

UNIVERSITÉ SORBONNE NOUVELLE

ED 625 – MAGIIE (Mondes Anglophones, Germanophones, Indiens, Iraniens et Études Européennes)

EA 4398 – PRISMES (Langues, Textes, Arts et Cultures du Monde Anglophone)

Thèse de doctorat en linguistique anglaise et sciences du langage

RÉSUMÉ SUBSTANTIEL

Loulou KOSMALA

Une étude multimodale et contrastive de la (dis)fluence à travers les langues et contextes : Vers une évaluation multidimensionnelle de l'inter-(dis)fluence

Thèse co-dirigée par Madame la Professeure Aliyah MORGENSTERN et

Madame la Professeure Maria CANDEA

Soutenue le 3 décembre 2021

Membres du jury :

Mme Maria, CANDEA, Professeure, Université Sorbonne Nouvelle

Mme Camille, DEBRAS, Maître de Conférence, Université Paris Ouest— Nanterre

Mme Gaëlle, FERRÉ, Professeure, Université de Poitiers

Mme Gaëtanelle, GILQUIN, Professeure, Université Catholique de Louvain

Mme Aliyah, MORGENSTERN, Professeure, Université Sorbonne Nouvelle

Mme Anne, SALAZAR ORVIG, Professeure, Université Sorbonne Nouvelle

Mme Eve, SWEETSER, Professeure, University of California, Berkeley

Une étude multimodale et contrastive de la (dis)fluence à travers les langues et contextes : Vers une évaluation multidimensionnelle de l'inter-(dis)fluence

Résumé

Ce travail de thèse porte sur les phénomènes dits de « disfluence », un domaine de recherche qui s'appuie traditionnellement sur l'annotation de formes a priori « disfluentes », telles que « uh » et « um », les silences, les réparations, les répétitions, etc., qui marquent une interruption ou une suspension de la chaîne parlée. Plus récemment, des chercheurs ont mis en avant une approche ambivalente de ces marqueurs, aussi connus sous le nom de « fluencemes » afin de dévoiler le potentiel qu'ont ces mêmes formes à avoir des emplois à la fois fluents et disfluents selon les contextes de production. La présente étude se situe dans la continuité de cette démarche, et intègre une approche multimodale et interactionnelle, en prenant en compte les différentes modalités qui participent à la construction du discours, tels que les gestes, le regard, les expressions faciales, ou l'utilisation d'objets. L'objectif de cette thèse est d'évaluer les degrés de fluence dans la séquentialité de l'interaction multimodale, via une échelle tridimensionnelle qui considère la parole, la gestualité, et l'interaction. Notre analyse porte plus particulièrement sur les caractéristiques temporelles, positionnelles, fonctionnelles, et visuo-gestuelles des fluencemes, en combinant des annotations quantitatives et micro analyses des données. A partir d'un corpus vidéo en français et en anglais comprenant des échanges entre étudiants universitaires dans différentes langues et contextes, cette étude montre que la notion de disfluence ne saurait se réduire à une difficulté cognitive sur le plan verbal, puisqu'elle incarne également des pratiques interactives multimodales récurrentes et pertinentes aux activités langagières des locuteurs.

Mots-clefs : disfluence, fluence, gestualité, interaction, acquisition langue seconde, multimodalité, registre

**A multimodal contrastive study of (dis)fluency
across languages and settings:
Towards a multidimensional scale of inter-(dis)fluency**

Abstract

The research presented in this thesis deals with so-called “disfluency” phenomena, a topic of study traditionally concerned with the annotation of a priori “disfluent” forms, such as “uh” and “um”, silences, repairs, repetitions, and the like, marking an interruption or a suspension in the verbal channel. More recently, a number of researchers have vouched for an ambivalent approach to these markers, also known as “fluencemes”, to uncover the potential for the same forms to serve both fluent and disfluent functions depending on the context. The present study is situated within this field of research, and offers an additional multimodal and interactional approach, taking into account the multiple modalities available to speakers in situated interactional practices, such as hand gestures, gaze, facial displays, or artefacts, used to build meaning in discourse. The purpose of this thesis is to go beyond production-oriented models of disfluency, and evaluate the degrees of fluency, fluidity, or flow, of face-to-face communication with a tridimensional scale, considering the levels of speech, gesture, and interaction. Our analysis targets more specifically the durational, positional, functional, sequential, and visual-gestural properties of fluencemes, and combines quantitative annotations with micro-analyses of the data. Based on a video dataset in French and English of university students engaged in different tasks across different settings and languages, this research shows that the construct of disfluency should not be restricted to the level of speech production, as it also exhibits recurrent interactive multimodal practices which are relevant to speakers’ language activities.

