la majorité sur les IPU. Si celles-ci constituent des unités formelles et automatisables permettant de rendre compte en partie d'une certaine forme d'organisation des tours de parole, du moins des changements de parole, l'amélioration dans la détermination des unités de tours à l'image de Ford & Thompson (1996), ou Degand & Simon (*BDU Basic Discourse Unit*) (2009) pour le français, se révèle indispensable pour mieux rendre compte de cette organisation. Ce travail est en cours sur le CID (mise en relation des unités syntaxique, discursive et prosodique, voir nos perspectives dans le chapitre 4).

### 3. Structuration prosodique et caractéristiques accentuelles du français

*PhonIACog: Rôle de l'accentuation initiale dans la structuration prosodique du français*

*-de la phonologie au traitement de la parole-*

C. Astésano (responsable, Laboratoire Lordat)

<http://www2.lpl-aix.fr/~phoniacog/>

2013-2016

Un autre aspect de ma recherche en prosodie concerne le travail mené en collaboration avec ma collègue C. Astésano. Je voudrais remercier Corine qui m'a permis de participer au projet qu'elle a coordonné entre 2013 et 2016, et pour lequel elle m'a fait l'honneur de m'en proposer la responsabilité scientifique pour le LPL. Ce projet a été l'occasion de discussions stimulantes, d'une rare intensité.

L'un des points principaux de ce projet résidait dans la volonté d'analyser des données issues à la fois de corpus les plus contrôlés jusqu'aux corpus les plus spontanés, cet aller-retour visant à éviter de produire des résultats impossibles à valider sur des données attestées. Cette précieuse collaboration m'a permis de renouer avec les aspects plus théoriques des modèles prosodiques du français, de m'interroger sur des aspects assez éloignés de mes axes de recherche antérieurs, d'offrir une opportunité de (re)travailler collectivement avec les collègues spécialistes de la prosodie et du cadre de l'AM, telles que Pauline Welby et Cristel Portes. Nos « discussions à 4 » sur les données ont été une réelle source d'inspiration et un vrai moment de plaisir. Enfin, cette ANR m'a offert l'opportunité d'envisager la prosodie « un peu autrement », avec un focus réel sur les implications théoriques des modèles prosodiques -dont Corine était indéniablement la vraie experte-, et en particulier son versant perceptif, dont j'étais assez peu familière.

#### 3.1 Objectif du projet

Ce projet avait pour ambition de décrire le système accentuel du français en l'abordant sous plusieurs angles : production, perception et neuro-imagerie. L'objectif visé consistait à décrire les caractéristiques accentuelles du français, à travers le patron bipolaire AI-AF (Accent Initial-Accent final) (Di Cristo 2000, Astésano 2001) pour en dégager la structure phonologique sous-jacente. Atteindre cet objectif ne pouvait se passer de l'étude d'une parole non seulement très contrôlée mais également d'une langue en usage. Ceci visait là encore deux buts principaux : d'une part permettre d'améliorer le traitement du français dans les modèles psycholinguistiques encore très imparfaits de ce point de vue selon Astésano. D'autre part, rendre le traitement automatique de la prosodie plus performant dans les corpus de taille conséquente, ce qui entrerait davantage dans mes prérogatives. Enfin, un autre défi dans ce projet consistait à combiner deux approches phonologiques complémentaires : la théorie métrique et fonctionnaliste (MF) dont se réclame C. Astésano, et la théorie métrique auto-segmentale (AM). Selon l'auteure, la MF posséderait un atout sur la première en ce sens qu'elle tient mieux compte des aspects rythmiques des discours.

#### 3.2 Frontières et prééminences

L'une des propositions principales du projet concernait un problème crucial pour le français qui nous a conduit à réinterroger les liens entre prééminences et frontières. Nous avons participé à ce travail à travers une étude perceptive dans laquelle nous avons montré que ces deux phénomènes ne sont pas, comme on l'a longtemps pensé en raison du syncrétisme en français (confusion entre l'accent final et le ton de frontière), un même phénomène sous-jacent. Notre analyse a porté sur les accents finals et initiaux en tant que marqueurs de structure droite et gauche. Nos résultats indiquent que les français, contrairement à une vision qui a longtemps perduré, ne sont pas « sourds » aux prééminences, qu'elles soient initiales ou finales, et qu'elles apparaissent indépendamment ou non des frontières prosodiques majeures. Par ailleurs, la perception des frontières et des prééminences indique un traitement distinct de ces deux phénomènes en français, ce qui nous permet de parler d'une possible dissociation partielle

de ces deux phénomènes très longtemps confondus (Astésano et al. 2012, 2013). Ces résultats semblent corroborer l'hypothèse proposée par Astésano (2016) selon laquelle ajouter un niveau supplémentaire dans la hiérarchie prosodique du français (correspondant au mot lexical (pw) et inférieur à l'ap) permettrait de mieux rendre compte du système prosodique du français.

En nous appuyant sur ces résultats, un travail conséquent a été également mené afin de dupliquer certaines de nos analyses. Après avoir identifié les différents syntagmes nominaux du corpus spontané, nous avons retenu les trois SN les plus fréquents. Nous avons mené un travail en perception dupliquant celui mené sur le corpus contrôlé. 73 sujets ont été testés sur 133 énoncés (de taille très variable, de 4 à 74 syllabes). Nous avons montré qu'ils étaient capables d'identifier à la fois les proéminences et les frontières sur des énoncés de parole spontanée très différents, et mieux que les scores des proéminences et des frontières étaient tout à fait compatibles avec les annotations expertes fondées sur une vision théorique très forte (Roux et al. 2016).

Les analyses sur les scores de proéminences menées tant en parole spontanée (Roux et al. 2016) qu'en parole contrôlée (Garnier et al. 2016) montrent ainsi une saillance plus importante pour les AI que pour les AF au niveau du mot, tendant à apporter à nouveau des arguments en faveur d'un système prosodique plus complexe qu'il n'y paraît.

Ce travail a eu d'importantes répercussions d'un point de vue théorique (Astésano 2016; Astésano & Bertrand 2016). Pour les tenants de la modélisation, le débat est toujours en cours, notamment dans la communauté psycho et neurolinguistique.

Poursuivre ces investigations nous semble par ailleurs important pour mieux rendre compte du système phonologique du français afin d'améliorer les systèmes automatiques de détection des événements prosodiques encore fort peu pourvus en ce sens (nous renvoyons également le lecteur aux travaux menés dans le projet PhraDiCo <https://uclouvain.be/fr/instituts-recherche/ilc/valibel/phradico-role-du-phrase-prosodique-pour-la-comprehension-du-discours.html>, Simon et al. 2016).

## Publications

- 
- Astésano, C. & **Bertrand**, R. (2016) "Accentuation et niveaux de constituance en français: enjeux phonologiques et psycholinguistiques", *Langue Française* 191, 11-30.
- Roux, G., **Bertrand**, R., Ghio, A., Astésano, C. (2016) "Naïve listeners' perception of prominence and boundary in French spontaneous speech", *Speech Prosody* 31 May-3 June, Boston (USA).
- Welby, P., **Bertrand**, R., Portes, C., Astésano, C. (2016) "Realization of the French initial accent : Stability and individual differences", *Proceedings of Tone and Intonation in Europe* (1er-3 sept 2016), Canterbury, Angleterre.
- Astésano, C., **Bertrand**, R., Espesser, R., Nguyen, N. (2013) "Dissociation between prominence and boundary phenomena in French : a perception study", *pS-prominenceS*, Viterbo, Italy, Dec. pages 12-14.
- Astésano, C.; **Bertrand**, R.; Espesser, R.; Nguyen, N. (2012) "Perception des frontières et des proéminences en français", In *Proceeding of JEP-TALN-RECITAL 2012*, Grenoble, 4-8 juin 2012, 353-360.

## Conclusion et nouvelles interfaces

Pour conclure sur ce dernier chapitre, je souhaiterais à nouveau insister sur ce qui apparaît désormais comme une réelle et nouvelle perspective dans laquelle le fossé entre conversationnalistes et prosodistes dont j'ai parlé en introduction semble se réduire. On observe un véritable changement de paradigme avec un intérêt croissant des uns et des autres, mais aussi des uns envers les autres, pour tenter de comprendre comment le sens se construit et s'élabore dans l'interaction, en envisageant les nombreuses interfaces dans une même perspective. Il est intéressant de constater que de plus en plus de travaux en linguistique interactionnelle observent ce qui est établi dans les modèles phonologiques et dans le même temps les prosodistes sont davantage enclins à regarder du côté des observations issues des études de corpus pour développer des paradigmes expérimentaux originaux et tester leurs hypothèses.

Afin d'illustrer cette tendance, il me paraît intéressant de présenter quelques pistes illustrant le développement de ces nouvelles interfaces, et qui ont notamment émergé à l'occasion de travaux d'étudiants que j'ai co-encadrés avec des collègues du laboratoire, spécialistes des aspects phonologiques de la prosodie. Ces travaux ont été réalisés dans le cadre de stages ou de Master.

- Prosodie et feedbacks appropriés en interaction médecin-patient

J'ai co-encadré avec ma collègue Caterina Petrone un travail effectué par Camilla Mastriani (étudiante Italienne, en stage durant un semestre). Son travail visait à mettre en évidence le rôle de la prosodie dans la production de réponses feedbacks appropriées, en fonction de leur contour intonatif, leur délai d'apparition après le tour précédent, et leur localisation dans l'activité en cours. Ce travail s'est inscrit dans le cadre du projet ANR *Acorformed* (coordonné par P. Blache, 2014-2017) dont l'objectif était d'améliorer la formation de médecins à l'annonce d'événements graves, en utilisant un agent artificiel pour le rôle du patient. Permettre à l'agent de produire les bonnes réponses s'avère crucial pour permettre au médecin d'améliorer ses performances. L'étude a été menée sur l'item "oui" qui est l'un des feedbacks les plus fréquents du corpus. Préalablement à l'étude perceptive principale destinée à évaluer la nature appropriée de l'item "oui", une étude sur corpus a permis de montrer que "oui" apparaissait préférentiellement dans deux phases identifiées dans l'activité d'annonce : celle de la *définition du problème* (DP) et celle *des implications futures* (IF) (Ochs et al. 2018). La phase DP apparaît préférentiellement au début de l'interaction considérée tandis que la phase IF se produit plutôt en fin. L'étude perceptive a ensuite été conduite sur 15 auditeurs. Elle a montré des résultats modérés pour le contexte d'apparition dans l'une ou l'autre phase, mais les résultats sur les contours intonatifs sont intéressants puisqu'ils modulent le score de jugement pour les femmes qui confirment nos hypothèses: le "oui" bouleversé est en effet jugé comme l'item le moins approprié dans la phase DP, tandis que c'est le "oui" questionnant pour les hommes. Un aspect intéressant montre que les hommes en revanche présentent un temps de réaction plus grand lorsqu'ils ont à juger le "oui" bouleversé en DP, alors que l'on s'attendrait à un délai plus court pour la production de la réponse dans cette phase. Ceci suggère que les hommes seraient plus sensibles au délai d'apparition plutôt qu'à l'intonation. Concernant la phase IF, les résultats ne sont pas significatifs à l'exception d'un petit effet sur le délai. Alors que nous attendions un délai plus long pour un "oui" bouleversé manifestant le trouble du patient, la faiblesse de nos effectifs ne nous permet pas de trancher. Cette première investigation conjuguant les contours intonatifs, les délais d'apparition et la localisation des items de feedbacks pour en déterminer la nature plus ou moins appropriée nécessite d'être poursuivie en élargissant à davantage de données ainsi qu'à d'autres feedbacks. L'implémentation de ces différents paramètres dans le patient virtuel sera alors cruciale car elle permettra d'améliorer les performances de l'agent artificiel qui, pour être un bon agent interactionnel, doit non seulement être pourvu des compétences pour produire son discours mais également pour montrer son écoute et sa compréhension.

### Publication

---

Mastriani, M.C., Petrone, C., **Bertrand**, R., Ochs, M. (2018) "Role of prosody on the perception of the "oui"/"yes" feedback in medical context", In *Proceedings 9th International Conference on Speech Prosody 2018*, 512-516, DOI: 10.21437/SpeechProsody.2018-104.

- Prosodie et affiliation

Un autre travail co-encadré avec C. Petrone a été effectué par Emilie Marty en Master 2, dans le cadre du même projet. Dans ce M2, l'objectif était de montrer que la notion interactionnelle de l'affiliation (Stivers 2009) détaillée dans le chapitre 2, consistant à produire le même point de vue émotionnel que le locuteur, selon qu'elle est produite depuis un rôle social différent (médecin versus ami), a un impact sur la manière dont on perçoit son interlocuteur, notamment en termes de compétence attribuée à ce dernier. L'affiliation a été étudiée par le biais de la prosodie émotionnelle. La majorité des études porte ordinairement sur l'impact perceptif de la prosodie émotionnelle des énoncés isolés. En étudiant l'affiliation prosodique, qui consiste à imiter la prosodie de son interlocuteur, le travail d'Emilie vise à

montrer comment la prosodie émotionnelle peut être utilisée comme indice de réussite de l'interaction. La question centrale du travail était la suivante : comment l'affiliation prosodique est-elle modulée par le rôle social des interactants? Et par ailleurs, s'affilier est-il perçu comme une preuve de compétence, étant entendu que paraître compétent pour le médecin est primordial pour instaurer une relation de confiance? L'étude a été menée sur une soixantaine d'énoncés constitués d'un énoncé cible suivi d'une hétéro-répétition. Le patient décrit ses symptômes en exprimant un stance "triste", et son interlocuteur répète cet énoncé soit en s'affiliant (triste), soit en exprimant l'énoncé de manière soit neutre soit surprise. L'hypothèse principale est que le cas d'affiliation prosodique, à savoir une prosodie triste par le médecin, pourrait le faire paraître moins compétent. Pour tester cette hypothèse, deux relations ont été comparées : la relation thérapeutique opposant le patient à son médecin, et une autre opposant le patient à un ami. Environ 60 auditeurs ont jugé du caractère affiliatif des énoncés selon qu'il s'agissait du médecin ou de l'ami, en lien avec le niveau de compétence qui leur était alors attribué. Les résultats ont montré que la perception de la compétence de l'interlocuteur en fonction de son affiliation prosodique (ou non) variait en fonction du rôle social. Comme attendu, le bon soignant est plutôt celui qui produit un énoncé neutre tandis que le bon ami est celui qui utilise la prosodie triste. Un autre aspect, moins attendu, montre que la prosodie neutre est perçue comme étant plus proche (plus affiliative) que la tristesse. Ceci suggère que les rôles sociaux pourraient avoir une incidence sur la perception même de la prosodie puisque pour un même stimulus audio, selon qu'il est produit par l'ami ou par le médecin, la perception de l'affiliation prosodique varie. Ces résultats encourageants nécessitent d'être poursuivis.

