Principes et outils pour l'annotation des corpus

Mary Amoyal, Roxane Bertrand, Brigitte Bigi, Auriane Boudin, Christine Meunier, Berthille Pallaud, Béatrice Priego-Valverde, S. Rauzy, Marion Tellier

TIPA. Travaux interdisciplinaires sur la parole et le langage

38 | 2022
Numéro spécial : Panorama des recherches au Laboratoire Parole et Langage

Principes et outils pour l'annotation des corpus

Principles and tools for corpus annotation

Mary Amoyal, Roxane Bertrand, Brigitte Bigi, Auriane Boudin, Christine Meunier, Berthille Pallaud, Béatrice Priego-Valverde, Stéphane Rauzy et Marion Tellier

Édition électronique
URL : https://journals.openedition.org/tipa/5424
DOI : 10.4000/tipa.5424
ISSN : 2264-7082

Éditeur
Laboratoire Parole et Langage

Référence électronique
Mary Amoyal, Roxane Bertrand, Brigitte Bigi, Auriane Boudin, Christine Meunier, Berthille Pallaud, Béatrice Priego-Valverde, Stéphane Rauzy et Marion Tellier, « Principes et outils pour l'annotation des corpus », *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/5424 ; DOI : https://doi.org/10.4000/tipa.5424

Principes et outils pour l'annotation des corpus

Principles and tools for corpus annotation

Mary Amoyal, Roxane Bertrand, Brigitte Bigi, Auriane Boudin, Christine Meunier, Berthille Pallaud, Béatrice Priego-Valverde, Stéphane Rauzy et Marion Tellier

1. Introduction

1.1 Prolégomènes à la constitution et l'enrichissement des corpus

1 La linguistique de corpus se donne pour objectif d'étudier la langue dans ses diverses manifestations en s'appuyant sur un éventail de données le plus large et le plus représentatif possible (écrit, oral, genres, régions, etc.), en appliquant des méthodes quantitatives pour l'essentiel. C'est dans ce cadre qu'ont donc émergé les projets de constitution de corpus de *référence* pour les différentes langues (cf. quelques corpus au niveau international dont de nombreux en langue anglaise, The *Santa Barbara Corpus of Spoken American English*, la *Maptask*, le *Français Parlé du GARS* (Aix-en-Provence), le corpus *Valibel* (Louvain), le *PFC* pour la phonologie, son corollaire *PAC* pour la langue anglaise, etc.). Ces projets de constitution de corpus se sont accompagnés d'un travail relatif à leur enrichissement, lequel a nécessité de mener une vraie réflexion sur la nature et les objectifs de cet enrichissement. Ce dernier se présente comme un jeu d'annotations spécifiques qui constitue un préalable à toute analyse linguistique. L'augmentation de la taille des corpus a par ailleurs nécessité le développement d'outils automatiques. Enfin, simultanément, une réflexion autour de l'archivage et de la mutualisation de ces corpus enrichis a été conduite et mise en œuvre au niveau national et international. Ceci permet de disposer de normes et de conventions relatives au bon usage autour de la création et de l'annotation des corpus (LDC, Ortolang, etc.).

2 Au LPL, les corpus ont toujours été au cœur des recherches menées sur la parole. Mais le projet de constituer des ressources orales enrichies a réellement débuté au début des années 2000 (projet OTIM Blache *et al.*, 2010a ; Blache *et al.*, 2017) notamment sur le CID (Bertrand *et al.*, 2008). L'objectif est alors de développer des corpus comportant des annotations sur les différents domaines de la linguistique (du niveau le plus fin de la phonétique jusqu'au niveau interactionnel, en passant par la prosodie, la morphologie, la syntaxe, le discours ainsi que le niveau mimo-gestuel). L'intérêt de disposer de tels corpus est de permettre l'analyse de chacun des domaines concernés mais aussi et surtout des liens entre chacun d'eux. Mais l'une des difficultés principales quand on cherche à rendre compte de la multimodalité de la parole concerne l'hétérogénéité des données considérées, qui requiert de penser l'annotation en amont, au niveau même de la représentation des informations. Une étape cruciale est donc l'élaboration d'un schéma d'annotation global permettant de considérer tous ces niveaux dans une seule et même approche formelle qui favorisera leur interrogation ultérieure.

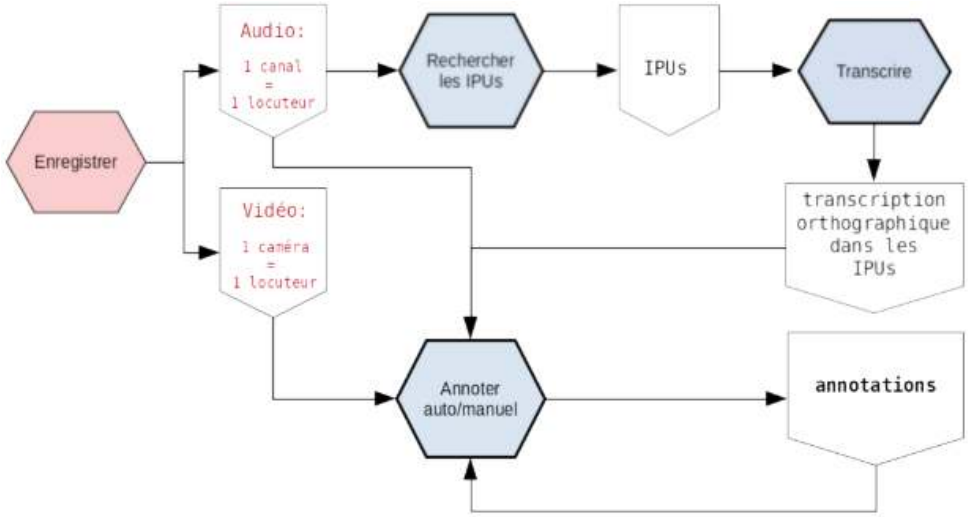
3 Le travail au sein de chaque niveau d'annotation est ensuite relativement similaire. Il s'agit d'établir un schéma permettant une annotation la plus reproductible et la plus robuste possible. Ce schéma est établi sur la base des connaissances théoriques et en vue de répondre aux questionnements de recherche. Une fois le schéma d'annotation établi, il est également possible de construire un guide d'annotation destiné à de potentiels annotateurs (experts / naïfs). Le plus souvent, les annotations sont réalisées en recourant à plusieurs annotateurs afin de rendre possible une évaluation de la consistance (accords inter-annotateurs)[1] (cf. Tellier, 2014 pour le gestuel ; Blache *et al.*, 2017 pour les différents domaines).

4 Des années de travail autour de la constitution et de l'enrichissement des corpus au LPL ont permis de mettre en œuvre des procédures désormais partagées au laboratoire et généralisables à tous les domaines de l'annotation linguistique. Ceci permet aux nouveaux concepteurs de corpus ou de nouvelles annotations de disposer d'un cadre et de suivre quelques étapes importantes facilitant l'analyse ultérieure d'interrogation des données enrichies. Simultanément, des outils, logiciels, ont été déployés pour faciliter ces différentes étapes.

1.2 Méthode générique de collecte et d'annotation proposée par le LPL

5 Dans cette section, nous présentons la méthodologie déployée au LPL pour collecter et annoter les corpus de parole multimodaux. Elle est désormais suffisamment générale pour s'appliquer à une large gamme de problématique de recherche. La figure 1 ci-après en résume les points principaux.

Figure 1 : Processus de recueil et étapes préalables à l'annotation



6 Cette figure illustre bien le fait que le processus de création d'un corpus et de ses annotations a lieu sous la forme d'une chaîne de traitements dont le point d'entrée sont les enregistrements. Bien que cet aspect ne soit pas le sujet de cet article, il est important de mentionner ici que la résolution et la qualité des appareils de capture (microphones, fréquence d'images, format de fichier, logiciel) vont exercer une influence déterminante sur la qualité du corpus, et donc sur les annotations à suivre. De même, le manque de normalisation des dits signaux impliquera que certaines annotations ne pourront pas être faîtes avec ces signaux, il est donc crucial d'apporter un très grand soin à ces aspects. Cependant, tous les corpus ne sont pas recueillis en laboratoire et pour les corpus de terrain (dits écologiques), la qualité des enregistrements entraîne parfois la perte d'une partie des données.

7 À partir des données primaires (audio), une première étape de segmentation de la parole consiste à déterminer les IPUs - Inter-Pausal Units (c.f. section suivante). Nous effectuons ensuite la transcription orthographique au sein de ces IPUs, qui dans le même temps, sont corrigées manuellement. Dès lors que la transcription orthographique a été obtenue et vérifiée, les étapes ultérieures d'annotations proprement dites en mots, catégories syntaxiques, phonèmes, syllabes, unités prosodiques, etc, sont effectuées. Par ailleurs, les aspects mimo-gestuels sont annotés sur la base de la vidéo mais ils peuvent aussi s'ancrer sur certains niveaux de l'annotation (tels que les mots par exemple).

8 Le procédé d'annotation est complexe et coûteux. Il nécessite donc une réflexion en amont et s'appuie quasiment toujours sur des modèles théoriques. Les choix opérés sont à la fois fondés sur les principes théoriques sous tendant l'objet et les questions de recherche, et doivent être suffisamment explicites et clairs pour que les annotateurs, qu'ils soient experts ou naïfs, puissent produire des annotations les plus fiables possibles. Les annotations semi-automatiques et automatiques sont le résultat de ce qui a été pensé en amont, ou annoté manuellement. Les annotations automatiques ne sont donc pas a-théoriques. Nous renvoyons le lecteur entre autres au développement du syllabeur (Bigi *et al.*, 2010) ou encore à la détection automatique des hétéro-répétitions (Bigi *et al.*, 2014). C'est donc bien le processus qui est automatique mais l'annotation quelle qu'elle soit dépend des présupposés théoriques du chercheur. Parmi les nombreux avantages de l'annotation automatique, la première consiste à permettre de

traiter des corpus de grande taille, à la différence de l'annotation manuelle qui demeure extrêmement coûteuse en temps notamment. Grâce au développement d'outils automatiques, les experts peuvent produire une annotation manuelle sur un corpus assez réduit, traditionnellement appelé *gold et* utilisé comme référence dans le cas d'annotations multiples. Ce dernier sert notamment à évaluer le système automatique à la fois quantitativement (par exemple à évaluer le taux d'erreur) et qualitativement (le type d'erreur). Ceci est favorisé par le fait que les annotations automatiques sont consistantes. « Highly reliable manually annotated resources are, naturally, more expensive to construct, rarer and smaller in size than automatically annotated data, but they are essential for the development of automated annotation tools and are necessary whenever the desired annotation procedure either has not yet been automated or cannot be automated. » (The Clarin User Guide).

9 L'enrichissement des corpus peut donc être totalement manuel ou totalement automatique. Le choix opéré au LPL (plus particulièrement avec le logiciel SPPAS présenté ci-après) est celui d'une annotation semi-automatique. Il est très irréaliste, selon nous, de penser que l'analyste humain peut être retiré du processus d'annotation. Généralement, ce type de système se construit de concert entre un expert linguiste et un informaticien spécialiste du traitement automatique de la parole et/ou du langage. L'annotation semi-automatique permet ainsi à l'humain d'inter-opérer, c'est-à-dire d'intervenir dans le processus d'annotation (notamment en termes de correction), ce qui favorise l'adaptation constante de l'outil utilisé.

10 Dans cet article, nous présentons donc la chaîne de traitement déployée sur nos différents corpus depuis la segmentation du flux de parole en blocs (IPU) bornés par des pauses silencieuses au sein desquels est effectuée la transcription à partir de laquelle les différents niveaux d'annotation seront déclinés. Chacune des sections présentées dans cet article tente de résumer les motivations, méthodes et outils (parmi eux SPPAS, Marsatag et HMAD) associés à chaque dimension considérée (phonétique, lexique, morphosyntaxe, interaction) ou à un phénomène étudié en particulier (transitions thématiques, sourires, feedbacks, humour, gestes manuels).

2. IPUs et Transcription

11 En raison de la taille des corpus de plus en plus importante, il s'avère nécessaire de préalablement segmenter le signal de parole. Le choix opéré au LPL, similairement à d'autres travaux existants tels que Koiso *et al.* (1998), a consisté à pré-découper automatiquement le signal en unités inter-pausales (IPUs). L'IPU est un bloc de parole borné par des pauses silencieuses, dont la durée varie selon les langues (durée minimale de 200 ms pour le français, dans le cas du CID). L'IPU présente plusieurs avantages dont ceux d'être une unité objective, formelle et repérable automatiquement. Elle permet notamment de disposer de petits blocs de parole au sein desquels l'étape de transcription ultérieure sera facilitée. L'IPU peut en outre être assimilée parfois aux tours de parole, lesquels sont encore très difficiles à déterminer puisqu'ils sont le résultat d'une conjonction d'indices syntaxique, prosodique et pragmatique (Ford et Thompson, 1996 ; Degand & Simon, 2009) qu'il reste encore aujourd'hui à caractériser.

12 Une fois segmenté en IPU, le corpus est transcrit selon des conventions très précises. Pour certains corpus au LPL tels que le CID, nous avons adopté une *transcription*

orthographique enrichie (*TOE*) (Bertrand *et al.* 2008). La motivation initiale de cette transcription est liée au manque encore important de travaux sur la parole spontanée dans les années 2000. Nous n'avons pas de certitudes sur le niveau de variabilité de cette parole, et donc de l'écart à la norme d'un tel jeu de données. Afin de ne pas rater des phénomènes, mais surtout pour optimiser les outils automatiques utilisés ultérieurement, nous avons donc privilégié cette *TOE*, qui comporte des éléments tels que les élisions, les prononciations ou liaisons non attendues, etc. (voir Bigi *et al.* 2012 pour plus de détails sur les bénéfices d'une telle *TOE* sur la phonétisation et l'alignement sur le signal audio).

13 Les conventions de transcription de la *TOE* sont inspirées de celles du GARS (Benveniste & Jeanjean, 1987). Cet ensemble de règles d'annotations peut être utilisé depuis n'importe quel logiciel d'annotation (Praat, Elan,...). La *TOE* constitue le point d'entrée de logiciels d'annotations automatiques développés au LPL ; nous avons constaté qu'elle en améliore les performances (Bigi *et al.*, 2012), en particulier en parole spontanée.

14 Ci-après, l'exemple d'une IPU issue d'un extrait de discours à l'Assemblée Nationale (Bigi *et al.*, 2013) :

> *euh les apiculteurs + et notamment b- on n(e) sait pas très bien + quelle est la cause de mortalité des abeilles m(ais) enfin il y a quand même + euh peut-êt(r)e des attaques systémiques*

avec les conventions de *TOE* suivantes :

> \+ : pause silencieuse inférieure à 200 ms
> b- : amorce
> (xxx) : élision

15 À partir de cette TOE, nous avons donc pu dériver automatiquement deux transcriptions, temporellement synchrones, l'une destinée aux domaines de l'organisation textuelle (morpho-syntaxe, syntaxe, discours) (cf. Marsatag section 5) et l'autre destinée aux domaines de l'organisation de la parole (phonétique, prosodie) (cf. SPPAS section 3) :

> tr1. *euh les apiculteurs + et notamment b on n sait pas très bien + quelle est la cause de mortalité des abeilles m enfin il y a quand même + euh peut-ête des attaques systémiques*
> tr2. *les apiculteurs et notamment on ne sait pas très bien quelle est la cause de mortalité des abeilles mais enfin il y a quand même peut-être des attaques systémiques*

16 Désormais, la question relative à la pertinence de recourir à la *TOE* se pose dans la mesure où les dictionnaires peuvent recenser de plus en plus de formes de l'oral grâce notamment aux nombreuses études menées durant ces dernières années. Il s'avère donc indispensable de repenser son usage, et de limiter l'enrichissement éventuel de l'orthographe standard à des phénomènes qu'il n'est pas encore possible de retrouver automatiquement. En revanche, cet enrichissement s'avère pertinent pour les langues dont les dictionnaires ne recensent que les formes standard. Cette convention a également permis des études spécifiques, notamment en ce qui concerne la fréquence des phénomènes transcrits. Nous avons par exemple comparé la fréquence des rires, des hésitations et des bruits dans différents corpus français (Bigi & Meunier, 2018). Par ailleurs, la transcription étant réalisée dans les IPUs, nous avons pu observer que ces trois phénomènes sont présents dans 3,3 % des IPUs en parole lue contre 20 % à 36 % dans les divers corpus de parole spontanée.

17 Enfin, si la TOE[2] améliore l'alignement automatique, elle peut parfois être un obstacle à la recherche de certains événements. Nous parlons ici des segments phonétiques non réalisés dans la parole spontanée ou encore des phénomènes de coalescence, plus

couramment appelés « réduction de la parole » [voir l'article *Conversation* dans ce numéro]. Les transcripteurs étant en mesure d'identifier certaines élisions, comme cela a été mentionné plus haut, on pourrait s'attendre à ce que la TOE nous fournisse un inventaire des élisions d'un corpus. Il serait ainsi possible d'inventorier les élisions d'un locuteur en recherchant tous les segments transcrits entre parenthèses (élision de /E/ et de /R/ dans la phrase « *m(ais) enfin il y a quand même + euh peut-êt(r)e des attaques* »). Cet inventaire nous fournit alors la liste des élisions perçues par les transcripteurs. Malheureusement, cette liste est très loin d'être exhaustive dans la mesure où les réductions phonétiques sont très fréquentes en parole spontanée (Johnson, 2004) et la plupart ne sont pas perçues par les auditeurs. Un recueil plus exhaustif des réductions doit donc passer par une procédure automatique non fondée sur la perception humaine. Les outils automatiques s'appuient sur les « erreurs » de l'alignement pour rechercher les réductions : lorsqu'une séquence est réduite, l'aligneur doit attribuer à une portion de signal un nombre important d'unités phonétiques qui ne sont, en réalité, pas produites. Chaque unité aura donc la durée minimale de la trame (souvent autour de 30 ms). C'est la contiguïté de ces segments extra-courts qui permettra d'identifier plus facilement les séquences réduites (Wu & Adda-Decker, 2021). Ainsi, il est probablement de nombreux aspects pour lesquels les erreurs de l'alignement automatique sont une source d'information importante, tout particulièrement pour ce qui concerne les corpus de parole spontanée.

3. Annoter automatiquement avec SPPAS

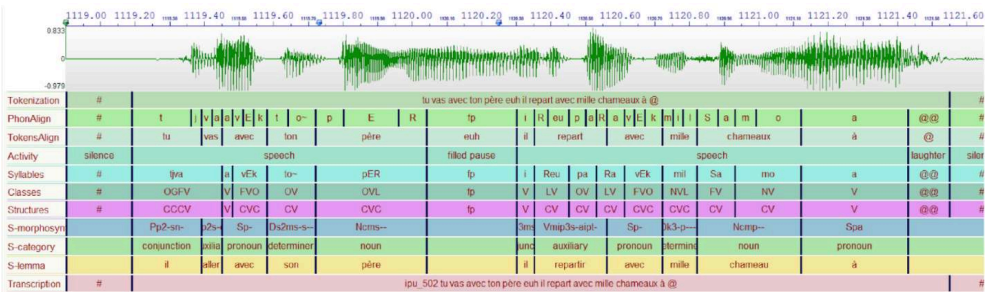
18 Le logiciel *SPPAS - the automatic annotation and analysis of speech* (Bigi, 2012 ; Bigi, 2015), développé au LPL, propose l'enrichissement du corpus avec des annotations qui sont produites **automatiquement** (voir la fiche technique [3] pour la liste exhaustive). Chacune de ces annotations a été élaborée pour répondre aux besoins d'un projet ou d'une collaboration. SPPAS est un support d'application qui résulte d'un processus de collaboration, d'un travail de mise en forme et de valorisation des résultats de la recherche. C'est donc un logiciel de type « Research Software » au sens de (Gomez-Diaz & Ricio, 2019) : son but principal concerne l'activité de recherche en elle-même, c'est-à-dire qu'il est conçu selon une démarche scientifique, et se destine à des scientifiques.