Keywords: **disfluency, fluency, gesture, interaction, second language acquisition, multimodality, register**

Introduction

Ce travail de thèse porte sur les phénomènes dits de « disfluence », un domaine de recherche qui s'appuie traditionnellement sur l'annotation de formes a priori « disfluentes », telles que « euh », « um », les réparations, les répétitions, les mots tronqués etc., qui marquent une interruption ou une suspension de la chaîne parlée (Pallaud et al., 2013; Shriberg, 1994). En production orale, les locuteurs et locutrices sont constamment amené·es à planifier leurs énoncés en réfléchissant à ce qu'ils ou elles vont dire et à la manière de le dire, tout en s'ajustant à l'autre. Ceci est constitutif de la parole dite « spontanée » : les locuteurs et locutrices traitent l'information contenue dans l'interaction en la codant et décodant de manière simultanée (Duez, 2001). Par conséquent, la « fluence », continuité, ou fluidité du discours se retrouve temporairement « perturbée » par la présence de marques de « disfluence », ou « d'hésitation ». Ces marqueurs ont fait l'objet de nombreuses recherches en pathologie du langage, psycholinguistique, acquisition des langues et linguistique computationnelle depuis une soixantaine d'années, et ont été principalement étudiés au niveau de la production verbale.

Plus récemment, certains chercheurs et chercheuses ont mis en avant une approche dite *ambivalente* de ces marqueurs, aussi connus sous le nom de « fluencemes » (Götz, 2013) afin de dévoiler leur dualité fonctionnelle. Cette approche de la « (dis)fluence », avec le « dis » entre parenthèses (Buysse, 2019; Crible, 2018, Crible et al., 2019; Kosmala, 2020) part du postulat que les mêmes formes peuvent à la fois avoir des emplois fluents et/ou disfluents selon les contextes de production. La présente étude se situe dans la continuité de cette démarche, et intègre une approche *multimodale et interactionnelle* (Couper-Kuhlen & Selting, 2001; Goodwin, 2010; Mondada, 2007; Morgenstern, 2014), en prenant en compte les différentes modalités qui participent à la construction du discours, tels que les gestes manuels, le regard, les expressions faciales, ou l'utilisation d'objets. L'objectif de ce travail de thèse est d'évaluer les degrés de fluence, fluidité, ou *flow*, au sein de l'interaction multimodale, via une échelle tridimensionnelle, qui considère la parole, l'interaction, et la posturo-mimo-gestuelle. Jusqu'à ce jour, très peu d'études sur la (dis)fluence ont considéré le rôle de la gestualité et de l'interaction, car un grand nombre d'entre elles portent essentiellement sur des analyses quantitatives orientées vers des mécanismes cognitifs de production verbale. Tout le long de ce travail de

thèse, les différentes dimensions de la fluence (parole, gestualité, et interaction) sont mises en exergue afin d'illustrer la complexité de ce phénomène à travers les langues, registres, et contextes de communication. Nous présentons dès lors deux apports majeurs de notre travail, soit la construction de notre modèle théorique interdisciplinaire (section I de ce résumé), regroupant la grammaire cognitive, la linguistique interactionnelle et les études gestuelles, et notre méthode mixte, (section II de ce résumé) qui repose sur des méthodes quantitatives et qualitatives. Une partie de ce travail est également le fruit de plusieurs publications et communications à des colloques (Kosmala, 2019, 2020a, 2020b, 2021; Kosmala et al., 2019), qui ont été reprises afin de les intégrer de manière enrichie à cette thèse.

La présente thèse comporte cinq chapitres. Le premier chapitre (synthétisé en section I) décrit les différents cadres théoriques dans lesquels s'insère l'étude de la (dis)fluence. Il expose les bases de notre travail. Le deuxième chapitre (synthétisé en section II) présente les corpus de la thèse et la méthodologie et outils employés. Il discute également du statut de la transcription et de l'apport des « petits » corpus en linguistique. Les chapitres 3, 4, et 5 (synthétisés dans les sections III, IV, et V de ce résumé) sont empiriques, et livrent les résultats de nos analyses sur les corpus étudiés. Ils reposent à la fois sur des analyses quantitatives et qualitatives, et répondent aux questions de recherches formulées tout le long de cette thèse.

