



**HAL**  
open science

## Etudier la conversation pour mieux comprendre le langage

Laurent Prevot, Roxane Bertrand, Philippe Blache, C. Meunier, Noël Nguyen,  
Berthille Pallaud

► **To cite this version:**

Laurent Prevot, Roxane Bertrand, Philippe Blache, C. Meunier, Noël Nguyen, et al.. Etudier la conversation pour mieux comprendre le langage. Travaux Interdisciplinaires sur la Parole et le Langage, 2022, Panorama des recherches au Laboratoire Parole et Langage, 38, 10.4000/tipa.6079 . hal-03917860

**HAL Id: hal-03917860**

**<https://hal.science/hal-03917860v1>**

Submitted on 2 Jan 2024

**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.



## TIPA. Travaux interdisciplinaires sur la parole et le langage

38 | 2022

Numéro spécial : Panorama des recherches  
au Laboratoire Parole et Langage

---

### Étudier la conversation pour mieux comprendre le langage

Laurent Prévot, Roxane Bertrand, Philippe Blache, Christine Meunier, Noël Nguyen et Berthille Pallaud

---



#### Édition électronique

URL : <https://journals.openedition.org/tipa/6079>

DOI : 10.4000/tipa.6079

ISSN : 2264-7082

#### Éditeur

Laboratoire Parole et Langage

#### Référence électronique

Laurent Prévot, Roxane Bertrand, Philippe Blache, Christine Meunier, Noël Nguyen et Berthille Pallaud, « Étudier la conversation pour mieux comprendre le langage », *TIPA. Travaux interdisciplinaires sur la parole et le langage* [En ligne], 38 | 2022, mis en ligne le 27 janvier 2023, consulté le 29 janvier 2023.

URL : <http://journals.openedition.org/tipa/6079> ; DOI : <https://doi.org/10.4000/tipa.6079>

---

Ce document a été généré automatiquement le 29 janvier 2023.



Creative Commons - Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International  
- CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

---

# Étudier la conversation pour mieux comprendre le langage

Laurent Prévot, Roxane Bertrand, Philippe Blache, Christine Meunier, Noël Nguyen et Berthille Pallaud

---

- 1 Cet article aborde les manifestations linguistiques de la conversation. Longtemps, les linguistes ont traité séparément chacun des domaines du langage oral (phonologie, lexique, syntaxe, discours), notamment à partir de corpus de parole préparée et non interactive. Les travaux sur la parole en conversation ont notamment mis en évidence l'interdépendance de ces domaines.
- 2 Nous proposons dans cet article d'aborder tout d'abord les particularités du langage conversationnel ainsi que leur impact sur les unités et les phénomènes généralement étudiés indépendamment par les linguistes : phonèmes, morphèmes, syntagmes, unités prosodiques, propositions, unités discursives (notons que la dimension mimo-gestuelle est bien entendu à intégrer dans ces phénomènes, cela fait l'objet d'un article distinct dans ce numéro [voir « Multimodalité(s) dans les interactions humaines »<sup>1</sup>]). Ces particularités proviennent notamment de paramètres communs à la plupart des situations communicatives conversationnelles : une certaine pression temporelle, la co-présence des participants, le caractère fugace des productions conversationnelles dont la forme est oubliée dès que le message a été transmis.
- 3 La pression temporelle donne des caractéristiques fortes au langage conversationnel liées à la gestion du temps (réductions lexicales, phonétiques, disfluences, etc.). La fugacité des productions autorise l'utilisation de stratégies particulières (*répétitions, redondance, reformulations*) qui sont généralement évitées dans les autres genres. La co-présence des participants, enfin, permet la mobilisation de toutes les modalités utilisées pour transmettre l'information, et sur lesquelles reposent les multiples mécanismes d'alignement entre participants. L'ensemble de ces particularités dues au langage conversationnel entraîne des ajustements dans tous les domaines de la production, mais de façon spécifique pour chacun d'entre eux. Ces ajustements ne sont pas toujours décrits dans chaque domaine, mais plus encore, ils sont rarement analysés

de façon conjointe. Il nous semble ainsi important de mettre ici en perspective les différents types de modifications linguistiques liées à la conversation.

- 4 De plus, comprendre ces questions générales nécessite la description précise des phénomènes propres à la conversation (cf. analyse conversationnelle et linguistique interactionnelle). Ces phénomènes incluent les tours de paroles, les signaux d'écoute et le feedback conversationnel, ou encore les énoncés co-produits. Les phénomènes sociaux liés aux dynamiques interpersonnelles sont traités dans ce numéro dans l'article « Alignement, affiliation et trajectoire interactionnelle dans la conversation »<sup>2</sup> et ils ne seront ici que brièvement évoqués pour la complétude de notre sujet. Ces phénomènes particuliers à la conversation doivent être mis en perspective avec la question de l'interaction existant entre les unités. Nous savons par exemple que les réductions phonétiques affectent plus spécifiquement les séquences dont la charge sémantique et informationnelle est faible. Il est donc nécessaire de décrire les conditions dans lesquelles ces interactions sont possibles. En particulier, il est nécessaire de les expliquer en les mettant en perspective par rapport à la question de la compréhension.
- 5 Cet article sera structuré de la façon suivante : 1/ la conversation sera discutée comme l'articulation entre représentation et communication par le langage en tant qu'action sociale dans un contexte partagé ; 2/ seront abordés ensuite les phénomènes propres à la conversation et qui ne sont pas identifiables dans d'autres types de parole (tours de parole, disfluences, feedback conversationnel, énoncés co-produits, énoncés non-phrastiques, énoncés non-lexicaux) ; 3/ une troisième partie visera à étudier l'impact de ces phénomènes sur les modèles linguistiques et sur les différentes unités linguistiques.

## 1. Langage dans la conversation : des représentations à la communication

### 1.1 Langage et Action

- 6 Avec l'émergence de l'Analyse Conversationnelle (CA-Conversation Analysis), puis de la Linguistique Interactionnelle (IL-Interactional Linguistics), l'interaction sociale est conçue comme un système, dont l'organisation méthodique repose sur les deux notions cruciales d'action et de séquence. Les ressources mobilisées (structure linguistique, formes grammaticales, patterns prosodiques) sont envisagées via les actions qu'elles permettent d'accomplir (offre, refus, etc) (Selting 1996, Couper-Kuhlen et Selting 2018), et ces actions sont étudiées les unes par rapport aux autres. Il s'agit donc d'étudier "l'organisation de l'action en contexte" (Mondada 2017), le langage -vu comme une ressource au même titre que d'autres- et l'action étant organisés de manière systématique et située. Les actions s'inscrivent dans des séquences et des activités produites de manière conjointe par l'ensemble des participants. Pour permettre la progressivité de l'interaction et son accomplissement le plus réussi, les actions nécessitent donc d'être identifiées et reconnues par les participants afin que ces derniers puissent y répondre de manière appropriée, dans le bon timing et au bon endroit.
- 7 L'objectif d'analyse de la CA vise à révéler la structure et l'organisation de l'interaction sociale en puisant dans diverses traditions de recherche (sociologie, philosophie,

linguistique, etc.). Si l'IL poursuit cet objectif, elle fait du langage son objet privilégié. La structure et l'organisation du langage tel qu'il est utilisé dans l'interaction quotidienne est au cœur de son projet. Les études en IL portent donc sur la manière dont les participants usent des différentes ressources langagières (syntaxe, prosodie, discours, mimo-gestualité) 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 ainsi à établir une grammaire de l'organisation de la conversation, en montrant comment le langage est déployé systématiquement et méthodiquement dans les interactions sociales. Au sein de l'IL, le courant de la grammaire en interaction concerne l'émergence de cette dernière qui serait le produit d'un double mouvement de l'interaction et du langage. Les auteurs de ce courant s'interrogent en effet sur la manière dont les participants à l'interaction exploitent les ressources grammaticales et inversement comment l'organisation interactive contribue à façonner les ressources linguistiques mobilisées (De Stefani et Horlacher 2017).

- 8 Nos travaux se situent dans cette perspective, proposant de prendre en compte simultanément l'ensemble des ressources disponibles dans la perspective de l'action visée. Il s'agit donc pour nous non seulement de décrire ces ressources, mais d'en expliquer l'interaction et ce dans un cadre global et si possible formalisé, ouvrant ainsi des perspectives computationnelles. Nous revenons dans le reste de cette section tout d'abord sur l'origine de ces ressources, décrites comme provenant des différents domaines de l'analyse linguistique et comment elles interagissent, typiquement pour construire du sens. Nous proposons ensuite un cadre théorique permettant de prendre en compte les spécificités liées à une interaction conversationnelle (données partielles, non canoniques, bruitées, etc.).

## 1.2 Distribution des informations dans le langage

- 9 Les théories linguistiques décrivent désormais le traitement du langage comme étant le résultat d'une interaction entre les informations provenant des différents domaines d'analyse (phonétique, prosodie, phonologie, morphologie, syntaxe, mais également discours). Mais ces théories diffèrent selon le moment où ces interactions entre les différents niveaux se font. Les approches classiques proposent une construction relativement indépendante de l'information pour chacun des domaines, en les décrivant séparément puis en définissant des mécanismes d'interface entre eux. Il s'agit en d'autres termes de ce qu'on appelle dans les approches traitant de la multimodalité d'une "fusion tardive". Ces approches reposent souvent sur une prééminence de la syntaxe que les autres domaines contribuent à informer et à partir de laquelle le sens peut être construit [Chomsky, 1981 ; Pollard & Sag, 1994]. Ces approches reposent sur plusieurs hypothèses. Tout d'abord, elles nécessitent pour chacun des domaines la possibilité de construire une représentation complète à partir desquelles les interfaces sont décrites (par exemple entre unités intonatives et unités syntaxiques, voir dans ce numéro l'article « La prosodie au Laboratoire Parole et Langage »<sup>3</sup>). De plus, elles reposent sur l'idée que les unités linguistiques de chacun des domaines sont porteuses d'un type d'information qu'il convient progressivement de composer pour parvenir à une interprétation finale, une compréhension de l'énoncé à traiter : il s'agit là de l'hypothèse de la **compositionnalité** selon laquelle le sens de

l'énoncé peut être calculé à partir du sens des unités qui le composent [Werning et al., 2012]. Elles s'appuient de fait sur une certaine conception de la modularité [Fodor, 1986] qui a été progressivement remise en cause notamment par les travaux décrivant l'interdépendance des unités linguistiques (voir [Van Valin & LaPolla, 1997] et surtout [Jackendoff, 1997]). Par ailleurs, et cet aspect est lié au précédent, ces approches reposent également sur l'idée que le traitement est **incrémental** et s'effectue mot à mot. Il s'agit d'une intuition classique et qui s'avère très pratique pour justifier les différentes architectures cognitives et computationnelles du traitement du langage : à chaque étape (à chaque mot nouveau), une série de traitements, toujours la même, est effectuée. De nombreux travaux se sont ainsi attachés à évaluer la charge cognitive liée à l'intégration d'un mot à la structure en cours de construction (voir par exemple [Hale, 2001]).