19 La réponse privilégiée lors de la conception d'une nouvelle annotation place les compétences et les connaissances des experts linguistes au cœur du système. Il combine ou utilise alternativement l'approche connexionniste ou l'approche symbolique de l'I.A. L'approche symbolique repose sur la logique formelle et l'exploitation de connaissances existantes pour résoudre des problèmes, elle s'appuie notamment sur des moteurs de règles et de faits. Elle permet notamment, mais pas seulement, de créer des systèmes experts, intégrés dans les systèmes d'aide à la décision. L'I.A. connexionniste utilise des méthodes plus empiriques basées sur l'observation et l'exploitation d'exemples, sur des méthodes statistiques, sur la reconnaissance de formes et ce qui est devenu le *machine learning* puis le *deep learning*. Les solutions d'I.A. connexionnistes exploitant de gros volumes de données cherchent notamment à découvrir les règles implicites contenues dans les données. Ses solutions s'expriment sous la forme d'un pourcentage de véracité, qui est jugé acceptable lorsqu'il est proche des erreurs humaines. Les annotations automatiques de SPPAS tentent de minimiser la quantité de données à observer pour apprendre un modèle/une

représentation et concevoir une approche méthodologique de l'annotation automatique qui intègre les besoins des chercheurs linguistes. L'autre aspect majeur des annotations de SPPAS réside dans son approche multilingue : les connaissances linguistiques sont externalisées dans des bases de connaissances que l'on appelle « ressources linguistiques ». Ces deux aspects conceptuels permettent d'une part de pouvoir traiter rapidement des langues peu dotées informatiquement (Bigi *et al.*, 2017), d'autre part d'ajouter de nouvelles langues ou variantes dans le système sans le modifier (Lancien *et al.*, 2020).

20 Quelle que soit l'annotation, lors de l'élaboration de l'annotation automatique, plusieurs questions se posent : d'une part, celle des étiquettes utilisées (décomposition, typologie, fonction, nature graduelle/catégorielle, etc.) ; d'autre part, celle de l'ancrage temporel de ces étiquettes (localisation et frontières). Ces questions doivent être pensées en fonction des objectifs de recherche. La figure 2 illustre certaines des annotations produites par SPPAS ainsi que celles produites par MarsaTag (section 5). À l'image du LPL, SPPAS est donc interdisciplinaire puisqu'il s'inscrit dans le domaine de la linguistique computationnelle appliquée (Intelligence Artificielle) et qu'il est utilisé par des chercheurs du domaine de la linguistique de corpus. Il s'intègre dans leur processus d'annotation de corpus comme en témoignent Ide *et al.* (2017 : 148) : « *New developments in providing automatic annotation for linguistic purposes are also appearing, and will lead to the development of new and more efficient workflow practices in this area (e.g. SPPAS [5]).* »

Figure 2 : IPUs transcrites manuellement (dernière ligne) et quelques-unes des annotations obtenues automatiquement à partir de celles-ci



De haut en bas, avec SPPAS, sont obtenues : normalisation de la transcription, phonèmes, mots, activité, syllabes, classes, structures syllabiques ; avec MarsaTag, sont obtenues : morpho-syntaxe, catégories morpho-syntaxiques, lemmes.

21 À titre d'exemple, nous présentons succinctement l'annotation automatique « Search for IPUs » qui recherche les segments audibles d'un signal audio selon deux paramètres qui sont fixés par l'utilisateur :

- la durée minimale des silences, qui dépend essentiellement de la langue. Nous avons observé qu'une durée de 200 ms (Bertrand *et al.* 2008) donne de bons résultats pour le français ;
- la durée minimale des IPUs, qui dépend essentiellement du type de corpus. En parole lue, une durée de 300 à 400 ms donne de bons résultats, tandis qu'en parole spontanée, une durée de 100 ms est plus appropriée afin d'assigner à des IPUs les « mh », « ouais » et autres feedbacks.

22 L'automatisation de notre système relève essentiellement de l'analyse des valeurs d'intensité observées afin de fixer un seuil de catégorisation silence/son optimal pour chaque fichier. Cette analyse s'effectue sur des segments de 10 à 20 ms qui sont ensuite

regroupés soit en silences, soit en IPUs, puis filtrés avec les deux critères de durées mentionnés ci-dessus.

23 Nous avons évalué les performances de ce système sur 5 dialogues du corpus conversationnel *Cheese* (Priego-Valverde, Bigi, & Amoyal, 2020). Pour ce faire, nous avons estimé le nombre d'actions requises par un annotateur pour corriger *a posteriori* la segmentation automatique (Bigi & Priego-Valverde, 2019). Le système a été paramétré de sorte qu'il trouve le plus grand nombre possible d'IPUs, **sans en oublier**. Ce réglage permet à l'utilisateur de n'écouter du signal que les IPUs trouvées par le système, sans devoir écouter également les silences à la recherche de segments non détectés. En revanche, un certain nombre des IPUs proposées devront être supprimées car non pertinentes.

4. Segments phonétiques et lexicaux

4.1 Annotation en phonétique

24 L'annotation de la parole de segments est indispensable pour (1) maîtriser le flux continu de la parole, difficile à appréhender et (2) faire des analyses (acoustiques, distributionnelles, quantitatives, etc.) sur ces segments. Sans un balisage, quel qu'il soit, il est difficile d'appréhender la parole pour en faire un produit interprétable. L'annotation/segmentation de la parole est également nécessaire pour de nombreuses expériences de perception dans lesquelles une partie du signal de parole doit être extraite. Enfin, ce balisage est également indispensable pour l'évaluation des outils d'alignement automatique.

25 La notion d'annotation en phonétique n'est venue que tardivement sous forme d'emprunt à la terminologie employée dans les approches d'enrichissement de corpus (syntaxique, lexicale, gestuelle, etc.). La notion de « segmentation » est effectivement caractéristique des problématiques liées au niveau phonétique. Segmenter phonétiquement le signal revient à apposer sur le signal de parole des étiquettes temporelles représentant approximativement le début ou la fin d'un segment phonétique. C'est évidemment dans le terme « approximativement » que se trouve tout le débat sur la segmentation des unités phonétiques. L'ensemble de nos connaissances sur la coarticulation (Farnetani, 1997) et sur la propagation des traits phonétiques sur de longues séquences (Nguyen *et al.*, 2004) a montré qu'il est illusoire de vouloir représenter les segments phonétiques comme une suite d'éléments contenus entre deux frontières. S'il est indéniable que la production des phonèmes est ordonnée dans le temps, les multiples articulateurs en jeu et leurs mécanismes complexes de synchronisation et d'anticipation suggèrent que la sortie acoustique observable comporte de multiples indices répartis au-delà de la manifestation saillante du phonème sur le signal de parole. Cette relation complexe entre niveaux linguistique, articulatoire et acoustique nous rappelle que les frontières de segments ne doivent pas être confondues avec des frontières de phonèmes (Fant, 1973). En effet, la notion abstraite de phonème ne peut être mise en relation directe avec des discontinuités du signal acoustique (Rossi, 1990). On admet donc que les étiquettes de début et de fin des unités phonétiques, qu'elles soient apposées automatiquement ou manuellement, ne représentent pas des frontières réelles de phonèmes. Il s'agit juste de repères

permettant de procéder à des analyses sur le signal de parole. La segmentation est donc un outil qui doit répondre à des questions préalables.

26 L'utilisation de grands corpus de parole ainsi que les outils développés pour leur exploitation ont rendu quasiment obsolète l'expertise manuelle pour ce qui concerne l'annotation des corpus dans des langues bien documentées. Le travail auparavant indispensable de segmentation manuelle du signal de parole (Meunier & Nguyen, 2013) évolue désormais vers un travail de correction manuelle de l'alignement automatique. Dans un avenir proche, il est donc probable que très peu (ou pas) de corpus seront encore segmentés manuellement dans leur ensemble, à l'exception peut-être de corpus très réduits. Toutefois, l'utilisation systématique d'outils du traitement automatique a pour conséquence que les phonéticiens n'appréhendent plus « manuellement » le signal de parole. Cette raréfaction de l'expertise humaine peut générer une vision incomplète des phénomènes spécifiques à la parole spontanée (omissions, réduction, fusion, métaplasmes, etc.) dans la mesure où nos hypothèses de recherche restent souvent fondées sur les phénomènes identifiés en parole plus contrôlée. Il reste donc nécessaire de « sonder » partiellement de façon manuelle le signal de parole même si les outils actuels peuvent nous en dispenser de façon très efficace. Les approches automatique et manuelle apparaissent ainsi comme complémentaires.

4.2 Segmentations automatiques

27 L'alignement automatique texte-son s'opère en trois tâches qui se succèdent : normalisation de la transcription orthographique, phonétisation du texte normalisé puis alignement forcé de la grammaire des phonèmes.

28 La normalisation de texte est le processus de segmentation d'un texte en unités lexicales. En principe, tout système qui traite des textes à des fins d'analyse requiert qu'ils soient normalisés. Cependant, la normalisation automatique de textes est d'une part principalement dédiée aux textes écrits dans la communauté du Traitement Automatique du Langage - TAL, d'autre part considérée comme une tâche qui dépend de la langue - donc un nouveau système est créé pour chaque nouvelle langue abordée. Il nous est apparu nécessaire de développer un nouveau système qui soit adapté aux divers corpus traités au LPL, à savoir des corpus de différents styles de parole (lue, conversationnelle, orientée tâche...) et de langues différentes. Ce système est intégré et distribué dans SPPAS (Bigi, 2014). Il supprime la ponctuation, passe le texte en minuscules, convertit les nombres dans leur forme textuelle (seul module interne qui dépend de la langue), remplace les symboles par leur forme écrite et effectue une segmentation en mots. Certaines de ces opérations s'appuient sur des listes de mots (externes au système) de la langue. En 2012, ce système automatique permettait de normaliser des transcriptions en français, anglais ou italien. Au fil du temps, d'autres langues ont été ajoutées ; en 2021, 17 langues peuvent être normalisées automatiquement avec ce système. À partir de la TOE, SPPAS peut ainsi dériver les deux annotations citées en exemple dans la section précédente.

29 La conversion graphème-phonème que nous avons également développée dans SPPAS utilise le texte normalisé (tr1) pour produire une grammaire des prononciations possibles (Bigi, 2016) de chaque mot (tr2) d'une IPU. La transcription phonétique d'un texte est une composante indispensable des systèmes de synthèse de la parole ; elle est aussi utilisée dans la modélisation acoustique pour la reconnaissance automatique de la

parole et dans d'autres applications de traitement du langage naturel. La transcription phonétique peut être mise en œuvre de nombreuses façons, selon des stratégies fondées sur des dictionnaires ou basées sur des règles, bien qu'il existe de nombreuses solutions intermédiaires. La solution que nous avons mise en œuvre utilise un dictionnaire de prononciation à la fois pour produire la prononciation des mots connus et celle des mots inconnus. Elle offre l'avantage de ne pas dépendre de la langue et donc de pouvoir être utilisée dès lors qu'un dictionnaire est disponible. En 2021, 16 dictionnaires sont disponibles sous licence libre et permettent à SPPAS d'effectuer la conversion graphème-phonème.

30 La grammaire obtenue par la phonétisation constitue le point d'entrée du système d'alignement forcé décrit dans (Bigi & Meunier, 2018). L'alignement forcé, aussi appelé segmentation phonétique, est le processus d'alignement de la parole avec sa transcription correspondante au niveau du phonème. Le système fonctionne en deux étapes : le système détermine d'abord la prononciation la plus probable de l'IPU en cours de traitement à partir de la grammaire puis il aligne temporellement chacun des phonèmes de cette prononciation. Cet alignement nécessite un modèle acoustique : c'est un fichier qui contient des représentations statistiques de chacun des sons distincts d'une langue. Une des particularités de SPPAS réside dans le fait qu'il permet d'aligner automatiquement les rires, les bruits, et les pauses remplies (selon la langue). De plus, puisqu'il s'appuie sur les deux étapes précédentes, cet alignement obtient des résultats de qualité équivalente quel que soit le style de parole : lue ou spontanée. En plus de l'alignement en phonèmes, le système produit l'alignement en mots. En 2021, 16 modèles acoustiques sont disponibles.

31 L'alignement temporel des phonèmes permet ensuite d'obtenir la segmentation en syllabes avec un système automatique à base de règles que nous avons développé (Bigi *et al.*, 2010). La syllabe est une unité linguistique largement utilisée pour des études phonologiques, phonétiques et/ou prosodiques (par exemple, association air-texte, alternances rythmiques, etc). Bien que la structure phonologique de la syllabe soit similaire dans différentes langues, les règles phonologiques et phono-tactiques de syllabation sont spécifiques à chaque langue. Les approches automatiques de la détection des syllabes doivent donc incorporer de telles contraintes pour localiser précisément les limites des syllabes. Ce système peut être utilisé pour traiter différentes langues : l'ajout d'une nouvelle langue nécessite d'établir les règles de cette segmentation avec un expert phonéticien. Pour l'heure, ce système fonctionne pour le français (17 règles), l'italien (12 règles) et le polonais (489 règles).

4.3 Correction de la segmentation automatique

32 L'annotation phonétique des corpus étant désormais essentiellement produite à l'aide d'outils d'alignement automatique, l'intervention de l'expert humain est-elle encore utile ? Il est possible de se passer complètement de l'expertise humaine pour ce qui concerne l'annotation phonétique. Lorsque les analyses phonétiques portent sur l'ensemble des segments et notamment sur leur durée globale ou sur des valeurs dont le positionnement peut rester approximatif (extraction de formants au centre d'une voyelle ou du bruit d'une fricative, par exemple), l'annotation fournie par l'annotation automatique (AA) semble satisfaisante et ne nécessite pas forcément de correction manuelle. En revanche, lorsque les analyses portent sur des zones transitoires dont les localisations doivent être précises (événements acoustiques tels que l'explosion d'une

plosive, des zones d'assimilation de voisement ou des phénomènes de coarticulation), l'annotation fournie par l'AA est trop approximative. Nous avons élaboré des critères de segmentations entre macro-classes phonétiques de façon à rendre accessible cette tâche parfois ardue pour les non spécialistes. Ces critères sont décrits dans Meunier & Nguyen (2013). Il s'agit de fournir des indices utilisés très couramment pour caractériser des frontières de segments que chacun pourra utiliser selon leur pertinence avec les objectifs posés.

33 Par ailleurs, la correction (ou l'adaptation) de l'annotation phonétique sera parfois indispensable dans les situations où la parole s'éloigne de sa production « typique ». Nous parlons ici des caractéristiques de la parole « pathologique » et notamment de la difficulté à annoter la parole lorsque les unités de production sont altérées. C'est le cas pour les différentes maladies caractérisées par une dysarthrie [voir l'article *Phonétique Clinique* dans ce numéro] impactant la motricité des gestes nécessaires à la production des sons de la parole. Si une attention particulière a été donnée depuis de nombreuses années au développement d'outils automatiques adaptés au traitement automatique de la parole pathologique (Fredouille & Pouchoulin, 2012 ; Laaridh *et al.*, 2018), il n'en reste pas moins que l'AA de ce type de parole doit parfois être corrigé. Dans ce cas, les critères classiques de segmentation phonétique doivent être adaptés et il est nécessaire d'intégrer des classes de segments spécifiques à ce type de parole (voir Meunier, 2014, pour un développement de ces aspects).

5. Syntaxe et morphosyntaxe

5.1 Annotations automatiques

34 Nous faisons ici le point sur la chaîne de traitement que nous avons développée au LPL dans le domaine de l'analyse de la syntaxe du français écrit ou parlé. Ces outils automatiques permettent d'annoter un texte ou une transcription en plusieurs étapes. L'entrée est tout d'abord segmentée en *tokens* (i.e. mot ou groupe de mots pouvant être considéré comme une unité syntaxique atomique), c'est la phase de segmentation (i.e. *tokenization*). La distribution hors contexte des catégories morphosyntaxiques possibles correspondant à la forme orthographique de chaque *token* est ensuite récupérée via un lexique morphosyntaxique. Une procédure de désambiguïsation est alors appliquée afin d'obtenir la catégorie morphosyntaxique associée au token en contexte dans l'énoncé, c'est l'étape d'étiquetage (i.e. part-of-speech *tagging*). L'étape finale consiste à structurer cette séquence de catégories en unités syntaxiques plus complexes, structures plates (i.e. *chunking*) ou structures d'arbres de constituants (i.e. *parsing*). Les relations liant ces différents constituants pourront aussi être décrites. Les ressources, la liste des catégories morphosyntaxiques ainsi que les constructions décrites dépendent de la langue analysée (ici le français) et du type de productions (e.g. écrit ou oral). Les raisons qui nous ont menés à développer notre propre système sont multiples. Nous avons notamment besoin dans d'autres champs d'étude d'avoir un accès direct et en temps réel à l'état du système pour, par exemple, réaliser une prédiction en temps réel des feedbacks d'un agent virtuel (Blache *et al.*, 2020) ou pour calculer pas à pas la complexité syntaxique d'une production (Blache & Rauzy, 2011 ; Rauzy & Blache, 2012). La description de notre chaîne de traitement pour le français écrit est présentée dans Rauzy & Blache (2009), nous spécifions dans cette section les

modifications apportées à notre système afin de traiter des transcriptions de l'oral spontané.

35 Dans un premier temps les transcriptions du corpus CID (Bertrand *et al.*, 2008), riche de 115 000 tokens, ont été traitées par l'étiqueteur du LPL (Rauzy & Blache, 2009). Il s'agit d'un étiqueteur de nature stochastique qui a été entraîné sur du français écrit. Il exécute trois opérations. La première consiste à segmenter le texte brut en entrée en une séquence de tokens correspondant à des unités syntaxiques atomiques (i.e. part-of-speech, POS). Un lexique morphosyntaxique permet ensuite de rechercher, compte tenu de la forme orthographique de chaque *token*, la distribution des catégories morphosyntaxiques associée à cette forme (voir tableau 1 pour un exemple). La dernière opération est la tâche de désambiguïsation. Il s'agit de sélectionner parmi la combinatoire des séquences de catégories générées par l'ambiguïté de chaque forme la séquence de probabilité maximum. La probabilité d'une séquence de catégorie est calculée en suivant un modèle stochastique dans le cadre du formalisme des Chaines de Markov Cachées (i.e. HMM). Le calcul est basé sur la distribution de probabilité des catégories conditionné par le contexte gauche (i.e. les catégories précédentes dans la séquence). Le modèle est en fait défini par une liste de « patrons morphosyntaxiques » définis par leur contexte gauche et la distribution de catégories qui leur est associée (voir un exemple tableau 2). Les patrons du modèle sont extraits du corpus GraceLPL, une version du corpus Grace/Multitag (Paroubek & Rajman, 2000) qui contient environ 700 000 tokens avec une étiquette morphosyntaxique qui suit le jeu de traits Multext (Ide & Véronis, 1994). La ressource GraceLPL est régulièrement corrigée et enrichie afin d'améliorer son étiquetage.

Tableau 1 : La distribution des catégories morphosyntaxiques pour la forme orthographique « bon » qui peut être utilisée comme un adjectif un adverbe ou un nom

Forme	Lemme	SAMPA	Catégorie	Traits morphosyntaxiques	Probabilité
bon	bon	bo~	Adjectif	Afpms	0.805
bon	bon	bo~	Adverbe	Rgp	0.190
bon	bon	bo~	Nom	Ncms	0.005

Les traits morphosyntaxiques rendent compte du codage précis de l'information morphosyntaxique, « Ncms » se lit « Nom commun masculin singulier ».