I. Apport théorique de la thèse : Construction d'un modèle théorique interdisciplinaire

L'étude de la disfluence s'inscrit dans plusieurs domaines, regroupant des cadres théoriques souvent très distincts les uns des autres, et qui ont très peu dialogué ces soixante dernières années. Cette partie présente un tour d'horizon de ces différentes approches comme point de départ, et propose un modèle intégratif visant à dépasser les visions souvent trop restrictives de la disfluence proposées dans la majorité des études. Les différentes étapes vers la construction de notre modèle théorique seront également illustrées au travers d'un exemple de notre corpus, afin de mettre en lumière l'application de notre méthodologie exposée en section II.

1.1. Les origines de la disfluence : définition et typologie

La disfluence, aussi connu sous le nom de « dysfluence » ou « disturbances » (Mahl, 1956; Meteer et al., 1995), fut d'abord abordée d'un point de vue pathologique. Bien que cette dimension ne soit pas le sujet de cette thèse, il est tout de même pertinent de retracer son origine. Ce terme est d'abord employé pour parler d'une « deviation » (*écart*) par rapport à l'« ideal delivery » (*élocution idéale*), soit un énoncé ne comportant ni pauses, ni interruptions, répétitions etc. (Clark, 2006). Chez les psycholinguistes, la disfluence souligne avant tout un temps d'arrêt, une rupture de la fluidité de la parole, mais il rejoint également l'idée de surcharge cognitive liée aux difficultés des activités langagières (Bortfeld et al., 2001; Finlayson & Corley, 2012; Fraundorf & Watson, 2014). Ces phénomènes sont donc plus présents lors de contextes de recherche lexicale (Beattie & Butterworth, 1979), de planification de longs énoncés (Oviatt, 1995; Shriberg, 1994), et d'incertitude (Schachter et al., 1991; Smith & Clark, 1993). Plusieurs chercheurs et chercheuses se sont donc intéressés à l'étude de la disfluence et des marqueurs de disfluence du point de vue de la production verbale, à travers notamment l'analyse de la vitesse d'élocution, et la mesure de fréquence, distribution et durée des pauses (Goldman-Eisler, 1968; Maclay & Osgood, 1959). Lorsqu'une « erreur » est détectée dans l'énoncé, elle constitue un point d'interruption qui appelle à être réparé, et qui peut s'analyser de manière structurelle, avec des indices, ou des « espaces », tels que « reparandum » « interregnum » et « reparans » (Clark & Wasow, 1998; Pallaud et al., 2019), illustrés dans l'exemple suivant, tiré du Corpus *SITAF*¹ (cf section II. 2.1. de ce résumé) :

Exemple 1 – structure disfluente d'un bout d'énoncé (décontextualisé)

dans une groupe euh qui:i qui discutons des choses politiques
dans une groupe (**Reparandum**)
euh qui qui:i (**Interregnum**)
discutons (**Reparans**) de choses politiques

Ce bout d'énoncé, analysé avec le système d'annotation de Pallaud et al., (2013, p. 22), délimite trois espaces qui illustrent les mécanismes d'interruption et de suspension de la fluence verbale. Le *Reparandum* fait référence à l'espace avant le point de rupture,

¹ Cet exemple est tiré de l'extrait 2B du Chapitre 3 de la thèse (Chap. 3, section II. 2.2.2.)

soit l'élément « indésirable » dans le segment fluent original (Ferreira & Bailey, 2004, p. 232), et l'*Interregnum* constitue le point d'interruption ou suspension de l'énoncé. Ce dernier contient des « indices » de disfluence (pauses remplies ou silencieuses, répétitions, mots tronqués etc.) où s'opère l'épisode interruptif. Enfin, le *Reparans* est la reprise fluente de l'énoncé, qui peut poursuivre, répéter ou modifier les éléments prononcés lors du *Reparandum*. Dans cet exemple plus précisément, l'*Interregnum* contient trois types de disfluences, soit une pause remplie (« euh »), une répétition du pronom relatif « qui », ainsi qu'un allongement de ce dernier (« qui:i »). Cet épisode dit « disfluent » comprend donc une combinaison de plusieurs marqueurs non-lexicaux qui suspendent la continuité de cet énoncé. Dans d'autres cas, les disfluences peuvent également interrompre un énoncé et provoquer des changements morphosyntaxiques, lorsque par exemple le *Reparans* répare une partie du *Reparandum* en rectifiant le choix d'un terme lexical et le remplaçant par un autre. Ce type d'étude souligne l'aspect structurel et formel des disfluences dans la chaîne parlée, ce qui a donné lieu à diverses typologies et classifications de marqueurs (cf Lickley, 2015, Shriberg, 1994).