### 1.3 Prendre en compte toutes les sources d'information pour décrire la conversation : quel cadre théorique ?

- 10 Cependant, ces hypothèses de compositionnalité et d'incrémentalité ne résistent pas à l'explication du traitement du langage pendant une interaction conversationnelle, en situation naturelle. Ce type de contexte de production est caractérisé par plusieurs phénomènes qui sont jugés par les approches classiques comme non canoniques : entassements, répétitions, reprises, unités incomplètes, etc. [Blanche-Benveniste, 1997]. Ces phénomènes sont incompatibles avec l'idée qu'à chaque étape (à chaque mot), une information nouvelle est ajoutée. De plus, nous sommes également confrontés à la nécessité de traiter des unités ou des structures incomplètes. Il est par exemple fréquent que les relations entre unités linguistiques ne soient pas formellement marquées par des relations syntaxiques, mais par des informations provenant d'un autre domaine, comme la prosodie. A titre d'illustration, les entassements ne sont pas identifiables par des relations syntaxiques à proprement parler, mais plutôt par des indices morphologiques et prosodiques. Il est donc nécessaire de décrire les unités linguistiques comme résultant de cette interaction entre les différentes sources d'informations. La représentation de l'interaction existant entre les différentes sources d'information se situe donc au cœur de la compréhension des phénomènes conversationnels. Au passage, de telles représentations offrent une façon directe de décrire le caractère fondamentalement multimodal de la conversation. Les *Grammaires de Construction* [Fillmore, 1988] proposent un cadre adapté pour cela. Elles reposent en effet sur l'idée que les phénomènes linguistiques doivent être décrits à partir de structures (appelées *constructions*) permettant d'établir une relation directe entre une forme et un sens. La forme d'une construction est décrite par un ensemble de propriétés caractéristiques du phénomène décrit, et qui peuvent provenir de sources diverses. Par exemple, une construction interrogative sera décrite par une forme contenant des informations morpho-syntaxiques (par exemple les particules interrogatives), syntaxiques (par exemple l'inversion sujet-verbe), prosodique (un contour final montant). Pour ce qui concerne la représentation du sens, les Grammaires de Construction s'appuient sur la *Frame Semantics* (voir par exemple [Fillmore, 2009]) qui définit le sens comme un ensemble de "frames" décrivant des situations, des événements, des référents, etc. Nous pourrions par exemple trouver un frame associé à l'événement "échange commercial" ou encore un autre associé à l'événement "élection". Il est donc possible grâce aux *Grammaires de Construction* de définir

directement des associations entre une forme détaillée et un sens complexe. Cette association est directe, et donc non compositionnelle. Elle est ainsi adaptée à la représentation de phénomènes typiquement non compositionnels comme les constructions idiomatiques. Les Grammaires de Construction constituent donc une réponse particulièrement adaptée pour la description des interactions entre les différentes sources d'information, en proposant une architecture de traitement non compositionnelle et non incrémentale, particulièrement adaptée à la description du langage dans ses usages conversationnels.

- 11 Par ailleurs, les travaux en linguistique computationnelle (surtout les plus récents s'appuyant sur l'apprentissage automatique et l'apprentissage profond) ainsi qu'en psycholinguistique reposent largement sur la prédictibilité du langage. Il s'agit d'un point essentiel pour la compréhension des mécanismes à l'œuvre pendant une conversation. Sans entrer dans les détails, les différentes techniques permettant d'évaluer cette prédictibilité s'appuient sur le calcul de la probabilité d'apparition d'une unité linguistique en fonction de son contexte. Ce point est essentiel pour notre vision des mécanismes permettant de comprendre le traitement du langage en situation naturelle. Nous pensons en effet que seule une approche "holistique" reposant sur l'idée que toutes les modalités, tous les domaines sont potentiellement porteurs d'information permet d'expliquer les phénomènes conversationnels. Mais nous disons en même temps que ces sources d'informations peuvent varier d'un contexte, d'une situation à une autre : dans certains cas, un geste permettra à lui seul de transmettre un message (par exemple un geste déictique dans une situation où le locuteur doit simplement désigner un référent). Dans d'autres cas, la prosodie peut jouer un rôle central (par exemple exprimant une émotion particulière). Cela signifie que les sources d'informations sont variées, mais que la quantité d'information transmise par chaque source varie également en fonction du contexte. L'intérêt d'une approche fondée sur des constructions reposant sur ces multiples sources d'informations est donc crucial pour la description du langage en interaction : la prédiction se fait non plus pour chaque unité linguistique, de chaque domaine, mais au niveau de la construction entière. Dès lors qu'une construction est "activée" (au sens de [Lewis & Vasishth, 2005]), son sens devient accessible et les contraintes pesant sur sa forme deviennent facultatives. En d'autres termes, l'espace de réalisation de chaque unité linguistique, quel que soit son domaine, est plus ou moins grand (i.e. les formes sont plus ou moins variables) selon la quantité d'information fournie par chacun des domaines.
- 12 En plus de cette prise de conscience de la nature dispersée des informations dans le flux langagier et de la nécessité de traiter de manière opportuniste toutes les sources d'informations, l'analyse du "langage" en tant que ressource, requiert une réflexion sur la nature des représentations en particulier sémantiques dans la conversation.

#### 1.4 Une représentation partagée comme socle de la communication

- 13 L'étude du sens, comme les autres domaines, s'est focalisée sur l'interprétation ou la construction du sens des phrases. C'est à dire sur la construction d'une représentation sémantique (par exemple dans un langage formel) à partir d'une séquence de formes linguistiques (voir juste au-dessus la question de la compositionnalité) ou dans l'autre sens, et plus spécifiquement dans le domaine du traitement automatique de la langue,

dans la *génération* d'une séquence de formes linguistiques sur la base d'une représentation formelle. Ces représentations, et leurs statuts, sont donc au cœur de l'entreprise sémantique. Elles sont vues selon les auteurs comme des représentations purement formelles, ou comme des représentations mentales plausibles. En tant que représentations mentales plausibles, elles correspondent aux représentations que les producteurs et les interprètes d'une séquence linguistique utilisent.

- 14 Dans la conversation, un même énoncé reçoit plusieurs interprétations concurrentes, a minima celle de son producteur et celle de l'interlocuteur. Ce fait crée la possibilité de décalages subtils voire de différences majeures entre les représentations associées à un même énoncé, en raison de problèmes de perception ou d'interprétation en particulier. Ainsi le contexte, l'ensemble des faits établis auparavant dans le discours, au lieu d'être un ensemble de propositions grandissant de manière ordonnée à l'aide des phrases articulées dans le monologue, devient un objet beaucoup plus complexe dans le cas du dialogue. En effet, dans le monologue même si l'on peut supposer que la représentation du locuteur et celle de l'audience peuvent diverger, ces dernières ne sont jamais confrontées en tout cas pas au cours de la production du monologue. Le dialogue et la conversation changent radicalement cela puisqu'à chaque énoncé, la correspondance (ou pas) de ces représentations peut donner lieu à des séquences, par exemple de clarification.
- 15 Par conséquent, le contexte linguistique, devient en conversation un *common ground* (Clark, 1996 ; Stalnaker, 2002) ; espace partagé entre les locuteurs pour lequel on requiert même pour l'analyse de certains phénomènes, comme les demandes de clarification ou les désaccords, d'avoir des parties privées et publiques (Ginzburg, 2012). Il existe de multiples définitions du terrain commun, l'idée générale est qu'il contient l'ensemble des informations partagées, et mutuellement reconnues comme telles par les interlocuteurs (outre de l'historique de la conversation ces faits communs peuvent être établis sur l'appartenance à une même communauté, ou une connaissance interpersonnelle des locuteurs).
- 16 Ces évolutions illustrent dans ces domaines, comment l'étude du lien entre les formes et les représentations s'est orientée vers une étude de la communication comme une co-construction du *common ground* à l'aide des ressources linguistiques à disposition des participants.

## 2. Phénomènes de la conversation

- 17 Dans cette section, nous faisons un inventaire non-exhaustif, mais basé sur nos travaux antérieurs des phénomènes propres à la conversation. Il s'agit d'un inventaire descriptif qui sera notamment réutilisé dans la section suivante traitant de l'impact de ces phénomènes sur les questions linguistiques.

### 2.1 Tours de parole

- 18 Depuis Sacks *et al.* (1974), on sait que les conversations reposent sur un système de construction et d'alternance de la parole. Dans ce modèle d'organisation, on tente de comprendre comment les transitions de tours peuvent être aussi fluides, douces (*smooth turn transitions*) alors que les principes qui les régissent ne sont pas explicitement appris. Cette alternance de tours repose en fait sur un mécanisme qui



s'appuie sur deux composantes : celle de construction des tours (construction des unités de tours) et celle d'allocation des tours (régulation et négociation).

- 19 Les unités de construction des tours (*Turn-Constructional Unit* = TCU) peuvent être une construction phrastique, une unité lexicale (1 seul mot : “salut”, “quoi”, etc). Ces TCU sont considérées comme achevées lorsqu'elles donnent lieu à des points de complétion possible rendant la transition de tour pertinente, bien que non obligatoire<sup>4</sup> :
- 20 Ces points de complétion possibles 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.
- 21 Bien qu'il existe de nombreux travaux sur l'organisation des tours de parole, il reste encore beaucoup à explorer concernant le rôle des différents indices dans la construction des unités de tours. Quelques travaux ont montré en effet que la construction des unités de tours était fondée sur la conjonction des paramètres non seulement syntaxiques mais également prosodiques (Degand & Simon 2009 : BDU Basic Discourse Unit) et pragmatiques (Ford & Thompson 1996).

## 2.2 Signaux d'écoute / Feedback / Backchannel

- 22 Un deuxième phénomène propre à la conversation est la présence massive d'énoncés de 'feedback conversationnel' correspondant aux productions de l'interlocuteur<sup>5</sup> et qui fournissent des indications sur son traitement du message et / ou de l'information qu'il véhicule (voir l'article « Principes et outils pour l'annotation des corpus, section “feedbacks” »<sup>6</sup> dans ce numéro). Ces énoncés de *feedback* sont la pierre angulaire du processus de *grounding* de l'information dans la conversation, ou comment les informations entrent dans le *common ground* de la conversation. Le feedback est fortement associé aux signaux d'écoute (ou *backchannels* (Yngve, 1970)) car ils sont le support préférentiel des énoncés de feedbacks (au moins de ceux concernant les premiers *niveaux de communication* exprimant que l'interlocuteur est attentif et / ou qu'il a bien entendu les paroles du locuteur principal). L'expérience des corpus conversationnels (Schegloff, 1982 ; Gravano *et al.*, 2011 ; Prévot *et al.*, 2016) a montré que ces énoncés de feedbacks se concentrent en deux lieux précis : (i) des énoncés courts et isolés, (ii) des unités situées dans la périphérie initiale d'énoncés plus complexes.

Table 1. Fréquence du Feedback Conversationnel dans 4 corpus du LPL

Corpus	# Tokens	#IPUs <sup>7</sup>	Durée	#Feedback <sup>8</sup>	Ratio IPU <sup>9</sup>
CID	125 619	13 134	7 h 34	3275	0,25
MTR	42 016	6 425	2 h 32	1652	0,26
MTX	36 923	5 830	2 h 33	1521	0,26
DVD	64 023	7 989	4 h 12	2138	0,27
CoFee (Tous)	268 581	33 378	16 h 51	8586	0,26

- 23 Après extraction des formes lexicales et quasi-lexicales (*mh, euh,...*) dominant ces positions isolées et initiales dans les données, il est possible de raffiner l'approche en examinant et en augmentant l'empan des domaines initiaux et finaux des formes observées à leur frontière. Dans l'exemple (CID), l'ensemble de la séquence initiale entre crochets est ainsi considéré au lieu de se limiter à un nombre fixe d'éléments. Un tel travail sur 4 corpus différents construits au LPL a permis d'identifier qu'un peu plus de 25 % des IPU's inclut une fonction de feedback explicite<sup>10</sup> et ce dans des corpus conversationnels de nature très différente (Voir Table 1.).

(CID) [oui oui oui oui ouais] mais il était en Arabie Saoudite c'est plus hard euh parce que les Emirats t'as d- t'as de l'alcool tout ça quoi

- 24 Des énoncés lexicalisés, donc plus complexes et plus longs que les items "génériques" (Bavelas *et al.* 2000) du type "mh", "ouais", envisagés ci-dessus, peuvent également être produits depuis le rôle d'interlocuteur. Il s'agit des feedbacks "spécifiques" (Bavelas *et al.* 2000) proches des "assessments" de Schegloff (1982). Ces énoncés renvoient à des formes lexicales variées, un ton de voix particulier, ou encore à certains éléments mimo-gestuels (sourire). Fréquemment produits sous forme de fragment, les feedbacks spécifiques renvoient à divers procédés interactionnels tels que des hétéro-répétitions ou des complétions (voir l'article « Principes et outils pour l'annotation des corpus, section "feedbacks" »<sup>11</sup> dans ce numéro). Nous avons pu montrer sur environ 150 narrations conversationnelles issues des 8 heures du CID que les feedbacks génériques et les feedbacks spécifiques étaient localisés dans des phases spécifiques, dépendant étroitement du niveau de common ground des participants. En effet, les feedbacks génériques apparaissent préférentiellement dans les premières phases de la narration durant lesquelles s'élaborent les connaissances partagées relatives au récit (personnages, événements, etc) tandis que les feedbacks spécifiques apparaissent préférentiellement en fin de narration, lorsque le niveau du common ground est suffisant et permet à l'interlocuteur de produire une évaluation et de devenir potentiellement, et ponctuellement, co-narrateur (Guardiola & Bertrand 2013 ; Bertrand & Espesser 2017). Ces résultats confirment et complètent ceux de Bavelas *et al.* (2000) sur des récits plus contrôlés.