Tableau 2 : Le patron morphosyntaxique identifié par le contexte gauche

Séquence contexte gauche	Catégorie suivante	Probabilité conditionnelle
Verbe Adverbe Déterminant	Adjectif	0.09
	Adverbe	0.01

	Nom	0.9

Le patron morphosyntaxique identifié par le contexte gauche « Verbe Adverbe Déterminant » peut être suivi par un adjectif avec une probabilité de 0.09, un adverbe avec une probabilité de 0.01 ou un nom avec une probabilité de 0.9. Les autres catégories (e.g. pronom, déterminant, préposition,...) ont une probabilité nulle d'apparaître dans ce contexte.

36 L'information morphosyntaxique a été organisée de manière *ad-hoc* en un jeu de 48 étiquettes : 2 types d'étiquettes pour les marqueurs de ponctuation, 1 pour les interjections, 2 pour les adjectifs, 2 pour les conjonctions, 1 pour les déterminants, 3 pour les noms, 8 pour les auxiliaires, 4 pour les verbes, 5 pour les prépositions, 3 pour les adverbes et 15 étiquettes pour les pronoms.

37 Le modèle stochastique à la base de notre étiqueteur est composé de 2 841 patrons de taille variable (le contexte gauche le plus long compte 8 catégories). L'évaluation de l'étiqueteur a été conduite en comparant les sorties de l'étiqueteur avec la référence GraceLPL. Pour le jeu de 48 catégories mentionné plus haut, la performance de l'étiqueteur (version 2011) atteint un score de F-mesure de 0.974. Cette valeur correspond à un taux d'erreur d'étiquetage de 2.56 %. Pour comparaison, le taux d'erreur obtenu en utilisant uniquement l'information sur les fréquences du lexique morphosyntaxique est de 10.7 % (voir Blache & Rauzy, 2008).

38 Nous avons opéré certaines modifications dans la chaîne de traitement afin d'étiqueter les transcriptions du corpus CID (Bertrand *et al.*, 2008). Dans un premier lieu, les annotations de la transcription orthographique enrichie ne présentant pas de contenu syntaxique ont été filtrées (pause remplie, hésitation, troncation de mot, liaison, rire, ...). L'étiqueteur a été aussi modifié afin de prendre en compte l'absence de marqueurs de ponctuation dans la transcription en entrée. D'une part la transcription a été fractionnée en segments de texte lorsqu'ils étaient séparés par des pauses d'une durée supérieure à 500 millisecondes. D'autre part, à l'intérieur de chacun de ces segments, nous laissons à l'étiqueteur la liberté d'insérer des marqueurs de ponctuation (ces marqueurs sont insérés lorsqu'ils augmentent la probabilité de la séquence en cours de traitement). Deux classes de marqueurs sont disponibles, les marqueurs de ponctuation forts qui correspondent à l'écrit aux ponctuations finales en fin de phrase et les marqueurs de ponctuation faibles associés aux virgules par exemple. Cette procédure permet de générer deux nouvelles unités définies uniquement sur la base de l'information syntaxique : des pseudo-phrases correspondant à des unités délimitées par deux marqueurs de ponctuation forts et des unités plus petites délimitées par les marqueurs de ponctuation faibles.

39 Dans un deuxième temps, nous avons conduit une analyse des erreurs sur cet étiquetage en sortie. Deux causes principales d'erreurs ont été identifiées, liées à l'absence de la forme orthographique dans le lexique ou l'absence de la catégorie morphosyntaxique appropriée associée à la forme. Ces constatations nous ont amenées à modifier notre lexique. Pour ce faire nous avons créé le lexique propre au corpus CID pour le comparer à son pendant créé à partir de corpus composés de textes écrits. Le lexique du CID contient 6 600 formes orthographiques différentes présentant un nombre d'occurrences variant de 1 à 3 130. Pour chaque forme, le rapport entre la fréquence orale mesurée dans le CID et la fréquence à l'écrit a été calculée. Un extrait du lexique du CID est présenté dans le tableau 3.

Tableau 3 : Pour chaque forme orthographique du CID, le nombre d'occurrences et le rapport entre fréquence orale et fréquence écrite

Forme	Nb d'occurrences dans le CID	Rapport oral *versus* écrit	Type
ouais	2916	26035.715	DM
mais	1429	3.483	DM
quoi	1175	56.26	DM
bon	677	18.774	DM
tu sais	635	119.266	DM
...	...	...	...
beh	59	infini	manquante
appart	11	infini	manquante
...	...	...	...

Les formes ont été classifiées par type, DM pour « Marqueur de discours » et « manquante » lorsque la forme n'est pas dans le lexique de l'écrit.

40 Les formes spécifiques au français parlé apparaissent dans cette table avec un rapport infini (i.e. leur fréquence dans le lexique de l'écrit est nulle). Nous avons listé ces formes et nous les avons ajoutées à notre lexique initial. Dans le corpus CID, les phénomènes les plus fréquents sont les réductions (e.g. « appart » pour « appartement », « exo » pour « exercice »,...), les versions régionales ou étrangères (e.g. « cagole », « strange »,...) et les onomatopées (e.g. « beh », « mh »,...).

41 Les formes qui montrent un rapport de fréquence oral versus écrit important sont potentiellement intéressantes. Ceci pourrait indiquer que les locuteurs utilisent ces formes dans un but différent et complémentaire à celui utilisé lors des productions écrites. Parmi ces formes, les marqueurs de discours ressortent du lexique du CID. Les marqueurs de discours à l'oral occupent des fonctions variées à des niveaux différents de l'analyse linguistique. Le marqueur de discours « ouais » (voir tableau 3) est par exemple massivement produit dans le CID par le locuteur secondaire comme un backchannel pendant la production du locuteur principal. D'autres marqueurs de discours (e.g. « mais », « quoi », « bon »,...) peuvent être utilisés par le locuteur principal à l'intérieur des fillers avec une fonction de planification du discours. Dans ces exemples, ces marqueurs de discours n'occupent pas la fonction syntaxique qu'ils occupent habituellement dans les productions écrites. Au niveau de l'étiquetage, nous avons décidé de leur associer l'étiquette « Interjection » lorsqu'un tel usage est fait de ces marqueurs. Ce choix n'est pas guidé par des raisons linguistiques mais plutôt motivé par des raisons techniques. En effet, nous avions à proposer une étiquette déjà présente dans le jeu d'étiquettes existant et qui devait de plus présenter des propriétés syntaxiques relativement neutre, ce qui est le cas de l'étiquette « Interjection ». Nous avons ainsi modifié le lexique pour ajouter l'étiquette « Interjection » pour les formes

orthographiques concernées et nous avons de plus modifié les distributions de probabilité associées à ces formes. Un exemple est proposé dans le tableau 4. Cette modification, si elle concerne moins de 40 formes du lexique, est d'une grande importance pour le traitement général. En effet, environ 10 % des *tokens* du corpus dans sa globalité sont impactés par cette modification.

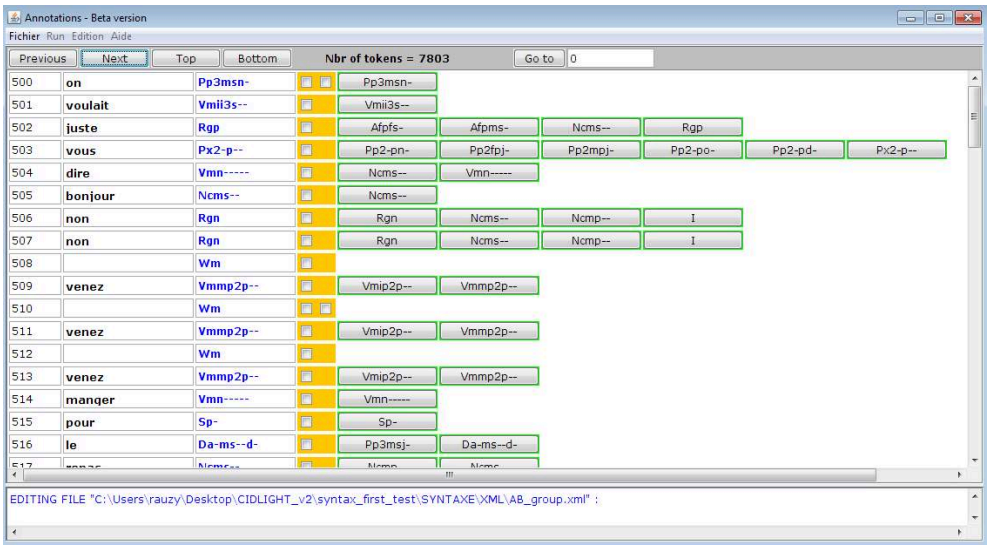
Tableau 4 : La modification de la distribution des étiquettes pour la forme « bon »

Forme	Lemme	SAMPA	Catégorie	Traits morphosyntaxiques	Probabilité
bon	bon	bo~	Interjection	I	0.95
bon	bon	bo~	Adjectif	Afpms	0.04025
bon	bon	bo~	Adverbe	Rgp	0.00950
bon	bon	bo~	Nom	Ncms	0.00025

Une entrée additionnelle a été ajoutée avec l'étiquette Interjection correspondant à l'utilisation de cette forme comme marqueur de discours. Les probabilités ont donc été redéfinies en prenant en compte le rapport entre l'usage à l'oral et à l'écrit (voir tableau 1 pour une comparaison).

42 Nous avons ensuite appliqué la nouvelle version de l'étiqueteur aux transcriptions en entrée. À partir de ce nouvel étiquetage automatique, une correction manuelle a été effectuée sur l'ensemble des 115 000 tokens du CID. Les données ont été partagées en 3 parties et chacune a été corrigée par un annotateur différent. Cette tâche a nécessité 200 heures de correction, en incluant le temps passé à se familiariser avec l'encodage du jeu de traits morphosyntaxiques et avec l'interface que nous avons développée pour effectuer cette opération (voir figure 3). Ceci équivaut à une durée de 25 heures d'annotation pour 1 heure de dialogue, soit une vitesse de traitement de 600 tokens à l'heure environ.

Figure 3 : L'interface utilisée pour la correction des étiquettes morphosyntaxiques



La transcription est présentée horizontalement, une ligne par *token*. Chaque ligne contient la forme orthographique du *token*, la solution de l'étiqueteur (deuxième colonne), et la liste des étiquettes possibles associées à la forme via le lexique. Pour corriger une étiquette, l'annotateur doit cliquer sur l'étiquette correcte (ou saisir le code du trait si elle n'est pas proposée).

43 La version corrigée manuellement a ensuite été utilisée pour évaluer les performances de notre étiqueteur adaptée au traitement des transcriptions du français parlé. Nous obtenons un score de F-mesure de 0.948, soit un taux d'erreur d'étiquetage de 5.2 %. Une analyse plus détaillée montre que 23 % des erreurs sont dues à une mauvaise classification des marqueurs de discours. Nous avons aussi examiné la localisation de ces erreurs pour une partie du corpus pour laquelle l'annotation des disfluences était disponible. Les données montrent que 20 % des erreurs d'étiquetage se produisent dans les zones de disfluences.

44 Notre étiqueteur forme le cœur du logiciel MarsaTag (Rauzy *et al.*, 2014), (voir la fiche technique Marsatag[4]). MarsaTag[5] effectue le traitement du français écrit ou des transcriptions du français parlé et propose un vaste choix d'options dans le format des entrées et des sorties.

45 Avec un score de F-mesure d'environ 0.95, les performances de notre étiqueteur sont acceptables. Ceci signifie en particulier que la sortie automatique de notre étiqueteur peut être utilisée par les études qui acceptent un taux d'erreur de 5 % sur l'information morphosyntaxique. C'est le cas par exemple lorsqu'il s'agit de faire la distinction entre mot outil et mot fonction comme dans Meunier & Espesser (2011) ou lorsqu'on utilise le compte de catégories morphosyntaxiques spécifiques comme indicateur de certaines propriétés comme dans Moreau *et al.* (2016) ou dans Bertrand & Espesser (2017). Pour d'autres études, et notamment pour proposer une analyse syntaxique en constituants de l'oral spontané, ce taux d'erreurs est trop important. Pour le CID par exemple, seul 40 % des structures sont identifiés par notre analyseur profond. L'amélioration de notre outil passe par la modélisation des phénomènes de disfluences (voir la section 5.2 sur les disfluences) et un traitement plus précis des marqueurs de discours. Ainsi nous sommes dans ce cas obligés d'abandonner la description sur un seul niveau linguistique et de combiner l'information provenant de plusieurs niveaux (ici disfluences, syntaxe et unité de discours) pour faire progresser notre modèle.

5.2 Les disfluences

46 Lorsqu'on se propose d'étudier la syntaxe d'une langue dans ses performances orales, il est nécessaire d'avoir recours à une étape intermédiaire qui garde trace de ce qui est prononcé. Cet intermédiaire indispensable est la construction d'un texte écrit à partir du message oral, autrement dit il faut le transcrire. Dès 1987 Benveniste et Jeanjean décrivent les méthodes, les résultats et les perspectives de la transcription du Français parlé. Des recueils enregistrés et transcrits (*Corpus de Référence pour le Français Parlé*, *CRFP* puis *Corpaix*) sont constitués qui avaient pour objectif l'analyse syntaxique. Ces entretiens conduits dans la plupart des régions de France présentaient des particularités propres à l'oral qui, dans la grande majorité des cas, perturbent la linéarité du déroulement syntaxique standard (c'est-à-dire celui de l'écrit). Ces phénomènes qui sont également transcrits dans ces corpus sont nommés « disfluences », terme anglais retenu pour ce qu'il signifie sous forme abrégée : un dysfonctionnement de la fluence.

47 Ainsi, la constitution de grands corpus oraux comme le CID, en nécessitant la transcription précise des enregistrements, a révélé l'importance des variations dans la fluence verbale et plus particulièrement l'apparition de phénomènes étrangers aux énoncés écrits. Il s'agit le plus souvent d'amorces de mots et de répétitions décrites dès 1984 par Jeanjean. Ces auto-interruptions provoquent des insertions d'éléments propres à l'oral et rompent le déroulement du texte de l'énoncé. De plus, elles sont à l'origine de ce que Blanche-Benveniste a nommé dès 1979, le piétinement syntaxique qui n'est autre qu'un entassement sur l'axe paradigmatique (cf. plus loin). L'auto-interruption définit, donc, un avant et un après la rupture dans l'énoncé, ce que Shriberg (1994) a décrit sous les termes : Reparandum et Reparans, l'Interregnum étant le lieu d'insertion de phénomènes non reliés syntaxiquement au reste de l'énoncé (c'est-à-dire au Reparandum et au Reparans) :

48 Le Reparandum est le mot suivi d'une auto-interruption (//), il peut être prononcé complètement (ex.1) ou tronqué (ex.2)

1. YM_1506 un gosse j'en // j'en // moi j'en voulais pas
2. YM_1739 ça commençait fé- // en février ou un truc comme ça

49 L'Interregnum est l'espace qui suit le Reparandum (il n'est pas lié syntaxiquement au Reparandum et au Reparans) ; il peut être vide comme dans les exemples précédents (ex. 1 et 2) ou rempli (ex. 3 et 4) :

3. YM_1977 quand il est parti il nous a laissé tous ses // **euh** tous ses épices
4. AP gpd_246 918.81 : un collèg- // **enfin** un mec

50 Le Reparans est ce qui suit l'Interregnum et manifeste que le déroulement syntaxique se poursuit. Dans un certain nombre de cas, l'auto-interruption « provoque » l'arrêt définitif du déroulement syntaxique ; la phrase est alors laissée inachevée ; il n'y a pas de Reparans (ex.5)

5. AG_1540 normalement on a la cr-// enfin c'est c'est O- // septembre c'est OK

51 Les données montrent que le locuteur a plusieurs possibilités pour poursuivre son énoncé. Elles correspondent à un entassement paradigmatique (Benveniste, 1979). Le

Reparans prend alors plusieurs formes qu'il est possible de regrouper en trois catégories :

- 1. une simple complétude du lexique commencé et interrompu (autrement nommée : **amorce de mot complétée**)
- CM gpd_46 euh qui était complètement ah voilà + dés-euh + désynchronisé d'une situation en fait
- 3. une reprise plus ou moins partielle de l'énoncé déjà prononcé (ex. 8) Cela peut aller jusqu'à des éléments précédents le syntagme contenant l'auto-interruption (ex.7) :

6. EDF, 8,38 mais l'eau est de vingt mè- l'entrant est de vingt mètres cubes
7. CM gpd_33 oui où tu perds un peu euh + tu perds un peu

- 3. La reprise de l'énoncé prononcé peut aussi être l'occasion, non seulement de répéter mais également d'apporter des modifications lors de cette répétition (Ex. 9,10,11) :

8. AG_1493 j'ai ma belle-// il y a ma belle-mère qui
9. YM_1124 chaque fois tu ma-//tu devais marquer
10. EDF 12,5 ça serait vraiment un endet- un gros endettement pour la France

52 Il est des cas où l'auto-interruption ne provoque ni inachèvement ni reprise de l'énoncé mais une simple poursuite de l'énoncé (ex. 12, 13, 14) :

12. Y M _2432 tu le fais avec // euh ce que tu veux
13. 1833 YM Bernard c'est vers là-bas // non qu'il habite
14. P gpd_242 907.934 le // euh mari

53 Afin de rendre compte de la fluence des énoncés oraux à un niveau de granularité très fin, nous avons eu pour objectif d'examiner (de suivre « pas à pas ») comment, au niveau morphosyntaxique, l'énoncé se construit (Pallaud et Bertrand, 2020). Nous avons donc identifié toutes les auto-interruptions réalisées par des indices variés et de nature différente (prosodique, discursive, syntaxique). L'interruption n'est considérée comme disfluente que si, lors du Reparans, elle est suivie d'une perturbation morphosyntaxique (répétition d'items avec ou non des modifications, voire un syntagme laissé inachevé). Les autres interruptions sont considérées comme simplement poursuivies, donc sans reprise ou répétition d'items ; elles ne sont donc pas disfluentes (Tableau 5).

Tableau 5 : Structures des auto-interruptions disfluentes et suspensives

Interruption		**Reparandum**	**Interregnum ou Break**	**Reparans**	
disfluente	Quant à elle	elle ne veut	euh ouais	elle ne veut	rien dire
suspensive	Quant à elle	elle ne veut	euh ouais	rien dire	

54 Une méthode de l'annotation de ces phénomènes a été publiée (Pallaud *et al.*, 2019). Toutes les auto-interruptions (ainsi définies) et leurs effets morpho-syntaxiques ont été intégralement annotés dans les huit dilogues du CID (Pallaud, 2014) et sont à la disposition des membres du LPL. Ces études se sont attachées, d'une part à identifier de façon exhaustive toutes les auto-interruptions dans un énoncé et d'autre part à décrire les relations existant entre ces auto-interruptions, leurs indices et leurs effets morphosyntaxiques (en particulier le piétinement syntaxique sur l'axe syntagmatique), ce qui a conduit à préciser la notion de disfluence.

5.3. Le système d'annotation des auto-interruptions et des disfluences morphosyntaxiques

55 Dans l'étude de Pallaud (2014) qui porte sur les transcriptions des huit dilogues (16 locuteurs), l'alignement de l'oral sur les tokens (assimilables aux mots) a été requis. Afin de décrire la totalité des interruptions dans les énoncés, nous avons employé successivement deux méthodes de détection, l'une semi-automatique et l'autre manuelle. Les deux méthodes utilisent le logiciel Praat (Boersma & Weenink, 2001) comme outil d'identification, d'annotation et de description. La méthode semi-automatique a consisté à repérer systématiquement tous les indices d'interruptions, qui ont en commun de ne pas avoir de lien syntaxique avec le reste de l'énoncé : pauses silencieuses ou remplies, marqueurs de discours (ou marqueurs illocutoires, Lapointe, 2017), énoncé parenthétique. Cette méthode semi-automatique permet donc d'identifier au total 79 % de toutes les interruptions dans la fluence verbale soit la grande majorité. La méthode manuelle a consisté à faire une lecture-écoute sémantique des transcriptions qui, à l'aide de paramètres prosodiques, sémantiques et/ou syntaxiques, révèle les ruptures restantes.