En bref, la disfluence a pu être définie de manière « traditionnelle » chez les psycholinguistes comme étant une manifestation disruptive ou « suspensive » (Pallaud et al., 2013) de la fluence verbale, par le biais de différentes formes « non-lexicales » sans contenu propositionnel (pauses remplies et silencieuses, allongements de syllabe, répétitions, réparations, troncations etc.). Ces formes vocales et morphosyntaxiques ont pu également être détectées dans des corpus oraux de manière automatique dans le cadre de la reconnaissance de la parole, et certaines technologies permettent même de les supprimer complètement afin de faciliter la transcription des données (Kaushik et al., 2010). Mais que serait le discours oral sans ces formes, constitutives de la parole spontanée ? En quoi l'élocution idéale, dite *fluente*, représenterait une base de normalisation ? Comme le souligne Lickley (2015), les études de corpus de données orales ont montré qu'une telle « fluence » représenterait une exception plutôt que la règle. Alors pourquoi employer le terme « disfluence » ? Certains chercheurs et chercheuses préfèrent au contraire parler de marqueurs de « fluence » (e.g. Tottie, 2011), ou optent même pour des termes radicalement différents et innovants (cf section 1.2.) afin de s'éloigner de la connotation négative suggérée par le terme « disfluence ». Notre travail de thèse a également pour but

d'apporter sa contribution à ce débat terminologique, et d'approfondir les questions théoriques qui le sous-tendent.

1.2. Entre « disfluence » « fluence », et « hésitation », un débat terminologique et théorique

Comme nous le verrons dans la présente section, la question de la terminologie joue un rôle majeur dans le choix d'approche théorique adopté. Cette section présente brièvement les différents termes employés dans la littérature (tirée du Chap. 1, section II de la thèse) et propose le choix de terminologie effectué dans le cadre de cette thèse.

Nous l'avons vu, l'opposition binaire entre les termes « disfluence » et « fluence » est problématique car elle identifie la fluence à un idéal de parole orale abstraite et idéalisée. Dans le domaine de la recherche en acquisition des langues secondes (L2), la notion de fluence (*fluency*) est souvent synonyme de compétence langagière, avec des expressions du type « I am fluent in English » pour dire « Je parle couramment l'anglais » (e.g., Lennon, 1990²). Le but des apprenants serait donc d'atteindre l'élocution « fluente » d'un natif, et l'évaluation de cette maîtrise de langue passerait par l'analyse d'un certain nombre de variables temporelles, telles que la durée et fréquence des pauses, ou la vitesse d'élocution (De Jong, 2018; Riggensbach, 1991). Autrement dit, la présence de disfluences dans les productions L2 semblerait systématiquement associée à une mauvaise maîtrise de la langue. Or, comme nous l'avons vu, ces phénomènes sont également constitutifs de la parole native, donc ce type d'analyse présente quelques limitations. Bien que plusieurs études aient montré une tendance chez les apprenants et apprenantes à produire des énoncés comportant davantage de disfluences dans leur L2 que leur L1 (e.g., Deschamps, 1980; Fehringer & Fry, 2007; Gilquin, 2008; Tavakoli, 2011), leur compétence langagière ne peut se réduire à des différences de variables temporelles. Le problème vient à nouveau du choix de la terminologie et de l'opposition des termes *fluence* et *disfluence* qui suggère qu'un énoncé est soit fluide ou interrompu, qu'un locuteur/locutrice doit être considéré-e fluent-e ou disfluent-e, ou qu'une forme est catégorisée comme étant lexicale ou non-lexicale. Tout le long de ce travail de thèse, nous montrerons que les phénomènes de fluence et disfluence ne devraient pas être réduits à l'emploi d'une

² Il convient de noter que Lennon parle plus précisément de définitions « restreintes » (*narrow*) et « larges » (*broad*) de la fluence, cf Chap. 1, section II, 2.1.1., de la thèse pour plus de détails.

seule étiquette ou d'un seul terme, et que leur description ne devrait pas se limiter aux outils d'un seul type d'approche.