Emilie Marty a obtenu un contrat doctoral.

- Prosodie et indices prédicteurs des feedbacks

Toujours dans le cadre du projet Acorformed, des travaux d'étudiants ont été menés en collaboration avec Magalie Ochs (LIS) et Philippe Blache ces dernières années.

J'ai co-encadré avec ma collègue informaticienne M. Ochs un travail de Master 2 (Chris Porhet) portant sur les feedbacks produits par le patient. Le postulat de base était que la prise en compte des feedbacks est une condition incontournable pour améliorer l'interaction considérée et permettre au médecin d'améliorer sa communication au patient. Des règles liées à l'apparition des feedbacks ont donc été extraites afin d'implémenter ces signaux de feedback (nature, timing) dans un avatar face auquel le médecin pourrait s'entraîner (Porhet et al. 2017).

Ce travail a été poursuivi et a donné lieu à de nouvelles règles (Penteado et al. 2019).

### Publications

Penteado, B.E., Ochs, M., **Bertrand**, R. & Blache, P. (2019) "Evaluating Temporal Predictive Features for Virtual Patients Feedbacks", *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 88-90.

Porhet, C., Ochs, M., Saubesty, J., de Montcheuil, G. & **Bertrand**, R. (2017) "Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks", *International Conference on Multimodal Interaction (ICMI)*, Glasgow, Scotland.

Le dernier travail, initié en octobre 2020, est mené par Auriane Boudin dans son M2 en sciences cognitives. Il est co-encadré par M. Ochs, P. Blache et moi-même. L'objectif est de mettre en évidence l'ensemble des indices prédicteurs des feedbacks (via une grammaire sémantique) afin de les implémenter dans un système de dialogue qui sera ensuite lui-même intégré au système de l'agent artificiel. Le travail d'Auriane vise à établir un ensemble le plus exhaustif possible d'indices prédicteurs de tous les feedbacks présents, qu'ils soient verbaux, vocaux ou mimo-gestuels. Cette étude est l'une des premières qui cherche à prédire les feedbacks selon qu'ils sont génériques ou spécifiques. Ces derniers sont par ailleurs eux-mêmes distingués selon leur polarité (positive / négative) et leur caractère attendu / non attendu. En s'appuyant sur ces deux seules dimensions, nous avons pu constater que nous couvrons l'ensemble des usages des feedbacks analysés. Les feedbacks génériques appartiennent à la liste fermée établie dans Cofee (Prévot et al. 2015, 2016) concernant les items

verbo-vocaux. Du point de vue mimo-gestuel, nous retenons les hochements de tête et les sourires. Enfin, nous considérons les rires, certains énoncés plus complexes (“c’est clair oui”, “oh mon dieu”, etc) qui relèvent de la catégorie des spécifiques, ainsi que certaines hétéro-répétitions. Ce travail est mené sur un large éventail de données, composé du corpus Acorformed (corpus médecin-patient) et de corpus de conversations qui varient entre elles par le degré de familiarité entre les participants (cf CHEESE, Priego-Valverde et al. 2020; PACO, Amoyal et al. 2020 pour un total de 25 dyades environ). Le recours à ces différents corpus vise à établir un système qui soit le plus générique possible. Les indices examinés reposent à la fois sur une revue exhaustive de la littérature existante ainsi que sur des observations plus empiriques. Sont étudiés les catégories morphosyntaxiques, les indices prosodiques (intonation et durée syllabique), la durée des IPU, la présence des pauses silencieuses, les marqueurs discursifs, la polarité des lexèmes, entre autres. Un traitement statistique est en cours pour permettre de dégager le rôle de ces différents indices dans la prédiction d’un feedback générique/spécifique, unimodal ou bimodal. A partir de ces traitements statistiques, nous pourrions définir les règles de la grammaire sémantique qui sera implémentée dans le système de dialogue développé par P. Blache au LPL. Le modèle sera ensuite évalué dans un agent artificiel (M. Ochs, LIS).

Depuis la revue de littérature de Cole (2015) qui regrettait le manque d’étude de la dimension prosodique dans les diverses questions relatives aux aspects discursifs et interactionnels, on peut désormais lister de plus en plus de travaux concernant l’organisation des tours de parole, la production des feedbacks, les phénomènes de convergence et d’imitation prosodique, mais aussi ceux concernant le rôle des émotions dans l’interaction, entre autres. Ces travaux s’inscrivent dans les recherches récentes des travaux expérimentaux sur le dialogue qui suggèrent que les patrons prosodiques sont multidimensionnels et que leur interprétation est modulée également par des facteurs contextuels et sociaux (Prieto 2015). Nos travaux et les développements que nous envisageons trouvent ainsi toute leur place dans cette nouvelle perspective.

## Chapitre IV. Conclusion et Perspectives

---

Sur la base des connaissances acquises sur l'interaction en face à face aux différents niveaux linguistiques et esquissées dans cette synthèse, nous envisageons de poursuivre nos investigations sur plusieurs points que nous déclinons dans trois projets principaux. De la même manière que pour nos études antérieures, nous développerons nos études dans le cadre de projets formalisés (projets ANR par exemple), ou lors de l'encadrement d'étudiants en master ou doctorat.

### Projet 1: Alignement et affiliation

#### Analyse séquentielle des hétéro-répétitions (other-repetitions OR)

Les analyses menées sur les réponses spécifiques des interlocuteurs (chapitre 2) ont confirmé plusieurs hypothèses que nous souhaitons tester dans la suite de notre travail. D'abord, nous affirmons que la similarité, quelle qu'elle soit (lexicale, prosodique, posturale, etc) n'est pas suffisante pour faire émerger une séquence interactionnelle convergente. Notre premier objectif est de mener une étude séquentielle dans laquelle nous analyserons les différentes OR en lien avec les concepts d'alignement et d'affiliation. Nous observerons également les trajectoires interactionnelles qu'elles entraînent, relativement à leurs fonctions et aux activités dans lesquelles elles sont produites.

Outre la collection de 400 répétitions déjà disponibles et annotées fonctionnellement sur le CID, nous avons lancé l'annotation automatique des OR sur les deux autres corpus Cheese et Paco. Ceci nous permettra de disposer de contextes conversationnels encore différents, et nous pourrions également estimer si le fait de se connaître ou non, pour les participants (Cheese vs Paco) a une incidence non seulement sur l'apparition des OR, mais surtout sur le type et les fonctions qu'elles jouent.

Nous espérons par ailleurs pouvoir avancer également sur les questions relatives au processus plus ou moins automatique relatif à la production des OR. Bien avant les travaux de Pickering & Garrod et leur modèle d'alignement automatique, la linguiste Tannen (1980) souligne le fait que la répétition permettrait de produire du langage d'une manière plus efficace et à moindre coût, ceci plaidant pour une forme d'automatisme dans la production des répétitions. Si s'aligner automatiquement consiste à reprendre les mêmes termes, les mêmes patrons prosodiques ou les mêmes patrons mimo-gestuels, qu'en est-il des répétitions qui sont utilisées pour se démarquer, se désaligner ?

En menant une analyse séquentielle très fine en lien avec les notions d'alignement et d'affiliation, en tenant compte des activités dans lesquelles sont produites les OR, leur fonction, leur localisation et le timing auquel elles sont soumises, ainsi que leurs corrélats prosodiques, notre ambition est d'affiner les connaissances relatives aux règles et aux contraintes qui pèsent sur les participants et qui expliqueraient les trajectoires interactionnelles qui conduisent ou non à une séquence convergente. Nous espérons également pouvoir avancer des arguments en faveur d'un processus plus ou moins automatique. Certaines répétitions ne font que reprendre les mêmes mots, comme pour créer un lien, assurer la cohérence du discours, et par là montrer un alignement de son propre modèle de situation avec celui de l'autre ; or, si l'on s'attache à la suite du discours, on peut observer que le locuteur revient, discute, nuance voire nie le contenu même de la répétition. Ceci pourrait être un argument en faveur de la nature automatique de la répétition. Mais dans d'autres cas, on observe également des cas dans lesquels on rebondit via la répétition pour immédiatement réfuter, se distancer, etc, ce qui semblerait relever alors davantage d'un processus de nature stratégique. Pour l'un ou l'autre cas, nous nous interrogerons sur la question relative à ce que Pickering & Garrod qualifie de common ground complet, et quand il semble y avoir nécessité de se représenter explicitement le CG de son interlocuteur? Plus globalement, ce postulat très fort chez P & G selon lequel il y aurait un alignement automatique des états mentaux et des niveaux linguistiques qui ne pourrait se réaliser que sous

réserve d'une très forte collaboration, ne se résumerait-il pas, souvent, à une simple collaboration de surface, précisément réalisée par ce type de procédé de répétitions ?

## Projet 2 : Les feedbacks

Le travail sur les feedbacks<sup>9</sup> dans la littérature mais aussi dans notre propre recherche est abondant mais il reste encore de nombreuses voies à explorer.

### Caractéristiques formelles

Dans le projet CoFee mené par L. Prévot, un travail conséquent a été effectué pour définir les fonctions communicatives des feedbacks. Mais des études supplémentaires permettraient de mieux les caractériser formellement, d'un point de vue prosodique notamment. Des études prenant en compte les feedbacks mimo-gestuels restent à mener également. Grâce à des outils tels qu'OpenFace, largement utilisé aujourd'hui au laboratoire, nous pouvons plus facilement identifier et annoter des items mimo-gestuels pour les caractériser plus finement et les considérer de manière plus intégrée dans un modèle de l'alternance des tours de parole. En ce sens, le travail mené dans le cadre du mémoire de Master d'Auriane Boudin notamment (que je co-encadre avec M. Ochs et P. Blache), relatif aux indices prédictifs des feedbacks multimodaux, visant à être intégrés dans un système de dialogue puis implémentés dans un agent virtuel, constitue un de nos projets actuels très important. Depuis la rédaction de cette synthèse, le travail d'Auriane a été soumis et accepté à la conférence *Text, Speech and Dialogue* (TSD21), et sera publié sous la forme d'un recueil de chapitres publié par Springer Verlag.

Par ailleurs, Auriane a été classée 1<sup>ère</sup> du Master2 de Sciences Cognitives et a obtenu une bourse de doctorat auprès de l'ILCB (classée 1<sup>ère</sup> également).

### Publication

---

Boudin, A., **Bertrand**, R., Rauzy, S., Ochs, M., & Blache, P. (accepté) "A Multimodal Model for Predicting Conversational Feedbacks", Springer Nature Switzerland AG 2021 K. Ekstein et al. (Eds.): *TSD 2021*, LNAI 12848, 1-13

Parmi ces phénomènes de feedbacks, une classe spécifique (les *grunts*, comportant les grognements, les soupirs, etc) a été annotée sur le CID en collaboration avec Laurent Prévot. Cette annotation disponible mérite d'être examinée plus spécifiquement, et faire l'objet d'un mémoire de master.

### Intégration des aspects informatiques et neuronaux dans l'étude de l'interaction

Dans le cadre du projet ANR Acorformed coordonné par P. Blache, nous avons développé un axe de recherche qui s'ancre très fermement sur les résultats de nos travaux précédents tout en ouvrant des perspectives qui nous permettent de tester des hypothèses linguistiques très intéressantes en collaborant avec des collègues en informatique et en sciences cognitives.

Le projet PEPS BRAIN-IHM (coordonné par M. Ochs) concerne l'utilisation d'un avatar pour tester certaines hypothèses qu'il ne serait pas possible de tester avec des humains. Ainsi, nous cherchons à comprendre le fonctionnement des feedbacks, plus spécifiquement à travers l'impact de la nature et du délai d'apparition de ces derniers dans la poursuite et la réussite de l'interaction.

Dans ce projet, nous nous interrogeons donc sur les corrélats cérébraux associés à la production d'un feedback non préféré (inapproprié). Pour ce faire, nous avons développé un axe plus centré sur les aspects perceptifs (vs études en production) des interactions.

---

<sup>9</sup> Outre les hétéro-répétitions qui, pour certaines, entrent également dans cette catégorie

Nous avons d'ores et déjà constitué un corpus (cf projets Acorformed et PEPS, CV) qui, à notre connaissance, n'a pas d'équivalent au niveau international. Nous présentons ci-après un résumé du protocole expérimental.