56 Chacun des espaces de la structure de l'auto-interruption varie quant à la quantité d'items (verbaux ou non) qu'il contient. Un codage a été établi (cf. la table 6 suivante) afin de décrire et préciser le plus finement possible les diverses catégories de ces auto-interruptions repérées :

- 1° le Reparandum (R ou I) le fragment de mot ou de syntagme qui précède le point de rupture et qui est simplement poursuivi, repris, répété, modifié (codage R) ou laissé inachevé (codage I).
- 2° l'Interregnum (ou Break interval codage B)
- 3° le Reparans (RA) est la partie potentielle de l'énoncé prononcé qui peut poursuivre, répéter ou modifier ce qui a été dit lors du Reparandum. Lorsque l'énoncé est définitivement interrompu, il n'y a pas de Reparans RA

  EX : CM 46 euh qui était complètement **R,P,tw** *ah voilà* + **B,dc,sp** dés- **RA,nr,co**
  R,W,lw *euh* + **B,fp,sp** désynchronisé **RA,wr,wc** d'une situation en fait

57 L'annotation peut donc se limiter à identifier les trois éléments de la structure (Reparandum, Interregnum et Reparans) ou décrire également chacun des types d'éléments qui les composent et dont la description peut se résumer par le tableau 6 suivant (Blache *et al.*, 2010b).

Tableau 6 : Système d'annotation des interruptions dans la fluence verbale

Reparandum		
Reparandum : Type	R	Interruption temporaire
	I	Interruption définitive
Reparandum : catégorie	W	Reparandum : mot
	P	Reparandum : syntagme, proposition
Type de lexique	tw	Mot outil

	lw	Mot lexical
Break_type B		
	no	Pas d'intervalle
	sp	Pause silencieuse (> 200 ms)
	fp	Pause remplie
	dc	Marqueur de discours
	ps	Incise parenthétique
	rt	Répétition d'amorce de mot
Reparans RA		
Reparans : localisation de la reprise	nr	Pas de reprise
	wr	Reprise d'un mot
	dr	Reprise jusqu'au déterminant
	pr	Reprise jusqu'au début du syntagme
	or	Reprise plus importante
Reparans : type	co	Achèvement de l'item
	wc	Reprise du mot sans modification
	rp	Reprise au-delà du mot sans modification
	rc	Reprise du mot avec modification
	rm	Reprise avec de multiples changements

58 La fréquence de ces auto-interruptions (Pallaud, 2004) suggère qu'on peut les considérer comme un soutien à l'élocution (Jeanjean, 1984 ; Boula de Mareüil *et al.*, 2005 ; Pallaud, 2004) mais aussi à la compréhension de l'énoncé (effet de redondance et moment cognitif[6]). Ce rythme particulier à l'oral semble créer les conditions d'une interaction optimale dans la mesure où, en provoquant une réorganisation de l'énoncé et en intercalant des moments de silence ou d'interjections, il allège pour le récepteur la charge informationnelle de l'énoncé.

59 L'analyse des phénomènes de disfluence (interruptions et répétitions essentiellement) a porté sur des locuteurs standard n'ayant pas de troubles d'élocution. Une description fine des disfluences a été faite (Pallaud, 2005). Elle peut ainsi servir de document contrôle pour l'étude des ratages dans la prise de parole chez des sujets présentant des difficultés de langage. C'est ce qui a été fait sur les disfluences présentes dans des corpus de paroles chez des sujets bègues (Pallaud & Xuereb, 2008). Un des résultats de

cette étude concerne l'élocution standard versus le bégaiement : chez un locuteur standard, un mot tronqué involontairement n'est complété qu'après une reprise au moins du début du mot (parfois du déterminant également, voire du début du syntagme, Pallaud, 2005). En revanche, des personnes bègues peuvent compléter sans reprise un mot tronqué (Pallaud & Xuereb, 2008).

6. Dimensions interactionnelles et multimodales

60 Dans cette dernière partie, nous avons sélectionné quelques exemples d'annotation différents qui s'inscrivent cependant tous dans l'étude plus générale de l'interaction interindividuelle. Pour chaque annotation, nous décrivons les objectifs de recherche qui lui sont associés mais surtout la procédure spécifique de chacune qui, bien que spécifique, repose sur les mêmes principes (ancrage temporel des annotations, élaboration du schéma d'annotation, guide et campagne d'annotation, nombre et type d'annotateur expert/naïf). La première section concerne l'organisation thématique de la conversation, la seconde traite de l'annotation des sourires et du développement d'un outil pour en faciliter l'annotation, la troisième concerne les items multimodaux de feedbacks produits par celui qui ponctuellement est en position d'écoute, la quatrième concerne le phénomène de l'humour, et la dernière aborde l'annotation des gestes manuels.

6.1 L'organisation thématique des conversations

61 Dans le cadre d'une thèse sur les sourires dans les transitions thématiques (Amoyal M., *Rôle du sourire dans les transitions thématiques de conversations en français*), notre analyse concerne un niveau linguistique discursif/pragmatique à savoir l'organisation thématique d'interactions conversationnelles. Afin d'analyser ce niveau linguistique, il est indispensable de repérer les segments thématiques ainsi que les transitions qui les lient et de les annoter de manière systématique. Afin de réaliser cette annotation manuelle, nous nous sommes appuyés sur plusieurs méthodologies existantes (Jefferson, 1984 ; Riou, 2015 ; Crow, 1983 ; Mondada & Traverso, 2005), sachant qu'il n'existe pas de méthodologie universelle et consensuelle. Comme toutes annotations, il faut faire des choix de catégories, d'étiquetage et de délimitation temporelle de ces catégories. Ces choix sont cruciaux dans la mesure où ils guideront ensuite toutes les analyses. Pour analyser l'organisation thématique et les transitions qui composent les interactions conversationnelles, nous avons réalisé un protocole d'annotation[7] en plusieurs étapes sous le logiciel PRAAT (Boersma & Weenink, 2018) :

1. L'annotation des thèmes conversationnels,
2. L'annotation des transitions qui séparent ces thèmes,
3. Le découpage en phases des transitions,
4. Le contre-codage de l'ensemble de ces niveaux d'annotations.

62 **Étape 1 :** Le thème est une notion pertinente pour analyser et décrire la collaboration des interactants. Le thème conversationnel ou segment thématique peut être défini comme « un bloc d'échanges reliés par un fort degré de cohérence sémantique et/ou pragmatique » (Kerbrat-Orecchioni, 1990 : 218). L'annotation des thèmes nécessite au préalable la transcription du signal de parole. À l'appui du signal audio ainsi que de la transcription, l'annotateur détermine les segments thématiques co-développés par les

interlocuteurs. Un thème est défini par trois critères : il est le centre de l'attention partagée, il est spécifique aux participants et il est co-construit (Riou, 2015).

63 **Étape 2 :** Les transitions thématiques sont des portions d'interactions qui relient un segment thématique à un autre. Les transitions sont plus ou moins explicitement énoncées par les interlocuteurs, soit par un énoncé tel que (« je ne t'ai pas raconté l'interview » Extrait de AWCG dans Cheese) ou de manière moins explicite en abordant le prochain sujet de discussion sans « pre-act » (Crow, 1983). Selon la typologie de Crow (1983) qui a mené son étude sur des conversations écologiques entre couples, les transitions thématiques peuvent être décrites selon plusieurs catégories. Une transition peut être de type :

- Enchaînement : une transition qui connecte/relie le thème précédent au suivant,
- Initiation : la première occurrence d'un thème, sans lien avec le(s) précédent(s),
- Insertion : transition qui donne lieu à une digression,
- Renouveau : la transition convoque un thème déjà évoqué précédemment dans la conversation (au moins deux thèmes précédents).

64 Cette typologie a été utilisée pour labelliser l'ensemble des transitions qui séparent des segments thématiques sur les corpus Cheese ! [Voir la fiche technique[8], Priego-Valverde & al., 2020] et PACO [Voir la fiche technique[9], Amoyal & al., 2020].

65 **Étape 3 :** Une fois les transitions systématiquement annotées, nous annotons les 3 phases qui les composent :

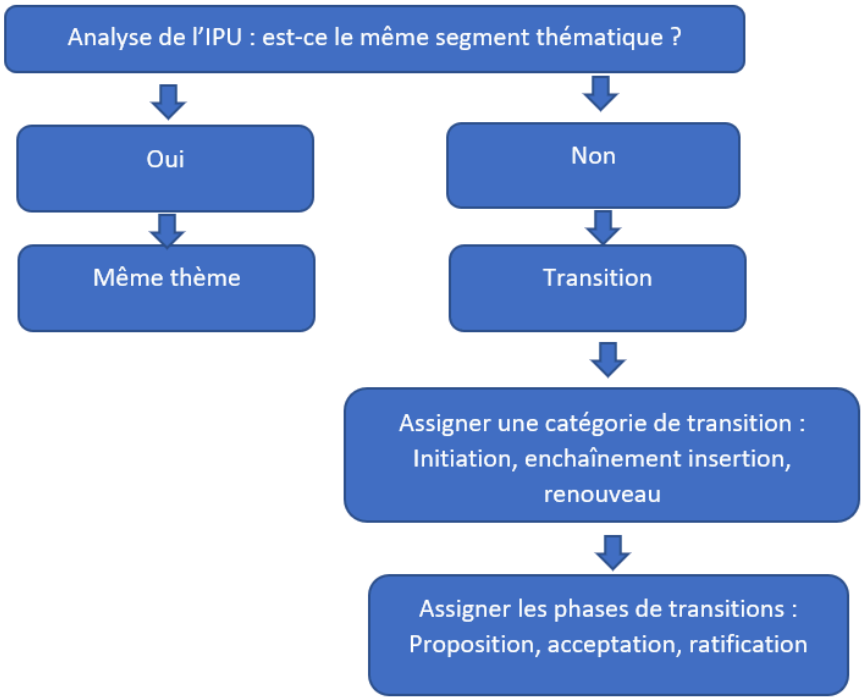
- La proposition : Le locuteur introduit le thème proposé,
- L'acceptation : L'interlocuteur accepte le thème proposé,
- La ratification : Le locuteur valide cette acceptation.

66 Ce sous-découpage en phases permet une analyse fine des processus et ressources linguistiques mobilisés par les interlocuteurs dans la co-construction des transitions thématiques.

67 **Etape 4 :** le contre-codage et l'évaluation de celui-ci pour validation de la fiabilité des annotations.

68 L'étape cruciale de l'annotation manuelle (quel que soit le niveau) est le contre-codage et l'évaluation de celui-ci. En effet, toute annotation manuelle nécessite un contre-codage pour en vérifier la fiabilité et l'objectivité. Ainsi, dans notre cas, le signal audio ainsi que la transcription des interactions ont été donnés au contre-codeur afin de réaliser les étapes d'annotation des thèmes, des transitions et des phases décrites plus haut. Il est donc nécessaire d'avoir bien établi un schéma ainsi qu'un protocole d'annotation pour pouvoir reproduire fidèlement les annotations. L'ensemble des annotations est ensuite soumis à une évaluation à l'appui d'un calcul d'accord inter-annotateur (l'indice kappa de Cohen, 1960). Le taux d'accord permettra de juger de la fiabilité des annotations et de leur validité. Une vigilance doit être apportée à ce résultat obtenu puisqu'un bon kappa n'indique pas forcément une extrême fiabilité des annotations (i.e. cela peut dépendre du nombre de juges, généralement 2 pour les gestes et au moins 3 pour les niveaux qui concernent le signal de parole), tout comme un mauvais kappa n'indique pas forcément de mauvaises annotations (i.e. cela peut être dû à de mauvaises instructions dans le guide d'annotation).

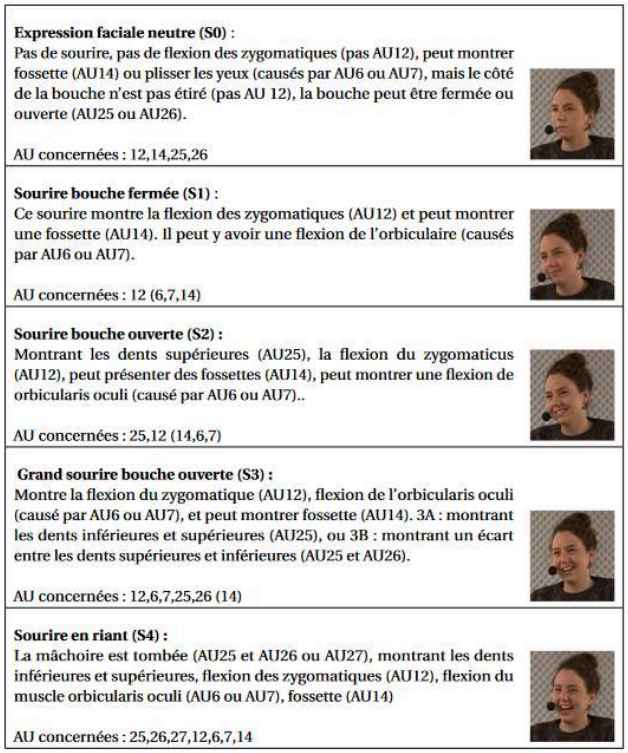
Figure 4 : Schéma d'annotation des transitions thématiques



6.2 Les sourires

69 Les interlocuteurs d'interactions conversationnelles utilisent plusieurs ressources multimodales [voir l'article *Multimodalité* dans ce numéro[10]]. Il est admis que la parole et les gestes forment « un tout intégré » (McNeil, 1992). Si la parole est étudiée à différents niveaux (voir les autres sections de cet article) les études sur la gestuelle se concentrent majoritairement sur les gestes des mains en sciences du langage (Kendon, 2004 ; Morgenstern *et al.*, 2010 ; Tellier, 2008). Qu'en est-il des expressions faciales ? Afin d'explorer cette question, il est intéressant d'étudier tout particulièrement une mimique : le sourire. Cette expression faciale est étudiée dans son rapport à ce qui est dit par les interlocuteurs. Puisque nous nous intéressons particulièrement au rôle du sourire dans l'organisation conversationnelle, nous souhaitons observer les fonctions du sourire dans la négociation des transitions thématiques. Comment les interlocuteurs sollicitent-ils leur sourire pour passer d'un thème conversationnel à un autre ? Afin de répondre à cette question, une annotation systématique des sourires des deux interlocuteurs est requise. Le sourire est une expression faciale complexe parce qu'il fait intervenir différents muscles du visage et graduelle puisqu'il existe différentes intensités du sourire. C'est pourquoi l'analyse ne saurait se contenter d'une annotation qui rend uniquement compte de sa présence ou de de son absence. Ainsi, pour décrire ce « geste facial » (Bavelas & Gerwing, 2007 ; Bavelas *et al.*, 2014) nous utilisons l'échelle du Smiling Intensity Scale (Gironzetti *et al.*, 2016) qui décompose le sourire en 5 niveaux (du visage neutre 0 au rire 4) en détaillant les différentes unités d'actions en cause dans chaque intensité (Ekman & Friesen, 1978). La figure ci-dessous détaille et illustre à l'appui d'images de notre corpus les différents niveaux de l'échelle du SIS.

Figure 5 : Description de chaque intensité de sourire selon le Smiling Intensity Scale (SIS) de Gironzetti *et al.* (2016)

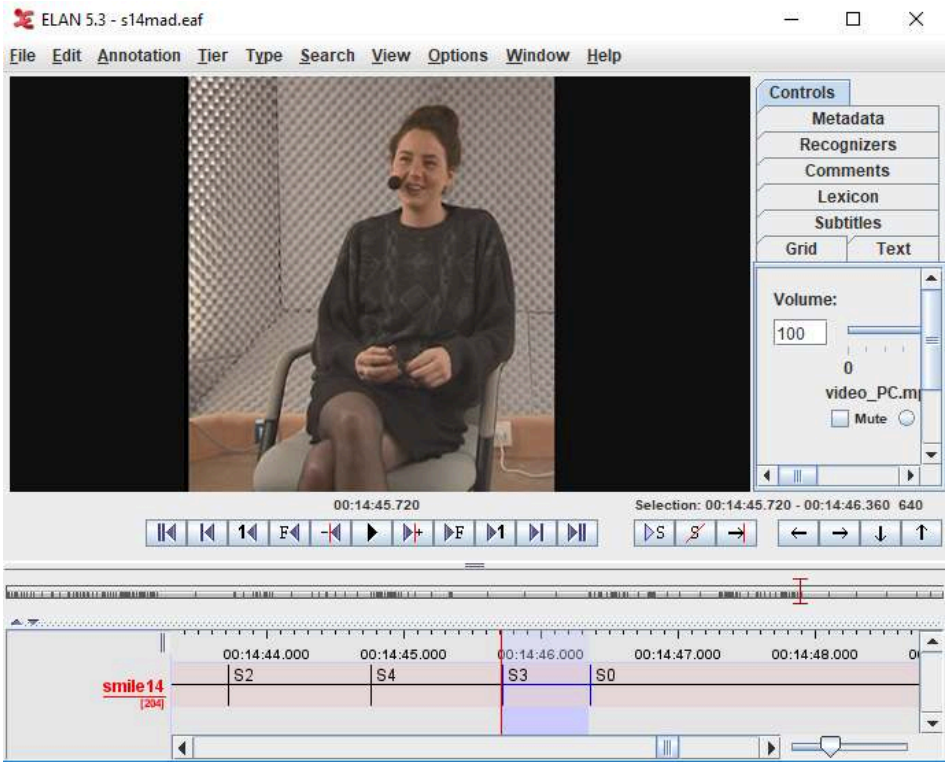


70 L'annotation des sourires à l'appui de cette échelle peut se réaliser selon plusieurs méthodes. La première méthode utilisée s'appuie sur une annotation manuelle des sourires sous le logiciel Elan (Brugman & Russel, 2004). Cette annotation perceptive a soulevé plusieurs problèmes. D'une part il parait difficile de poser une borne de début et de fin de sourire, d'autre part la comparaison des annotations par plusieurs codeurs s'est révélée difficile à réaliser du fait des bornes non stables. Nous avons donc décidé de découper en intervalles réguliers les corpus et d'attribuer ensuite une catégorie de sourire.

71 L'annotation manuelle de l'intensité des sourires selon l'échelle SIS de Gironzetti *et al.* (2016) est chronophage (i.e. 60 heures en moyenne par un codeur expérimenté, pour 1 heure de vidéo). Comme c'est le cas pour la majorité des annotations concernant la gestualité, ce coût en temps limite en pratique la taille des corpus annotés et par conséquent le nombre de phénomènes linguistiques accessibles à l'analyse quantitative. Pour atteindre notre objectif (i.e. l'analyse et donc l'annotation des sourires des 8 h de conversations qui composent nos corpus), nous avons proposé une méthode alternative.

72 Pour cela, nous avons conçu et développé un outil pour annoter automatiquement une vidéo selon l'échelle SIS. Le logiciel SMAD[11] (voir la fiche technique HMAD[12]) propose en sortie une séquence d'intervalles temporels adjacents portant chacun une des 5 étiquettes de l'échelle SIS (i.e. de l'expression neutre du visage S0 au rire S4). Cette annotation automatique se décline sous différents formats, notamment le format. eaf qui permet d'éditer et de corriger la sortie de SMAD avec le logiciel Elan (Brugman & Russel, 2004). Une illustration est présentée figure 6.