### 2.3 Marqueurs de discours, Phatiques

- 25 Les marqueurs de discours sont largement répandus dans d'autres genres que la conversation. Ils servent à lier et à articuler les phrases et des zones de discours plus larges entre elles.<sup>12</sup> Globalement, ces formes permettent d'exprimer des liens entre les propositions, portées à l'écrit par les phrases. Leur nature est par conséquent plus riche dans le genre conversationnel. Premièrement, nous venons de le voir, les locuteurs produisent en grande quantité des énoncés de feedback verbaux composés de formes lexicales simples et qui opèrent de manière similaire aux connecteurs sur un empan assez large. Des formes apparaissent également pour exprimer une attitude d'un locuteur (*e.g. ah / mais, bon*) envers une proposition. Enfin, certaines de ces marques ont un rôle plus interactionnel encore car dirigé vers l'interlocuteur (*ex tu\_vois, hein*).
- 26 Ces formes peuvent être approchées par l'examen des distributions de fréquences des éléments lexicaux situés sur ces positions initiales, finales et isolées. Par exemple, pour le corpus CID, la table 2 montre les mots plus fréquents du corpus, et ceux relativement plus fréquents (+20 % de fréquence brute) dans des positions isolées (IPUs de 3 mots ou

moins), initiales (début d'IPU de longueur > 4 mots) et finales (début d'IPU de longueur > 4 mots). Nous soulignons, bien que ces listes ne soient guère surprenantes, qu'aucune des formes apparaissant préférentiellement dans ces positions (à l'exception des conjonctions *et*, *mais* et *donc* et de l'adverbe *alors*) ne sont étudiées de manière intensive dans les travaux linguistiques non-focalisés sur la conversation ou l'oral spontané. Par ailleurs, même en ce qui concerne les formes étudiées il a été montré des glissements d'usage et de sens important pour ces contextes (Hansen, 1997). Enfin, tous ces termes appartiennent à la partie haute de la distribution de fréquence des mots du corpus ce qui souligne leur importance dans la prise en compte des phénomènes de la conversation.

Table 2. Mots plus relativement plus fréquents (+20 % de fréquence brute) dans les positions isolées, initiales et finales que dans le reste du corpus

Isolés	mh, ouais, ah oui, voilà, putain, d'accord, hum, eh, uh, ok
Initiaux	et, mais, ah, donc, non, alors, ben, oh, putain
Finaux	euh, quoi, hein, tu_vois, hum, pff

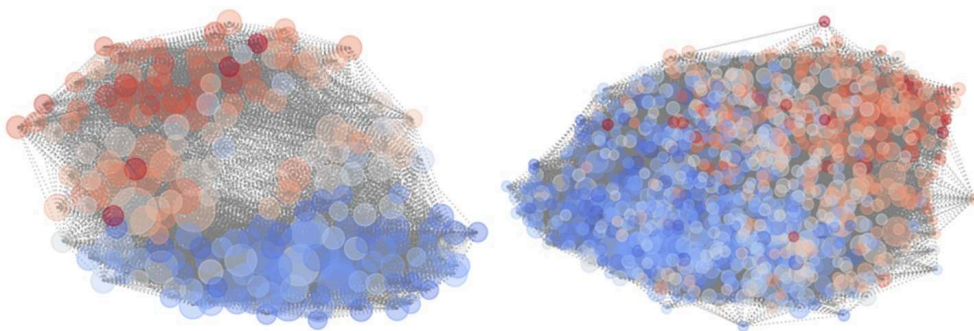
## 2.4 Fragmentation des énoncés

- 27 Dans le domaine syntaxique, les structures syntaxiques, étudiées principalement sur le langage écrit en ce qui concerne les langues les plus décrites, ne sont pas remises en cause mais plutôt exploitées dans ce contexte conversationnel. Il a été observé que leur distribution est radicalement différente entre l'écrit et l'oral conversationnel. Plus précisément, en ce qui concerne les structures des énoncés, les différents domaines de la syntaxe de l'oral (voir l'article « Le début des phrases en français parlé »<sup>13</sup> dans ce numéro) ainsi que des travaux de (Haselow 2017 ; Degand et Crible, 2021) pointent sur l'importance des domaines initiaux et finaux, travaillés aussi avec les notions de périphéries droite et gauche (voir l'article « Périphéries gauche et droite »<sup>14</sup> dans ce numéro) et des parenthétiques, autant de phénomènes étroitement associés avec la position syntaxique d'IP<sup>15</sup>-adjoind, parfois analysée comme une position discursive (Aurnague *et al.*, 2001). Par ailleurs, ces distributions varient au point que certaines structures disparaissent à l'oral ou que certaines collocations extrêmement fréquentes se lexicalisent (*e.g il\_y\_a, il\_y\_en\_a, mais\_bon, je\_sais\_pas...*) (Lambrecht, 1987) alors qu'elles sont inexistantes à l'écrit. Par ailleurs, l'ordonnancement des structures est perturbé, au moins en surface, par les disfluences, les interruptions et les complétions. Enfin, ces structures sont enrichies par d'autres modalités voire intégrées dans des constructions multimodales (voir l'article « Multimodalité(s) dans les interactions humaines »<sup>16</sup> dans ce numéro).
- 28 Concernant le niveau syntaxique et positionnel, un point qui change radicalement la prise en compte des données conversationnelles est la présence massive de matériel linguistique sur des positions syntaxiques non-régies fermement à un des éléments de la structure prédicative principale de l'énoncé oral (voir l'article « Le début des phrases en français parlé »<sup>17</sup> dans ce numéro). Un premier travail d'approximation pour atteindre le matériel linguistique situé sur ces positions consiste en l'analyse des débuts

et des fins d'unités. Une difficulté réside cependant dans la notion même d'unité de base de l'analyse. En effet, pour ne pas anticiper l'analyse linguistique nous ne travaillons pas l'oral conversationnel avec des transcriptions ponctuées.<sup>18</sup> Nous utilisons deux stratégies différentes, en fonction des corpus étudiés pour déterminer ces unités. Dans le cas où nous ne disposons que de la transcription orthographique brute alignée sur le signal, nous utilisons les Unités Inter-Pausales (IPUs)<sup>19</sup>. Dans certains cas, nous disposons d'une segmentation en unités de discours effectuée lors de campagne d'annotation (Prévoit et al, 2015 ; Prévoit *et al.*, 2021).

- 29 Une visualisation intéressante pour appréhender l'impact radical de la conversation sur les structures syntaxiques est proposée dans la figure 1. Pour la produire, nous avons tout d'abord traité et étiqueté morphosyntaxiquement deux corpus : écrit (Corpus Grace) vs. conversation (CID). Nous avons ensuite extrait toutes les séquences de 5 tags et nous avons gardé les N plus fréquents. Nous avons ensuite fusionné les sources (écrit et conversation) des séquences (en retenant leurs fréquences relatives dans les deux genres) et nous avons construit une matrice de distance pour toutes ces séquences en utilisant la distance de Levenshtein sur les séquences de tags. Sur cette base nous construisons le graphe des séquences (nœuds du graphe) dont la connectivité dépend de la distance entre deux séquences.<sup>20</sup> Enfin, nous avons attribué une couleur à chaque condition sur la base du ratio de fréquence conversation / écrit (plus bleu : plus fréquent à l'écrit ; plus rouge : plus fréquent en conversation). Cette figure nous montre que bien que la plupart des structures soient communes aux deux conditions, la fréquence des séquences est radicalement différente d'un genre à l'autre et la répartition dans le graphe de ces séquences illustre que ces différences sont valables pour des familles entières de séquences (et non quelques structures particulières)<sup>21</sup>. Ce travail a été reproduit pour l'anglais et l'on observe des résultats similaires (Pour plus détails voir Prévoit *et al.*, 2019).

Figure 1 Distance entre les 5-grammes d'étiquettes morphosyntaxiques les plus fréquents



Gauche : français ; Bleu = écrit (Grace) ; Rouge = conversation (CID) ; N = 200  
Droite : anglais am. ; Bleu = écrit (Brown) ; Rouge = conversation (Switchboard) ; N = 1 000

- 30 Un phénomène qui illustre particulièrement bien ces particularités des structures syntaxico-sémantiques de la conversation est la complétion d'un énoncé par un interlocuteur comme illustré en (1). Prises ensemble, les parties produites par le narrateur (AB\_203b) et l'interlocuteur (CM\_338), forment un tout cohérent grâce à l'anaphore (le pronom "les" correspondant à l'antécédent "ces bonnes soeurs") et le temps verbal (présent simple). Les structures impliquées dans chaque énoncé sont particulièrement incomplètes : (AB\_203b) est composée d'une première proposition

complète (*oui euh faites un truc*) d'une articulation par un marqueur de discours (de toute façon) puis d'un syntagme nominal ; L'interlocutrice reprend en complétant ce syntagme nominal par une construction verbale (précédée d'un feedback). Si l'analyse et l'interprétation des deux productions de manière conjointe ne posent aucune difficulté, ces complétions, banales même si elles ne sont pas fréquentes interrogent les modèles de production des énoncés et des structures syntaxiques.

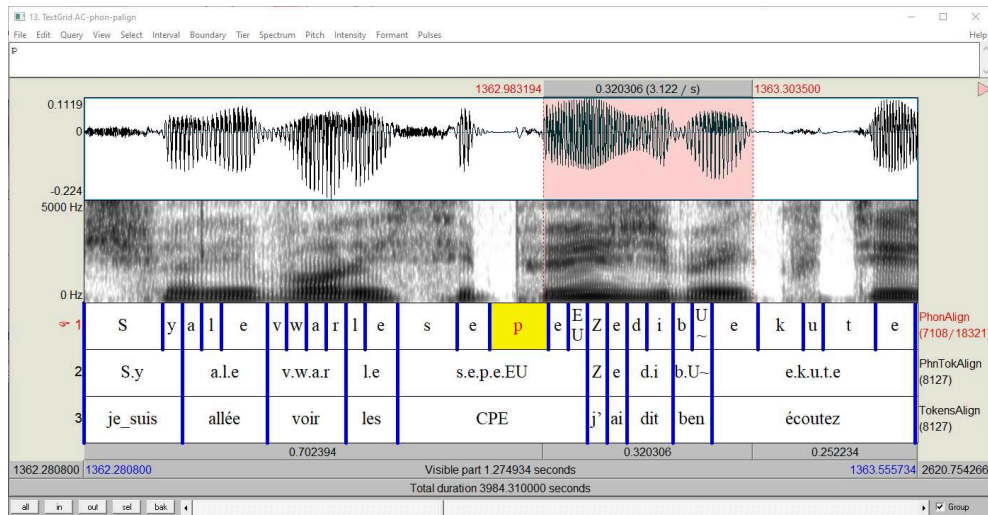
#### Exemple 1 (extrait de Guardiola et Bertrand 2013)

AB_203	et puis on avait dit \$on va faire un truc\$ etc elle avait <u>dit oui</u>
CM_336	<u>ah ouais</u>
AB_203b	<u>oui euh faites un truc de toute façon ces bonnes sœurs euh</u> (0.305)à
CM_337	ouais (0.523)
CM_338	<b>faut les bouger</b> @
AB_204	<u>et patin et couffin</u> etc faut les bouger etc puis
CM_339	mh
AB_204b	<u>façon tout le monde va croire que ce sont les carabins parce qu'à chaque</u> <u>fois qu'il y a des conneries qui sont faites</u>
CM_340	@
CM_341	ouais + <u>ouais</u>
AB_205	c'est la faute des carabins etc
CM_342	ouais

## 2.5 Réductions

- 31 Si la parole en conversation fait apparaître des phénomènes absents ou rares dans d'autres situations de parole (tels que les feedbacks, par exemple), elle en fait également disparaître certains. Notamment de nombreux phonèmes et syllabes tendent à s'affaiblir au point parfois de disparaître complètement. Si l'affaiblissement ou l'omission des phonèmes apparaît dans tous les types de productions orales, y compris les plus contrôlées, cela concerne essentiellement les phénomènes de schwa ou des simplifications phonotactiques de clusters consonantiques (« parce que » produit /pask/ par exemple). Mais les **réductions** en conversation sont d'une tout autre dimension et affectent un nombre considérable de mots. Johnson (Johnson, 2004) observe que dans un corpus de parole conversationnelle (Pitt et al., 2005) plus de la moitié des mots sont déviants par rapport à leur forme prototypique. Ce phénomène est très documenté dans un numéro spécial de *Journal of Phonetics* (Ernestus et Warner, 2011) qui fait état des différentes formes de réduction et des problèmes théoriques que cela soulève. L'étude des réductions de la parole a pris un essor considérable depuis une vingtaine d'années en prenant appui sur l'analyse des grands corpus de parole, le plus souvent spontanée. La réduction représente une diminution de l'information phonétique, compensée par l'information contextuelle très riche présente en parole conversationnelle. D'un point de vue phonétique, les réductions peuvent être caractérisées comme des omissions ou encore des assimilations, mais très souvent on trouve des phénomènes de coalescence dans lesquels il n'est pas possible de déterminer la nature des phonèmes qui sont réalisés dans la séquence réduite. Cet aspect est important car il signifie qu'en conversation les mots ne sont pas produits sous la forme de phonèmes distincts. Ces réalisations réduites peuvent alors prendre des caractéristiques très éloignées de ce que l'on trouve en parole lue.