#### Publication

Ochs, M., Blache, P., de Montcheuil, G., Pergandi, J. M., **Bertrand**, R., Saubesty, J., Francon, D. & Mestre, D. (2018) "The Acorformed Corpus: Investigating Multimodality in Human-Human and Human-Virtual Patient Interactions", In *Selected papers from the CLARIN Annual Conference 2018*, Pisa, 8-10 October 2018, 159, 108-115. Linköping University Electronic Press.

Le corpus est inspiré d'interactions médecin-patient.e (patiente en l'occurrence) dans lesquelles le médecin doit annoncer une mauvaise nouvelle à son interlocutrice. Plusieurs scénarios relatifs à différentes interventions chirurgicales ont été imaginés. Les rôles discursifs sont fixes : le médecin est le locuteur principal et la patiente est principalement dans une situation d'écoute. Les scénarios sont écrits. Afin d'assurer une certaine validité écologique à ces situations d'interaction produites dans un environnement expérimental très contraint, nous avons joué une première interaction et laissé la patiente produire le plus naturellement possible les feedbacks (quels qu'ils soient) qui lui semblaient les plus appropriés du point de vue sémantique et en termes de localisation. Sur cette base, nous avons ensuite intégré les feedbacks au scénario déjà existant pour le médecin.

3 feedbacks cible ont été retenus: *ah oui* (surprise et haussements de sourcils), *voilà* (confirmation et hochement de tête type nod), *ah non* (rejet et mouvement de tête type shake). Ce sont ces trois feedbacks qui seront ensuite déplacés pour créer des situations incongrues. Au total 18 feedbacks congrus/incongrus par vidéo (chacun produit 6 fois dans l'interaction = 6 X 3 pour les vidéos congrues, ou 3 X 3 congrus + 3 X 3 incongrus) seront examinés.

D'autres feedbacks seront également produits au cours de l'interaction afin de garantir le maintien et la poursuite de l'interaction: oui (bouleversé et mouvement de tête), d'accord (confirmation et nod), oui (acquiescement et nod), mhm.

Le protocole expérimental consiste en une interaction (contrôle) entre deux humains, à la suite duquel l'agent artificiel (en l'occurrence Greta) remplace la patiente humaine et produit les feedbacks de manière similaire à l'humain. La patiente, humaine ou agent, est alors intégrée dans un environnement virtuel similaire, créé au Centre de réalité virtuelle de Marseille (voir figures 25a et b). Un exemple de scénario est présenté en Annexe 1 (p. 108).

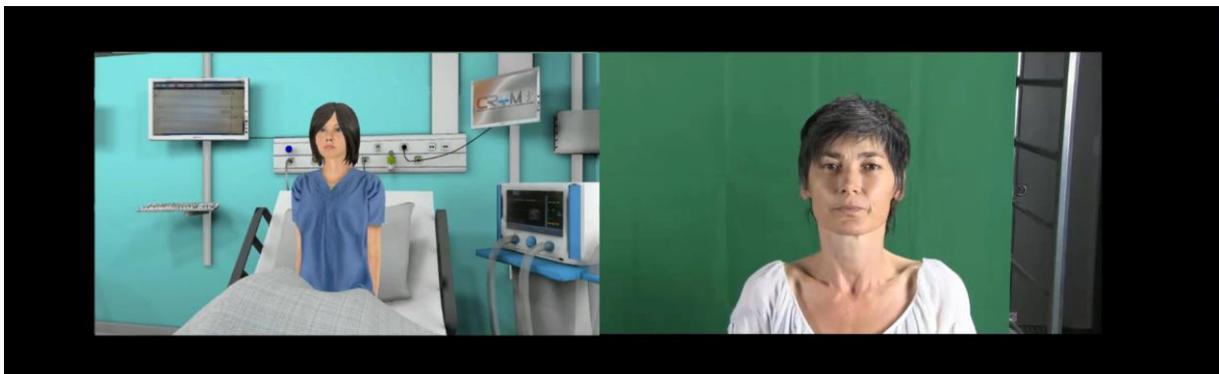


Figure 25a : pré-enregistrement des dialogues du corpus BRAIN-IHM : à droite la patiente humaine sur laquelle seront calqués les comportements de l'agent virtuel Greta (à gauche) dans son environnement virtuel

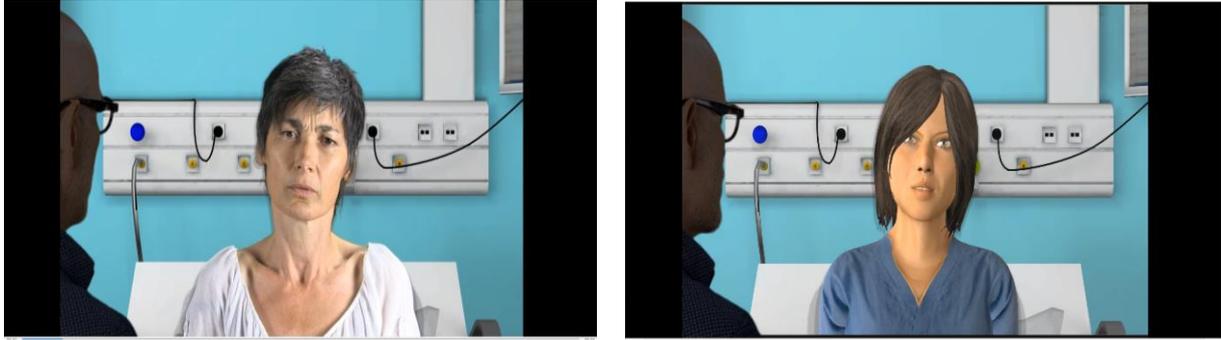


Figure 25b: patiente humaine (à gauche), et virtuelle (à droite) dans le même environnement virtuel et produisant le même feedback (ah non, avec froncement de sourcil et tête légèrement penchée à droite)

La dernière étape de l'expérience consiste à équiper des participants qui vont observer différentes interactions humain-humain et humain-machine. L'objectif est de mettre en évidence au niveau cérébral (potentiels évoqués) une trace potentielle du caractère inapproprié du feedback en question par rapport au feedback attendu (figure 26).



Figure 26 : préparatifs pour l'expérience EEG dans laquelle un participant visionne sur un écran les différentes interactions (humain-humain / humain-machine)

Les analyses concernant les corrélats cérébraux associés à la perception d'un feedback approprié versus inapproprié (en fonction de son type et de sa localisation), dans un dialogue humain-humain et humain-avatar, sont en cours. Des premiers résultats très encourageants font état notamment de l'existence de potentiels évoqués spécifiques (P600) notamment pour le sujet humain versus l'agent virtuel, en condition incongrue. Cette étude vise plus largement à interroger la nature plus ou moins automatique de la production des feedbacks en interaction.

### Avatar et persuasion

Depuis 2018, je suis impliquée dans le projet ANR Copains (*Cognitive Planning in Persuasive Multimodal Communication* (resp. E. Lorini, IRIT, Toulouse), en tant que responsable scientifique locale pour le LPL. L'objectif du projet est de contribuer à doter un agent virtuel de la capacité à persuader des personnes âgées à faire du sport, dans le but de garantir leur santé et leur bien-être. Côté LPL, nous sommes impliqués dans la constitution du corpus et dans l'identification et la caractérisation des indices multimodaux de la persuasion dans les dialogues humain-humain et humain-machine. Les difficultés liées à la crise sanitaire n'ont pas permis la création du corpus dans lequel nous étions impliqués avec des collègues de l'INSERM à Dijon. Au stade actuel, après avoir effectué une revue de la littérature la plus exhaustive possible, afin de développer notre schéma d'annotation futur, nous cherchons à caractériser sur des corpus existants les indices multimodaux susceptibles de favoriser une séquence persuasive afin d'appliquer et d'évaluer notre schéma d'annotation sur des données concrètes.

Parallèlement, depuis février 2021, Afef Cherni travaille dans le cadre de son post-doctorat avec M. Ochs et moi-même sur le corpus existant *POTUS* (cf Janssoone et al. 2020) pour lequel des annotations ont été utilisées comme classes d'apprentissage pour les algorithmes d'apprentissage automatique appliqués aux signaux sociaux. Pour rendre compte du niveau de persuasion du sujet étudié (en l'occurrence l'ancien président Obama), sont étudiées la dominance et l'amicalité et les indices multimodaux tels que les mouvements de tête, les unités d'action faciale, ou encore certaines mesures de f0. Les résultats basés sur divers algorithmes sont en cours d'étude.

Enfin, plus récemment, le corpus *TRUENESS* (Théâtre-forum virtuel pour lutter contre les discriminations, projet coordonné par M. Ochs & P. Blache) a été enregistré pour travailler sur les aspects de discrimination. Plus précisément, ce projet vise à développer un environnement de réalité virtuelle simulant le contexte d'un "théâtre-forum" utilisé dans le cadre de la lutte contre les discriminations sociales. Un tel développement nécessite la création d'un agent conversationnel animé jouant le rôle d'un individu à l'origine d'une discrimination, et capable ensuite de dialoguer avec le sujet humain témoin de la scène en question. Il va donc s'agir de développer un modèle de comportement verbal et mimo-gestuel de l'agent. La première étape a donc été de recueillir un corpus pour ce type de situation afin de les analyser, ce qui a été fait très récemment (juin 2021). Plus particulièrement, les scénarios conçus consistent en des scènes de discrimination directe dans lesquelles des acteurs jouent une scène de discrimination sexuelle ou raciale. D'autres scènes de confrontation suivent cette première scène, scènes au cours desquelles le témoin de la première scène tente de faire prendre conscience à l'auteur de la discrimination de son comportement, en vue, à terme, de le persuader d'en changer. L'objectif de ce projet consiste ainsi à reproduire les scènes de discrimination sur des agents virtuels animés. Les scènes de confrontation quant à elles nous permettront d'analyser plus particulièrement les caractéristiques comportementales multimodales des acteurs en situation de persuasion notamment (vitesse de mouvement, analyse des unités faciales, prosodie, etc) en fonction des différentes attitudes des acteurs (agressif, conciliant, froid) pour les mettre en lien notamment avec les premiers résultats obtenus sur le corpus *POTUS*. En Annexe 2 quelques détails supplémentaires sur ce corpus *TRUENESS*.

### Impact général sur les agents conversationnels animés

Dans différents travaux impliquant un avatar, nous avons pu bénéficier de la plateforme GRETA (De Sevin et al. 2010). Aux dires de ses principaux développeurs eux-mêmes, l'architecture actuelle de GRETA ne permet pas de tenir compte en temps réel de divers phénomènes inhérents et définitoires de l'interaction. Parmi eux, on peut citer les phénomènes très fréquents de disfluences, mais aussi les phénomènes d'alignement/convergence, ainsi que les réponses feedbacks qui, outre les génériques qui montrent l'écoute et la compréhension mutuelle, véhiculent aussi avec les feedbacks spécifiques des aspects davantage liés aux points de vue des locuteurs, à leur stance, leur état émotionnel, etc. L'un de nos objectifs et aspiration consiste en ce que les travaux présentés au cours de cette synthèse mais aussi ceux à venir contribuent à alimenter les connaissances sur l'interaction et soient utilisés par les concepteurs / développeurs d'agents virtuels pour en faire des agents encore plus et mieux « adaptés ». Le caractère adapté des avatars renvoie au fait que ces derniers soient capables de produire mais aussi de reconnaître les actions véhiculées par le langage via les différents niveaux linguistiques et les modalités impliqués. Cette capacité à reconnaître et produire en temps réel ce qui se déroule dans l'interaction en face à face est indispensable pour permettre aux agents virtuels d'assumer certains rôles qui leur sont de plus en plus destinés au niveau sociétal, en termes de formation ou de santé par exemple. Le projet ANR Acorformed auquel j'ai participé, et plusieurs fois évoqué dans cette synthèse, visait précisément à améliorer l'annonce d'une mauvaise nouvelle par un médecin. L'agent virtuel devait jouer le rôle du patient face auquel les médecins pouvaient s'entraîner. Ceci avait pour but de permettre un gain de temps mais aussi d'argent dans ces formations indispensables, selon la *Haute Autorité de la Santé*, étant entendu que jusqu'alors les formations sont effectuées grâce à des volontaires bénévoles. Le second projet ANR Copains dans lequel je suis

toujours impliquée en tant que responsable scientifique pour le LPL, vise également à dôter l'agent virtuel impliqué, de compétences sociales permettant de le rendre plus persuasif. La littérature existante (cf Granier, Bertrand & Ochs 2021 pour une revue) montre la variété des indices multimodaux mis en jeu dans une telle pratique, qu'il s'agit de mieux caractériser et formaliser afin de pouvoir les intégrer dans le système. A cet égard, il me semble important de souligner qu'il est bien entendu très important pour les chercheurs intéressés de près ou de loin par l'intelligence artificielle d'être extrêmement précautionneux sur les implications éthiques que pourrait avoir ce type de recherche. Il est de notre responsabilité, même si nous ne sommes pas directement impliqués dans le développement d'agents, de veiller à ce que nos résultats ne soient pas détournés du but initial qui doit toujours être la recherche du meilleur pour l'humain. Pour éviter de s'égarer, il est important de ne jamais négliger ces aspects éthiques et de leur donner la place qu'ils méritent (cf un petit article signé avec trois collègues dans *TheConversation* destiné au grand public et qui insiste sur ce point, Ochs et al. 2020).