Figure 6 : Exemple d'une sortie du logiciel d'annotation automatique des sourires SMAD au format eaf éditable sous Elan.



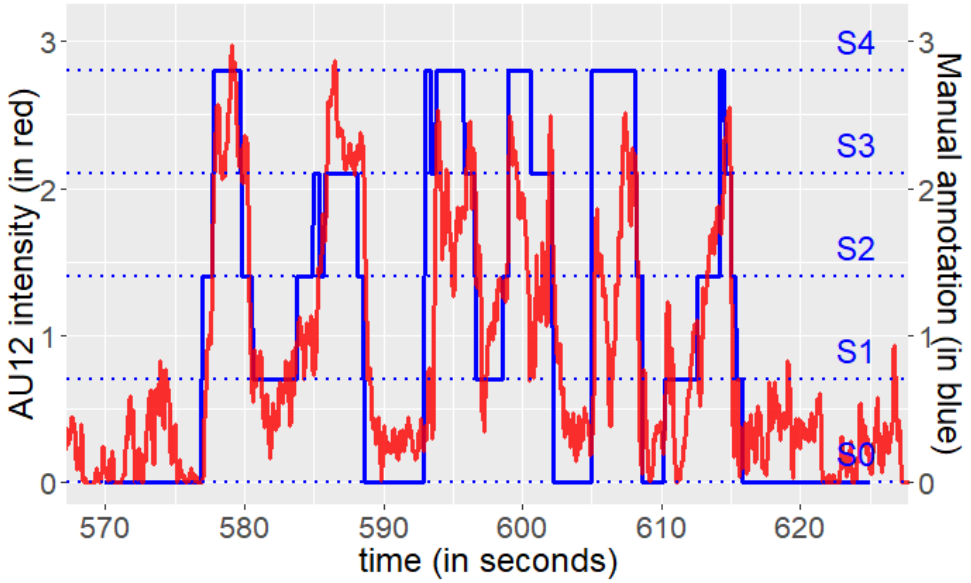
73 L'annotation automatique est obtenue en deux temps. Dans un premier temps la vidéo est traitée par le logiciel OpenFace (Baltrusaitis *et al.*, 2018) qui réalise la détection et le suivi des mouvements de tête du participant filmé. OpenFace mesure également les mouvements intra-faciaux comme la variation temporelle de la position de certains points de référence du visage (i.e. les landmarks) ainsi que la variation au cours du temps des unités d'action, nommées les Action Units (AU) du système FACS qui caractérisent l'état de certains muscles faciaux (Ekman & Friesen, 1978). La figure 7 présente la capture d'une image de la vidéo traitée par Open Face.

Figure 7 : La capture d'une image de la vidéo traitée par OpenFace montrant la position des landmarks et l'orientation spatiale de la tête (la projection du cube en bleu)

74 Dans un deuxième temps, la sortie d'OpenFace (notamment la mesure de la variation des intensités des AUs) est traitée par le modèle de sourires implanté dans SMAD et fournit une annotation automatique selon l'échelle SIS. Le modèle statistique sous-

jacent fait partie de la famille des HMMs. Il a été entraîné par apprentissage automatique sur un corpus d'environ 1 heure de vidéo annoté manuellement et contre-codée selon l'échelle SIS (4 extraits du corpus CHEESE !, 2 paires de participants en conversation pendant 15 minutes environ, annotés et contre-codés, correspondant à 1268 intervalles étiquetés de S0 à S4). Le modèle statistique capture les correspondances observées entre les mesures des intensités des AUs pour chacun des 5 niveaux de sourire de l'échelle SIS. La figure 8 illustre par exemple la corrélation temporelle entre la mesure OpenFace de l'intensité de l'AU12 (étirement des zygomatiques) et l'annotation manuelle selon l'échelle SIS. Le modèle implanté dans SMAD est dynamique : la prédiction du niveau de sourire à l'instant t dépend en effet de la mesure des AUs à cet instant et des états antérieurs du système.

Figure 8 : Sur une durée de 60 secondes, variation temporelle de l'intensité de l'AU12 telle que mesurée par OpenFace (en rouge) versus l'annotation manuelle des niveaux de sourire selon l'échelle SIS (en bleu)



L'AU12 est associée à la contraction du muscle zygomatique majeur.

75 Nous appliquons ensuite une procédure de correction manuelle sur ces sorties automatiques. Elle sera effectuée sous Elan (Brugman & Russel, 2004) en vérifiant à l'appui de la vidéo chaque intervalle de sourire. La correction concerne d'une part les labels des sourires annotés automatiquement et d'autre part éventuellement les bornes de ceux-ci. Concernant les labels, l'outil SMAD[13] annote avec un X tous les intervalles dont il a perdu le signal (soit parce que le locuteur bouge et qu'il n'est plus dans le champ de la caméra, soit parce que l'intervalle de confiance du logiciel indique qu'il n'a pas tous les critères pour indiquer une annotation fiable). Ces intervalles peuvent être laissés comme tel (X) lorsque l'annotateur ne voit effectivement pas le sourire du participant, ou étiqueté avec une intensité de sourire si elle est visible. Les autres intensités annotées peuvent également être erronées, il faudra alors corriger le label avec l'intensité de sourire perçue. Concernant les bornes, le logiciel peut indiquer qu'un sourire commence ou se termine avant ou après sa réalisation effective. Dans ce cas, il suffira de déplacer la borne pour que l'annotation corresponde au déploiement du sourire de la première contraction des muscles en cause au relâchement de ceux-ci.

Le visionnage image par image que permet Elan s'avère ici extrêmement utile afin d'être précis dans l'annotation.

76 L'évaluation de notre outil automatique peut être réalisée en appliquant la technique de validation croisée à notre échantillon d'apprentissage. Sur une comparaison image à image, nous obtenons un taux de bonnes prédictions (i.e. accuracy factor) de 68 % sur les 5 niveaux de l'échelle. Nous avons aussi évalué la performance de SMAD en comparant le temps passé à corriger manuellement les sorties de SMAD avec le temps mis pour annoter de façon entièrement manuelle les mêmes données. Les résultats obtenus sur environ 1 heure de corpus (4 extraits du corpus CHEESE ! différents de notre corpus de référence) mettent en évidence un gain de temps d'un facteur 10 avec le recours à SMAD suivi d'une correction manuelle par rapport à une annotation entièrement manuelle.

77 Nous avons réalisé cette procédure semi-automatique d'annotation des sourires sur les deux corpus étudiés : Cheese! (Priego-Valverde *et al.*, 2020) et Paco (Amoyal *et al.*, 2020) qui représentent en tout 8 h de conversations. La correction des sourires a été réalisée par deux codeurs entraînés à l'échelle du SIS. Le calcul de l'accord inter annotateur révèle un Kappa (Cohen, 1960) de 0.63 ce qui est très satisfaisant au regard des accords obtenus dans les études en gestuelle (Tellier *et al.*, 2012 ; Tellier, 2014). Cela nous permet de valider la relative objectivité des catégories proposées pour le découpage des sourires ainsi que la fiabilité des annotations effectuées. Ces annotations nous ont permis de quantifier le sourire dans la conversation : cette expression faciale est présente dans 33 % de nos corpus, contre 56 % de visage neutre et 11 % de rire (Amoyal, Priego-Valverde et Prévot, 2021).

78 La section suivante concerne les items de feedbacks, et parmi eux certains sourires, qui sont produits par l'interlocuteur en situation d'écoute pour montrer son écoute et sa participation au discours en cours.

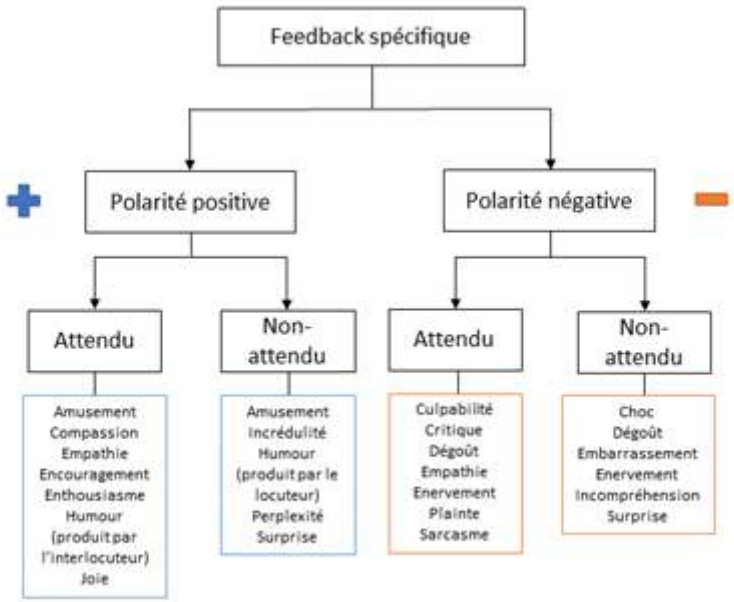
6.3 Les Feedbacks

79 L'interaction est une activité dynamique, élaborée par la collaboration entre le locuteur et l'interlocuteur. Si ces rôles changent activement au cours d'un échange, il est nécessaire pour le locuteur, quel qu'il soit, de recevoir des signaux d'écoute -appelés feedbacks (FB)- de la part de l'interlocuteur. Plus complexes que de simples réponses d'écoute, les feedbacks occupent différentes fonctions comme celle de montrer sa compréhension ou des réactions plus spécifiques étroitement liées au contexte sémantique et pragmatique. La notion de feedback a été introduite par Schegloff (1982) sous le terme de backchannel, basé sur les travaux de Ygnve (1970) et Duncan & Fiske (1977). Le rôle crucial des feedbacks a été mis en avant dans plusieurs modèles linguistiques et/ou psycholinguistiques du dialogue (Clark, 1996 ; Horton, 2017 ; Bavelas & al 2000 ; Pickering & Garrod, 2013), notamment via leur rôle dans l'établissement d'un savoir partagé et d'un alignement entre les participants. Depuis les travaux de Bavelas (2000), on distingue les FB génériques et spécifiques (proches des *continuer/ assessment* de Schegloff 2000). Nous adoptons dans notre thèse (A. Boudin, *Les feedbacks en conversation : modélisation et implantation*) cette distinction : les FB génériques sont essentiellement composés de courtes interjections (e.g. mhmh, d'accord) et/ou de hochements de tête. Les FB spécifiques, quant à eux, peuvent renvoyer à des réactions variées (évaluation, amusement, joie, horreur, perplexité, etc.). Ils peuvent être

produits dans des formes lexicalisées plus ou moins longues (e.g. « *non c'est pas vrai* »), mais aussi renvoyer à des mouvements de mains, de tête ou encore des mimiques faciales. Tous ces éléments pouvant être combinés, ce qui les rend plus variables et plus complexes à repérer que les FB génériques.

80 Pour l'annotation des FB, nous avons donc adopté la typologie générique/spécifique à laquelle nous avons ajouté deux sous-niveaux pour les spécifiques, qui, comme vu précédemment, peuvent avoir des fonctions et réalisations très variées. Les deux niveaux hiérarchiques ajoutés aux FB spécifiques et définis ci-dessous sont la polarité (positive ou négative) et le caractère attendu ou inattendu de l'information à laquelle le FB répond. Nous avons donc cinq catégories possibles : générique, positif-attendu, positif-inattendu, négatif-attendu, négatif-inattendu. Ainsi, comme le montre la figure 9, notre schéma d'annotation couvre un large ensemble de comportements avec cinq catégories.

Figure 9 : Classification des feedbacks spécifiques



Générique *VS* Spécifique

81 Les FB génériques sont des interventions « discrètes » de l'interlocuteur, généralement, ils prennent la forme d'une interjection, d'un hochement de tête, d'un sourire ou d'une combinaison/répétition de ces éléments (Stivers, 2008 ; Bertrand *et al.*, 2017). L'intonation associée est monotone, et le débit faible ou moyen. Les feedbacks spécifiques se distinguent par leur dépendance au contenu sémantique et peuvent prendre différentes formes dans leurs réalisations (intonation plus ou moins marquée, intensité, gestuelle, lexicalisation...), tout en n'apportant pas réellement d'information nouvelle dans l'interaction.

Positif *VS* Négatif

82 Les feedbacks spécifiques sont évalués en fonction de la polarité véhiculée dans le discours du locuteur. Les questions posées dans le guide d'annotation sont les suivantes :

> Q1 : Est-ce que le feedback permet de réagir à quelque chose de positif ou quelque chose de négatif ?
> Q2 : Le locuteur parle-t-il d'une expérience ou d'un événement qu'il évalue négativement en exprimant une critique, de la tristesse, de la colère, etc. ? Ou au contraire, son évaluation est-elle positive en exprimant de la joie, de l'humour, de l'enthousiasme, etc. ?

83 Nous faisons l'hypothèse que si les deux participants sont pleinement engagés dans l'interaction, cette polarité sera partagée et influencera le choix du feedback produit.

Attendu *VS* Inattendu

84 Le second niveau de classification concerne le caractère attendu ou inattendu de l'information pour l'interlocuteur (Prévot *et al.*, 2008). Dans le cas d'un feedback « attendu », le locuteur connaissait déjà l'information ou s'attendait à recevoir cette information, elle ne le surprend pas. Dans le cas d'un feedback « inattendu », l'interlocuteur prend connaissance d'une information qu'il n'avait pas, comprend quelque chose après une clarification ou demande une clarification. Un feedback inattendu peut également répondre à une production du locuteur non prévisible.

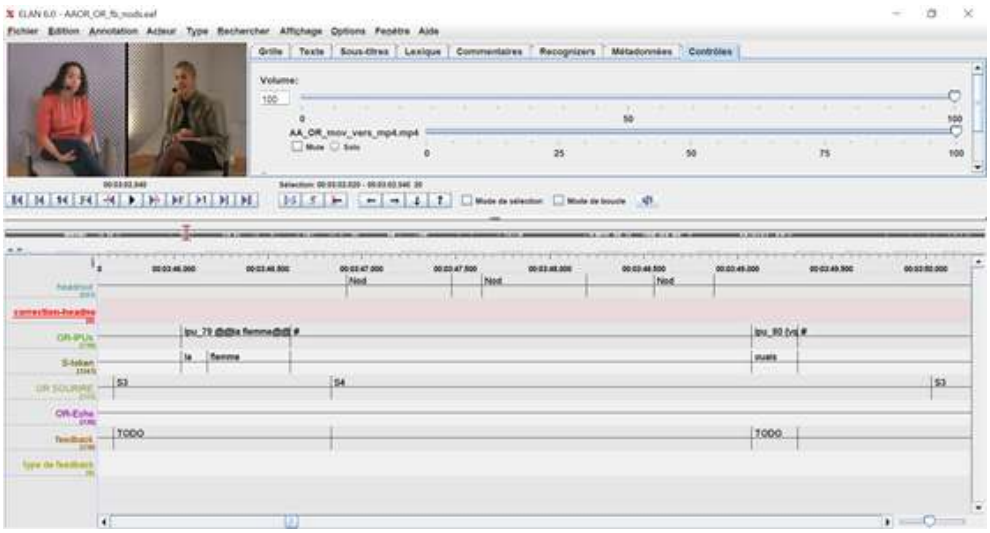
85 Les questions posées sont les suivantes :

> Q1 : Est-ce que l'interlocuteur s'attendait à recevoir cette information à ce moment-là ?
> Q2 : L'interlocuteur avait-il déjà connaissance de l'information à laquelle il réagit ?
> Q3 : La production de cette information par le locuteur à ce moment-là de l'interaction était-elle logique ou attendue ?

86 Les feedbacks ont été annotés manuellement sur Paco [Amoyal & al., 2020] et Cheese ! [Priego-Valverde & al., 2020] par trois annotateurs (actuellement, 13 dyades ont été annotées pour un total de 2380 feedbacks) avec le logiciel ELAN.

87 L'annotation manuelle étant une tâche chronophage, nous avons tenté de faciliter la tâche des annotateurs avec un niveau de pré-détection. Les éléments fréquemment utilisés pour produire les feedbacks (rires, sourires, répétitions, interjections) ont été sélectionnés automatiquement, sur la base des annotations existantes, par un script python et ajoutés sur une nouvelle tire d'annotation avec la mention « TODO ». La première étape pour les annotateurs était de changer le libellé « TODO » en « TRUE » si un feedback est bien réalisé ou en « FALSE » si l'élément ne correspond pas à un feedback. Le kappa obtenu pour cette tâche est de 0.78, ce qui correspond à un « accord fort ».

Figure 10 : Exemple d'une sortie de pré-détection des feedbacks, avant correction/annotation



88 La deuxième étape de l'annotation concerne le type de feedback réalisé. Lorsqu'un feedback a bien été produit, les annotateurs doivent indiquer le type de feedback produit sur la tier « type de feedback ». Cinq catégories sont donc possibles (générique ; positif-attendu ; positif-inattendu ; négatif-attendu ; négatif-inattendu). Lors de cette étape, les annotateurs doivent aussi ajouter les feedbacks non pré-détectés, essentiellement des hochements de tête (annotés en parallèle par les mêmes annotateurs) et du contenu lexical. Le kappa obtenu pour cette tâche est de 0,57 correspondant à un « accord modéré ».