Figure 2 : Séquence extraite d'une conversation dans le Corpus CID (Bertrand *et al.*, 2008) : « je suis allée voir les CPE j'ai dit ben écoutez »



32

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

Séquence totale « je suis allée voir les CPE j'ai dit ben écoutez » (figure 2).

33

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

Séquence sélectionnée en rose (figure 2) : « **ÉEU J'AI DIT BEN É** » [εOZεδιβE]ε. La transcription comporte 9 phonèmes mais dans la réalisation effective ces phonèmes ne sont pas identifiables de façon distincte.

- 34 La réduction phonétique est présente dans toutes les langues (Gendrot & Adda-Decker, 2007, Ernestus et Warner, 2011), mais la façon dont elle se manifeste dépend des propriétés linguistiques de chaque langue. Il a ainsi été montré que la réduction est liée à plusieurs facteurs linguistiques et communicatifs tels que la phonologie (Kemps *et al.*, 2004, Pharao, 2010, Meunier, 2016), la fréquence des mots (Pluymakers *et al.*, 2005), la situation de parole (Gendrot *et al.*, 2012), ou les marqueurs de discours (Bybee *et al.*, 2016).

## 2.6 Disfluences et Interruptions

- 35 Comme rappelé plus haut, la parole conversationnelle contient des séquences *disfluentes*, c'est-à-dire où l'enchaînement linéaire des unités est perturbé et laisse place à des silences, des pauses remplies et à du matériel lexical surnuméraire. Ces phénomènes créent des structures absentes du langage écrit mais aussi de la parole lue ou préparée car ils sont généralement interprétés comme un manque de compétence oratoire dans ces exercices. En parole spontanée, ils sont en revanche extrêmement fréquents. Pallaud *et al.* 2019 reporte qu'un mot sur 15 est disfluent dans le corpus CID et les estimations pour les corpus conversationnels en général varient entre 3 % et 6 % des mots du corpus.

- 36 Plus précisément les disfluences correspondent à une interruption de la fluence verbale au sein de l'énoncé, impactant généralement le domaine morpho-syntaxique. Cet ancrage morpho-syntaxique ne limite pas les domaines impliqués dans les disfluences étant donné qu'elles ont des effets manifestes sur la réalisation phonétique et sur la prosodie et que leur traitement complet requiert de prendre en compte la sémantique et la pragmatique. Ces interruptions peuvent être des énoncés purement *abandonnés* comme en (2), c'est à dire trop incomplets pour être interprétés comme tels et non-immédiatement "réparés". Une autre catégorie correspond à des suspensions de la fluence verbale mais malgré laquelle l'énoncé parvient à se dérouler sans impact sur la structure syntaxique (3). La dernière catégorie inclut une *réparation* de la séquence morpho-syntaxique à cause notamment de la présence d'un mot tronqué ou du matériel lexical surnuméraire (3). Dans ce dernier cas, (Shriberg, 1994) a proposé d'analyser ces disfluences en trois zones : le *Reparandum* ('r' dans les exemples) (la zone à réparer), l'*Interregnum* ('B' dans les exemples) (le point de *break*, qui peut être vide) et le *Reparans* ('R') (la zone réparante).

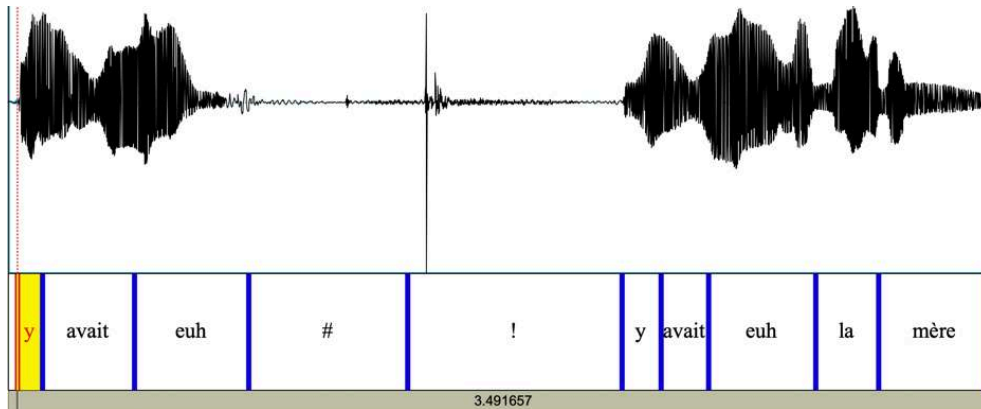
(2) [que j'avais envie d- enfin bref]<sub>abandonné</sub> # [et (#)S on l'a accueillie (b- a-)r on lui a rien demandé]

(3) (ou des)<sub>r</sub> (euh non)<sub>B</sub> (pas des)<sub>r</sub> (f-)<sub>B</sub> (pas des frustrations)<sub>R</sub> (des (#)s espèces de)<sub>r</sub> (euh # mhm # ouais)<sub>B</sub> (des)<sub>r</sub> (des vues différentes sur le boulot quoi)<sub>R</sub>

## 2.7 Formes pseudo-lexicales

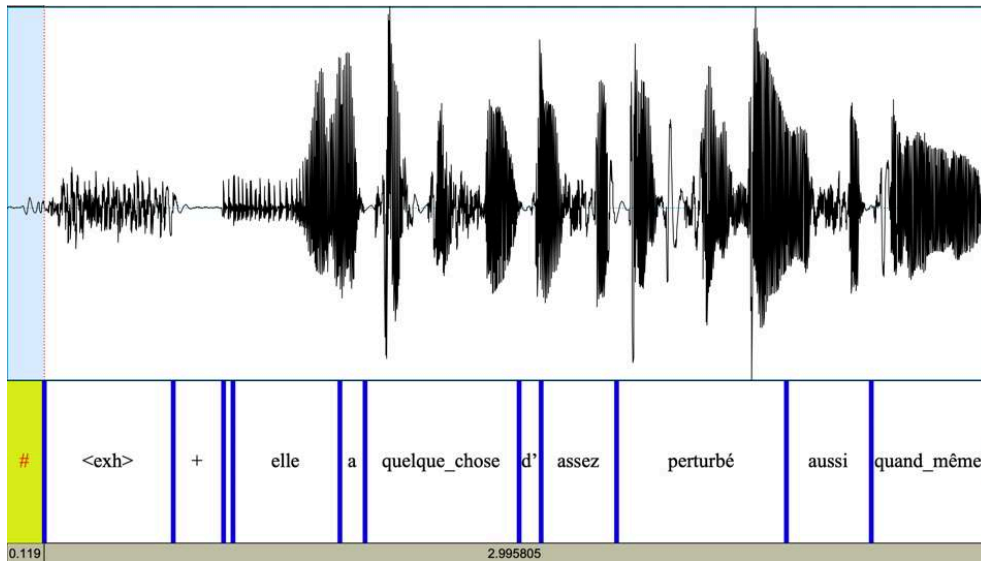
- 37 Certaines formes absentes dans le répertoire linguistique de la langue donnée peuvent se produire en conversation. Les cas les plus spectaculaires sont sans doute les études sur les clicks de l'anglais conversationnel réalisées par Richard Ogden et ses collègues (Ogden, 2013). Ils montrent, sur la base d'étude qualitative sur corpus que les clics en anglais conversationnels possèdent des distributions et des fonctions semblables à d'autres objets indiscutablement linguistiques. Ward (2006) parvient à une conclusion similaire dans son étude systématique des *Conversational Grunts*. Dans une étude de corpus non-publié (Prévot / Bertrand / Granier, 2018) nous avons annoté tous les événements de parole du corpus CID transcrits comme des bruits (symbole \* dans nos conventions). Les formes ré-annotées (N=863) étaient principalement composées d'inspirations sonores (512), de clicks (101), de raclement de gorges (96) et d'expirations sonores (31). Nous avons également annoté la fonction de ces items : il s'agissait principalement de *zone préparatoire* à la production d'énoncé (480) (en particulier 415 prises de souffle sonores), mais aussi de recherche d'idée ou de mot (70) (Figure 3, exemple sonore), ou encore d'expression d'une forme d'embarras (17) (Figure 4, exemple sonore), ou d'être pris au dépourvu (16) et quelques cas de feedback (Figure 5, exemple sonore) ou d'annonce d'auto-correction.

Figure 3 Click constituant le break d'une disfluençe, en position de recherche de mot



Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

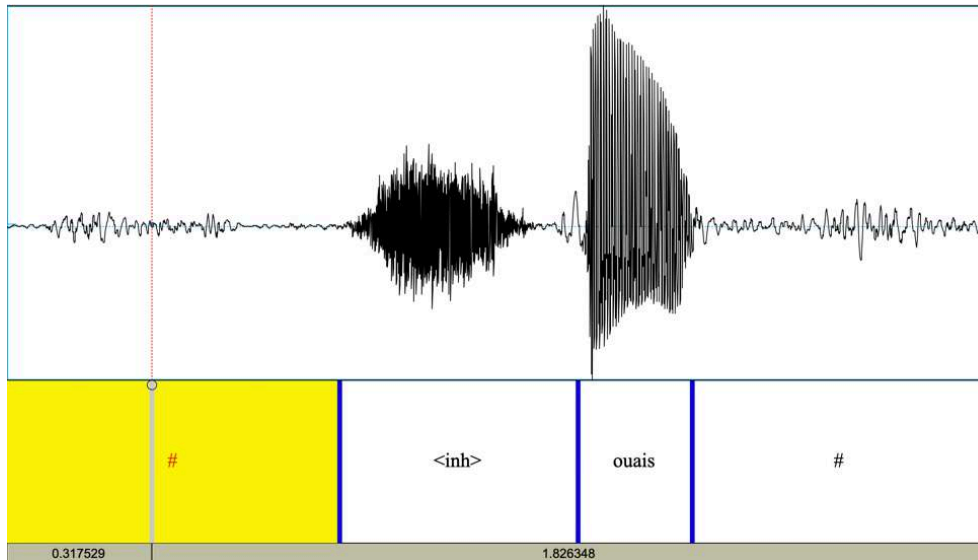
Figure 4 Expiration sonore signalant une forme d'embaras



Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>



Figure 5 Inspiration sonore constituant un feedback conversationnel



40

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

- 41 Au-delà des formes sonores sans matériel segmental, nous observons l'apparition de formes segmentales qui semblent résulter de phénomènes de réduction et de lexicalisation (en cours) extrêmes. Ainsi, *bouais*, *mouais*, *vouais* sont des formes peu fréquentes mais présentes dans les corpus alors que d'autres formes proches comme *\*touais*, *\*louais*, *\*chouais*, etc sont introuvables et semblent interdites. Cette différence semble pouvoir s'expliquer par l'agrégation puis la réduction des séquences : *bon / ben + ouais*, *mh + ouais*, *voilà + ouais*. Comme nous l'avons dit plus haut, tous ces items sont fréquemment utilisés comme marqueurs de feedbacks. Ils apparaissent donc de manière conjointe et fréquemment leur valeur sémantique tend à fusionner.
- 42 Plus simplement, un item comme *mh* dont nous avons parlé à plusieurs reprises déjà, ne fait pas partie du répertoire phonologique (ni lexical) du français. Il n'en est pas moins une des formes les plus fréquentes de l'oral conversationnel.<sup>22</sup>
- 43 Enfin ces formes qui semblent correspondre au niveau lexical, ont fréquemment valeur d'un énoncé complet (une unité discursive complète), comme peuvent l'avoir certains gestes ou expressions mimo-gestuelles.