Nous souhaitons mentionner enfin un dernier point relatif à l'utilisation d'agents artificiels, d'avatars ou encore de tête parlante, en un mot d'intelligence artificielle dans la recherche en sciences du langage. Les agents artificiels peuvent également s'avérer intéressants dans la mesure où ils peuvent nous offrir l'opportunité d'examiner des phénomènes interactionnels quelquefois difficiles à obtenir dans l'interaction humain-humain. Typiquement, c'est ce que nous avons tenté de faire dans le projet PEPS présenté ci-dessus, en réinjectant dans la parole de l'agent des feedbacks inappropriés en vue de mesurer l'impact qu'ont ces derniers selon qu'ils sont produits par un humain ou un agent artificiel, ce qui nous permettra d'affiner nos connaissances sur la nature des processus en jeu dans la compréhension des dialogues.

### Projet 3 : Détermination des Unités conversationnelles

Malgré là encore une abondante littérature sur les questions relatives aux tours de parole, l'identification et la caractérisation des unités de tours sont encore très loin d'être atteintes. Or, bénéficier de ces unités de tours, conversationnelles, permettraient de mieux comprendre l'alternance de la parole. Actuellement, les études sont souvent menées sur les IPU qui restent des unités relativement grossières, qui ne permettent pas l'examen minutieux des tours de parole. Les unités conversationnelles nécessitent une définition précise des unités prosodiques, syntaxiques et discursives (pour le travail fondateur, voir Ford & Thompson 1996 et Selting 1998). Pour disposer de ces unités, il faut donc non seulement les avoir caractérisées préalablement d'un point de vue théorique, mais aussi les avoir annotées. Une autre tâche consiste à analyser les liens entre les unes et les autres, liens qui ne sont pas toujours aisés à comprendre.

Globalement la littérature recense de nombreuses études relatives aux liens entre syntaxe et prosodie, en particulier syntaxe et intonation. Celles qui nous intéressent plus spécifiquement renvoient aux travaux qui ont tenté d'intégrer la dimension discursive/pragmatique des discours, laquelle se fonde notamment sur la dimension syntaxique. Pour le français, Degand & Simon 2009, Simon & Degand 2011, ou Lacheret et al. (2011) ancrent pour une part leur analyse sur les aspects syntaxiques, avec pour les premières un lien entre ces structures et les représentations sémantiques (un « context model ») tandis que pour les secondes les unités syntaxiques sont basées sur l'approche macro et micro syntaxique ainsi que sur la dimension illocutoire des énoncés. Grâce au travail conséquent qui a été mené tant au niveau théorique qu'au niveau des annotations concrètes sur les 8h du CID, nous disposons aujourd'hui d'annotations extensives au niveau discursif, prosodique, syntaxique mais également au niveau des disfluences qui ont été annotées très précisément sur l'ensemble du corpus. En nous fondant sur une approche relativement similaire à celle de Degand & Simon, notre objectif vise donc à mieux définir les unités de construction des tours, qui s'avèrent essentielles pour mieux comprendre comment s'instancie la compréhension mutuelle favorisant la réussite de l'interaction.

Au sein de ce projet, plusieurs axes sont envisagés.

- Unités discursives, prosodiques et disfluences  
Dans un premier temps, en collaboration avec L. Prévot et S. Rauzy, nous avons pour objectif d'exploiter ces données pour mettre en lumière l'impact que les disfluences ont sur les différents niveaux de segmentation. Ceci nous permettra de comprendre certains mismatches potentiels. En effet, les disfluences sont une grande source de difficultés pour chacun des niveaux si elles ne sont pas clairement distinguées. Elles peuvent opacifier par exemple les unités prosodiques dans la mesure où certains paramètres de disfluences sont également des paramètres impliqués dans l'identification des unités prosodiques.  
Une fois que nous aurons mis en lumière et résolu certaines difficultés liées à l'identification des différentes unités, nous pourrions déterminer avec plus d'exactitude et de certitude les unités conversationnelles qui restent totalement sous-spécifiées à ce jour. Pour le français, nous nous inscrivons dans l'approche déjà évoquée ci-dessus représentée par les études de Degand & Simon (2009) ou Lacheret & Kahane (2020). Ces derniers défendent une approche intéressante pour la mise en relation des différents niveaux examinés, qui s'oppose par exemple à une approche trop centrée syntaxe ou pragmatique et qui pourrait favoriser l'émergence de constructions ou de dispositifs prosodiques nouveaux :

*Le traitement intonosyntaxique qui fonde la démarche consiste à appréhender la prosodie et la syntaxe comme des processus dynamiques d'assemblages plus ou moins complexes d'éléments (...), qui reposent sur différents modes de synchronisation temporelle entre les deux niveaux d'organisation du message. Ces assemblages conduisent à l'émergence en temps réel d'unités prosodiques dont il s'agit dans un premier temps d'établir les propriétés formelles, pour en proposer ensuite une lecture fonctionnelle et enfin préciser les variables situationnelles, en particulier les genres de discours, qui sous-tendent ces constructions (2020: 4)*

Depuis la rédaction de cette synthèse, nous avons soumis un travail qui sera présenté à la conférence DISS visant à mieux cerner l'impact des disfluences sur le (mis)matching entre unités prosodiques et unités discursives.

#### Publication

---

Prévot, L., **Bertrand**, R., & Rauzy, S. "Investigating Disfluencies contribution to Discourse-Prosody Mismatches in French Conversations, Proceeding of DiSS 2021, 25-27 August 2021, Paris 8 University, France

- Mieux caractériser les unités prosodiques en spontané  
Dans un second temps mais dans la continuité du sous-projet précédent, nous avons pour projet de reprendre sur le même jeu de données la caractérisation d'une partie des unités intonatives établies sur le CID. Avec C. Portes, nous avons proposé que les unités phonologiques puissent être interrompues, ce qui n'a pas été, à notre connaissance, proposé par ailleurs. Or ce point est crucial pour renforcer la nature indéniablement phonologique de ces unités. Grâce aux disfluences annotées dont nous disposons désormais, qui sont un élément pouvant contribuer à expliquer notre proposition, nous envisageons de réexaminer empiriquement les unités concernées afin de préciser le cas échéant notre hypothèse.
- Marqueurs discursifs, unités prosodiques et implications interactionnelles  
Enfin, un travail sur certains marqueurs discursifs pourra être conduit grâce aux nombreux niveaux d'annotations dont nous disposons. Avec C. Meunier, qui s'intéresse aux phénomènes de réduction dans la parole spontanée, nous souhaiterions mener une analyse sur les corrélats phonétiques de certains marqueurs discursifs pour les mettre en lien avec leurs implications discursives et interactionnelles. Ces phénomènes de marqueurs discursifs notamment sont également cruciaux car ils sont souvent associés à des frontières prosodiques. Leur examen pourrait donc également contribuer à améliorer l'identification des unités prosodiques.

# Bibliographie

---

- Aijmer, K., Rühlemann, C. (2015) *Corpus Pragmatics. A Handbook*. Cambridge University Press
- Amoyal, M., Priego-Valverde, B., & Rauzy, S. (2020) "PACO: A corpus to analyze the impact of common ground in spontaneous face-to-face interaction", *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 621-626. Marseille
- Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., MC Allister, J., Miller, J., Sotillo, C., Thompson, H., Weinert, R. (1991) "The HCRC map task corpus", *Language and Speech* 34, 351-366
- Astésano, C. (2016) "The Prosodic Characterization of Reference French", In Detey, S., Durand, J., Laks, B., Lyche, Ch. (Eds.), *Varieties of Spoken French: a source book* Oxford University Press, 68-85
- Astésano, C. (2001) *Rythme et accentuation en Français : invariance et variabilité stylistique*, Paris : L'Harmattan
- Astésano, C. & Bertrand, R. (2016) "Accentuation et niveaux de constituance en français: enjeux phonologiques et psycholinguistiques", *Langue Française* 191, 11-30
- Astésano, C., Bertrand, R., Espesser, R., Nguyen, N. (2013) "Dissociation between prominence and boundary phenomena in French: a perception study". *pS-prominenceS*, Viterbo, Italy, Dec.12-14
- Astésano, C., Bertrand, R., Espesser, R., Nguyen, N. (2012) "Perception des frontières et des proéminences en français", *Actes Journées d'études sur la Parole* (2012 juin 4-8: Grenoble, FRANCE), 8 pages
- Auer, P. (2005) "Projection in interaction and projection in grammar", *Text* 2, 7-36. <http://doi.org/10.1515/text.2005.25.1.7>
- Auran, C., Bertrand R., Chanet, C., Colas, A., Di Cristo, A., Portes, C., Régnier, A., Vion, M. (2006) *Proceedings of Discourse-Prosody Interface Symposium - IDPO5* (Eds), Cederom
- Authier-Revuz, J. (1982) "Hétérogénéité montrée et hétérogénéité constitutive, éléments pour une approche de l'autre dans le discours", *DRLAV*, 26, 91-115
- Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018) "Openface 2.0: Facial behavior analysis toolkit", *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 59-66
- Bard, E.G., Anderson, A.H., Sotillo, C., Aylett, M., Doherty-Sneddon, G. & Newlands, A. (2000) "Controlling the Intelligibility of Referring Expressions in Dialogue", *Journal of Memory and Language* 42, 1-22
- Baude, O., Benveniste, C.B., Calas, M.F., Cappeau, P., Cordereix, P. et al. (2006) *Corpus oraux, guide des bonnes pratiques 2006*, CNRS Editions, Presses Universitaires Orléans, 203, 2006. (hal-00357706)
- Bavelas, J.B., Coates, L. & Johnson, T. (2000) "Listeners as Co-narrators", *Journal of Personality and Social Psychology* 79 (6): 941-952
- Benus, S., Gravano, A. & Hirschberg, J. (2011) "Pragmatic aspects of temporal accommodation in turn-taking", *J. Pragmat.* 43 (12), 3001--3027
- Benzitoun, C., Debaisieux, J.M. & Deulofeu, H.J. (2016) "Le projet ORFEO : un corpus d'étude pour le français contemporain", *Corpus* [En ligne], 15 | 2016, <http://journals.openedition.org/corpus/2936> ; DOI : <https://doi.org/10.4000/corpus.2936>
- Bertrand, R. & Espesser, R. (2017) "Storyteller and listener's behavior in French conversational storytelling", *Journal of Pragmatics* 111, 33-53
- Bertrand, R. & Goujon, A. (2017) "(Dis)aligning for improving mutual understanding in talk-in-interaction", *Revue Française de Linguistique Appliquée (RFLA)*, vol XXII-2, 53-70
- Bertrand, R. & Priego-Valverde, B. (2017) "Listing practice in French conversation. From collaborative achievement to interactional convergence", *Discours* (en ligne)
- Bertrand, R. & Priego-Valverde, B. (2013) "Does prosody play a specific role in conversational humor? ", In S. Attardo, M. M. Wagner and E. Urios-Aparisi, *Prosody and Humor*. USA: John Benjamins Publishing Company, 143-165
- Bertrand, R., Ferré, G. & Guardiola, M. (2013) "French face-to-face interaction: Repetition as a multimodal resource", In N. Campbell & M. Rojc (eds) *Coverbal Synchrony in Human-Machine Interaction*. USA: Science Publishers, 30 pages
- Bertrand, R. & Portes, C. (2012) "Pour une approche phonologique de la prosodie dans l'interaction", *Langue française*, vol. 2012/3, 175, 19-36