Figure 11 : Exemple d'annotation d'un feedback négatif-inattendu

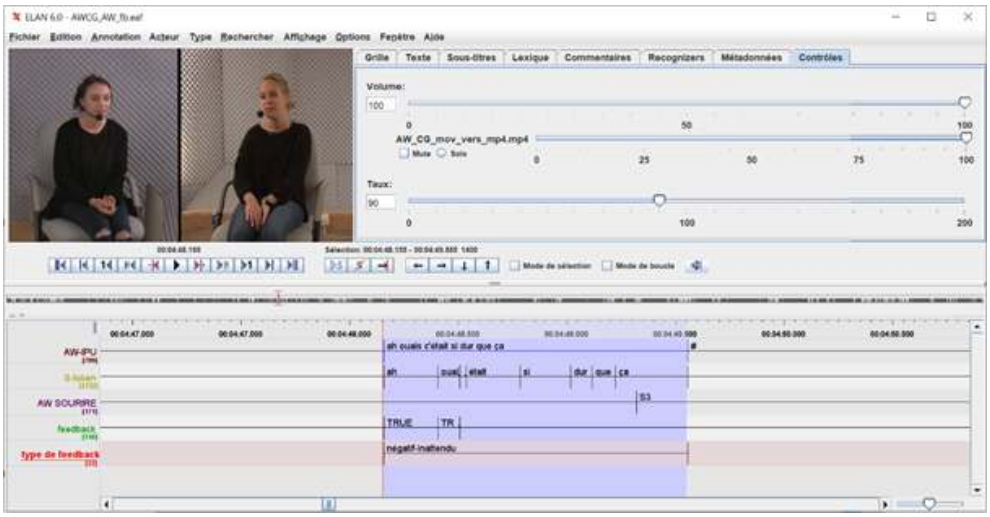
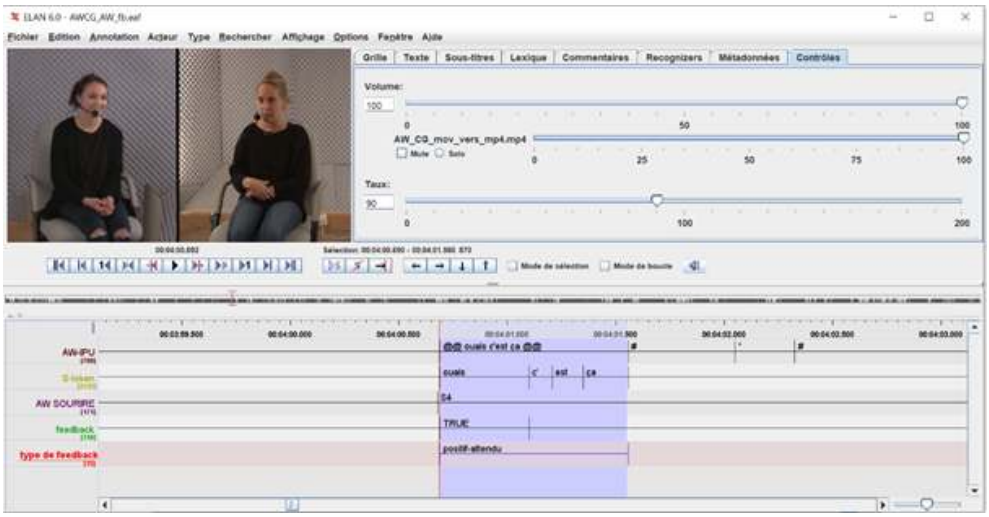


Figure 12 : Exemple d'annotation d'un feedback positif-attendu



89 Le guide d'annotation a été détaillé et illustré d'exemples afin que les annotateurs puissent comprendre la distinction entre les différents types de feedbacks. Les annotateurs pouvaient également se référer à un ensemble de questions pour les cas ambigus. Afin de limiter les ambiguïtés lors de l'annotation, un ensemble de cas ont été décrits comme n'étant pas des feedbacks :

- Lorsque l'interlocuteur tente de prendre le tour de parole, il commence une phrase mais ne peut pas la produire entièrement car le locuteur garde le tour de parole/lui coupe la parole. L'interlocuteur semble vouloir réagir en apportant un élément nouveau, en donnant son avis, mais on n'a pas réellement accès à l'intention de son message. Exemples : « *mais du coup...* » ; « *et la dernière fois...* » ; « *ouais et...* ».
- Les interjections et les pauses pleines montrant une hésitation, par exemple lorsque le locuteur ou l'interlocuteur cherche ses mots, réfléchit à la formulation de son idée, tic de langage, etc.
- Les réponses à des questions. Parfois l'interlocuteur peut produire un « *oui* », il n'y a pas nécessairement de changement de tour de parole mais l'interlocuteur répond à une demande explicite du locuteur. Exemple :
  Locuteur : « *tu as une sœur toi non ?* »
  Interlocuteur : « *oui* »
- Les demandes de clarification de plus de 4 tokens, par exemple un « *ah ouais ?* » ou « *quoi ?* » sont considérée comme un feedback (inattendu), mais pas les exemples suivants :
  « *oui puis j'allais dire c'est quand ton anniversaire ?* »
  « *celle qui se sacrifie là ?* »
- Tour de parole de moins de 10 tokens. Il faut distinguer les tours de parole courts, avec une alternance rapide entre les deux locuteurs, des productions qui correspondent à du feedback.
- Les éléments de feedback qui sont suivis par un tour de parole (plus de 10 tokens) sans pause (noté # ou d'une durée minimale de 130 ms) entre les deux. Exemple : « *ouais et tu sais les bruits aussi parce que comme c'était des enfants* ».

90 Pour plus de détails, nous reportons le lecteur à l'article qui présente les premiers résultats concernant les indices de prédiction des feedbacks (Boudin *et al.*, 2021).

91 La section suivante est consacrée aux difficultés que pose l'annotation de l'humour.

6.4 L'humour

92 L'humour, lorsqu'il apparaît dans la conversation, est un phénomène non seulement protéiforme, mais également subjectif, tant pour les participants que pour les annotateurs (Priego-Valverde, 2003). À partir de là, un problème épineux apparaît : est-ce qu'une occurrence doit être annotée comme humoristique si elle est perçue comme telle par les participants mais ne fait pas rire les annotateurs ? Inversement, est-ce qu'une occurrence doit être annotée comme humoristique si elle est perçue comme telle par les annotateurs mais non par les participants ? Tout l'enjeu de l'annotation est alors de dépasser, autant que possible, cette subjectivité en dégageant des critères solides d'identification.

93 Analyser l'humour dans une conversation ne se résume pas à analyser les procédés linguistiques de l'humour (comme les figures de style employées) qui, par ailleurs, peuvent être mobilisés quel que soit le cadre dans lequel il apparaît (cadre professionnel, thérapeutique… ou même un sketch). En revanche, analyser l'humour dans une conversation nécessite de prendre en compte à la fois le fonctionnement de l'humour lui-même, et le fonctionnement de la conversation en tant qu'activité hautement ordonnée (Sacks, Schegloff & Jefferson, 1974) pour montrer à quel point elle peut façonner à la fois la forme de l'humour qui y apparaît, et sa trajectoire interactionnelle.

94 Ainsi, l'annotation de l'humour – nécessairement manuelle – peut s'élaborer en trois étapes. La première est une annotation du contexte dans lequel l'humour apparaît. Quatre niveaux peuvent être dégagés : un niveau séquentiel (ce qui a été dit avant l'occurrence humoristique), un niveau plus large qui permet d'identifier l'activité conversationnelle dans laquelle les participants sont engagés (narration, explication, argumentation…), un niveau qui permet de déterminer le registre de communication (le « frame » (Bateson, 1953), sérieux ou déjà humoristique) dans lequel l'occurrence annotée apparaît. Un quatrième niveau peut aussi avoir son importance à condition qu'il soit disponible pour les annotateurs : celui, plus macro, qui réfère à l'histoire conversationnelle des participants.

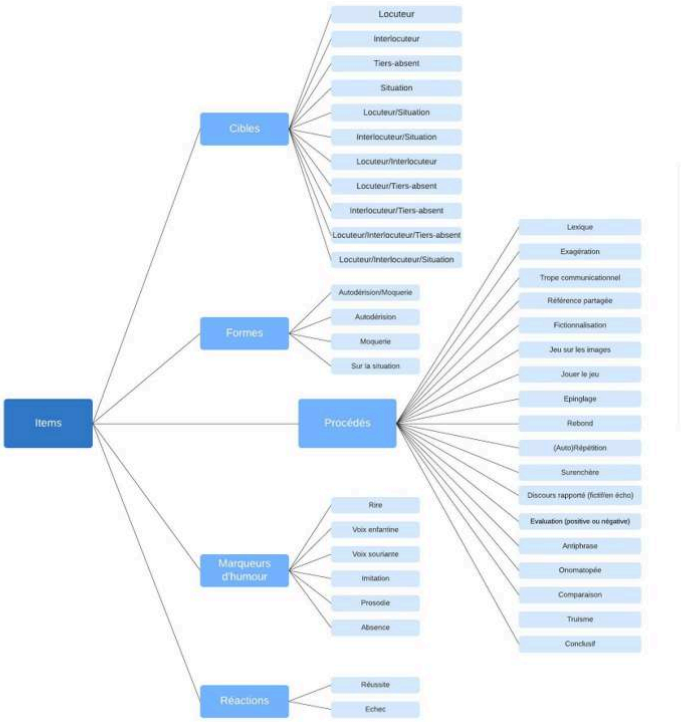
95 La deuxième étape consiste à annoter les procédés et/ou marqueurs de l'humour comme les jeux sur différents niveaux linguistiques (prosodique, lexical, syntaxique…), ou discursifs (discours rapportés, discours référant à un savoir partagé…) (Haugh, 2014 ; Priego-Valverde, 2016 ; Mullan & Béal, 2018 ; Attardo, 2000).

96 Enfin, la troisième étape d'annotation s'effectue en observant la ou les réactions à l'occurrence humoristique (Drew, 1987 ; Hay, 2001 ; Attardo, 2008 ; Bell, 2009a ; 2009b ; Priego-Valverde, 2009). Deux types de réactions sont alors annotées : les réactions positives (rires, commentaires évaluatifs, surenchère…), et les réactions négatives (ignorer l'humour, y répondre de manière sérieuse, ou le rejeter de manière explicite). Cette annotation permet de catégoriser l'humour en échec ou réussite.

97 Comme pour tout autre phénomène analysé, le schéma d'annotation de l'humour doit être adapté à la question de recherche. Les trois étapes complémentaires présentées ci-dessus peuvent être plus ou moins développées selon l'objectif de l'étude. Ainsi, dans une étude mettant l'accent sur le rôle du (dés)alignement et / ou de la (dés)affiliation des participants dans l'échec de l'humour (Priego-Valverde, à paraître[14]), le schéma d'annotation permettait de mettre en avant le contexte séquentiel dans lequel l'humour apparaissait : nature du frame et de l'activité dans laquelle l'humour est

inséré, type de réactions obtenues (positives ou négatives). En revanche, sur la base d'une comparaison des deux corpus « Cheese! » et « Paco », deux autres études sont en cours. La première porte sur la manière dont l'humour est produit selon que les participants se connaissent ou non[15]. Le schéma d'annotation qui a été élaboré met donc davantage l'accent sur les procédés et marqueurs de l'humour. Ce schéma se présente comme suit :

Figure 13 : Schéma d'annotation du corpus « Cheese! »



98 A ce jour, ces annotations portent sur les 11 interactions du corpus « Cheese! », ce qui représente 670 items humoristiques. L'ensemble de ces annotations ont été effectuées par une seule annotatrice.

99 Sur la base de ces annotations, la seconde étude (en cours elle aussi) porte sur le rôle du sourire dans la production et la perception de l'humour. L'accent est mis ici, d'une part sur les rôles interactionnels que les participants occupent (locuteur/interlocuteur) lors de phases humoristiques, et d'autre part, sur la manière que ces participants ont (ou non) de construire conjointement ces phases. Pour cela, une étape d'annotation s'est avérée nécessaire pour distinguer les items produits isolément de ceux produits « en cascade ». Nous avons donc regroupé en « segment » les items humoristiques qui étaient séparés de moins de 5 secondes[16]. Chaque segment humoristique a fait l'objet d'une double annotation permettant le calcul de l'accord inter-annotateur (de 0.22 à 0.94 selon les interactions). Le corpus « Cheese! » compte 339 segments humoristiques. Ce sont sur ces unités que porte l'analyse des sourires.

6.5 Annoter les gestes

100 Annoter les gestes manuels coverbaux (McNeill, 1992) pose les mêmes problématiques que les autres annotations évoquées précédemment, à savoir : la segmentation (des gestes), la création d'un schéma et d'un guide d'annotation détaillé et la création de catégories en fonction des questions de recherche. Les gestes dont il est question ici sont ceux qui accompagnent spontanément la parole et forment un seul et même système avec elle (McNeill, 1992). Dans la plupart des études sur la gestuelle menée au LPL, on écartera donc les gestes dits extra-communicatifs (Cosnier, 1982) c'est-à-dire les gestes nerveux ou d'auto-contact (bien que certaines études se soient penchées sur ces phénomènes y compris dans notre laboratoire).

101 Tout d'abord, la question de la segmentation du geste se traite en regardant les vidéos du corpus image par image. On annote un geste depuis la préparation de celui-ci (les mains commencent à bouger, l'image devient floue) jusqu'à la rétraction du geste (retour des mains en position de repos ou immobilité) ou l'enchaînement sur un autre geste. Dans ce second cas, identifier le point de bascule lorsque les mains entament un nouveau geste sans passer par la position de repos peut s'avérer complexe (voir Tellier *et al.*, 2012). Examiner le geste image par image permet cependant de repérer le changement d'orientation des mains, témoin d'une nouvelle unité gestuelle. À noter que lorsque l'on travaille sur des corpus écologiques avec des locuteurs qui se déplacent (comme dans des films de classe), la visibilité de certains gestes peut être occultée et nécessite une reconstitution interprétative (Azaoui 2014a).

102 Ensuite, se pose la question de la constitution du schéma d'annotation et des catégories qui le composent.

103 Il n'existe pas de schéma d'annotation universel pour la gestuelle (Calbris, 2011) et il convient donc de définir des catégories en fonction de ce que l'on cherche à explorer. Nous citerons ici différents exemples à partir de recherches effectuées au LPL pour indiquer ce qui a été annoté et comment cela répondait à des questions de recherche.

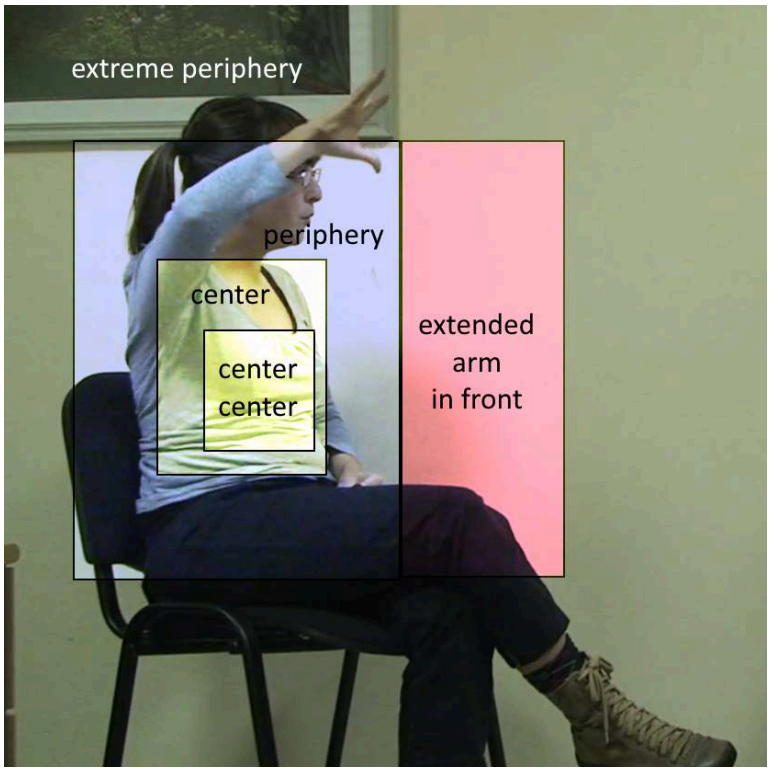
104 Un premier exemple de type d'annotation concerne la manualité *(handedness*). On peut vouloir identifier si le locuteur gestualise avec une main et laquelle ou avec les deux mains de manière symétrique (en parallèle) ou asymétrique (les deux mains font des mouvements différents). Dans ce cas, on distinguera 4 étiquettes dans le schéma d'annotation : main gauche, main droite, deux mains symétriques, deux mains asymétriques. Ce schéma a été utilisé pour étudier les réponses faites par un député à l'Assemblée Nationale alors qu'il était interrompu pendant son discours (Bigi *et al.*, 2012). Cette annotation nous a permis de constater que le député utilisait des gestes de la main gauche pour illustrer son discours préparé et que, lorsqu'il était interrompu, il maintenait sa main gauche suspendue et utilisait sa main droite pour répondre aux interruptions d'autres membres de l'hémicycle.

105 Un deuxième type d'annotation que l'on retrouve dans de nombreuses études au LPL est basé sur les dimensions élaborées par David McNeill (1992, 2005) pour les gestes coverbaux, à savoir iconique, déictique, métaphorique et battement. Cette catégorisation a souvent été enrichie avec d'autres types comme les gestes emblèmes (gestes culturels), les gestes avortés (abandonnés), les Butterworth (gestes de recherche lexicale) ou encore les interactifs de Bavelas *et al.* (1995). À titre d'exemple, on peut citer les études de Azaoui (2014b), Bigi *et al.* (2012), Holt *et al.* (2015), Tan *et al.* (2010) ou encore Tellier *et al.* (2021).

106 Pour développer un exemple, citons Tellier *et al.* (2021) qui avaient pour objectif de déterminer si des futurs enseignants de Français Langue Étrangère utilisaient des gestes plus illustratifs en fonction du niveau de compétence en langue de leur interlocuteur (natif vs apprenant). Il était donc pertinent d'annoter les dimensions gestuelles et de distinguer des gestes à fort degré d'iconicité (comme les iconiques), de geste à faible degré comme les battements [voir fiche corpus GTT[17]].

107 Dans cette même étude, l'espace gestuel a également été annoté afin de déterminer si les futurs enseignants produisaient des gestes plus grands avec des non natifs. Le schéma d'annotation de l'espace gestuel de McNeill (1992) a ainsi été adapté pour ce corpus (voir Figure 14). Ce genre d'annotation sur la taille du geste a également été utilisée par Bouget et Tellier (2015) pour déterminer si des aidants utilisaient des gestes plus grands lorsqu'ils s'adressaient à des personnes âgées.

Figure 14 : Schéma d'annotation gestuelle



Tellier, Stam et Ghio (2021)

108 Un troisième type d'annotation gestuelle est l'annotation fonctionnelle qui consiste à déterminer ce que le geste permet de faire dans l'interaction. Par exemple, pour étudier le geste pédagogique (produit par des enseignants), on peut analyser la fonction pédagogique du geste en se basant sur la typologie de Tellier (2008) à savoir informer, animer et évaluer. Cela permet d'analyser la façon dont le geste sert l'action pédagogique. Cette typologie a par exemple été réemployée dans Azaoui (2014b) ou encore Tellier (2016). Une autre typologie fonctionnelle a été élaborée par Tellier *et al.* (2013) afin de déterminer et comparer les fonctions des gestes produits dans les pauses du discours dans des interactions natif/natif vs natif/non natif. Il nous semblait en effet que dans les interactions exolingues (natif/non natif), les pauses étaient fréquentes

ainsi que les gestes produits dans ces moments de silence. Il nous a donc fallu créer un schéma d'annotation *ad hoc* pour étudier ce phénomène et son rôle dans l'interaction.

109 Nous n'avons indiqué ici que quelques exemples afin de mettre au jour la diversité des annotations possibles en gestuelle et de montrer combien il est nécessaire de les relier aux questions de recherche. En guise de conclusion, il est nécessaire de rappeler qu'afin de réduire la subjectivité de ces annotations qui reposent parfois sur de l'interprétation, on peut mettre en place un certain nombre de filtres méthodologiques tels que la rédaction d'un guide d'annotation précis et le contre-codage des données en aveugle par plusieurs annotateurs (voir Tellier, 2014 pour des conseils méthodologiques sur ces aspects).

7. Conclusion et perspectives

110 Si les premiers corpus étaient essentiellement constitués d'une langue très prototypique et dépourvue de contexte (principalement écrits), la nature des corpus s'étant modifiée, l'essor des recherches sur le langage en interaction a entraîné une vision plus complexe des éléments linguistiques à appréhender et, surtout, de la multiplicité des niveaux d'analyse du langage oral.

111 L'objectif de cet article est de présenter les principes généraux de constitution et d'enrichissement de corpus multimodaux/conversationnels. Nous avons choisi de présenter ici différents types d'annotations possibles aux niveaux phonétique, morpho-syntaxique, discursif, interactionnel, posturo-mimo-gestuel qui font appel, malgré leur diversité, à un certain nombre d'éléments méthodologiques communs qui caractérisent les travaux sur corpus au LPL.

112 Parmi ces éléments méthodologiques, l'élaboration de schémas d'annotations, commune à tous les niveaux, dépend des questions de recherche et se construit en se fondant sur la littérature, les modèles existants, et/ou des observations préliminaires du corpus. Par ailleurs, un schéma d'annotation, mis à l'épreuve du corpus, doit être souvent modifié et adapté. Ce processus montre ainsi que l'annotation est un préalable nécessaire à l'exploitation ultérieure des données, mais qu'elle est elle-même le résultat d'un processus d'analyse. Enfin, la rédaction d'un guide d'annotation et le contre-codage à plusieurs mains permettent de valider le schéma d'annotation élaboré, et sont nécessaires pour réduire la subjectivité des annotations. Cette procédure de constitution de schéma et de guide d'annotation est commune à tous les niveaux linguistiques et quel que ce soit le mode d'annotation (manuel, automatique, semi-automatique).