## 2.8 Manifestations de la convergence inter-locuteurs

- 44 Différentes manifestations de la convergence ont été étudiées (du niveau phonétique au niveau mimo-gestuel en passant par les divers niveaux syntaxique, prosodique, et discursif). L'un de nos objectifs est de décrire la convergence en rassemblant toutes les modalités et sources d'information dans un modèle unique. Différents types de mesures sont utilisés pour évaluer la convergence : la répétition, la similarité, les modèles conjoints (coproduction unitaire), la dynamique (alignement des modèles de l'interlocuteur), etc... Notre modèle sera basé sur des caractéristiques fournissant des mesures issues de ces différents domaines. Concernant la prosodie, nous avons déjà montré comment les locuteurs imitent certaines caractéristiques du son de la parole de

leur interlocuteur en comparant les formes de F0 et leurs moyennes (Aubanel & Nguyen, 2020). Nous proposons de passer à l'échelle supérieure en appliquant le même type de mesure au niveau d'une conversation (similarité de forme, moyenne de F0, débit de parole). Notons qu'une telle mise à l'échelle nécessite une étude de l'interaction entre d'autres dimensions telles que la prise de parole, l'interaction sociale, l'engagement (De Looze *et al.*, 2014), les activités/pratiques au sein des conversations, etc.

- 45 Sur ce dernier point, nous avons plus particulièrement étudié la convergence à travers différentes pratiques interactionnelles telles que les listes (Bertrand & Priego-Valverde 2017), ou le discours rapporté (Guardiola & Bertrand 2013), catégorisées comme des feedbacks spécifiques (cf. section feedbacks ci-dessus). Nous avons montré que la convergence interactionnelle dépendait des attentes générées par l'activité et les rôles qui lui sont associés. Lorsque l'interlocuteur produit un feedback générique au début d'une activité pour montrer qu'il accepte d'entrer dans cette dernière, il s'aligne à cette activité (au sens de Stivers 2008). Il peut aussi s'affilier en adoptant le même point de vue que son partenaire. La convergence interactionnelle requiert cet alignement et cette affiliation mais n'est pas suffisante. En effet, une séquence convergente requiert 3 mouvements (énoncé 1 par le locuteur, énoncé 2 = FB spécifique par interlocuteur, énoncé 3 : ratification de ce FB par le locuteur). La convergence ainsi définie est donc ponctuelle, locale, et peut entraîner un changement, voire une inversion, des rôles discursifs (voir l'article « Alignement, affiliation et trajectoire interactionnelle dans la conversation »<sup>23</sup> dans ce numéro).

### 3. Impact sur les modèles linguistiques

- 46 Dans cette section, nous revenons sur certains des phénomènes évoqués au-dessus et nous montrons en quoi ils impactent radicalement les travaux linguistiques, sur les questions fondamentales des formes linguistiques et des unités linguistiques.

#### 3.1 Formes linguistiques méconnues

- 47 Comme nous l'avons vu dans la section 2.5, les conversations hébergent, pour une langue donnée, des formes linguistiques absentes des autres genres. Cela a plusieurs conséquences importantes pour les théories et modèles linguistiques.
- 48 Premièrement, ces formes illustrent la capacité de mobilisation par les locuteurs de tous les signaux envisageables pour atteindre leurs buts cognitifs. Cela ouvre la question de la créativité pour non seulement créer de nouvelles formes mais surtout pour exploiter des formes afin d'atteindre certains effets communicatifs dans des contextes précis. Ce phénomène est généralement accepté pour les études gestuelles mais beaucoup moins connu en ce qui concerne les formes sonores. Ces phénomènes semblent se positionner sur un continuum en termes de conventionnalité vs. créativité des formes et des associations forme-fonction et pour lequel le contexte et l'usage jouent un rôle prépondérant. Ceci montre à quel point les différentes modalités, dans le contexte conversationnel, sont proches en termes de fonctionnement. Ces observations rappellent à quel point un modèle linguistique doit être un modèle multimodal du langage et de la communication.

- 49 Par ailleurs, c'est sans doute cette flexibilité présente en permanence et dans tous les domaines linguistiques, qui permet l'apparition des variantes notamment lors de dynamiques inter-personnelles de convergence comme illustré dans la section 2.8 et développé en section 3.4.
- 50 Deuxièmement, certaines de ces formes non-standard sont en réalité parmi les plus fréquentes du genre conversationnel. Au niveau lexical, c'est le cas des formes *mh* et *eh* qui occupent une place de choix dans les distributions de fréquences. Dans des modèles de traitement où la fréquence joue un rôle important, la présence de ces formes dans ces listes impacte significativement les modèles de langues associés. Ce phénomène ne se réduit pas à la dimension lexicale et concerne aussi la dimension phonémique.
- 51 La distribution des phonèmes dans les corpus est en partie liée à la fréquence des mots. On observe ainsi une surreprésentation de certains phonèmes due à leur présence dans des mots ou pseudo-mots très utilisés en conversation. C'est le cas de /ɛ/ dans 'ouais' ou de la nasale /m/ dans 'mh'. La réduction phonétique est souvent évoquée pour la production de la parole en conversation, mais en réalité, il est plus question d'une très forte variabilité temporelle allant des segments extra-courts (inférieurs à 30 ms) à des segments extra-long (voyelles de 2 sec.). Ces allongements phonétiques sont le plus souvent distribués dans des interjections.

### 3.2 Implications des phénomènes de réduction phonétique

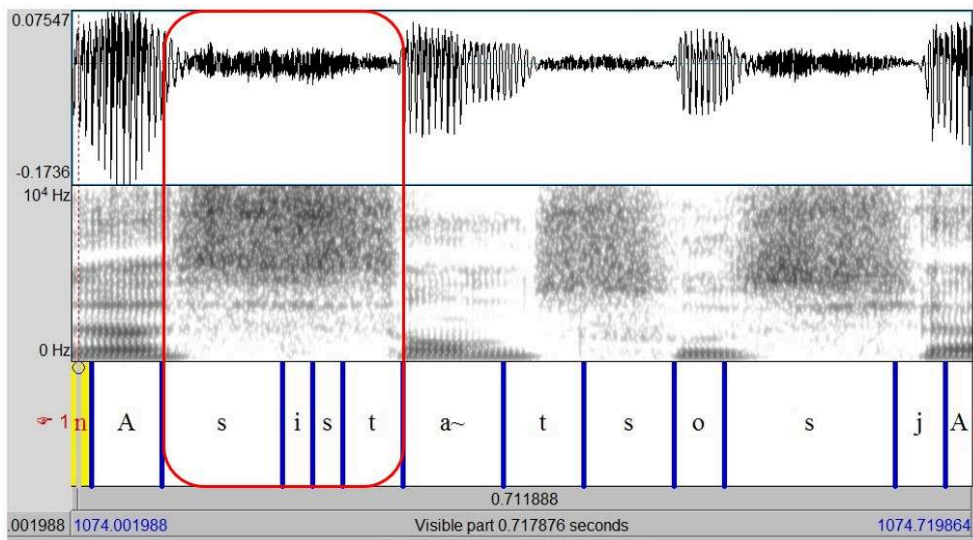
- 52 L'observation accrue des phénomènes de réductions phonétiques amène à poser les questions suivantes :
1. Pourquoi les réductions sont-elles aussi importantes en conversation ?
  2. Comment fonctionnent-elles ?
  3. Comment parvenons-nous, malgré tout, à percevoir les mots ?
- 53 A la première question il est tentant de répondre que cela répond à la nécessité de réduire notre effort articulatoire. C'est probablement le cas dans la mesure où des contraintes bio-mécaniques nous imposent une optimisation des gestes articulatoires en fonction du débit de notre parole. Or le débit de parole en conversation est plus élevé qu'en parole plus contrôlée. Toutefois, si l'augmentation du débit engendrait systématiquement et physiologiquement une réduction de nos gestes articulatoires, nous devrions observer une réduction phonétique homogène et constante sur la parole en conversation. Or, nous ne réduisons pas « n'importe où ». Nos efforts sont également contraints par la communication et son succès (Lindblom, 1990). Ainsi, les réductions affectent plus spécifiquement les mots à faible contenu sémantique (mots fonction : pronoms, déterminants, conjonctions, auxiliaires, etc., Johnson, 2004 ; Ernestus, 2014, Meunier & Bigi, 2021). Enfin, l'importance des réductions est proportionnelle à la richesse de l'information contextuelle : plus l'information est conséquente, dispersée et donc redondante (comme c'est le cas en conversation), plus il est possible de sous-spécifier le niveau phonétique sans que l'information globale soit altérée. A l'inverse, si le contexte informationnel est pauvre, le support phonétique doit être exhaustif sans quoi le message oral peut être inintelligible.
- 54 La deuxième question est probablement celle qui a le plus attiré notre attention. Sachant que les réductions n'apparaissent pas au hasard, quelles sont les contraintes

qui régissent leur apparition ? Mais d'une certaine façon la question peut être posée de façon inverse : pourquoi certaines parties de nos conversations ne sont-elles pas réduites ? La réponse qui nous vient en premier est que ces parties contiennent des informations importantes pour la communication avec l'interlocuteur. Ces informations seraient suffisamment importantes pour que le support phonétique nécessite de rester intact. La conclusion est ainsi inversée pour les réductions : les informations contenues dans les mots ne nécessitent pas une forme phonétique complète pour que les mots soient perçus. Comme nous l'avons vu plus haut, il a été montré que les mots de contenus étaient généralement moins réduits que les mots fonction, moins porteurs d'information. Toutefois, la réduction ne se projette pas uniquement dans le cadre du mot. Les formes de réduction les plus documentées concernent des mots ou des séquences de mots très fréquents dans le langage courant comme illustré dans la figure 6. Des exemples existent dans de nombreuses langues.

Figure 6 Exemples extraits d'Ernestus & Warner (2011)

- English: "a little" produced /ɔlə/ "yesterday" produced /jeʃeɪ/
- French: "je sais pas" produced /ʃepa/ "c'était" produced /ste/
- Dutch: "ongever" produced /ɔfer/ "koninklijk" produced /konɪk/.

55 Ces formes spécifiques sont généralement identifiables par les auditeurs (et transcripteurs, [voir l'article « 30 ans de recherches en phonétique clinique au LPL »<sup>24</sup> dans ce numéro]) capables de détecter les formes réalisées : « je ne sais pas » ou « je sais pas » ou « ché pas » ou « chpa ». Ces réductions *stéréotypées* peuvent être considérées comme différents exemplaires d'une même représentation lexicalisée [voir 2.4 Fragmentation des énoncés]. Elles apparaissent plusieurs fois dans une conversation et sont généralement produites par tous les locuteurs. Cependant, certaines réductions plus *opaques* apparaissent également en conversation. Elles ont été moins étudiées car i/ elles sont très difficiles à percevoir dans le flux de parole (figure 7) et ii/ elles n'affectent pas des mots ou des séquences spécifiques (Meunier, 2013), iii/ la séquence réduite n'apparaît éventuellement qu'une seule fois dans une conversation. En tant que telles, ces réductions sont extrêmement difficiles à collecter (car non perçues et peu reproductibles) et sont dispersées dans de longues unités, telles que la prosodie ou le discours, et peuvent affecter n'importe quelle séquence ou mot du moment qu'ils surviennent à un point critique du flux de parole. C'est justement la distribution de ces réductions opaques, probablement dépendantes d'une organisation supra-lexicale, qu'il nous reste à mieux comprendre désormais. Ainsi la distribution des réductions phonétiques est vraisemblablement régie par l'ensemble des strates linguistiques (phonologie, lexicale, syntaxe, prosodie, discours, etc.).

Figure 7 : Séquence extraite du Corpus CID (Bertrand *et al.*, 2008) : « assistante sociale »

56

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

Séquence totale « assistante sociale » (figure 7).