- Bertrand, R., Ader, M., Blache, P., Ferré, G., Espesser, R. & Rauzy, S. (2009) "Représentation, édition et exploitation de données multimodales: le cas des backchannels du corpus CID", *Cahiers de Linguistique*, Numéro spécial Grands corpus de français parlé, S., Bruxelles, L., Mondada, A.C., Simon, V., Traverso (eds), 33/2, 183-212. ISSN 0771-6524
- Bertrand, R., Portes, C. & Sabio, F. (2008) "Distribution syntaxique, discursive et interactionnelle des contours intonatifs du français dans un corpus de conversation", *Travaux Neuchâtelois de Linguistique*, 47, 59-77. ISSN 1010-1705
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. (2008) "Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle", *Traitement Automatique des Langues*, 49-3, 105-134
- Bertrand, R., Ferré, G., Blache, P., Espesser, R. & Rauzy, S. (2007) "Backchannels revisited from a multimodal perspective", *Proceedings of Auditory-visual Speech Processing* (2007 août 31-septembre 3: Hilvarenbeek, THE NETHERLANDS), 1-5. Cederom
- Bertrand, R. & Chanet, C. (2005) "Fonctions pragmatiques et prosodie de *enfin* en français spontané", *Revue de Sémantique et Pragmatique*, 17, 41-68
- Bertrand, R. (2003) "Etre soi avec les mots d'autrui", In J.M. Merle (Ed) *Le sujet*, 269-280
- Bertrand, R. & Espesser, R. (2002) "Voice diversity in conversation: a case study", In B. Bel; I. Marlien (eds.) *Proceedings of Speech Prosody 2002*. Aix-en-Provence: Laboratoire Parole et Langage, 171-174
- Bertrand, R. & Espesser, R. (2002) "Prosodic Cues of Speaker Changes in Speech Overlap Phenomena in French", *Rapport non publié*
- Bertrand R. & Espesser R. (2001) "About Speech Overlaps: Prosodic Cues Contribution in Predicting a Change of Speaker", In Puppel, S. and Demenko, G. (eds), *Proceedings of Prosody 2000*, Adam Mickiewicz University, Pologne, 29-35
- Bertrand, R. (1999) De l'Hétérogénéité de la Parole: Analyse Énonciative de Phénomènes Prosodiques et Kinésiques dans l'Interaction Interindividuelle. *Thèse de Doctorat*, Université de Provence
- Bertrand, R. & Espesser, R. (1998) "Prosodie et discours rapporté: La mise en scène des voix", In Verschueren J. (ed.) *Pragmatics in 1998*. Antwerp, Belgium: Int. Pragmatics Association, vol.2, 45-56
- Bertrand, R. & Espesser, R. (1997) "Gestion dialogique et dialogale d'une mise en scène énonciative spécifique: prosodie du discours rapporté en conversation", *Actes du Congrès International des Linguistes* (26 : 1997 juillet 20-25 : Paris, France). Cederom
- Bertrand, R. & Casolari, F. (1996) "Approche prosodique et pragmatique des modulations", *Actes des Journées d'Etude sur la Parole (JEP)* (Avignon, France), 179-182
- Bertrand, R., Boyer, J., Cavé, C., Guaitella, I., Santi, S. (1995) "Relationship between gestures and voices in verbal interaction: prosodic and kinesic aspects of back-channel signals", *Proceedings of the XIII<sup>th</sup> ICPhS 95* (Stockholm, Suède), 746-749
- Bigi, B. & Bertrand, R. (2016) "Laughter in French spontaneous conversational dialogs", 10th *edition of the Language Resources and Evaluation Conference* (LREC), 23-28 May 2016, Portorož (Slovenia)
- Bigi, B., Bertrand, R. & Guardiola, M. (2014) "Automatic detection of other-repetition occurrences: application to French conversational speech", *Proceedings of The 9th edition of the Language Resources and Evaluation Conference* (2014 mai 26-31: Reykjavik, ICELAND)
- Bigi, B., Péri, P. & Bertrand, R. (2012) "Orthographic Transcription: Which Enrichment is required for Phonetization?" *The 8th International Conference on Language Resources and Evaluation* (2012 mai 23-25: Istanbul, TURKEY), 1756-1763, ISBN 978-2-9517408-7-7
- Bigi, B., Péri, P. & Bertrand, R. (2012) "Influence de la transcription sur la phonétisation automatique de corpus oraux", *Journées d'Étude sur la Parole*, pages 449-456, Grenoble (France)
- Bigi, B., Meunier, C., Nesterenko, I., Bertrand, R. (2010) "Automatic detection of syllable boundaries in spontaneous speech", *Proceedings of Language Resource and Evaluation Conference* (2010 mai 19-21: La Valette, MALTA), 3285-3292. Cederom
- Bigi, B., Meunier, C., Bertrand R., Nesterenko, I. (2010) "Annotation automatique en syllabes d'un dialogue oral spontané", *Actes XVIIIèmes Journées d'Étude sur la Parole* (2010 mai 24-28: Mons, BELGIUM), 4 pages. Cederom
- Bigi, B., Bertrand, R. & Guardiola, M. (2010) "Recherche automatique d'hétéro-répétitions dans un dialogue oral spontané", *Actes XVIIIèmes Journées d'Étude sur la Parole* (2010 mai 24-28: Mons, BELGIUM), 4 pages. Cederom

- Blache, P., Abderrahmane, M., Rauzy, S., Bertrand, R. (2020) "An integrated model for predicting backchannel feedbacks", *20th ACM International Conference on Intelligent Virtual Agents* (acceptance rate 50%), 19<sup>th</sup>- 23<sup>rd</sup> of October 2020, University of Glasgow, Glasgow, United Kingdom
- Blache, P., Bertrand, R., Ferré, G., Pallaud, B., Prévot, L., Rauzy, S. (2017) "The Corpus of Interactional Data: a Large Multimodal Annotated Resource", in *Handbook of Linguistic Annotation* N. Ide & J. Pustejovsky (eds), Springer bookseries Text, Speech, and Language Technology, 1323-1356
- Blache, P., Bertrand, R., Bigi, B., Bruno, E., Cela, E., Espesser, R., Ferré, G., Guardiola, M., Hirst, D., Magro, E.-P., Martin, J.-C., Meunier, C., Morel, M.-A., Murisasco, E., Nesterenko, I., Nocéra, P., Pallaud, B., Prévot, L., Priego-Valverde, B., Seinturier, J., Tan, N., Tellier, M., Rauzy, S. (2010) "Multimodal Annotation of Conversational Data", *Proceedings of Linguistic Annotation Workshop* (2010 juillet 15-16: Uppsala, SWEDEN), 186-191
- Blache, P., Bertrand R., Guardiola, M., Guénot, M.-L., Meunier, C., Nesterenko, I., Pallaud, B., Prévot, L., Priego-Valverde, B., Rauzy, S. (2010) "The OTIM formal annotation model: a preliminary step before annotation scheme", *Proceedings of Language Resource and Evaluation Conference* (2010 mai 19-21: La Valette, MALTA), 3262-3267
- Blache, P., Bertrand, R. & Ferré, G. (2009) "Creating and exploiting multimodal annotated corpora: the ToMA project", In Kipp M. (eds.) *Multimodal Corpora*. Berlin: Springer-Verlag, vol. LNAI 5509, 16, 38-53
- Blache, P. & Rauzy, S. (2008) "Influence de la qualité de l'étiquetage sur le chunking: une corrélation dépendant de la taille des chunks", In *Actes de Traitement Automatique des Langues Naturelles*, 290-299, Avignon, France
- Blache, P., Bertrand, R., Ferré, G. (2008) "Creating and exploiting multimodal annotated corpora", *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, (2008: Marrakech, MOROCCO). Marrakech: ELDA, 6 pages. Cederom
- Blache, P., Ferré, G., Rauzy, S. (2007) "An XML Coding Scheme for Multimodal Corpus Annotation", *Corpus Linguistics*, Jul 2007, Birmingham, United Kingdom, 1-17
- Blöndal, Þ. (2005) "Feedback in conversational storytelling", In: *Feedback in Spoken Interaction*. Nordtalk, Gothenburg Papers in Theoretical Linguistics, 1-17
- Bögels, S. & Torreira, F. (2015) "Listeners use intonational phrase boundaries to project turn ends in spoken interaction", *Journal of Phonetics*, 10.1016/j.wocn.2015.04.004
- Boersma, P. & Weenink, D., (2009) *Praat: Doing Phonetics by Computer* (Version 5.1.05) [Computer Program]
- Bolly, C.T., Christodoulides, G. & Simon, A.C. (2016) "Disfluences et vieillissement langagier. De la base de données VALIBEL aux corpus outillés en français parlé", *Corpus* [En ligne], 15 | 2016, <http://journals.openedition.org/corpus/3006>; DOI : <https://doi.org/10.4000/>
- Boudin, A., Bertrand, R., Rauzy, S., Ochs, M. Blache, P. (2021). A multimodal Model for predicting Conversational Feedbacks, Springer Nature Switzerland AG 2021 K. Ekstein et al. (Eds.): *TSD 2021*, LNAI 12848, 1-13
- Brennan, S. E. & Clark, H. H. (1996) "Conceptual Pacts and Lexical Choice in Conversation", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482-1493
- Bunt, H. (1994) "Context and dialogue control", *Think Quarterly*, 3(1):19-31
- Cavé, C., Guaitella, I., Santi, S., Bertrand, R., Harlay, F., Espesser, R. (1996) "About the relationship between eyebrow movements and Fo variations", *Proceedings of International Conference on Spoken Language Processing (ICSLP)* (4: 1996: Philadelphia, USA), 2175-2178
- Chafe, W. L. (1988) "Linking intonation units in spoken English", In Haiman, J. Thompson, S. A. (Eds), *Clause combining in Grammar and Discourse*, Amsterdam, John Benjamins, 1-27
- Chen, H., Prévot, L., Bertrand, R., Priego-Valverde, B., Blache, P. (2012) "Toward a Mandarin-French Corpus of Interactional Data", *SemDial 2012 (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue* (2012 septembre 19-21: Paris, FRANCE)
- Clark H. H. (1996) *Using Language*, Cambridge University Press, Cambridge
- Clark, H. H. & Brennan, S. E. (1991) "Grounding in communication", In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (p. 127-149). American Psychological Association. <https://doi.org/10.1037/10096-006>

- Clark, H. H. & Schaefer, E. F. (1987) "Concealing One's Meaning from Overhearers", *Journal of Memory and Language*, 26, 209-225
- Cole, J. (2015) "Prosody in context: a review", *Language, Cognition and Neuroscience*, 30:1-2, 1-31, doi: 10.1080/23273798.2014.963130
- Corminboeuf, G., & Horlacher, A. S. (2016) "La projection en macro-syntaxe et en linguistique interactionnelle: dimensions théoriques et empiriques", *Langue française*, 192, 15-36
- Couper-Kuhlen, E. (2011) "Grammaticalization and Conversation", In Narrog, H. & Heine, B. (eds) *The Oxford Handbook of Grammaticalization*, Oxford, OUP, 424-437
- Couper-Kuhlen, E. & Selting, M. (2018) *Interactional Linguistics. Studying Language in Social Interaction*, Cambridge: Cambridge University Press
- Couper-Kuhlen, E. & Selting, M. (1996) *Prosody in Conversation*, Cambridge: Cambridge University Press
- Couper-Kuhlen, E. & Ford, C. E. (2004) *Sound Patterns in Interaction. Cross-linguistic studies from conversation*, Amsterdam, John Benjamins Publishing Company
- Couper Kuhlen, E. & Thompson, S. (2008) "On assessing situations and events in conversation 'Extrapolation' and its relatives", *Discourse studies*, 10 (4), 443-467
- Danet, B. (1980) "'Baby' or 'Fetus?' Language and the Construction of Reality in a Manslaughter", *Trial. Semiotica*, 32, 187-219
- Degand, L. & Simon, A.C. (2009) "On identifying basic discourse units in speech: theoretical and empirical issues", *Discours 4*
- Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S.-A. Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R. & Yoo, H. (2015) "Intonational Phonology of French: Developing a ToBI system for French", In Frota, S. & Prieto, P. (eds) *Intonation in Romance*, chapter 3, Publisher: Oxford University Press
- Delais-Roussarie, E. & Post, B. (2011) "Quand frontières prosodiques et frontières syntaxiques se rencontrent", *Langue française 2* (170), 29-44
- Delais-Roussarie, E. (2005) "Phonologie et grammaire : études et modélisation des interfaces prosodiques", *Mémoire d'HDR*. Université de Toulouse Le Mirail
- de Stefani, E. & Horlacher, A.S. (2017) "Une étude interactionnelle de la grammaire : la dislocation à droite évaluative dans la parole en interaction", *Revue Française de linguistique appliquée*, XXII-2, 15-32
- Delattre, P. (1966) "Les dix intonations de base du français", *The French Review*, XI (1): 1-14
- Deppermann, A. (2010) "Future prospects of research on prosody: the needs for publicly available corpora: comments on Margret Selting 'Prosody in Interaction: State of the art'", in D. Barth-Weingarten, E. Reber & M. Selting (eds), *Prosody in Interaction*, Amsterdam: John Benjamins, 41-47
- Di Cristo, A. (2016) *Les musiques du français parlé*, vol 1, De Gruyter
- Di Cristo, A. (2013) *La prosodie de la parole*, Marseille, De Boeck-Solal
- Di Cristo, A., Auran, C., Bertrand, R., Chanet, C., Portes, C., Regnier, A. (2004) Outils prosodiques et analyse du discours, in A.C. Simon, A. Auchlin et A. Grobet (eds), *Cahiers de Linguistique de Louvain 30/1-3*, Louvain-la-neuve : Peeters, vol. 28, 27-84
- Di Cristo, A., Auran, C., Bertrand, R., Chanet, C., Portes, C. (2003) "An integrative approach to the relations of prosody to discourse: towards a multilinear representation of an interface network", in A. Mettouchi & G. Ferré (eds), *Interfaces Prosodiques*, Nantes, 29-34
- Di Cristo, A. (2000) "Vers une modélisation de l'accentuation en français", *Journal of French Language Studies* 10, 27-44
- Di Cristo, A. (1998) "Intonation in French", in Hirst & Di Cristo (1998 eds) *Intonation Systems*, Cambridge University Press, 88-103
- Di Cristo, A. & Hirst, D.J. (1996) "Vers une typologie des unités intonatives du français", *XXIème JEP* (Avignon, France), 219-222