113 Enfin, l'ensemble des données annotées dans un même formalisme permet leur mise en relation ultérieure et favorise ainsi l'analyse multi-niveaux d'un phénomène donné.

114 Les outils logiciels pour l'annotation que nous avons présentés dans cet article sont déposés sous des licences libres et sont facilement accessibles ; ils sont décrits plus en détails dans des fiches techniques de ce numéro. Les corpus et annotations mentionnés dans cet article sont également accessibles *via* la forge « ortolang » et font l'objet de fiches corpus de ce numéro.

BIBLIOGRAPHIE

Amoyal, M., Priego-Valverde, B., Prévot, L. (2021) Changer de sourire pour changer de thème conversationnel, In *ISGS France*. ⟨hal-03299742⟩.

Amoyal, M., Priego-Valverde, B., & Rauzy, S. (2020) PACO: A corpus to analyze the impact of common ground in spontaneous face-to-face interaction, *Language Resources and Evaluation Conference*, p. 621-626.

Attardo, S. (2000) Irony markers and functions: Towards a goal-oriented theory of irony and its processing, *Rask*, 12, 1, p. 3-20.

Attardo, S. (2008) Semantics and Pragmatics of Humor, *Language and Linguistics Compass*, 2, 6, p. 1203-1215.

Azaoui, B. (2014a) Segmenter les gestes à partir d'un film de classe, In M. Tellier & L. Cadet (Eds.) *Le corps et la voix de l'enseignant : mise en contexte théorique et pratique*, Paris : Éditions Maison des Langues p. 177-180.

Azaoui, B. (2014b) Multimodalité des signes et enjeux énonciatifs en classe de FL1/FLS, In M. Tellier & L. Cadet (Eds.) *Le corps et la voix de l'enseignant : mise en contexte théorique et pratique* Paris : Éditions Maison des Langues, p. 115-126.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018) Openface 2.0 : Facial behavior analysis toolkit, *In 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, p. 59-66, IEEE.

Bavelas, J., Chovil, N., Coates, L. & Roe, L. (1995) Gestures specialized for dialogue, *Personality and social psychology bulletin*, 21, p. 394-405.

Bavelas, J. B., & Gerwing, J. (2007) Conversational hand gestures and facial displays in face-to-face dialogue, *Social communication*, p. 283-308.

Bavelas, J., Gerwing, J., & Healing, S. (2014) Including facial gestures in gesture-speech ensembles. *From gesture in conversation to visible action as utterance: Essays in honor of Adam Kendon*, p. 15-34.

Bavelas, J. B., Coates, L., & Johnson, T. (2000) Listeners as co-narrators, *Journal of personality and social psychology*, 79, 6, p. 941.

Bateson, G. (1953) The position of humor in human communication, in Heinz von Foerster (ed.) *Cybernetics*, Ninth Conference, Josiah Macey Jr Foundation, New York, p. 1-47.

Bell, N. (2009a) Responses to failed humor, *Journal of Pragmatics*, 41, p. 1825-1236.

Bell, N. (2009b) Impolite responses to failed humor, in Delia Chiaro, Neal Norrick (eds) *Humor in Interaction*, Amsterdam: John Benjamins, p. 143-163.

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*, ATALA, 2008, 49, 3, p. 105-134

Bertrand, R. & Espesser, R. (2017) Co-narration in French conversation storytelling: A quantitative insight, *Journal of Pragmatics*, Elsevier, 111, p. 33-53.

Bigi, B., Meunier, C., Nesterenko, I. & Bertrand, R. (2010) Automatic detection of syllable boundaries in spontaneous speech. In *Language Resource and Evaluation Conference* (LREC), p. 3285-3292, La Valetta, Malta.

Bigi, B., Péri, P. & Bertrand, R. (2012) Orthographic Transcription: which enrichment is required for phonetization? In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey, p. 1756-1763.

Bigi, B. (2012) SPPAS: a tool for the phonetic segmentation of speech, In *Proceedings of the Eight International Conference on Language Resources and Evaluation*, Istanbul, Turkey, p. 1748-1755.

Bigi, B., Portes, C., Steuckardt, A. & Tellier, M. (2013) A multimodal study of answers to disruptions, *Journal on Multimodal User Interfaces*, 7, 1-2, p. 55-66.

Bigi, B. (2014) A Multilingual Text Normalization Approach. *Human Language Technology Challenges for Computer Science and Linguistics, Lectures Notes in Artificial Intelligence*, *8387*, p. 515-526, ISBN 978-3-319-14120-6.

Bigi, B. Bertrand, R. & Guardiola M. (2014) Automatic Detection of Other-Repetition Occurrences: Application to French Conversational Speech, *In Proceedings of the Ninth International Conference on Language Resources and Evaluation,* Reykjavik, Iceland, p. 836-842.

Bigi, B. (2015) SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech, *The Phonetician,* 111-112, p. 54-69.

Bigi, B. (2016) A phonetization approach for the forced-alignment task in SPPAS, *Human Language Technology. Challenges for Computer Science and Linguistics*, LNAI-9561, Springer Berlin Heidelberg, p. 397-410, ISBN 978-3-319-43807-8.

Bigi, B., Caron, B. & Oyelere, A.-S. (2017) Developing Resources for Automated Speech Processing of the African Language Naija (Nigerian Pidgin), *In 8th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland, p. 441-445.

Bigi, B. & Meunier, C. (2018) Automatic segmentation of spontaneous speech, *Revista de Estudos da Linguagem. International Thematic Issue: Speech Segmentation,* 26, 4, e-ISSN p. 2237-2083.

Bigi, B. & Priego-Valverde, B. (2019) Search for Inter-Pausal Units: application to Cheese! corpus, *In 9th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland, p. 289-293.

Blache, P. & Rauzy, S. (2008) Influence de la qualité de l'étiquetage sur le chunking : une corrélation dépendant de la taille des chunks, *Traitement Automatique des Langues Naturelles,* Avignon, France. p. 1-10.

Blache, P., Bertrand, R., Bigi, B., Bruno, E., Cela, E., Espesser, R., Ferré, G., Guardiola, M., Hirst, D., Magro, E.-P., Martin, J.-C., Meunier, C., Morel, M.-A., Murisasco, E., Nesterenko, I., Nocera, P., Pallaud, B., Prévot, L., Priego-Valverde, B., Seinturier, J., Tan, N., Tellier, M. & Rauzy, S. (2010a) Multimodal Annotation of Conversational Data, *In The Fourth Linguistic Annotation Workshop (LAW)*, Uppsala, Sweden, p. 186-191.

Blache P., Bertrand, R., Guardiola M., Guénot M.L., Meunier C., Nesterenko I, Pallaud B., Prévot L., Priego-Valverde B., Rauzy S. (2010b) The OTIM formal annotation model: a preliminary step before annotation scheme, *Proceedings of LREC: Workshop on Multimodal Corpora, Valetta, Malta*, p. 3262-3267.

Blache, P. & Rauzy, S. (2011) Predicting Linguistic Difficulty by Means of a Morpho-Syntactic Probabilistic Model, *PACLIC-2011,* Dec 2011, Singapore, Singapore. p. 160-167.

Blache, P., Bertrand, R., Ferré, G., Pallaud, B., Prévot, L., Rauzy, S. (2017) The Corpus of Interactional Data: a Large Multimodal Annotated Resource, In Ide N. & Pustejovsky J. (eds), *Handbook of Linguistic Annotation*, Springer bookseries Text, Speech, and Language Technology, p. 1323-1356

Blache, P., Abderrahmane, M., Rauzy, S. & Bertrand, R. (2020) An integrated model for predicting backchannel feedbacks, *ACM International Conference on Intelligent Virtual Agents (IVA'20)*, Oct 2020, Glasgow, United Kingdom.

Blanche Benveniste C., Borel B., Deulofeu J., Durand J., Giacomi A., Loufrani C., Mezziane B. & Pazery N. (1979) Des grilles pour le français parlé, *Recherches sur le Français Parlé*, 2, p. 163-208.

Blanche-Benveniste C. & Jeanjean C. (1987) *Le français parlé. Transcription et édition*, Didier Erudition, Paris.

Boersma, P. & Weenink, D. (2018) *Praat: doing phonetics by computer [Computerprogram], Version 6.0.37,* retrieved 14 March 2018 from http://www.praat.org

Boudin, A., Bertrand, R., Rauzy, S., Ochs, M. Blache, P. (2021) A multimodal Model for predicting Conversational Feedbacks, Springer Nature Switzerland AG 2021 K. Ekstein K. *et al.* (Eds.): *International Conference on Text, Speech, and Dialogue (TSD)*, 2021, Olomouc, Czech Republic.

Bouget, C. & Tellier, M. (2015) Use of gesture space in gestural accommodation, In Ferre, G. & Tutton, M. (Eds) *Gesture and Speech in Interaction - 4th edition*p. Nantes, 2-4 September 2015., p. 37-42. Disponible sur : https://hal.archives-ouvertes.fr/hal-01195646/document

Boula de Mareüil P., Habert B., Bénard F., Adda-Decker M., Barras C., Adda G., Paroubek P. (2005) A quantitative study of disfluencies in French broadcast interviews, *ISCA TR Workshop on Disfluency in Spontaneous Speech (DISS)*, Aix-en-Provence, p. 27-32.

Brugman, H., Russel, A. (2004) Annotating Multimedia/Multi-modal resources with ELAN, In *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, p. 2065-2068.

Calbris, G. (2011) *Elements of Meaning in Gesture*, Amsterdam, NL: John Benjamins Publishing.

Clark, H. H. (1996) *Using language*, Cambridge university Press.

Cohen, J. (1960) A coefficient of agreement for nominal scales, *Educational and psychological measurement*, 20, 1, p. 37-46.

Cosnier, J. (1982) Communications et langages gestuels, In Cosnier, J., Berrendonner, A., Coulon, J. & Orecchioni, C. (Eds.) *Les voies du langage : communications verbales gestuelles et animales,* Paris : Bordas, p. 255-303.

Crow, B. K. (1983) Topic shifts in couples' conversations, In *Conversational coherence: Form, structure, and strategy*, Sage Beverly Hills, CA, p. 136-156.

Drew, P. (1987) Po-faced receipts of teases, *Linguistics*, 25*, p.* 219-253.

Duncan, S., & Fiske, D. W. (2015) *Face-to-face interaction: Research, methods, and theory*. Routledge.

Ekman, P. & Friesen, W. V. (1978) *Facial Action Coding System: Manual. 1 et 2,* Consulting Psychologists Press.

Fant, G. (1973) *Speech, Sounds and Features,* Massachusetts: the MIT Press, Cambridge, 1973.

Fredouille, C., & Pouchoulin, G. (2012) Détection automatique de zones de déviance dans la parole dysarthrique : étude des bandes de fréquences, In *Actes de la conférence JEP-TALN-RECITAL 2012*, vol. 1, p. 377-384, JEP, Grenoble, France.

Farnetani, E. (1997) Coarticulation and connected speech, In *The Handbook of Phonetic Sciences*, Blackwell, Oxford, 371-404, Hardcastle W.J., Laver J., (eds), 1997.

Gironzetti, E., Attardo, S., & Pickering, L. (2016) *Smiling, gaze, and humor in conversation. Metapragmatics of Humor: Current research trends, p.* 237-256.

Gomez-Diaz, T. & Recio, T. (2019) *On the evaluation of research software: the CDUR procedure*. F1000Research, 8, 1353. Disponible sur : https://doi.org/10.12688/f1000research.19994.2

Haugh, M. (2014) Jocular Mockery as Interactional Practice in Everyday Anglo-Australian Conversation, *Australian Journal of Linguistics*, 34, 1, p. 76-99.

Hay, J. (2001) The pragmatics of humor support, *Humor,* 14, 1, p. 55-82.

Holt, B., Tellier, M. & Guichon, N. (2015) The use of teaching gestures in an online multimodal environment: the case of incomprehension sequences, In Ferre, G. & M. Tutton (Eds) *Gesture and Speech in Interaction - 4th edition*. p. 149-154, Nantes, 2-4 September 2015. Disponible sur : https://hal.archives-ouvertes.fr/hal-01195646/document

Horton, W. S. (2018) Theories and approaches to the study of conversation and interactive discourse, In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes,* p. 22-68, Routledge/Taylor & Francis Group.

Ide, N. & Véronis, J., (1994) MULTEXT: Multilingual Text Tools and Corpora*, Proceedings of the 15th. International Conference on Computational Linguistics (Coling 94), 1994*, Kyoto, Japan. p. 588-592.

Ide, N., & Pustejovsky, J. (2017) (Eds.) *Handbook of linguistic annotation* (Vol. 1), Berlin: Springer. 148 p.

Ide, N., Calzolari, N., Eckle-Kohler, J., Gibbon, D., Hellmann, S., Lee, K. *et al* (2017) Community Standards for Linguistically-Annotated Resources, In Ide, N., & Pustejovsky, J. (Eds.) *Handbook of linguistic annotation* (Vol. 1), Berlin : Springer p. 113-165.

Jeanjean C. (1984) Les ratés c'est fa- fabuleux. Etude syntaxique des amorces et des répétitions, *LINX,* 10, *Syntaxe et discours*, p. 171-177.

Jefferson, G. (1984) On stepwise transition from talk about a trouble to inappropriately next-positioned matters, In *Structures of social action: Studies in conversation analysis,* Cambridge: Cambridge University Press, p. 191-222.

Johnson, K. (2004) Massive reduction in conversational American English, In K. Yoneyama, & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis*, Tokyo: The National International Institute for Japanese Language, p. 29-54.

Kendon, A. (2004) *Gesture: Visible action as utterance,* Cambridge University Press.

Kerbrat-Orecchioni, C. (1990) *Les Interactions verbales, vol. 1, approche interactionnelle et structure des conversations*, A. Colin : Paris.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., Den, Y., (1998) An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs, *Language and Speech*, 41 (3-4), 323-350.

Laaridh, I., Meunier, C. & Fredouille, C. (2018) Perceptual Evaluation for Automatic Anomaly Detection in Disordered Speech: Focus on Ambiguous Cases, *Speech Communication*, 105, p. 23-33.

Lancien, M., Côté, M.-H. & Bigi, B. (2020) Developing Resources for Automated Speech Processing of Quebec French, In *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille, France, p. 5323-5328.

McNeill, D. (1992) *Hand and Mind: What gestures reveal about thought*, Chicago : The University of Chicago Press.

McNeill, D. (2005) *Gesture and thought*, Chicago: The University of Chicago Press.

Meaume, M. (2020) L'humour sous toutes ses formes : Analyse de l'humour dans la conversation entre personnes qui se connaissent, mémoire de Master 1, Aix-Marseille Université.

Meunier, C. & Espesser, R. (2011) Vowel reduction in conversational speech in French: the role of lexical factors, *Journal of Phonetics*, 39, 3, p. 271-278.

Meunier, C., Nguyen, N. (2013) Traitement et analyse du signal de parole, In Nguyen Noël, Adda-Decker Martine (eds.) *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, Traité IC2, série Cognition et traitement de l'information, Hermès, p. 85-119.

Meunier, C. (2014) *Variation de la parole : contraintes linguistiques et mécanismes d'adaptation*, Thèse, Habilitation à Diriger des Recherches, Université Lyon 2.

Mondada, L. & Traverso, V. (2005) (Dés) alignements en clôture. Une étude interactionnelle de corpus de français parlé en interaction, *Lidil. Revue de linguistique et de didactique des langues,* 31, p. 35-59.

Moreau, N., Rauzy, S., Viallet, F., Champagne-Lavau, M. (2016) Theory of Mind in Alzheimer Disease: Evidence of Authentic Impairment During Social Interaction, *Neuropsychology, American Psychological Association,* 30, 3, p. 312-321.

Morgenstern, A., Caët, S., Collombel-Leroy, M., Limousin, F., & Blondel, M. (2010) From gesture to sign and from gesture to word: Pointing in deaf and hearing children, *Gesture,* 10, 2-3, p. 172-202.

Mullan, K. & C. Béal (2018) Conversational humor in French and Australian English: What makes an utterance (un)funny?, *Intercultural Pragmatics*, 15, 4, p. 457-485.

Nguyen, N., Fagyal, S. & Cole, J. (2004) Perceptual relevance of long-domain phonetic dependencies, In *Actes des Journées d'Etudes Linguistiques*, Nantes, France, p. 173-178.

Pallaud B. & Henry S. (2004) Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé, In *Le poids des mots. Actes des 7e Journées Internationales d'Analyse statistique des Données Textuelles*, Louvain-la-Neuve, 10-12 mars 2004. Louvain : PUL, 2, p. 848-858.

Pallaud B., 2005, Les amorces de mots et leur contexte droit en français parlé spontané, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence,* 24, p 117-138.

Pallaud B., Xuereb R. (2008) Les troncations et les répétitions de mots chez un locuteur bègue, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA), Laboratoire Parole et Langage*, 26, p. 93-113.

Pallaud B., Rauzy S. & Blache P. (2014) Identification et annotation des auto-interruptions et des disfluences dans le corpus du CID. Disponible sur : http://LPL/Filer\agora\OTIM\Annotations\Disfluences

Pallaud B., Bertrand B., Prevot L, Blache P, & Rauzy S. (2019) Suspensive and Disfluent Self Interruptions in French Language Interactions, In Degand L., Gilquin, G., Meurant, L. & Simon, A-C. (Eds) *Fluency and Disfluency across Languages and Language Varieties. Corpora and Language in Use – Proceedings* 4, p. 109-138, Corpora and Language in Use. ffhal-02096964f

Pallaud B. & Bertrand B, (2020) Espaces interruptifs, interruptions suspensives et disfluentes en français parlé dans les huit dilogues du CID, In : Hirsch F., Didirkova I., Dodane C. (Eds) *Manuel de pausologie. Recueil de recherches sur les pauses présentes dans la parole et le discours,* Paris : L'Harmattan, p. 111-130, Langue et Parole, 978-2-343-19621-3.

Paroubek, P. & Rajman, M. (2000) MULTITAG, une ressource linguistique produit du paradigme d'évaluation, *Actes de Traitement Automatique des Langues Naturelles, 2000,* Lausanne, Suisse, p. 287-306.

Pickering, M. J. & Garrod, S. (2013) An integrated theory of language production and comprehension, *Behavioral and brain sciences*, 36, 4, p. 329-347.

Prévot, L., Gorisch, J., & Bertrand, R. (2016) A cup of cofee: A large collection of feedback utterances provided with communicative function annotations. Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, May 2016

Priego-Valverde, B. (2003) *L'humour dans la conversation familière : description et analyse linguistiques*, Paris : L'Harmattan.

Priego-Valverde, B. (2009) Failed humor in conversation: a double voicing analysis, in Neal Norrick & Delia Chiaro, D. (eds.), *Humor in interaction*, Amsterdam, Philadelphia: John Benjamins. p. 165-183.

Priego-Valverde, B. (2016) Teasing in casual conversations: an opportunistic discursive strategy, in Leonor Ruiz-Gurillo (ed). *Metapragmatics of Humor*, Amsterdam, Philadelphia: John Benjamins, p. 215-233.

Priego-Valverde, B., Bigi, B., & Amoyal, M. (2020) *Cheese!:* a Corpus of Face-to-face French Interactions. A Case Study for Analyzing Smiling and Conversational Humor, In *Proceedings of the 12th Language Resources and Evaluation Conference*, p. 467-475.

Priego-Valverde, B. (à paraître) Failed humor in conversation: disalignment and (dis)affiliation as a type of interactional failure, *Humor*.

Rauzy, S. & Blache, P. (2009) Un point sur les outils du LPL pour l'analyse syntaxique du français, *Journée ATALA Quels analyseurs syntaxiques pour le français ?, Oct. 2009,* Paris, France. p. 1-6.