Un autre terme populaire dans la littérature pour discuter de ces phénomènes est « hésitation » (Duez, 2001; Guaitella, 1993; Holmes, 1988; Maclay & Osgood, 1959). Il est à la fois utilisé pour parler d'incertitude (Jokinen & Allwood, 2010) de choix lexical (Beattie, 1979), mais aussi de caractéristiques acoustiques (Betz, 2020; Guaitella, 1993; Shriberg, 2001). Ce terme est notamment critiqué par Candea (2000) qui le considère comme étant à la fois trop large et trop spécifique. En effet, le problème avec le terme « hésitation » c'est qu'il présuppose un sentiment d'indécision ou d'incertitude, et il est par ailleurs très souvent associé à des marques spécifiques comme « euh », ce qui est à nouveau réducteur. De plus, son emploi est assez disparate dans la littérature, car il est tantôt confondu, tantôt exclu des autres marqueurs de disfluence (e.g. Merlo & Mansur, 2004). Pour cette raison, Candea (2000) a opté pour le terme « travail de formulation », reprenant celui de Morel & Danon-Boileau (1998). D'autres termes innovants ont également été introduits dans la littérature, comme « planner » employé par Jucker (2014) et Tottie (2014), « communication management » (Allwood et al., 2005), ou « confluence » (McCarthy, 2009). Tous ces termes offrent une approche nouvelle de ces phénomènes, et se concentrent sur des aspects plus pragmatiques, discursifs et communicationnels. Toutefois, bien qu'ils offrent des alternatives novatrices aux notions de disfluence et d'hésitation, ces termes restent tout de même assez marginaux, et peu reconnus dans la littérature.

En bref, on trouve au sein de la littérature deux approches dominantes sur ces phénomènes en question (Nicholson, 2007) : d'une part, ils reflètent une surcharge cognitive due à des difficultés de production, perturbant la fluidité des énoncés (Bortfeld et al., 2001; Hartsuiker & Notebaert, 2009) ; d'autre part, ils fonctionnent comme des signaux communicatifs, permettant de rétablir la fluence communicationnelle (Allwood, 2017; Tottie, 2016). Cette dualité, que l'on trouve à la fois dans l'opposition des termes « fluence » et « disfluence » et dans leurs fonctionnements, est mise en lumière dans le travail de Crible et al. (2019). Cette équipe de chercheuses souligne l'idée que les mêmes formes, appelées « fluencemes » (Götz, 2013) peuvent, selon leur durée, leur place dans l'énoncé, ou leur combinaison avec d'autres formes tels que les marqueurs de discours, être *polyvalentes*, soit être à la fois des marqueurs de fluence ou de disfluence selon de nombreux facteurs contextuels. Elles proposent le terme *(dis)fluence* avec le « dis » entre parenthèses

pour dépasser la dichotomie fluence/disfluence et symboliser la nature dynamique et flexible de ces phénomènes. Ces notions d'ambivalence et de dynamisme représentent un aspect capital de ce travail de thèse qui se situe dans la continuité de ces études. En outre, cette thèse propose d'aller plus loin en intégrant d'autres dimensions du discours, soit l'interaction et la gestualité. Le cœur du problème, nous le soulignerons, ne vient pas seulement de la terminologie, mais du cadre théorique employé. La plupart des études sur la (dis)fluence en psycholinguistique, linguistique de corpus et acquisition des langues secondes repose essentiellement sur l'analyse verbale, que ce soit au niveau syntaxique ou discursif, et par conséquent explore très peu la dimension interactionnelle (à part quelques exceptions, Allwood, 2017; Kjellmer, 2003; McCarthy, 2009; Tottie, 2011). Pour ce qui est de la dimension gestuelle, elle est peu étudiée, mis à part quelques études portant sur la synchronisation de la suspension de la parole et des gestes (Chui, 2005; Graziano & Gullberg, 2018; Stam & Tellier, 2017; Seyfeddinipur, 2006) mais qui explorent à nouveau très peu la dimension interactionnelle. Par contraste, en analyse conversationnelle (Sacks et al., 1974), ces phénomènes, connus sous le nom de « repairs » (réparations) sont étudiés en détail dans le cadre de micro-analyses d'extraits explorant l'orientation des participants et leurs relations sociales, la progressivité de l'échange, et les fonctionnements des unités de tours de parole. Nous reviendrons sur ce point dans la prochaine section.