- D'Imperio, M., Bertrand, R., Di Cristo, A., Portes, C. (2007) "Investigating phrasing levels in French: Is there a difference between nuclear and prenuclear accents?" In J. Camacho, V. Deprez, N. Flores, L. Sanchez, *Selected Papers from the 36th Linguistic Symposium on Romance Languages* (LSRL). New Brunswick: John Benjamins Publishing Company, 97-110
- D'Imperio, M., Bertrand, R., Di Cristo, A., Portes, C. (2006) "The phonology and phonetics of prenuclear and nuclear accents in French", *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics* (2006 août 28-30: Athens, Greece), 121-124
- Dubois, J.W., Chafe, W.L., Meyer, C., Thompson, S.A. (2000-2005) *Santa Barbara Corpus of Spoken American English*. Part I. <https://doi.org/10.35111/s2q7-gq73>
- Duncan, S. (1972) "Some signals and rules for taking speaking turns in conversations", *Journal of Personality and Social Psychology*, 23, 283-292. doi:10.1037/h0033031
- Engdahl, E. (2005) *Information packaging in questions*, talk given at the Colloque de syntaxe et sémantique à Paris (CSSP), Paris: France, sept 29th – oct 1st 2005.
- Ferré, G., Bertrand, R., Blache, P., Espesser, R. & Rauzy, S. (2007) "Intensive Gestures in French and their Multimodal Correlates", *Proceedings of Interspeech* (2007 août 27-31: Antwerp, BELGIUM), 690-693. Cederom
- Ford, C. E. Fox, B.A. & Thompson, S.A. (2002) Contiguency and the grammar of turn increments, In Ford, C. E. Fox, B.A., Thompson, S.A. (eds), *The language of Turn and Sequence*, Oxford, OUP, 14-38
- Ford, C. E. & Thompson, S. A. (1996) "Interactional Units in Conversation : syntactic, intonational and pragmatic resources for the management of turns", In *Interaction and Grammar*, E. Ochs, E. A. Schegloff & S. A. Thompson (eds), 134-184, Cambridge UP
- Fougeron, C. & Jun, S.-A. (1998) "Rate Effects on French Intonation: Phonetic Realization and Prosodic Organization", *Journal of Phonetics* 26, 45-70
- Fox, B.A & Henemann, T. (2016) "Rethinking format : An examination of requests", *Language in Society*, 45, 499-531
- FoxTree, J.E. (1999) "Listening in on monologues and dialogues", *Dis. Process.* 27, 35–53. doi:10.1080/01638539909545049
- Fuscone, S. (2020) "Une approche computationnelle pour caractériser les dynamiques interpersonnelles dans les conversations naturelles", *Thèse de doctorat*, AMU
- Garnier, L., Baqué, L., Dagnac, A., & Astésano, C. (2016) "Perceptual investigation of prosodic phrasing in French", In *Proceedings of Speech Prosody*, 1153-1157
- Garrod, S. & Pickering, M. (2007) "Alignment in dialogue", *Oxford Handbook of Psycholinguistics*. 10.1093/oxfordhb/9780198568971.013.0026
- Garrod, S. & Pickering, M. J. (2004) "Why is conversation so easy?", *Trends in Cognitive Sciences* 8:8-11
- Garrod, S. & Anderson, A. (1987) "Saying what you mean in dialogue: A study in conceptual and semantic co-ordination", *Cognition*, 27, 181-218
- Gorisch, J. & Prévot, L. (2014) *Audio-visual condition of Aix Map Task*, Laboratoire Parole et Langage - UMR 7309 (LPL, Aix-en-Provence FR), Speech and Language Data Repository (SLDR/ORTOLANG), ISLRN oai:sldr.org:sldr000875
- Goodwin, C. (1981) *Conversational organization: Interaction between speakers and hearers*. New York, NY: Academic Press
- Goodwin, C. (1984) "Notes on story structure and the organization of participation", In: Atkinson, J.M., Heritage, J. (Eds.), *Structures of Social Action: Studies in Conversation Analysis*, Cambridge University Press, London, 225-246
- Goujon, A. (2019) "Indices d'incompréhension et séquences de réparation dans l'interaction en face-à-face : une analyse multimodale", *Thèse de doctorat*, Université d'Aix Marseille
- Goujon, A., Bertrand, R. & Tellier, M (2015) "Eyebrows in French talk-in-interaction? ", *Proceedings of Gesture and Speech in Interaction (GESPIN)*, Nantes, September 2-4
- Goujon A., Bertrand R. & Tellier M. (2017) "Eyebrow Movements: What Occurs in Sequences of Misunderstanding in a Maptask Dialogue", *IGesto International Conference on Gesture and Multimodality*, 02-03/02/17, University of Porto, Portugal

- Granier, O., Bertrand, R. & Ochs, M. (2021) "La persuasion dans les interactions humain-humain et humain-machine : une revue de la littérature", *Workshop sur les Affects, Compagnons Artificiels et Interactions*, 13-15 octobre 2021, Ile d'Oléron, France
- Gravano, A. & Hirschberg, J. (2010) "Turn-taking cues in task-oriented dialogue", *Computer Speech and Language* 25, 601-634.
- Grosz, B.J. & Sidner, C.L. (1986) "Attention, intentions, and the structure of discourse", *Computational Linguistics* 12, 175-204
- Guaïtella, I., Santi, S., Cavé, C., Bertrand, R., Boyer, J., Faraco, M., Lagrue, B., Mignard, P., Paboudjian, C., Purson, A. (1998) "Les relations voco-gestuelles dans la communication interpersonnelle: émergence d'une problématique et carrefour interdisciplinaire", In Santi, Guaïtella, Cavé & Konopczynski (eds.) *Oralité et Gestualité, communication multimodale, interaction*. Paris: L'Harmattan, 13-24
- Guardiola, M. & Bertrand, R. (2013) "Interactional convergence in conversational storytelling: when reported speech is a cue of alignment and affiliation", *Frontiers in Cognitive Science*, 30 pages
- de Gaulmyn, M.-M. (1992) "Grammaire du français parlé. Quelques remarques autour du discours rapporté", in *Actes du Congrès de l'ANEFLE Grammaire et français langue étrangère*, Joussaud & Pétrissans, Grenoble, 22-23
- Hakulinen, A. & Selting, M. (2005) *Syntax and Lexis in conversation*, John Benjamins Publishing Company
- Heldner, M. & Edlund, J. (2010) "Pauses, gaps and overlaps in conversations", *Journal of Phonetics*, 38:555-568
- Hirschberg, J. & Ward, G. (1995) "The Interpretation of the High-Rise Question Contour in English", *Journal of Pragmatics* 24, 407-412
- Hirst, D.J. & Di Cristo, A. (1998) *Intonation systems*, Cambridge, Cambridge University Press
- Hirst, D.J. & Di Cristo, A. (1984) "French Intonation: A parametric approach", *Die Neueren Sprachen*, 5: 554-569
- Horton, W.S. & Gerrig, R.J. (2005), "Conversational common ground and memory processes in language production", *Discourse Processes*, 40, 1-35
- Janssoone, T., Bailly, K., Richard, G. & Clavel, C. (2020) "The POTUS Corpus, a database of weekly addresses for the study of stance in politics and virtual agents", In *Proceedings of The 12th Language Resources and Evaluation Conference*, 1546-1553
- Jefferson, G. (1987) "Notes on "latency" in overlap onset". In G. Button, P. Drew & J. Heritage (Eds.), *Interaction and language use*, Special issue of *Human Studies*, 9, 153-183
- Jefferson, G. (1978) 'Sequential Aspects of Storytelling in Conversation', in J. Schenkein (ed.), *Studies in the Organization of Conversational Interaction*, 219-48. New York: Academic Press
- Jun, S.-A. & Fougeron, C. (2002) "Realizations of accentual phrase in French intonation", *Probus* 14, 147-172
- Jun, S.-A. & Fougeron, C. (2000) "A Phonological model of French intonation", in Botinis, A. (ed.) *Intonation : Analysis, Modeling and Technology*, Dordrecht : Kluwer Academic Publishers, 209-242
- Koiso, H., Horiuchi Y., Tutiya S., Ichikawa A. & Den Y. (1998) "An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogues", *Language and Speech*, 41 (3-4), 295-321
- Kurtic, E., Brown, G. J. & Wells, B. (2009) "Fundamental frequency height as a resource for the management of overlap in talk-in-interaction", In: *Where Prosody Meets Pragmatics. Studies in Pragmatics*, 8. Emerald Group Publishing Limited, Bingley, UK ISBN 978-1-84950-631-1
- Labov, W. & Waletzky, J. (1966) "Narrative analysis: oral versions of personal experience", In: J. Helm (ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. Seattle (Univ of Washington Press), 12-44
- Lacheret, A. & Kahane, S. (2020) "Unités syntaxiques et unités intonatives majeures en français parlé: inclusion, fragmentation, chevauchement", In *SHS Web of Conferences* (Vol. 78, p. 14005). EDP Sciences
- Ladd, R. (1996, 2008) *Intonational Phonology*, Cambridge Studies in Linguistics 79, Cambridge: Cambridge University Press
- Lerner, G.H. (1996) "On the semi-permeable character of grammatical units in conversation: Conditional entry into the trun space of another speaker", In Ochs, E., Schegloff, E.A. & Thompson, S (eds), *Interaction and Grammar*, 238-276, Cambridge University Press

- Local, J. (2007) "Phonetic detail in talk-in-interaction: on the deployment and interplay of sequential context and phonetic resources", *Nouveaux cahiers de linguistique française*, vol. 28, 67-86
- Local, J. & Walker, G. (2005) "[Methodological imperatives for investigating the phonetic organisation and phonological structures of spontaneous speech](#)", *Phonetica*, vol. 62, no. 2-4, 120-130
- Local, J. & French, P. (1986), "Prosodic features and the management of interruptions", In *Intonation in Discourse* (Catherine Johns-Lewis, ed.), London, Croom Helm, 157-180
- Local, J., Kelly, J. & Wells, B. (1986) "[Towards a phonology of conversation: turn-taking in Tyneside English](#)", *Journal of Linguistics*, vol. 22, no. 2, 411-437
- Local, J. & French, P. (1983) "[Turn competitive incomings](#)", *Journal of Pragmatics*, vol. 7, 1, 17-3
- Mann, W.C. & Thompson, S.A. (2001) "Deux perspectives sur la théorie de la structure rhétorique (RST)", *Verbum* XXIII/1, 9-29
- Marandin, J.-M. (2004) "Contours as constructions", *ICCG-3 : the third international conference on construction grammars*, 7-10 juillet 2004, Marseille
- Marandin, J.-M., Beyssade, C., Delais-Roussarie, E., Rialand, A. & De Fornel, M. (2004) "The meaning of final contour in French", <http://www.lif.cnrs.fr/fr/Marandin/>, September 15th
- Maschler, Y., Pekarek Doehler, S., Lindström, J. & Keevallik, L. (2020) *Emergent syntax for Conversation*, John Benjamins Publishing Company
- Mastriani, M.C., Petrone, C., Bertrand, R., Ochs, M. (2018) "Role of prosody on the perception of the "oui"/"yes" feedback in medical context", *Proc. 9th International Conference on Speech Prosody 2018*, 512-516, DOI: 10.21437/SpeechProsody.2018-104
- Matsumoto, K. (2003) "Unit Linking in conversational Japanese", *Language Sciences*, Vol. 25, 5, 433-455
- Michelas, A. (2011) "Caractérisation phonétique et phonologique du syntagme intermédiaire en français: de la production à la perception", *Thèse de doctorat*, Université de Provence - Aix-Marseille I
- Meyer, A.S., Phillip, A., Decuyper, C., Knudsen, B. (2018) "Working Together: Contributions of Corpus Analyses and Experimental Psycholinguistics to Understanding Conversation", *Frontiers in Psychology* 9, doi : 0.3389/fpsyg.2018.00525
- Mondada, L. (2011) "Understanding as an embodied, situated and sequential achievement in interaction", *Journal of Pragmatics*, 43(2), 542-552, doi:10.1016/j.pragma.2010.08.019
- Mondada, L. (1999) "L'organisation séquentielle des ressources linguistiques dans l'élaboration collective des descriptions", *Langage et société*, 89,9-36, doi : 10.3406/lso.1999.2882
- Müller, F.E. (1996) "Affiliating and disaffiliating with continuers", In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in Conversation* (pp. 131-176). Cambridge University Press
- Nesterenko, I., Rauzy, S. & Bertrand, R. (2010) "Prosody in a corpus of French spontaneous speech: perception, annotation and prosody ~ syntax interaction", *Proceedings of Speech Prosody* (5. 2010 mai 11-14: Chicago, United States of America), 4 pages. Cederom
- Norrick, N.R., (2012) "Listening practices in English conversation: the responses responses elicit", *J. Pragmat.* 44, 566-576
- Norrick, N.R. (2010) "Incorporating listener evaluation into stories", *Narrat. Inq.* 20 (1), 183-204
- Norrick, N.R. (2008) "Negotiating the reception of stories in conversation: teller strategies for modulating response", *Narrat. Inq.* 18 (1), 131-151
- Norrick, N.R. (2007) "Conversational storytelling", In: Herman, D. (Ed.), *The Cambridge Companion to Narrative*. Cambridge University Press, 127-141.
- Norrick, N.R. (1987) "Functions of repetition in conversation", *Text - Interdisciplinary Journal for the Study of Discourse*, 7(3): 245-264
- Ochs, E., Schegloff, E.A. & Thompson, S.A. (1996) *Grammar and Interaction*. Cambridge: Cambridge University Press
- Ochs, M., Bertrand, R., Goujon, A., Bolger, D., Dubarry, A.S., et al. (2020) "Le jeu de données Brain-IHM: une nouvelle ressource pour l'analyse des bases cérébrales des conversations humain-humain et humain-agent-virtuel", *Workshop sur*