57

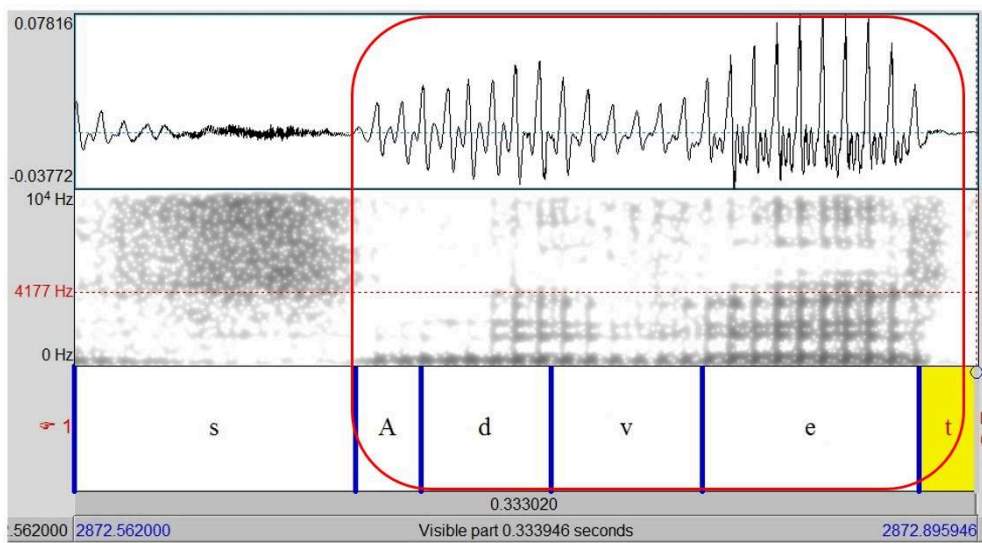
Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

Séquence « sist » [sist] (entourée en rouge) transcrite avec 4 phonèmes alors qu'un seul segment bruité est visible sur le signal de parole.

- 58 La troisième question nous renvoie aux modèles de perception et à la façon dont ceux-ci formalisent le stockage phonétique des mots. Qu'il s'agisse des modèles abstractionnistes (Lahiri & Reetz, 2002), des modèles à exemplaires (Goldinger, 1998), ou encore des modèles hybrides (Pierrehumbert, 2002), le mot est toujours le cadre dans lequel se structure le niveau phonétique. Ainsi dans l'ensemble des modèles les formes réduites sont considérées comme des variantes de production des mots. Les modèles hybrides peuvent rendre compte du phénomène de réduction, même s'ils ne peuvent expliquer l'ensemble du phénomène (Ernestus, 2014). Mais il reste que certaines formes de réductions (les réductions opaques) rendent inopérantes l'accès au mot via une entrée phonétique. En effet, ces réductions dépassent le cadre du mot et doivent être interprétées dans un modèle de compréhension et non seulement de perception. Notamment, il semble nécessaire d'envisager que l'accès à la totalité des mots n'est pas indispensable pour l'interprétation d'un énoncé en conversation. Cela ne signifie pas que ces mots n'ont pas besoin d'être produits, mais plutôt que leur forme très approximative et incomplète est un support suffisant pour comprendre l'ensemble de l'énoncé.
- 59 Une autre hypothèse, qu'il nous reste à explorer, est basée sur la pertinence de certains détails phonétiques fins (*Fine Phonetic Details*, Hawkins, 2003). Considérer les réductions comme des productions phonétiques sous-spécifiées c'est admettre que leur réalisation pourrait être remplacée par n'importe quel support phonétique sans que cela nuise à la

transmission de l'information. Or, en sommes-nous si certains ? L'observation de certaines réductions nous amène à soutenir une autre hypothèse. Dans certains cas, nous avons pu noter que le processus de réduction tendait à préserver des caractéristiques phonétiques porteuses d'information. Dans l'exemple ci-dessous (figure 8) la voyelle /a/ est transcrite mais n'est pas réalisée, la production réelle est /sdveet/. Habituellement, dans ce contexte, on devrait trouver une assimilation de voisement entre /s/ et /d/, ce qui n'est pas le cas. Ici, l'absence d'assimilation entre les deux consonnes témoigne de la *trace* de la voyelle /a/, pourtant non réalisée. Ce phénomène permet la perception non lacunaire de la séquence. Notre hypothèse est que, dans de nombreux cas, les réductions ne correspondent pas systématiquement à l'élision simple d'un segment mais sont caractérisées par des processus articulatoires complexes tendant à préserver des indices pertinents qui permettent une récupération optimale de l'information phonétique, et donc du message linguistique.

Figure 8 : Séquence extraite du Corpus CID (Bertrand *et al.*, 2008) : « ça devait être »



Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

60

Séquence totale « ça devait être » (figure 8).

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/tipa/6079>

61

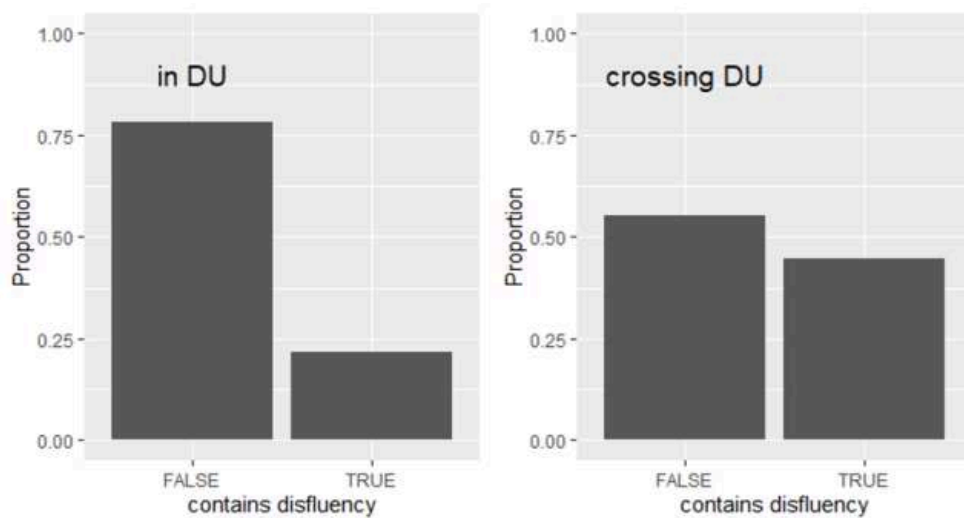
Séquence entourée en rouge (figure 8 [ɑðvɛ]). L'écoute permet de constater l'absence du /a/ pourtant perçu lors de l'écoute de la séquence totale (avec /s/).

### 3.3 Impact des disfluences sur les unités

62 Le résultat de notre étude (Prévoit *et al.*, 2021) sur l'impact des disfluences sur les unités prosodiques et discursives (et leur association) a confirmé des choses connues dans ce domaine. Premièrement, les disfluences impactent fortement les unités prosodiques et discursives (une fois les faux départs exclus). Deuxièmement, les disfluences expliquent

une partie significative des dissociations entre unités prosodiques et unités discursives qui sont délicates à analyser du seul point de vue de l'interface discours-prosodie (Figure 9).

Figure 9. Comparaison de l'articulation entre unités prosodiques et discursives selon la présence (ou pas) d'une disfluente dans l'unité prosodique



"In DU" : Unité prosodique interne à l'unité discursive ; "crossing DU" : Unité prosodique à cheval entre deux unités discursives

Figure tirée de Prévot *et al.*, 2021

- 63 L'étude des disfluences en lien avec le phrasé prosodique, et notamment des items produits durant les breaks (parmi lesquels les pauses silencieuses et remplies ou les marqueurs discursifs par exemple), reste donc une voie à explorer pour mieux appréhender notamment les phénomènes de frontière prosodique. Plus particulièrement en ce qui concerne les pauses silencieuses, la majorité des travaux en prosodie les considèrent en effet comme une marque de frontière majeure. Or, les pauses silencieuses peuvent être utilisées à d'autres fins (en lien avec la structure informationnelle, pour créer un effet stylistique, etc.). Par ailleurs, dans notre tentative de caractérisation des unités intonatives en parole spontanée (Portes et Bertrand 2011), nous avons proposé de considérer que certaines pauses silencieuses ne sont pas toujours la marque d'une frontière majeure mais que l'unité intonative considérée est tout simplement interrompue, au même titre que les autres unités syntaxique ou discursive. Ainsi, même en présence d'une pause silencieuse, on pourrait considérer qu'il n'y a qu'une seule et même unité prosodique. A notre connaissance, cette proposition qui fait l'objet d'un travail en cours, n'a pas été avancée par ailleurs. Or ce point est crucial pour renforcer la nature indéniablement phonologique des unités prosodiques, et pourrait avoir un réel impact sur les modèles du phrasé prosodique du français (voir l'article : "La prosodie au Laboratoire Parole et Langage" dans ce numéro<sup>25</sup>).

### 3.4 Persistance des convergences

- 64 Si elle s'inscrit par définition dans le cadre d'une interaction, l'imitation a sur le plan langagier des conséquences qui s'étendent bien au-delà des limites temporelles de cette

interaction. Studdert-Kennedy (2002) et Goldstein (2003) considèrent ainsi qu'elle remplit une fonction centrale dans le développement langagier, en contribuant à permettre à l'enfant de s'approprier le système phonologique de sa langue. L'imitation ferait également partie des mécanismes présidant à la mise en place et à l'évolution des langues humaines (Nguyen & Delvaux, 2016). Par ailleurs, l'imitation chez l'humain revêt une portée qui dépasse bien sûr de très loin le domaine du langage lui-même. Piaget en faisait ainsi une étape clé du développement de l'enfant, entre le stade sensori-moteur et la formation des premières images mentales. En montrant que l'enfant a dès la naissance la capacité de mimer des mouvements faciaux, Meltzoff et Moore (1997) ont avancé l'hypothèse que l'imitation joue un rôle crucial dans la construction de soi, par l'intermédiaire et dans son rapport à autrui. Selon van Geert et Steenbeek (2005), l'imitation constitue un comportement que les êtres humains ont tendance à adopter de manière automatique en situation d'interaction, et qui serait inscrit dans notre patrimoine génétique en formant l'un des mécanismes de base pour la transmission des caractères acquis entre les générations.

- 65 Il est depuis longtemps supposé qu'à l'intérieur d'une interaction conversationnelle, les phénomènes d'imitation s'étendent aux caractéristiques phonétiques des énoncés produits par les interactants. Pardo (2006) a recueilli des données compatibles avec cette hypothèse, au travers d'une expérience dans laquelle différents locuteurs ont été amenés à produire les mêmes mots avant, pendant et après une conversation. Des auditeurs ont jugé que la similarité des locuteurs dans les formes prononcées des mots s'était accrue au fil de l'interaction et après que celle-ci eut pris fin. Ces résultats ont d'importantes implications pour le débat sur la manière dont le locuteur/auditeur se représente la forme sonore des mots, entre l'approche dite abstractionniste, d'une part, et la perspective très différente plus récemment offerte par les modèles à exemplaires, d'autre part. Selon les tenants de l'approche abstractionniste, le signal de parole est rapidement converti par l'auditeur en un ensemble d'unités phonologiques abstraites et indépendantes du contexte, par le biais d'un processus de normalisation visant à minimiser les variations intra- et inter-individuelles observables dans la forme de surface des mots. Une démarcation très nette est établie dans cette approche entre la forme phonétique de surface d'un mot, et la représentation phonologique sous-jacente associée à ce mot, conçue comme étant à la fois unique et conservée de manière permanente dans la mémoire à long terme (Lahiri, 2002). Dans les modèles à exemplaires (Goldinger, 1998), au contraire, il est supposé que les mots et les constructions grammaticales de haute fréquence sont représentés en mémoire sous la forme de listes d'exemplaires, dont chacun est enregistré avec tous les détails relatifs aux circonstances particulières dans lesquelles il a été produit et/ou entendu (caractéristiques sensori-motrices, sémantiques, pragmatiques, identité du locuteur, situation d'occurrence, etc.). Ces modèles entrent en contradiction avec l'idée selon laquelle chaque mot est formé à partir d'un ensemble de primitives phonologiques pré-établies et indépendantes du contexte, et ils donnent à penser que les représentations associées aux différentes entrées lexicales, dans ce qui se présente comme un lexique mental très élargi, contiennent une large variété de détails phonétiques fins (Pierrehumbert, 2002).
- 66 De manière plus générale, la division entre formes sous-jacentes et formes de surface est fortement remise en question dans le cadre de la linguistique dite orientée usage. Selon les tenants de cette approche (voir par ex. Barlow & Kemmer, 2000), l'usage que



les locuteurs font du langage exerce une influence directe sur la manière dont le langage est mentalement représenté. Pour ne citer que l'un des exemples les plus souvent associés à cette hypothèse, on s'attache aujourd'hui à montrer que la fréquence d'utilisation d'un mot a un impact sur la forme phonologique rattachée à ce mot dans le lexique mental (Bybee, 2006). De profondes conséquences méthodologiques en découlent dans la mesure où les études sur le langage sont appelées à replacer celui-ci dans son « site d'occurrence premier », c'est-à-dire en situation d'interaction sociale (Couper-Kuhlen & Ford, 2004 ; Local, 2003), par opposition aux recherches traditionnellement conduites sur de la parole lue par exemple. Sur le plan théorique, certains auteurs (par ex. Bybee & McClelland, 2005) évoquent un véritable changement de paradigme, en s'écartant de ce qui a pourtant formé jusqu'à présent l'un des principes de base de l'analyse linguistique, c'est-à-dire l'idée selon laquelle le langage possède une structure combinatoire, en plaçant résolument l'accent sur la dimension sociale du langage, par opposition à sa dimension référentielle par exemple, et en affirmant que les représentations mises en jeu dans la production et le traitement du langage ne préexistent pas entièrement aux interactions pouvant prendre place entre deux locuteurs, mais sont au contraire sujettes à un processus de co-construction dans lequel les deux interactants sont engagés.