En somme, les phénomènes dits de « disfluence » « hésitation » ou « réparation » touchent à différentes disciplines en sciences du langage, soit la psycholinguistique, la linguistique de corpus, l'acquisition des langues secondes, la linguistique computationnelle, et l'analyse conversationnelle, mais ces domaines marquent souvent une conception restreinte, liée à un modèle théorique unique. Ce travail de thèse propose de développer une conception alternative de la (dis)fluence en intégrant plusieurs domaines théoriques afin d'approcher la richesse et la complexité de ces phénomènes à travers plusieurs niveaux d'analyse. Dans la lignée du travail de Crible et al. (2019) nous emploierons les termes *(dis)fluence* et *inter-(dis)fluence* tout le long de cette thèse. La notion de « fluence » ici n'est pas uniquement synonyme de compétence langagière, elle incarne à la fois les notions de continuité, fluidité, et progressivité à des niveaux d'analyses différents, soit la parole, l'interaction et la posturo-mimo-gestuelle. Notre choix du suffixe *inter-* a également son importance car il souligne l'*interrelation* et l'*interaction* entre ces différents niveaux d'analyse, qui s'ajustent les uns aux autres pour construire la fluence du discours. Cette étude de

l'inter-(dis)fluence émane dès lors de plusieurs cadres théoriques, décrits dans la section suivante.

1.3. Vers une étude de l'inter-(dis)fluence : contributions de la grammaire cognitive, la linguistique interactionnelle, et les études gestuelles

Afin d'offrir une conception plus large et innovante de la notion de « disfluence », nous nous repons sur plusieurs modèles théoriques qui offrent à la fois des concepts et outils adaptés à notre analyse multidimensionnelle de l'inter-(dis)fluence. Étudions à nouveau l'exemple 1 introduit en section 1.1 :

Exemple 1 contextualisé et situé au sein d'une séquence multimodale³

- 1 *FRAN: eum mais (1.278) en même temps on peu:ut vraiment
 si [/] si:i (1.033) dans une groupe euh

 ((main gauche suspendue; regard vers le haut e.))
 qui:i [/] qui discutons de:es des choses <politiques>.

 ((main gauche en rotation))
- *ELEN: +< mm mm
 ((hauchement de tête))
- 2 *FRAN: ou des choses euh (1.655) <quoi tu [/] tu>.

 ((main gauche suspendue e.)) ((paume gauche ouverte vers ELEN f.))



- 3 *ELEN: <religieuses politiques> les [/] les choses

 ((main gauche en rotation + regard sur FRAN g.))
 un peu:u [/] un peu tabou.

³ Cet extrait est transcrit avec notre système de transcription décrit dans le Chapitre 2, section 1.5.3. de la thèse, synthétisé dans la section II. 2.2. de ce résumé. * représente la production d'un geste synchronisé avec la parole, et _ quand le geste est suspendu.



4 *FRAN: +< oui tout ça.

 ((haussement d'épaule, paumes des deux mains vers le haut g.))

Dans la section 1.1., nous avons présenté un bout d'énoncé décontextualisé afin de décrire sa structure disfluente, suivant la tradition des études en (dis)fluence, mais sans rendre compte de son usage en contexte, ni des caractéristiques visuo-gestuelles qui l'accompagnent (direction du regard, geste manuel etc.). Ce type d'analyse décontextualisée est assez fréquent dans les études en linguistique de manière générale, et a été notamment critiqué par Cienki (2012) dans le domaine de la grammaire cognitive. Dans ses travaux, Cienki (2017, 2015, 2012) propose une vision de la grammaire dynamique et adaptative, où les locuteurs interagissent selon « a scope of relevant behaviors » (*un éventail de conduites pertinentes*, traduction de Morgenstern, 2019) et où une sélection s'opère entre les différentes ressources sémiotiques disponibles au sein de l'activité langagière (parole, intonation, gestes, posture, expressions faciales etc.) en contexte. On retrouve également des principes similaires en anthropologie linguistique, notamment chez Goodwin (2000) qui parle de « contextual configuration » (*configuration contextuelle*), et en études gestuelles, avec Müller & Tag (2010) qui proposent la théorie de « dynamic focus on attention » (*centre d'attention dynamique*). Ces principes, issus de domaines théoriques différents (grammaire cognitive, linguistique interactionnelle et études gestuelles) convergent et apportent des perspectives tout à fait pertinentes à notre étude de l'inter-(dis)fluence en interaction multimodale, illustré dans notre exemple. Ce dernier est tiré d'un échange entre un locuteur américain (Francis) et une locutrice française (Elena) dans un cadre tandem en français, où Francis parle sa deuxième langue, et Elena sa première. Ici Francis parle plus particulièrement des types de sujets de conversation un peu sensibles qui peuvent resurgir lors de discussions entre amis, mais il ne trouve pas exactement ses mots. Le bout d'énoncé analysé initialement en section 1.1 est situé à la ligne 1, et il est suivi d'un autre énoncé à la ligne 2 qui comporte