Rauzy, S. & Blache, P. (2012) Robustness and processing difficulty models. A pilot study for eye-tracking data on the French Treebank. *Workshop on Eye-tracking and Natural Language Processing at The 24th International Conference on Computational Linguistics (COLING), Dec 2012*, Mumbai, India. p. 1-15.

Rauzy, S., Montcheuil, G., Blache, P (2014) MarsaTag, a tagger for French written texts and speech transcriptions, *Second Asian Pacific Corpus linguistics Conference, Mar 2014,* Hong Kong, China. p. 220-220.

Riou, M. (2015) A methodology for the identification of topic transitions in interaction. Discours, *Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics,* 16.

Rossi, M. (1990) Segmentation automatique de la parole : pourquoi ? Quels segments ?, *Traitement du signal*, 7, 4, p. 315-26.

Sacks, H., Schegloff E., Jefferson G. (1974) A Simplest Systematics for the Organisation of Turn-Taking for Conversation, *Language*, 50, p. 696-735.

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.

Shriberg, E.E. (1994) *Preliminaries to a Theory of Speech Disfluencies*, PhD thesis, University of California at Berkeley.

Sloetjes, H., & Wittenburg, P. (2008) Annotation by category-ELAN and ISO DCR, In *6th international Conference on Language Resources and Evaluation (LREC 2008)*.

Stivers, T. (2008) Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation, *Research on language and social interaction*, 41, 1, p. 31-57.

Tan, N., Ferré, G., Tellier, M., Cela, E., Morel, M.-A., Martin, J.-C. & Blache, P. (2010) Multi-level Annotations of Nonverbal Behaviors in French Spontaneous Conversation, In Kipp, M., Martin, J.-C., Paggio, P. & Heylen, D. (Eds) *Procedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality.* Workshop held at the 7th International Conference for Language Resources and Evaluation (LREC 2010). P. 17-23 May 2010, Malta, p. 74-79. [ligne] Disponible sur : http://www.multimodal-corpora.org/mmc10.html.

Tellier, M. (2008) Dire avec des gestes, *Le français dans le monde. Recherches et applications*, 44, p. 40-50.

Tellier, M., Azaoui, B., & Saubesty, J. (2012) Segmentation et annotation du geste : méthodologie pour travailler en équipe, In *JEP-TALN-RECITAL 2012 : Atelier DEGELS 2012 (Défi GEste Langue des Signes)*, p. 41-55.

Tellier, M., Stam, G. & Bigi, B. (2013) Gesturing while pausing in conversation: Self-oriented or Partner-oriented?, *Proceedings of TIGER- Tilburg Gesture Research Meeting Conference.* [En ligne] : Disponible sur : http://tiger.uvt.nl/list-of-accepted-papers.html

Tellier, M. (2014) Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés, *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, 15. Disponible sur : http://journals.openedition.org/discours/8917

Tellier, M. (2016) Prendre son cours à bras le corps. De l'articulation des modalités kinésiques avec la parole, *Recherches en didactique des langues et des cultures,* 13,1, Disponible sur : http://rdlc.revues.org/474 ; DOI : 10.4000/rdlc.474

Tellier, M., Stam, G. & Ghio, A. (2021) Handling Language: How future language teachers adapt their gestures to their interlocutor, *Gesture*, 20,1, p. 30-62.

Wu, Y., & Adda-Decker, M. (2021) Réduction des segments en français spontané : apports des grands corpus et du traitement automatique de la parole, Corpus, 22. Disponible sur : https://doi.org/10.4000/corpus.5812

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

NOTES

1. Cette évaluation permet d'évaluer la relative fiabilité des annotations posées ainsi que la robustesse des catégories annotées.
2. La convention complète de la *TOE* désormais utilisée dans de nombreux corpus au LPL ainsi que les recommandations pour son usage avec SPPAS peuvent être téléchargées à cette adresse : https://hdl.handle.net/11403/sldr000873.
3. https://journals.openedition.org/tipa/pdf/5745
4. https://journals.openedition.org/tipa/pdf/5735
5. MarsaTag peut être téléchargé librement (url : https://hdl.handle.net/11403/sldr000841).

6. Ceci renvoie au moment où le locuteur hésite, réfléchit, mais ne dit rien (différent de la redondance où il répète et sans doute se donne ainsi le temps de réfléchir aussi)
7. https://hdl.handle.net/11403/paco/v1
8. https://journals.openedition.org/tipa/pdf/5525
9. https://journals.openedition.org/tipa/pdf/5773
10. http://journals.openedition.org/tipa/5167
11. Acronyme signifiant Smile Movement Automatic Detection
12. https://journals.openedition.org/tipa/pdf/5728
13. La documentation et les scripts de notre outil SMAD sont téléchargeables sur la page du projet open source HMAD https://github.com/srauzy/HMAD.
14. Etude à partir du Corpus « Cid »
15. Cette étude et le schéma d'annotation sont réalisés dans le cadre du travail de Master de Manon Meaume.
16. En ligne avec (Gironzetti et al., 2016), cet intervalle correspond à 2 secondes avant et 3 secondes après l'item humoristique. Il permet d'analyser le sourire lors de la production de l'item humoristique proprement dit, mais également d'observer son évolution avant et après sa production.
17. https://journals.openedition.org/tipa/pdf/5790

RÉSUMÉS

La linguistique de corpus, c'est à dire les recherches sur le langage portant sur un matériel linguistique écrit ou oral recueilli et conservé, s'est considérablement développée au cours des dernières décennies. Ce développement s'est fait à l'aide de corpus constitués de plus en plus nombreux et de plus en plus importants. L'augmentation de la taille des corpus a nécessité, pour leur analyse, le développement d'outils automatiques, mais aussi une vraie réflexion sur la nature et les objectifs de l'annotation des corpus. L'enrichissement de corpus par un jeu d'annotations spécifiques est alors apparu, la plupart du temps, comme un préalable à toutes analyses linguistiques.
Annoter un corpus consiste à ajouter des informations pertinentes pour son exploitation. L'intérêt de disposer de corpus annotés, c'est-à-dire enrichis à différents niveaux linguistiques, est de pouvoir étudier chacun de ces derniers ainsi que les liens mutuels entre les uns et les autres. Les travaux menés au LPL sur ces questions d'enrichissement des corpus ont été effectués initialement pour rendre possible l'étude de la multimodalité, à savoir la prise en compte des niveaux de granularité les plus fins (phonèmes) jusqu'aux niveaux mimo-gestuels, en passant par les niveaux syntaxique, discursif, prosodique, et interactionnel. Il s'est donc avéré nécessaire de penser l'annotation en amont, au niveau même de la représentation des informations. Un schéma d'annotation global permet en effet de considérer tous ces niveaux dans une seule et même approche formelle qui favorise leur interrogation ultérieure.
Quel que soit le niveau d'annotation, plusieurs questions se sont posées : d'une part, celle des étiquettes utilisées (décomposition, typologie, fonction, nature graduelle/catégorielle, etc.) ; d'autre part, celle de l'ancrage temporel de ces étiquettes (localisation et frontières). Pour certains niveaux d'annotation il sera question de décrire les niveaux de dépendance entre les différentes étiquettes. Ces questions doivent être pensées en fonction des objectifs de recherche. Le travail au sein de chaque niveau d'annotation est ensuite relativement similaire. Il s'agit

d'établir un schéma d'annotation permettant une annotation la plus constante et la plus robuste possible. Ce schéma est établi sur la base des connaissances théoriques et en vue de répondre aux questionnements des recherches. Une fois le schéma d'annotation établi, il est également possible de construire un guide d'annotation destiné à de potentiels annotateurs (experts / naïfs). Le plus souvent, les annotations sont effectuées en recourant à plusieurs annotateurs afin de rendre possible une évaluation de la consistance (accords inter-annotateurs). La question transversale de l'hétérogénéité des annotations humaines sera traitée dans ce chapitre.

Dans ce chapitre, nous développons quelques-unes des principales étapes d'enrichissement qui ont été mises en œuvre pour annoter manuellement ou automatiquement les corpus, ainsi que les problématiques de recherche qui leur sont associées. Ces étapes sont listées ci-après :

- Recherche automatique des IPUs et transcription orthographique

À partir des données primaires collectées, sont recherchées automatiquement les IPUs - *Inter-Pausal Units*, qui nous permettent d'obtenir une segmentation en blocs de silences versus blocs sonores. Nous effectuons ensuite la transcription orthographique au sein de ces IPUs. Cette étape de transcription est cruciale dans la mesure où elle constitue la ligne (*tier*) sur laquelle se développeront les autres niveaux d'annotation. Là encore les choix effectués en termes de transcription (convention choisie) ont une incidence sur la mise en lien des niveaux d'annotation. Une fois la transcription orthographique effectuée - et alignée sur le signal au niveau des IPUs, de nombreuses annotations peuvent être obtenues, soit manuellement, soit automatiquement, soit semi-automatiquement.

- Annotation phonétique et lexicale

Nous développons, distribuons et enrichissons régulièrement un logiciel d'annotation automatique -SPPAS, qui permet notamment de normaliser le texte transcrit, c'est à dire d'obtenir les *tokens*. À partir de ces *tokens* au sein des IPUs, SPPAS peut effectuer la conversion graphèmes-phonèmes sous la forme d'une grammaire des prononciations possibles de chaque IPU. Enfin, SPPAS fournit l'alignement temporel des phonèmes qui, désormais, est rarement réalisée manuellement. Toutefois, les aspects manuels et automatiques de l'annotation phonétique relèvent de processus différents mais complémentaires. Ainsi, la parole spontanée engendre des réalisations phonétiques (réductions) difficilement gérables au niveau de l'alignement automatique. En conséquence 1/ il peut être nécessaire de corriger manuellement certaines parties de l'alignement automatique : 2/ il est possible d'utiliser les erreurs d'alignement pour localiser ces réalisations phonétiques spécifiques. Nous aborderons dans ce chapitre les questions liées à ces deux aspects. D'autres annotations peuvent ensuite être obtenues de cette segmentation en phonèmes. Notamment, ils permettent d'obtenir automatiquement l'alignement des *tokens* ; un système à base de règles permet de regrouper les phonèmes en syllabes.

- Annotation syntaxique

L'annotation syntaxique vient s'ancrer sur les *tokens*. S'il existe des analyseurs syntaxiques automatiques disponibles pour l'écrit, l'analyse syntaxique du français parlé reste encore un défi. Nous présentons ici la méthodologie que nous avons adoptée pour adapter notre étiqueteur de l'écrit afin de traiter les transcriptions de l'oral spontané. Si les performances de notre étiqueteur MarsaTag sont d'ores et déjà acceptables, l'amélioration de notre outil nécessitera une modélisation multi-niveaux incluant les phénomènes de disfluences (voir ci-dessous) et le traitement plus précis des marqueurs de discours.

- Annotation des disfluences

Les énoncés oraux comportent de nombreuses variations de la fluence verbale et, cela, à plusieurs niveaux (par exemple, le débit de prononciation des mots, des syntagmes ou des propositions). Mais ces variations peuvent se manifester également aux niveaux acoustiques et phonétiques. Sur les plans morphologiques et syntaxiques, certaines de ces variations se traduisent par de véritables auto interruptions qui suspendent le déroulement syntagmatique

dans l'émission verbale. Nos analyses de corpus ont prévu de conserver (en plus des pauses remplies ou non, éléments discursifs, interjections) les traces d'élaboration des énoncés que sont, entre autres, les amorces ou fragments de mots et les ruptures de syntagme. Cette stratégie a permis d'envisager une description fine et exhaustive de ces phénomènes désignés sous le terme de disfluence.

- Annotation du discours et des interactions

À partir du signal de parole et de sa transcription, il est également possible d'envisager une annotation de plusieurs niveaux pragmatiques tels que l'organisation thématique d'interactions conversationnelles. Plusieurs niveaux d'annotations seront donc décrits dans ce chapitre : l'annotation des thèmes conversationnels, des transitions thématiques (i.e. les mouvements conversationnels qui permettent de passer d'un sujet à un autre), ainsi que les phases de ces transitions. D'autres phénomènes seront également décrits, tels que les items de feedbacks et les séquences humoristiques. Nous présenterons le protocole d'annotation associé à ces différents phénomènes ainsi que les méthodes d'évaluation choisies pour évaluer la fiabilité de ces annotations.

- Annotation mimogestuelles

À partir du signal vidéo, il est possible d'envisager une annotation mimo-gestuelle (les expressions faciales ou les gestes manuels coverbaux par exemple). Cela peut se faire soit de façon manuelle soit semi-automatisée. Nous présenterons dans ce chapitre le protocole d'annotation semi-automatique des sourires que nous avons élaboré afin d'annoter deux corpus conversationnels. Tout d'abord, nous présenterons l'outil SMAD qui permet d'annoter automatiquement les sourires. Puis nous exposerons, le protocole de correction de ces annotations. Enfin nous décrirons la méthode d'évaluation choisie afin d'évaluer la robustesse des données annotées. Nous évoquerons également l'annotation manuelle des gestes coverbaux ainsi que les problématiques méthodologiques inhérentes telles que les schémas et guide d'annotation, les typologies et la segmentation. Nous donnerons des exemples d'études réalisées au LPL qui proposent différentes approches pour l'annotation des gestes.

Corpus linguistics (i.e. research on language based on written or oral linguistic material that has been collected and saved) has been considerably developed over the last decades. This development has occurred using more numerous and larger corpora. The increased size of corpora has required the development of automatic tools for their analysis, but also a real reflection on the nature and the objectives of the annotation of corpora. The increased size of the corpora has required the development of automatic tools for their analysis, but also a real reflection on the nature and the objectives of the annotation of the corpora. The enrichment of corpora by a set of specific annotations has emerged, in most cases, as a preliminary to any linguistic analysis.

Annotating a corpus consists in adding relevant information for its exploitation. The interest of having annotated corpora (i.e. enriched at different linguistic levels) is to be able to study each annotated levels and the mutual links between them. The work carried out at the LPL on these issues of corpus enrichment was initially meant to make possible the study of multimodality, such as the finest levels of granularity (phonemes) up to the mimo-gestural levels, passing through the syntactic, discursive, prosodic, and interactional levels. It was therefore necessary to think about annotation early on, at the level of information representation. A global annotation scheme allows to consider all these levels in a single formal approach that facilitates their subsequent interrogation.

Whatever the level of annotation, several questions have arisen: on the one hand there were questions about the labels used (e.g. decomposition, typology, function, gradual/categorical nature); on the other hand there were questions about the temporary embedment of these labels (location and boundaries). For certain levels of annotation, it will be necessary to describe the

levels of dependence between the different labels. These questions must be considered in relation to the research objectives. The work within each annotation level is then relatively similar. It is a question of establishing an annotation scheme that allows the most consistent and robust annotation possible. This scheme is established based on theoretical knowledge and in order to answer research questions. Once the annotation scheme is established, it is also possible to build an annotation guide for potential annotators (expert/naive). Most often, annotations are performed using several annotators to make possible an evaluation of the consistency (inter-annotator agreements). The transversal issue of heterogeneity in human annotations will be addressed in this chapter.

In this chapter, we develop some of the main annotation steps that have been performed to annotate corpora manually or automatically, as well as the research issues associated with them. These steps are listed below:

- **Automatic search of IPUs and orthographic transcription**

From the collected primary data, we automatically search for IPUs - Inter-Pausal Units - which allow us to obtain a segmentation into silence blocks versus sound blocks. We then perform the orthographic transcription within these IPUs. This transcription step is crucial as it constitutes the tier from which the other annotation levels will be developed. Here again, the choices made in terms of transcription (chosen convention) have an impact on the links between annotation levels. Once the orthographic transcription is done - and aligned with the signal on IPUs - many annotations can be obtained, either manually, automatically, or semi-automatically.

- **Phonetic and lexical annotation**

We develop, distribute, and regularly enrich an automatic annotation software -SPPAS, which also allows to normalize the transcribed text, which means to obtain the tokens. From these tokens within the IPUs, SPPAS can perform the grapheme-phoneme conversion based on a grammar of the possible pronunciations of each IPU. Finally, SPPAS provides the temporal alignment of phonemes which is now rarely performed manually. However, the manual and automatic aspects of phonetic annotation are different but complementary processes. Thus, spontaneous speech generates phonetic realizations (reductions) that are difficult to manage at the level of automatic alignment. Consequently 1/it may be necessary to manually correct some parts of the automatic alignment: 2/it is possible to use the alignment errors to locate these specific phonetic realizations. In this chapter, we will address the issues related to these two aspects. Other annotations can then be obtained from this phoneme segmentation. They allow to automatically obtain the alignment of tokens; a rule-based system allows to group phonemes into syllables.

- **Syntactic annotation**

Syntactic annotation is based on tokens. If there are automatic syntactic analyzers available for written language, syntactic analysis of spoken French remains a challenge. We present here the methodology we have adopted to adapt our writing tagger to handle spontaneous spoken transcripts. If the performances of our MarsaTag tagger are already acceptable, the improvement of our tool will require a multi-level modeling including the phenomena of disfluencies (see below) and the more precise treatment of discourse markers.

- **Annotation of disfluencies**

Oral utterances contain many variations in verbal fluency at several levels (e.g. the rate of pronunciation of words, phrases, or clauses). But these variations can also occur at the acoustic and phonetic levels. On the morphological and syntactic levels, some of these variations are translated by real self-interruptions which suspend the syntagmatic flow in the verbal emission. In our corpus analyses, we have planned to keep (in addition to filled or unfilled pauses, discourse elements, interjections) the evidence of the discourse elaboration which are, among other things, initiations or fragments of words and the syntagms' breaks. This strategy made it possible to envisage a detailed and exhaustive description of these phenomena designated under

the term of "disfluency".

- **Annotation of speech and interactions**

From the speech signal and its transcription, it is also possible to consider an annotation of several pragmatic levels such as the thematic organization of conversational interactions. Several levels of annotation will be described in this chapter: the annotation of conversational themes, thematic transitions (i.e. conversational movements that allow to go from one topic to another), and the phases of these transitions. Other phenomena will also be described, such as feedback items and humorous sequences. We will present the annotation protocol associated with these different phenomena as well as the evaluation methods chosen to assess the reliability of these annotations.

- **Mimogestual annotation**

From the video signal, it is possible to consider a mimogestual annotation (facial expressions or coverbal manual gestures for example). This can be done either manually or semi-automatically. In this chapter, first we will present the semi-automatic annotation protocol of smiles that we have developed in order to annotate two conversational corpora. We will present the SMAD tool which allows to automatically annotate smiles. Then, we will describe the protocol of correction of these annotations. Finally, we will discuss the evaluation method chosen to assess the robustness of the annotated data. We will also present the manual annotation of coverbal gestures as well as the inherent methodological issues such as annotation schemes and guides, typologies and segmentation. We will give examples of studies carried out at LPL that propose different approaches for gesture annotation.

INDEX

Keywords : oral corpus, conversational interactions, annotation schema, phonetics, lexicon, morpho-syntactics, disfluences, mimogestuality, automatic tools

Mots-clés : corpus oral, interactions conversationnelles, schéma d'annotation, phonétique, lexique, morpho-syntaxe, disfluences, discours, mimogestualité, outils automatiques

AUTEURS

MARY AMOYAL

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
mary.amoyal@univ-amu.fr

ROXANE BERTRAND

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

BRIGITTE BIGI

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
brigitte.bigi@univ-amu.fr

AURIANE BOUDIN

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
auriane.boudin@etu.univ-amu.fr

CHRISTINE MEUNIER

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

BERTHILLE PALLAUD

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

BÉATRICE PRIEGO-VALVERDE

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
beatrice.priego-valverde@univ-amu.fr

STÉPHANE RAUZY

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
stephane.rauzy@univ-amu.fr

MARION TELLIER

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
marion.tellier@univ-amu.fr