## 4. Conclusion

67 Cet article aura été l'occasion de rappeler la pertinence des mouvements anciens de faire de la conversation un objet d'étude de première classe (Sacks et al, 1973) et d'étudier l'oral spontané (voir l'article « Le début des phrases en français parlé »<sup>26</sup> dans ce numéro). La linguistique interactionnelle (Couper-Kuhlen et Selting, 2018) poursuit le travail de description et d'analyse, et développe la notion des formes langagières en tant que ressources pour l'interaction, en particulier la dimension multimodale de ces ressources (voir l'article « Multimodalité(s) dans les interactions humaines »<sup>27</sup> dans ce numéro). Nous avons de notre côté d'une part illustré que (i) la conversation comprenait des phénomènes propres et ce pour tous les domaines de l'analyse linguistique, (ii) que ces phénomènes étaient représentés de manière massive dans les données conversationnelles et (iii) qu'ils ne pouvaient pas être réduits ou écartés si l'on souhaite comprendre comment fonctionne le langage dans ce contexte. Par ailleurs, nous avons rappelé la nature dispersée de la distribution des informations dans le flux langagier et la fugacité des formes qui les véhicule. Ces propriétés requièrent un modèle linguistique, articulé avec ce que l'on sait de l'architecture cognitive de traitement de production du langage, qui demande encore à être construit. Nous avons commencé dans la dernière section de cet article par montrer comment certains phénomènes conversationnels décrits impactent les objets linguistiques et par conséquent les théories linguistiques sous-jacentes.

---

## BIBLIOGRAPHIE

- Aubanel, V., & Nguyen, N. (2020) Speaking to a common tune: Between-speaker convergence in voice fundamental frequency in a joint speech production task, *PLoS ONE*, 15(5), e0232209.
- Barlow, M., & Kemmer, S., editors (2000) *Usage-Based Models of Language*, Center for the Study of Language and Information, Stanford, CA.
- Bavelas, J.B., Coates, L. & Johnson, T. (2000) Listeners as Co-narrators, *Journal of Personality and Social Psychology*, 79 (6), p. 941-952.
- 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), p. 105-134.
- Bertrand, R. et Espesser, R. (2017) Storyteller and listener's behavior in French conversational storytelling, *Journal of Pragmatics*, 111, p. 33-53.
- Bertrand, R. & Priego-Valverde, B. (2017) Listing practice in French conversation. From collaborative achievement to interactional convergence, *Discours* (en ligne).
- 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*, Dordrecht: Springer, p. 1323-1356.
- Bybee, J. (2006) *Frequency of Use and the Organization of Language*, Oxford, UK: Oxford University Press.
- Bybee, J., & McClelland, J. L. (2005) Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition, *The Linguistic Review*, 22, p. 381-410.
- Clark, Herbert H. (1996) *Using language*, Cambridge university press.
- Couper-Kuhlen, E., Ford, C. E., editors (2004) *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*, Amsterdam, The Netherlands: John Benjamins.
- Couper-Kuhlen, E. & Selting, M. (2018) *Interactional Linguistics. Studying Language in Social Interaction*, Cambridge: Cambridge University Press.
- Degand, L. & Simon, A.C. (2009) On identifying basic discourse units in speech: theoretical and empirical issues, *Discours*, 4.
- Degand, Liesbeth & Crible, Ludivine (2021) Discourse markers at the peripheries of syntax, intonation and turns, *Pragmatic Markers and Peripheries*, vol. 325, p. 19-48.
- 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, p. 15-32.
- Ernestus, M. (2014) Acoustic reduction and the roles of abstractions and exemplars in speech processing, *Lingua*, 142, p. 27-41.
- Ford, C. E., & Thompson, S. A. (1996) Intonational, and pragmatic resources for the management of turns, in E. Ochs, E.A. Schegloff & S. A. Thompson (eds) *Interaction and grammar*, (13), 134-184.

- Gerdes, K., & Kahane, S. (2009) Speaking in piles: Paradigmatic annotation of French spoken corpus, In *Fifth Corpus Linguistics Conference*, p. 1-15.
- Ginzburg, Jonathan (2012) *The interactive stance*, Oxford University Press.
- Goldinger, S. (1998) Echoes of echoes? An episodic theory of lexical access, *Psychological Review*, vol. 105, p. 251-279.
- Goldstein, L. (2003) Emergence of discrete gestures, In *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, Spain, 3-9 August 2003, p. 85-88.
- Guardiola, M. & Bertrand, R. (2013) Interactional convergence in conversational storytelling: when reported speech is a cue of alignment and/or affiliation, *Frontiers in Psychology*, 4:705.
- Haselow, A. (2017) *Spontaneous spoken English: An integrated approach to the emergent grammar of speech*, Cambridge University Press.
- Gravano, Agustín, Hirschberg, Julia, et Beňuš, Štefan (2011) Affirmative cue words in task-oriented dialogue, *Computational Linguistics*, vol. 38, n° 1, p. 1-39.
- Hansen, Maj-Britt Mosegaard (1997) Alors and donc in spoken French: A reanalysis, *Journal of pragmatics*, vol. 28, n° 2, p. 153-187.
- Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding, *Journal of Phonetics*, 31(3-4), p. 373-405.
- Johnson, K. (2004) Massive reduction in conversational American English. Spontaneous speech: Data and analysis, *Proceedings of the 1st session of the 10th international symposium*, p. 29-54.
- Lacheret-Dujour, A., & Kahane, S. (2020) Unités syntaxiques et unités intonatives majeures en français parlé : inclusion, fragmentation, chevauchement, in *SHS Web of Conferences*, vol. 78, EDP Sciences, p. 14005.
- Lahiri A., Reetz H. (2002) Underspecified recognition, in Gussenhoven C., Warner N. (Eds.), *Papers in Laboratory Phonology VII*, p. 637-675, Berlin, Germany: Mouton de Gruyter.
- Lindblom, B. (1990) Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Éds.) *Speech Production and Speech Modelling*, Netherlands: Springer, p. 403-439.
- Local, J. (2003) Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, 31, p. 321-339.
- Meltzoff, A. N., & Moore, M. K. (1977) Imitation of facial and manual gestures by human neonates, *Science*, 198(4312), p. 75-78.
- Meunier C., Bigi B. (2016) Répartition des phonèmes réduits en parole conversationnelle. Approche quantitative par extraction automatique, *Actes des Journées d'Etude sur la Parole*, Paris (France), juillet 2016.
- Meunier, C. & Bigi, B. (2021) Variations temporelles des phonèmes en parole conversationnelle : propriétés phonétiques et facteurs lexicaux, *Studii de lingvistică*, vol. 11.
- Meunier, C. (2012) Contexte et nature des réalisations phonétiques en parole conversationnelle, *Actes des Journées d'Etude sur la Parole*, Grenoble (France), juin 2012, p. 1-8.
- Meunier, C., Espesser, R. (2011) Vowel reduction in conversational speech in French: The role of lexical factors, *Journal of Phonetics*, vol. 39, Issue 3, July 2011, p. 271-278.

- Meunier, C. ; Meynadier, Y. ; Espesser, R. (2008) Voyelles brèves en parole conversationnelle, *Actes des Journées d'Etude sur la Parole (JEP)*, 27 : 2008 juin 9-13, Avignon, France, p. 97-100.
- Mondada, L. (2017) Nouveaux défis pour l'analyse conversationnelle : l'organisation située et systématique de l'interaction sociale, *Langage et société*, vol. 160-161, 2, p. 181-197.
- Nguyen, N., & Delvaux, V. (2015) Role of imitation in the emergence of phonological systems, *Journal of Phonetics*, 53, p. 46-54.
- Ogden, Richard (2013) Clicks and percussives in English conversation, *Journal of the International Phonetic Association*, vol. 43, n° 3, p. 299-320.
- Pallaud, B., Bertrand, R., Prevot, L., Blache, P. & Rauzy, S. (2019) Suspensive and Disfluent Self Interruptions in French Language Interactions, in Liesbeth Degand; Gaëtanelle Gilquin; Laurence Meurant & Anne Catherine Simon (Eds) *Fluency and Disfluency across Languages and Language Varieties. Corpora and Language in Use -Proceedings 4*, p. 109-138.
- Pardo, J. S. (2006) On phonetic convergence during conversational interaction, *The Journal of the Acoustical Society of America*, 119(4), p. 2382-2393.
- Pierrehumbert, J., (2002) Word-specific phonetics. In Gussenhoven, C., Warner, N., Rietveld, T. (Eds.), *Phonology & Phonetics [Laboratory Phonology 7]*, Berlin: Mouton, p. 101-140.
- Pitt, M. A., 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(1), p. 89-95.
- Portes, C. & Bertrand, R. (2011) Permanence et variation des unités prosodiques dans le discours et l'interaction, *Journal of French Language Studies*, p. 97-110
- Prévot, Laurent, Tseng, Shu-Chuan, Peshkov, Klim, et al. (2015) Processing units in conversation: A comparative study of French and Mandarin data, *Language and Linguistics*, vol. 16, n° 1, p. 69-92.
- Prévot, Laurent, Gorisch, Jan, & Bertrand, Roxane (2016) A CUP of CoFee: A large Collection of feedback Utterances Provided with communicative function annotations, In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016. p. 3180-3185.
- Prevot, Laurent, Magistry, Pierre & Lison, Pierre (2019) Should we use movie subtitles to study linguistic patterns of conversational speech? A study based on French, English and Taiwan Mandarin, In *Third International Symposium on Linguistic Patterns of Spontaneous Speech*, Taipei, 2019.
- Prévot, Laurent, Bertrand, Roxane & Rauzy, Stéphane (2021) Investigating Disfluencies Contribution to Discourse-Prosody Mismatches in French Conversations, In *The 10<sup>th</sup> Workshop on Disfluency in Spontaneous Speech*, Paris, 2021.
- Sacks, H., Schegloff, H. & Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation, *Language*, 50, p. 696-735, <https://doi.org/10.2307/412243>
- 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, p. 71-93.
- Selting, M. (1996) On the interplay of syntax and prosody in the constitution of turn-constructive units and turns in conversation, *Pragmatics*, 6(3), p. 371-388.
- Shriberg, E. E. (1994) *Preliminaries to a Theory of Speech Disfluencies*, Ph.D. dissertation, University of California at Berkeley.
- Stalnaker, R. (2002) Common ground, *Linguistics and philosophy*, 25(5/6), p. 701-721.

Stivers, T. (2008) Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation, *Research on Language & Social Interaction*, 41(1), p. 31-57, <https://doi.org/10.1080/08351810701691123>.