également un nombre de fluencemes (pause remplie, pause silencieuse et répétition). Mais ce que cette transcription nous révèle c'est que ce locuteur n'est pas simplement « disfluent », confronté à des difficultés de production, mais qu'il participe de manière collaborative à l'activité langagière en cours en se positionnant en tant que locuteur non-natif qui invite sa partenaire native à prendre part à l'échange. Il exécute cette action non pas par le biais de la parole, mais par son regard et sa gestuelle : à la fin de son tour (l.2), il déploie sa paume gauche vers le haut, orientée vers Elena, tandis qu'il pose le regard sur elle. Cette manifestation gestuelle, connue et répertoriée sous le nom de *Palm-Up-Open-Hand* (Kendon, 2004; Müller, 2004) a ici une fonction interactive qui opère un changement de tour : en effet, Elena prend la parole à la ligne 3 pour participer à la co-construction de l'échange. Elle propose une suite d'adjectifs (« religieuses » « politiques » « tabous ») qu'elle présente de manière consécutive avec un mouvement cyclique de la main droite, et qui sont par la suite « validés » par Francis qui acquiesce en répondant « oui tout ça » (l.4) accompagné d'un haussement d'épaules, les deux paumes levées vers le haut⁴. En bref, ce que cet extrait nous révèle, c'est que malgré la présence de formes marquant une « disfluence » au niveau verbal (i.e. la fluence verbale est temporairement perturbée), ces mêmes formes, accompagnées de signaux visuo-gestuels, marquent également une *fluence*, une fluidité, ou progressivité sur le plan de *l'interaction multimodale* (i.e. co-construction du discours, regards mutuels, continuité gestuelle).

La (dis)fluence n'est donc pas autonome par rapport aux autres systèmes linguistiques et ne relève pas uniquement de processus mentaux decontextualisés. Elle est profondément située et imbriquée dans l'interaction sociale, elle-même déterminée par des contraintes situationnelles. Ce courant de pensée est mobilisé dans le cadre de la *Linguistique Interactionnelle*, *l'Analyse Conversationnelle*, et *l'Anthropologie Linguistique* (Couper-Kuhlen & Selting, 2001; Goodwin, 2000; Mondada, 2007; Pekarek Doehler, 2006; Streeck et al., 2011), mais aussi en *Linguistique Cognitive* (Cienki, 2017; Geeraerts, 2006; Langacker, 1987). Ces domaines théoriques offrent des perspectives nouvelles au champ de la (dis)fluence, car ils permettent de considérer les notions de fluence et disfluence non pas uniquement sur un plan verbal, mais également sur un plan interactionnel et social. Un énoncé considéré « disfluent » d'un point de vue strictement verbal peut au contraire être très « fluent » sur le plan

⁴ Connu sous le nom de « shrug » (Debras, 2017; Jokinen & Allwood, 2010; Streeck, 2009), décrit plus précisément dans le chapitre 1 de la thèse (section III. 3.3.1)

intersubjectif ou interpersonnel, si l'on prend en compte sa place au sein de l'échange, ainsi que les attentes interpersonnelles de la situation (Grosman, 2018). De plus, la dimension visuo-gestuelle, jusque-là très peu étudiée dans ce domaine, et que l'on retrouve au coeur des *études gestuelles* (Alibali, 2005; Gullberg, 2008; Kendon, 2004; McNeill, 2000; Morgenstern, 2014; Streeck, 2009) offre une perspective multimodale, considérant également la fluence comme flux ou continuité gestuelle. Ce travail de thèse s'est employé à faire dialoguer ces différents niveaux d'analyse, dans la lignée de Morgenstern et al. (2021), pour offrir une conception multidimensionnelle de l'inter-(dis)fluence, illustrée ci-dessous (tirée du schéma 15 de la thèse, Chap. 1, section IV.4.2).