- les Affects, Compagnons artificiels et Interactions*, CNRS, Université Toulouse Jean Jaurès, Université de Bordeaux, Saint Pierre d'Oléron, France. ([hal-02933479](https://hal.archives-ouvertes.fr/hal-02933479))
- Ochs, M., Bertrand, R., Goujon, A., Bolger, D., Dubarry, A.S., et al. (2020) "The Brain-IHM Dataset: a New Resource for Studying the Brain Basis of Human-Human and Human-Machine Conversations", *Language Resources and Evaluation Conference (LREC)*, Marseille, France. ([hal-02893898](https://hal.archives-ouvertes.fr/hal-02893898))
- Ochs, M., Blache, P., de Montcheuil, G., Pergandi, J.M., Bertrand, R., Saubesty, J., Francon, D. & Mestre, D. (2019) "The Acorformed Corpus: Investigating Multimodality in Human-Human and Human-Virtual Patient Interactions", *Selected papers from the CLARIN Annual Conference 2018*. Linköping Electronic Conference Proceedings 159: 113–120
- Ochs, E., Schegloff, E.A. & Thompson, S.A. (1996) *Grammar and Interaction*. Cambridge: Cambridge University Press.
- Ogden, R. (2004) "Non-modal voice quality and turn-taking in Finnish", In Couper-Kuhlen, E. & Ford, C.E. (eds), *Sound Patterns in Interaction*, Amsterdam, Benjamins, 29-62
- Ogden, R. (2001) "Turn transition, creak and glottal stop in Finnish talk-in-interaction", *Journal of the International Phonetic Association*, 31, 139-152
- Pallaud, B. & Bertrand, R. (2020) "Auto-interruptions et disfluences à l'oral. C'était euh tu vois complètement loufoque comme si ouais euh comme situation", In Hirsch, F., Didirková, I. et D. Christelle (Eds.). *Manuel de Pausologie*. Recueil. Paris: L'Harmattan, coll. Langue et Parole, 21-49
- Pallaud, B., Bertrand, R., Blache, P., Prévot, L. & Rauzy, S. (2019) "Suspensive and disfluent self interruptions in French language interactions", In L. Degand, G. Gilquin, L. Meurant & A.-C. Simon (eds.) *Fluency and Disfluency across Languages and Language Varieties*. Corpora and Language in Use 4, Louvain-la-Neuve: Presses universitaires de Louvain, 109-138
- Pekarek Doehler, S. (2019) "At the Interface of Grammar and the Body: Chais pas ("dunno") as a Resource for Dealing with Lack of Recipient Response", *Research on Language and Social Interaction*, 52:4, 365-387, DOI: [10.1080/08351813.2019.1657276](https://doi.org/10.1080/08351813.2019.1657276)
- Pekarek S. (2016) "More than an epistemic hedge: French je sais pas 'I don't know' as a resource for the sequential organization of turns and actions", *Journal of Pragmatics* <http://dx.doi.org/10.1016/j.pragma.2016.06.014>
- Pekarek-Doehler, S., De Stefani, E. & Horlacher, A.-S. (2015) *Time and Emergence in Grammar. Left-dislocation, right-dislocation, topicalization and hanging topic in French talk-in-interaction*. Amsterdam: Benjamins
- Penteado, B.E., Ochs, M., Bertrand, R., Blache, P. (2019) "Evaluating Temporal Predictive Features for Virtual Patients Feedbacks", *ACM International Conference on Intelligent Virtual Agent (IVA)*, Jul 2019, Paris, France. ([10.1145/3308532.3329438](https://hal.archives-ouvertes.fr/hal-02355386)). ([hal-02355386](https://hal.archives-ouvertes.fr/hal-02355386))
- Perrin, L., Deshaies, D. & Paradis, C. (2003) "Pragmatic functions of local diaphonic repetitions in conversation", *Journal of Pragmatics*, 35:1843–1860
- Peshkov, K., Prévot, L. & Bertrand, R. (2013) "Evaluation of automatic prosodic segmentation", *Proceedings of Prosody-Discourse Interface*, Leuven. September 11-13
- Peshkov, K., Prévot, L. & Bertrand, R. (2013) "Prosodic phrasing evaluation: measures and tools", *Proceedings of Tools and Resources for Analysing Speech Prosody (TRASP)*, Aix en Provence, août 2013
- Pickering, M. J., & Garrod, S. (2006) Alignment as the Basis for Successful Communication", *Research on Language and Computation*, 4(2), 203–228 <https://doi.org/10.1007/s11168-006-9004-0>
- Pickering, M.J. & Garrod, S. (2004) "Toward a Mechanistic Psychology of Dialogue", *Behavioral and Brain Sciences*, 27, 169-225
- Pierrehumbert, J. & Hirschberg, J. (1990) "The Meaning of Intonational Countours in the Interpretation of Discourse", in Cohen P. et al. (eds). *Intentions in Communication*, Cambridge: MIT Press, 271-311
- Pierrehumbert, J. (1980) The phonology and phonetics of English intonation, *PhD dissertation*, MIT
- Pitt, M., Johnson, K., Hume, E., Kiesling, S., Raymond, W. (2005) "The Buckeye Corpus of Conversational Speech: Labeling Conventions and a Test of Transcriber Reliability", *Speech Communication*, 45, 90-95. <http://buckeyecorpus.osu.edu/pubs/BuckeyeCorpus.pdf>
- Porhet, P., Ochs, M., Saubesty, J., de Montcheuil, G. and Bertrand, R. (2017) "Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks", In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17)*. ACM, New York, NY, USA, 6 p. <https://doi.org/10.1145/311361755.31356816>

- Portes, C. & Beyssade, C. (2015) "Is intonational meaning compositional ? ", *Verbum XXXVII*, 2, 207-233
- Portes, C., Beyssade, C., Michelas, A., Marandin, J.-M., Champagne-Lavau, M. (2014) "The Dialogical Dimension of Intonational Meaning: Evidence from French", *Journal of Pragmatics*, 74. 15–29, 10.1016/j.pragma.2014.08.013
- Portes, C. & Bertrand, R. (2011) "Permanence et variation des unités prosodiques dans le discours et l'interaction", *Journal of French Language Studies*. 2011, 97-110 <http://hal.archives-ouvertes.fr/hal-00576851>
- Portes, C., Bertrand, R. & Espesser, R. (2007) "Contribution to a grammar of intonation in French. Form and function of three rising patterns", *Nouveaux Cahiers de Linguistique Française*, 2, 155-162
- Portes, C. & Bertrand, R. (2006) "Some cues about the interactional value of the «continuation» contour in French", *Actes Discours et Prosodie comme Interface Complexe (IDP)* (1: 2005 octobre 8-9: Aix-en-Provence, France). Cederom (14 pages)
- Post, B. (2000) *Tonal and phrasal structures in French intonation*, The Hague : Holland Academic graphics.
- Prévoit, P., Gorisch, J. & Bertrand, R. (2016) "A CUP of CoFee: A large Collection of feedback Utterances provided with communicative function annotations", 10th edition of the *Language Resources and Evaluation Conference (LREC)*, 23-28 May 2016, Portorož (Slovenia)
- Prévoit, L., Gorisch, J. & Bertrand R. (2015) "CoFee project Corpus Collection". Laboratoire Parole et Langage - UMR 7309 (LPL, Aix-en-Provence FR), Speech and Language Data Repository (SLDR/ORTOLANG), ISLRN oai:sldr.org:sldr000911
- Prévoit, L., Bertrand, R., Peshkov, K., Rauzy, S., Blache, P. (2015) "Prosody, Discourse and Syntax in French Conversations: Resource creation and evaluation", *Rapport non publié*
- Prévoit, L., Gorisch, J., Bertrand, R., Gorene, E., Bigi, B. (2015) "A SIP of CoFee: A Sample of Interesting Productions of Conversational Feedback", *Proceedings of 16<sup>th</sup> Annual SIGdial Meeting on Discourse and Dialogue*, Prague, Sept 2-4
- Prévoit, L., Bertrand, R., Priego-Valverde, B., Blache, P. (2010) "Discourse Interaction in French Conversation: A Case Study for Interoperable Semantic Annotation", *Proceedings of Workshop on Interoperable Semantic Annotation* (2010 janvier 15-17: Hong-Kong, HONG KONG), 39-48
- Priego-Valverde B. (1999) "L'humour dans les interactions conversationnelles : jeux et enjeux". *Thèse de Doctorat, Université de Provence*
- Priego-Valverde, B., Bigi, B. & Amoyal, M. (2020) "Cheese! : a corpus of face-to-face French interactions", *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 460–468, Marseille
- Prieto, P. (2015) "Intonational meaning", *Wiley Interdisciplinary Reviews: Cognitive Science*. 6. 10.1002/wcs.1352
- Purson, A., Santi, S., Bertrand, R., Guaïtella, I., Boyer, J., Cavé, C. (1999) "The relationships between voice and gesture: eyebrows movements and questioning", *Proceedings of the VI<sup>th</sup> European Conference of Speech Communication and Technology* (Budapest), 1735-1739
- Purson, A., Bertrand, R., Guaïtella, I., Boyer, J., Cavé, C., Santi, S., Harlay, F. (1998) "Mouvements de sourcils dans le questionnement : relation entre geste et voix", In Santi, Guaïtella, Cavé & Konopczynski (eds.) *Oralité et Gestualité, communication multimodale, interaction* Paris: L'Harmattan, 561-566
- Rabatel, A., (2005) "Analyse énonciative et interactionnelle de la confidence", *Poétique* 2005/1, 141, 93-113
- Rauzy, S. & Amoyal, M. (2020) "SMAD: A tool for automatically annotating the smile intensity along a videorecord", In *HRC2020*, 10th Humour Research Conference
- Robinson, J.D., (2013) "Overall structural organization", In Sidnell, J., Stivers, T. (Eds.), *The Handbook of Conversation Analysis*, Blackwell, 257-280
- Roger, D. (1989) "Experimental studies of dyadic turn-taking behaviour", In D. Roger & P. Bull (Eds.) Peter Bull Editors, *Conversation*, 75-95, Multilingual Matters LTD
- Roulet, E., Filliettaz, L. & Grobet, A. (2001) *Un modèle et un instrument d'analyse de l'organisation du discours*, Bern : Peter Lang
- Rossi, M., Di Cristo, A., Hirst, D., Martin, P. & Nishinuma, Y. (1981) *L'intonation: de l'acoustique à la sémantique*, Paris: Klincksieck
- Roux, G., Bertrand, R., Ghio, A., Astésano, C. (2016) "Naïve listeners' perception of prominence and boundary in French spontaneous speech", *Speech Prosody* 31 May-3 June, Boston (USA)

- Rühlemann, C. & Gries, S. (2015) "Turn order and turn distribution in multi-party storytelling", *Journal of Pragmatics*. 87. 171-191. 10.1016/j.pragma.2015.08.003
- de Ruiter, J.P.D., Mitterer, H. & Enfield, N.J. (2006) "Projecting the end of a speaker's turn: A cognitive corner stone of conversation", *Language*, 82, 515-535
- Sacks, H., Schegloff, H. & Jefferson, G. (1974) "A simplest systematics for the organization of turn-taking for conversation", *Language* 50, 696-735. doi:10.2307/412243
- Schegloff, E. A. (2007) *Sequence organization in interaction: Volume 1: A primer in conversation analysis*. Cambridge University Press.
- Schegloff, E. (2000) "Overlapping Talk and the Organization of Turn-Taking for Conversation", *Language in Society*, 29. 10.1017/S0047404500001019.
- Schegloff, E. A. (1996) "Turn organization: One intersection of grammar and interaction", In *Interaction and Grammar*, E. Ochs, E. A. Schegloff & S. A. Thompson (Eds), Cambridge: Cambridge University Press, 52-133
- Schegloff, E. (1992) "Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation", *American Journal of Sociology - AMER J SOCIOL.* 97. 10.1086/229903.
- Schegloff, E.A. (1982) "Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences", In D. Tannen (Ed.), *Analyzing discourse: Text and talk*, Washington, DC: Georgetown University Press, 71-93.
- Schegloff, E. & Sacks, H. (1973) "Opening Up Closings", *Semiotica*. 8. 289-327. 10.1515/semi.1973.8.4.289.
- Schober, M.F. & Clark, H.H. (1989) "Understanding by addressees and overhearers", *Cognitive Psychology*, 21, 211-232
- Selting, M. (2007) "Lists as Embedded Structures and the Prosody of List Construction as an Interactional Resource", *Journal of Pragmatics* 39 (3): 483-526
- Selting, M. (2000) "The Construction of "Units" in Conversational Talk", *Language in Society* 29 (4): 477-517
- Selting, M. (2005) "Syntax and Prosody as methods for the construction and identification of turn-constructive units in conversation", In Hakulinen, H. & Selting, S. (Eds), *Syntax and lexis in conversation: Studies on the use of linguistic resources in talk-in-interaction*, 17-44, Amsterdam: Benjamins <http://doi.org/10.1075/sidag.17.04sel>
- Selting, M. (1996) "On the interplay of syntax and prosody in the constitution of turn-constructive units and turns in conversation", *Pragmatics* 6(3), 371-388
- de Sevin, E., Niewiadomski, R., Bevacqua, E, Pez, A.M., Mancini, M. & Pelachaud, C. (2010) "[Greta, une plateforme d'agent conversationnel expressif et interactif](#)", 751-776, Technique et Science Informatiques, *Revue des sciences et technologies de l'information*, ISSN-L (papier) : 0752-4072 · ISSN (électronique) : 2116-5920
- Sidnell, J., & Stivers, T. (2013) *The Handbook of Conversation Analysis*, Chichester, West Sussex, UK: Wiley-Blackwell
- Simon, A.-C. & Christodoulides G. (2016) "Perception of Prosodic Boundaries by Naïve Listeners in French", In *Speech Prosody* 2016 Boston, <http://hdl.handle.net/2078.1/171133>
- Simon, A.C. & Degand, L. (2011) "L'analyse en unités discursives de base : pourquoi et comment ? ", *Langue française* 170, 45-59
- Stalnaker, R.C. (1978), "Assertion", In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics*, 315-332, New York : Academic Press
- Stivers, T., Mondada, L. & Steensig, J. (2011) "Knowledge, morality and affiliation in social interaction", In *The Morality of Knowledge in Conversation*. 10.1017/CBO9780511921674.002
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... others (2009) "Universals and cultural variation in turn-taking in conversation", *Proceedings of the National Academy of Sciences*, 106(26), 10587-10592
- Stivers, T. (2008) "Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation", *Research on Language & Social Interaction*, 41(1), 31-57, doi:10.1080/08351810701691123
- Stivers, T., & Robinson, J.D. (2006) "A preference for progressivity in interaction", *Language in Society*, 35 (3), 367-392
- Swerts, M. & Hirschberg, J. (1998) Prosody and Conversation: An Introduction. *Language and Speech*, 41(3-4):229-233 doi:[10.1177/002383099804100401](https://doi.org/10.1177/002383099804100401)
- Szczepek Reed, B. & Raymond, G. (2013) *Units of Talk - Units of Action*, John Benjamins