Studdert-Kennedy, M. (2002) Mirror neurons, vocal imitation and the evolution of particulate speech, In M.I. Stamenov & V. Gallese (Eds), *Mirror Neurons and the Evolution of Brain and Language*, Amsterdam: John Benjamins, p. 207-227.

Van Geert, P., & Steenbeek, H. (2005) Explaining after by before: Basic aspects of a dynamic systems approach to the study of development, *Developmental review*, 25(3-4), p. 408-442.

Ward, Nigel (2006) Non-lexical conversational sounds in American English, *Pragmatics & Cognition*, vol. 14, n° 1, p. 129-182.

Yngve, Victor H. (1970) On getting a word in edgewise, In *Chicago Linguistics Society, 6<sup>th</sup> Meeting*, 1970, p. 567-578.

## NOTES

1. <https://journals.openedition.org/tipa/5167>
2. <https://journals.openedition.org/tipa/4819>
3. <https://journals.openedition.org/tipa/5210>
4. [...] on their (TCUs) possible completion, transition to a next speaker becomes relevant (although not necessarily accomplished) (Schegloff, 1996: 55)
5. Le participant qui n'est pas à ce moment-là en position de locuteur principal.
6. <https://journals.openedition.org/tipa/5424#tocto2n11>
7. Unités Inter-Pausales (Inter-Pausal Unit, IPU)
8. Ce compte correspond aux IPU pré-sélectionnées automatiquement mais ensuite validées manuellement par plusieurs annotateurs (élimination des faux positifs). Le nombre réel d'IPUs de feedback est donc très légèrement supérieur en réalité, en raison des faux négatifs de la pré-sélection automatique.
9. #Feedback / #IPU
10. En pragmatique, on considère que tout nouvel énoncé cohérent avec ce qui a été produit par l'interlocuteur porte une forme de feedback positif implicite (Allwood, 1995 ; Clark, 1996).
11. <https://journals.openedition.org/tipa/5424#tocto2n11>
12. On parle dans ce cas de connecteurs de discours.
13. <https://journals.openedition.org/tipa/5179>
14. <https://journals.openedition.org/tipa/5104>
15. Inflection Phrase
16. <https://journals.openedition.org/tipa/5167>
17. <https://journals.openedition.org/tipa/5179>
18. Ponctuer une séquence d'oral conversationnel requiert une compréhension fine de la séquence et demande l'analyse (implicite) d'aspects pragmatiques et interactionnels.
19. Nous utilisons une valeur seuil de pause de 200 ms qui correspond à la fenêtre temporelle dans laquelle un changement de locuteur tend à se produire (Stivers *et al.*, 2009).
20. Seules les séquences très proches (distance = 2 ou 3) sont connectées dans ces graphes.
21. Au moment de la création du graphe, la position des deux genres n'est pas contrainte, ce sont uniquement les distances entre les séquences (plus conversationnelles ou écrites) qui créent les "zones" du graphe.
22. MTR : 11ème forme la plus fréquente ; fréquence relative 1,9 % (*Corpus sans contact visuel*) ; CID : 17ème forme la plus fréquente avec une fréquence relative de 1,2 % ;

DVD : 37ème forme la plus fréquente ; fréquence relative 0,7 % ;

MTX : 48ème forme la plus fréquente ; fréquence relative 0,5 %.

23. <https://journals.openedition.org/tipa/4819>

24. <https://journals.openedition.org/tipa/5854>

25. <http://journals.openedition.org/tipa/5210>

26. <https://journals.openedition.org/tipa/5179>

27. <https://journals.openedition.org/tipa/5167>

---

## RÉSUMÉS

Ce chapitre aborde les manifestations linguistiques de la conversation. Longtemps, les linguistes ont traité séparément chacun des niveaux du langage oral (phonologie, lexique, syntaxe, discours), notamment à partir de corpus de parole préparée et non interactive. Les travaux sur la parole en conversation permettent de mettre en évidence ce que les théories linguistiques modernes cherchent à intégrer : l'interdépendance de ces dimensions. En outre, au-delà de la description de variantes, ces travaux nous permettent d'étudier de façon spécifique les processus de production et de compréhension du langage, y compris en les considérant de façon conjointe, en nous appuyant sur les théories les plus récentes de l'interaction.

Nous proposons dans ce chapitre d'aborder tout d'abord les particularités du langage conversationnel ainsi que leur impact sur les unités et les phénomènes généralement étudiés par les linguistes : phonèmes, morphèmes, syntagmes, unités prosodiques, propositions, unités discursives. Ces particularités proviennent notamment de paramètres communs à la plupart des situations communicatives conversationnelles : une certaine pression temporelle, la co-présence des participants, le caractère fugace de ces productions conversationnelles dont la forme est oubliée dès que le message a été transmis. La pression temporelle donne des caractéristiques fortes au langage conversationnel liées à la gestion du temps (réductions lexicales, phonétiques, disfluences, etc.). La fugacité des productions autorise l'utilisation de stratégies particulières (répétitions, redondances, reformulations) qui sont généralement évitées dans les autres usages langagiers. La co-présence, enfin, permet l'utilisation de toutes les modalités impliquées dans une interaction, et sur lesquelles reposent les multiples mécanismes d'alignement entre participants. Il faut ajouter que l'étude de ces particularités est profondément impactée par les notions de contexte et de connaissance partagée qui fondent une interaction. L'ensemble de ces particularités dues au langage conversationnel entraîne des ajustements à tous les niveaux de la production, mais de façon spécifique pour chacun d'entre eux. Ces ajustements ne sont pas toujours décrits à chaque niveau, mais plus encore, ils sont rarement analysés de façon conjointe. Il nous semble ainsi important de mettre ici en perspective les différents types de modifications linguistiques liées à la conversation.

De plus, comprendre ces questions générales nécessite la description précise des phénomènes propres à la conversation (cf. analyse conversationnelle et linguistique interactionnelle). Ces phénomènes incluent les tours de paroles, les signaux d'écoute et le feedback conversationnel, ou encore les énoncés co-produits.

Ces phénomènes particuliers à la conversation doivent être mis en perspective avec la question de l'interaction existant entre les unités. Nous savons par exemple que les réductions phonétiques affectent plus spécifiquement les séquences dont la charge sémantique et informationnelle est faible. Il est donc nécessaire de décrire les conditions dans lesquelles ces

interactions sont possibles. En particulier, il est nécessaire de les expliquer en les mettant en perspective par rapport à la question de la compréhension.

Les participants à une conversation cherchent à atteindre des buts externes à l'activité communicative, mais qui passent par la réalisation de leurs buts communicatifs. Pour atteindre ces buts, ils / elles doivent articuler des éléments informatifs grâce aux structures du langage étudiées traditionnellement et qui permettent de construire et d'articuler des représentations. La conversation est un lieu d'observation privilégié de ces phénomènes. Les interactions entre participants se construisent généralement sur la base d'un échange d'informations. Celui-ci n'est pas un mécanisme figé comme l'a longtemps expliqué la linguistique. Il peut reposer sur des traitements très superficiels pendant lesquels chaque participant produit et perçoit des bribes d'information. La façon dont celles-ci s'agrègent pendant une conversation ne repose sans doute pas toujours sur un processus compositionnel, mais consiste plutôt à permettre à tous les participants de construire un ensemble partagé de connaissances. Nous décrirons dans ce chapitre les indices observables de ce mécanisme, en abordant en particulier la question de l'alignement des unités linguistiques que nous observons au cours d'une conversation. Cette approche de l'interaction permet de fournir un cadre théorique pour l'explication des phénomènes conversationnels cités précédemment.

Ce chapitre sera structuré de la façon suivante : 1/ la conversation sera discutée comme l'articulation entre représentation et communication par le langage en tant qu'action sociale ; 2/ une deuxième partie visera à dresser un inventaire des « traces » de l'activité de conversation (but communicatif, pression temporelle, co-présence, charge cognitive, pression sociale) sur les différentes unités linguistiques (phonèmes, lexique, syntaxe, prosodie) ; 3/ seront abordés ensuite les phénomènes propres à la conversation n'étant pas identifiables dans d'autres types de parole (tours de parole, disfluences, feedback conversationnel, énoncés co-produits, énoncés non-phrastiques, énoncés non-lexicaux, etc.) ; 4/ nous développerons pour finir quelques perspectives liées à l'ensemble de ces travaux.

This chapter deals with the linguistic manifestations of conversation. For a long time, linguists have been interested in the levels of oral language (phonology, lexicon, syntax, discourse) separately, particularly based on prepared and non-interactive speech corpora. Research on conversational speech highlights what modern linguistic theories seek to integrate: the interdependence of these dimensions. Moreover, beyond the description of variants, this work allows us to study in a specific way the processes of language production and comprehension, including by considering them jointly, by relying on the most recent theories of interaction.

In this chapter, we propose to first discuss the particularities of conversational language and their impact on the units and phenomena generally studied by linguists: phonemes, morphemes, syntagms, prosodic units, propositions, discourse units. These particularities stem from common parameters to most conversational communication situations: temporal pressure, co-presence of participants, fleeting nature of these conversational productions whose form is forgotten as soon as the message has been transmitted. Temporal pressure provides strong characteristics to conversational language linked to time management (lexical and phonetic reductions, disfluencies, etc.). The transience of productions allows the use of particular strategies (repetitions, redundancies, reformulations) which are generally avoided in other language uses. Finally, co-presence allows the use of all the modalities involved in an interaction, and on which the multiple mechanisms of alignment between participants are based. Moreover, the study of these particularities is profoundly impacted by the notions of context and shared knowledge that underlie an interaction. All these particularities due to conversational language suggest adjustments at all levels of production, but in a specific way for each of them. These adjustments are not always described at each level, but even more so, they are rarely analysed together. An important point is to put into perspective the different types of linguistic modifications linked to

conversation.

Moreover, understanding these general issues requires a precise description of conversation phenomena (cf. conversational analysis and interactional linguistics). These phenomena include turns of speech, listening signals and conversational feedback, or co-produced utterances.

These specific phenomena of conversation must be put into perspective with the question of the interaction existing between units. We know, for example, that phonetic reductions more specifically affect sequences with a low semantic and informational load. It is therefore necessary to describe the conditions under which these interactions are possible. In particular, it is necessary to explain them by putting them in perspective with the question of comprehension.

The participants in a conversation seek to achieve goals that are external to the communicative activity, but which involve the achievement of these participants' communicative goals. In order to achieve these goals, the participants must articulate informative elements through the traditionally studied structures of language which allow the construction and articulation of representations. Conversation is a privileged place to observe these phenomena. Interactions between participants are generally built on an exchange of information. This is not a fixed mechanism as linguistics has long explained. It can be based on very superficial processing during which each participant produces and perceives bits of information. The way in which these are aggregated during a conversation is probably not always based on a compositional process, but rather consists in allowing all participants to construct a shared body of knowledge. In this chapter, we will describe the observable evidence of this mechanism, addressing in particular the question of the alignment of linguistic units that we observe during a conversation. This approach to interaction provides a theoretical framework for explaining the conversational phenomena mentioned above.

This chapter will be structured as follows: 1/ conversation will be discussed as the articulation between representation and communication through language as a social action; 2/ a second part will aim at drawing up an inventory of the "traces" of the activity of conversation (communicative goal, temporal pressure, co-presence, cognitive load, social pressure) on the different linguistic units (phonemes, lexicon, syntax, prosody); 3/ we will then address the phenomena specific to conversation which are not identifiable in other types of speech (speech turns, disfluencies, conversational feedback, co-produced utterances, non-phrastic utterances, non-lexical utterances, etc.); 4/ we will finally develop some perspectives related to all these works.

## INDEX

**Mots-clés** : conversation, feedbacks, réductions, disfluences, alignement, interaction

**Keywords** : conversation, feedbacks, reductions, disfluences, alignement, interaction

## AUTEURS

### LAURENT PRÉVOT

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
laurent.prevot@univ-amu.fr

### ROXANE BERTRAND

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
roxane.bertrand@univ-amu.fr



**PHILIPPE BLACHE**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
philippe.blache@univ-amu.fr

**CHRISTINE MEUNIER**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
christine.meunier@univ-amu.fr

**NOËL NGUYEN**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
noel.nguyen-trong@univ-amu.fr

**BERTHILLE PALLAUD**

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France  
berthille.pallaud@orange.fr