- Tan, N., Ferré, G., Tellier, M., Cela, E., Morel, M.A., et al. (2010) "Multi-level Annotations of Nonverbal Behaviors in French Spontaneous Conversation", *International Conference for Language Resources and Evaluation*, May 2010, Malta, 74-79
- Tannen, D. (2007, 1989) "Talking voices: Repetition, dialogue, and imagery in conversational discourse", *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. 1-233. Cambridge University Press DOI: [10.1017/CBO9780511618987](https://doi.org/10.1017/CBO9780511618987)
- Tannen, D. (1987) "Repetition in conversation: Toward a poetics of talk", *Language* 63(3), 574-605
- Tannen, D. (1980) "A comparative analysis of oral narrative strategies", In *The pear stories*, Chafe, W. (ed), 51-87. Norwood, NJ: Ablex.D
- Tolins, J. & Fox Tree, J.E. (2014) "Addressee Backchannels Steer Narrative Development", *Journal of Pragmatics* 70: 152-164
- Tomlinson, J. & Fox Tree, J.E. (2011) "Listeners' comprehension of uptalk in spontaneous speech", *Cognition* 119, 58-69
- Traverso, V. (1998) *La conversation familière: Analyse pragmatique des interactions*, Presses Universitaires de Lyon.
- Vincent D. & Dubois S., (1996) *Le discours rapporté au quotidien*, Nuit Blanche
- Vion R. (1995) "La gestion pluridimensionnelle du dialogue", in *Cahiers de Linguistique Française*, 17, 179-203.
- Ward, N. & Tsukahara, W. (2000) "Prosodic features which cue back-channel responses in English and Japanese", *Journal of Pragmatics*, 23, 1177-1207
- Ward, N. (1996) "Using prosodic clues to decide when to produce back-channel utterances", *Proceedings ICSLP 96*, Philadelphie, 3-4, 1728-1731
- Welby, P., Bertrand, R., Portes, C., Astésano, C. (2016) "Realization of the French initial accent: Stability and individual differences", *Proceedings of Tone and Intonation in Europe* (1er-3 sept 2016), Canterbury, Angleterre
- Wells, B & Macfarlane, S. (1998) "Prosody as an interactional resource: turn-projection and overlap", *Language and Speech*, 41, 265-298

## Table des figures

---

Figure 1a : no speaker change (C=P) .....	14
Figure 1b : speaker change (C ≠ P) .....	14
Figure 2: the five selected f0 localization points.....	15
Figure 3 : Dispositif du CID pour deux dialogues.....	19
Figure 4 : Plan du traitement automatique.....	22
Figure 5: exemple d'annotation sous Praat .....	22
Figure 6 : annotation automatique des syllables et de leur structure interne .....	23
Figure 7 : illustration de la sortie du système automatique .....	24
Figure 8 : schéma d'annotation des discours rapportés directs .....	25
Figure 9 : annotation des discours rapportés directs du CID .....	26
Figure 10a : annotations expertes et naïves du phrasé prosodique .....	27
Figure 10b : annotations du niveau morphosyntaxique, des unités discursives et des narrations .....	27
Figure 11a : annotations expertes et naïves du phrasé prosodique .....	28
Figure 11b : annotations du niveau morphosyntaxique, du niveau discursif et des narrations.....	28
Figure 12 : évolution temporelle des IPU en fonction du rôle discursif (narrateur/interlocuteur) ....	49
Figure 13 : distribution des catégories morphosyntaxiques (POS) et des rires en fonction du rôle discursif (narrateur/interlocuteur).....	50
Figure 14 : évolution temporelle des trois catégories retenues (mot de contenu, mot outil et interjections dont feedbacks) pour caractériser la production des narrateurs et des interlocuteurs .	50
Figure 15 : Schéma des places.....	67
Figure 16: Prosodie et Discours rapporté.....	70
Figure 17: exemple de grille multilinéaire.....	74
Figure 18 : Les 10 intonations de base du français.....	77
Figure 19: The semantics of French intonation.....	78
Figure 20 : exemple du CID présentant un contour continuatif majeur (MC ou LH%) composé de deux contours mineurs (mc ou LH*) .....	79
Figure 21 : liste comportant plusieurs items.....	80
Figure 22 : pause silencieuse interrompant une seule unité intonative potentielle .....	82
Figure 23: très longue pause silencieuse séparant le syntagme nominal sujet du syntagme verbal suivant .....	82
Figure 24 : unités intonatives et disfluences.....	83
Figure 25a : pré-enregistrement des dialogues du corpus BRAIN-IHM .....	91
Figure 25b: patiente humaine (gauche) et virtuelle (droite) dans le même environnement virtuel.....	92
Figure 26 : préparatifs pour l'expérience EEG.....	92

## Liste des abréviations et sigles les plus fréquents

---

AM : auto-segmentale métrique  
AP : accentual phrase  
BC : Backchannels  
CA : Conversation Analysis  
CID : Corpus of Interactional Data  
Cobra : Conversation Brains  
CG: Common Ground  
COFEE : Conversational Feedback  
CRDO: Centre de Ressources des Données Orales  
DR(D) : discours rapporté (direct)  
FB : Feedbacks  
IL : Interactional Linguistics  
ILCB : Institut Language, Communication, Brain  
ILF: Institut de Linguistique Française  
IP : intonational phrase  
Ip : intermediate phrase LDC: Linguistic Data Consortium  
IPU : Interpausal units  
IRCOM: Corpus oraux et multimodaux  
MC : major continuative  
Mc : minor continuative  
OR : Other-Repetition  
SLDR: Speech and Language Data Repository  
TOE : Transcription Orthographique Enrichie  
TCU : turn-construction unit  
TRP : transition-relevance place

# Annexe 1: l'expérience Brain-IHM

---

## Exemple de scénario congru (+ quelques FB incongrus à titre d'illustration)

### Phase 1

Bonjour Mme Provost / Bonjour docteur

Votre nom se prononce-t-il bien comme ça ? / "oui + nod"

Je suis le chirurgien, c'est moi qui ait effectué l'opération / "d'accord + nod"

### Phase 2

Nous avons eu un petit souci pendant l'intervention / "ah oui" doute

J'ai une mauvaise nouvelle à vous annoncer / "ah non !" rejet (congru) vs "voilà" de confirmation (incongru)

L'opération ne s'est pas passée comme prévu / "ah oui" surprise

on vous a malheureusement perforé l'intestin / "ah non !" rejet

Ça veut dire que vous avez maintenant un petit trou à cet endroit / "oui" de suivi

C'est une complication qui est assez rare / "d'accord + nod"

Ce n'est pas anodin / "mhm"

C'est assez sérieux / "d'accord + nod"

Ça arrive parfois lorsque le polype est gros / "d'accord + nod"

Est-ce que vous comprenez ? / "oui + nod"

Les infirmiers m'ont dit que vous avez mal / "oui" bouleversé

Vous vous êtes plainte au réveil de douleurs à l'endroit de l'opération / "voilà" de confirmation

Depuis, vous vous sentez nauséuse et un peu fébrile / "voilà" de confirmation

Comme si vous aviez de la fièvre / "voilà" de confirmation (congru) vs "ah oui" de surprise (incongru)

Ces symptômes sont normaux compte tenu de la situation / "d'accord + nod"

Les douleurs sont dues aux complications suite à l'opération / "mhm"

Ça ne signifie pas une aggravation de votre état / "d'accord + nod"

On ne peut malheureusement pas traiter cette complication avec des médicaments / "ah oui" surprise

On va tenter de rendre la douleur supportable / "d'accord + nod"

### Phase 3

La seule solution est de vous opérer à nouveau / "ah non" rejet

Cette nouvelle opération est urgente et doit avoir lieu aujourd'hui / "ah oui" surprise

On doit opérer tout de suite pour réparer la perforation et retirer les liquides écoulés dans l'abdomen / "ah non" rejet, "mhm"

Mais il faut que vous soyez rassurée / "d'accord + nod"

C'est la seule solution / "d'accord + nod"

Cette situation est assez rare, elle est sérieuse / "mhm"

Mais c'est une chirurgie très simple et rapide / "d'accord + nod"

Vous serez ressortie aussitôt du bloc / "d'accord + nod"

Et les douleurs disparaîtront très rapidement après l'intervention / "mhm"

Aussitôt que nous aurons rebouché la perforation / "d'accord + nod"

Tout va redevenir normal très vite / "d'accord + nod"

Vous pourrez sortir après quelques jours après l'opération / "ah oui" surprise (congru) vs "voilà" de confirmation (incongru)

Mais il faut tout de même vous garder quelques jours / "ah non !" rejet

Et oui ! on voudrait surveiller l'évolution de votre état et s'assurer que tout va bien avant de partir / "d'accord" de confirmation

Comme prévu pour la première opération, vous recevrez un traitement post-opératoire pour prévenir toute infection potentielle / "d'accord + nod"

Même si les risques sont minimes. Les soins seront faits à domicile / "ah oui" surprise

Mais on vous expliquera tout ça avant votre sortie de l'hôpital / "d'accord" de confirmation

Concernant l'opération, ce n'est pas moi qui la ferai / "ah oui ?" surprise (congru) vs "ah non" de rejet (incongru)

C'est un confrère qui la fera / "ah oui ?" surprise

Il se prépare en ce moment au bloc / "d'accord + nod"

et on viendra vous chercher lorsque tout sera prêt / "d'accord + nod"

Dans tous les cas, ça sera avant ce soir / "ah oui ?" surprise

Dans les moments qui viennent / "d'accord + nod"

En attendant, essayez de vous reposer / "mhm"

Vous voulez peut-être prévenir vos proches ? / "oui" de confirmation + nod

Très bien, nous allons nous en occuper, en attendant l'infirmière va vous donner des antidouleurs pour atténuer les douleurs dont vous souffrez / "mhm"

Juste avant l'opération et la préparation de l'opération / "d'accord + nod"

#### Phase 4

Est-ce que tout vous paraît clair ? “oui” de confirmation + nod

Avez-vous des questions ? / “mh non + headshake”

Très bien, dans cas, au revoir Madame Provost, ne vous inquiétez, tout va bien se passer. / d'accord, merci docteur, au revoir

Les douleurs sont dues aux complications suite à l'opération / Ça ne signifie pas une aggravation de votre état / “d'accord + nod”

On ne peut malheureusement pas traiter ces douleurs avec des médicaments / “ah oui” surprise

On va tenter de les rendre supportables / “d'accord + nod”

#### Phase 3

La seule solution est de vous opérer à nouveau / “ah non” rejet Cette nouvelle opération doit avoir lieu aujourd'hui / “ah oui” surprise

On doit opérer tout de suite pour réparer la perforation et retirer les liquides écoulés dans l'abdomen / “ah non” rejet C'est la seule solution / “oui” bouleversé

Mais il faut que vous soyez rassurée / “d'accord + nod” Cette situation est assez rare, elle est sérieuse, la seule solution est l'opération / “mhm”

Mais c'est une chirurgie très simple et rapide / “d'accord + nod”

Vous serez ressortie aussitôt du bloc / “d'accord + nod” Et les douleurs disparaîtront très rapidement après l'intervention / “mhm”

Tout va redevenir normal très vite / “d'accord + nod”

Vous pourrez sortir après quelques jours /

Mais il faut tout de même vous garder quelques jours pour surveiller l'évolution de votre état / “ah non” rejet et s'assurer que tout va bien avant de partir / “voilà” de confirmation

Comme prévu pour la première opération, vous recevrez un traitement post-opératoire pour prévenir toute infection potentielle / “d'accord + nod”

Même si les risques sont minimes. Les soins seront faits à domicile / “ah oui” surprise

Mais on vous expliquera tout ça avant votre sortie de l'hôpital / “voilà” de confirmation

Je ne ferai pas moi-même l'opération / “ah non” rejet C'est un confrère qui la fera / “d'accord + nod” Il se prépare en ce moment au bloc ^

et on viendra vous chercher lorsque tout sera prêt / “voilà” de confirmation avant ce soir / —

En attendant, essayez de vous reposer /

Vous voulez peut-être prévenir vos proches ? / “oui “ Très bien, en attendant l'infirmière va vous donner des médicaments pour atténuer la douleur.

#### Phase 4

Est-ce que tout vous paraît clair ? Avez-vous des questions ?

Très bien, dans ce cas, au revoir Madame, tout va bien se passer.

ne vous inquiétez pas

## Annexe 2 : le corpus TRUENESS

---

Dispositif expérimental du corpus TRUENESS enregistré en juin 2021

Il s'agit de récolter des données à la fois audio, vidéo et de motion capture d'acteurs enregistrés en interaction interindividuelle. On constate le dispositif extrêmement lourd qui entoure les acteurs qui sont eux-mêmes équipés de combinaisons qui permettent la capture de leurs mouvements (posture, tête, torse). L'une des difficultés de ce type de corpus est de parvenir à synchroniser les différentes sources de données (voix, gestes, postures) pour les mettre en lien ultérieurement. La motion capture quant à elle permettra de doter très rapidement les agents virtuels animés des comportements multimodaux captés sur les participants humains.

La finalité principale de ce projet est de développer une plateforme de réalité virtuelle permettant de sensibiliser les individus aux comportements sociaux discriminants et de les former à faire face à de telles situation lorsqu'ils en sont témoins.

Dispositif expérimental avec les acteurs





Visualisation de la motion capture des interactions humain-humain