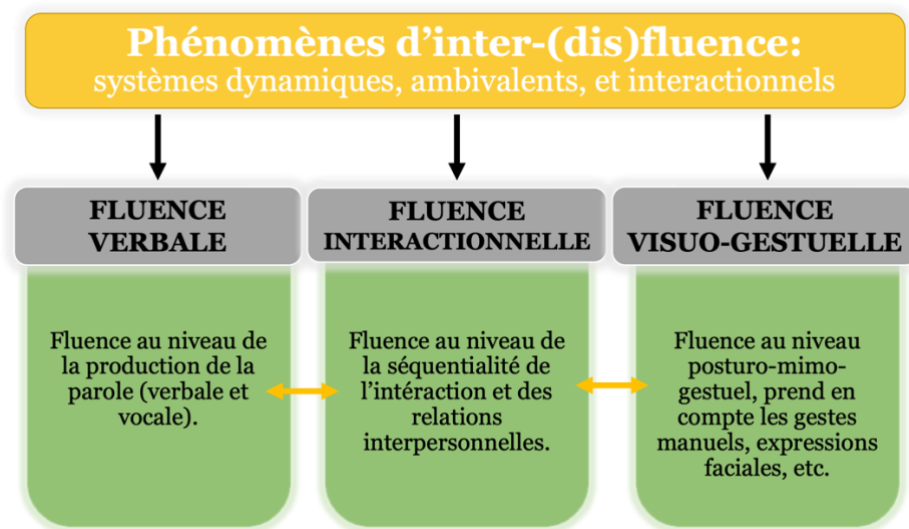


Fig. 1. Modèle multi-dimensionnel de l'inter-(dis)fluence

Ce modèle, inspiré de Götz, (2013), Grosman (2018) et Segalowitz (2016) propose que la fluence s'opère à plusieurs niveaux du discours. Dans le premier niveau, on s'intéresse aux formes étudiées dans le domaine « traditionnel » de la (dis)fluence décrit en section 1.1. (voir Chap. 1, section I de la thèse) et de leur place au sein de l'énoncé. Dans le deuxième, on considère la fluence comme processus interactionnel, avec les notions de progressivité (*progressivity*, voir Schegloff, 2007), intersubjectivité (*intersubjectivity*, voir Sterponi & Fasulo, 2010) et structure préférentielle (*preference structure*, voir Yule, 1996). Enfin, dans le troisième niveau, on s'intéresse à la construction de la fluence par le biais de plusieurs ressources sémiotiques, plus précisément les mouvements du corps comme les mains, les bras, le buste, le visage etc. Durant cette recherche nous avons tâché d'articuler ces différents niveaux et d'évaluer la multifonctionnalité des phénomènes d'inter-(dis)fluence à

travers les langues et contextes. Nos hypothèses et questions de recherches générales, issues de la construction de notre cadre théorique interdisciplinaire, peuvent se résumer ainsi (voir Chap. 1, section IV.4.3.) :

- Comment dépasser la notion restrictive de la « disfluence », en adoptant une approche multidimensionnelle de la fluence, couvrant les notions de fluidité, progressivité, et flux ?
 - Nous partons du postulat que les fluencemes sont des formes fluides et adaptatives, qui représentent plusieurs degrés de convention, lexicalisation, et de fluence. Elles peuvent également constituer une interruption ou une disruption, qui s'opère à plusieurs niveaux (interruption de la fluence verbale, de l'activité interactionnelle, langagière, ou gestuelle) selon leur emploi dans le contexte. Par contexte nous entendons à la fois le co-texte, la structure séquentielle de l'interaction, et la situation d'énonciation.
 - On retrouve également au sein de ces phénomènes une dualité fonctionnelle : d'une part ils relèvent de processus cognitifs auto-centrés (travail de formulation, recherche lexicale etc.) ; d'autre part, ils participent à l'organisation et l'intersubjectivité de l'échange (processus interactionnels, co-construction du discours, alignement, affiliation etc.). Cette dualité peut être notamment évaluée par le biais de marques visuo-gestuelles (direction du regard, fonction du geste, positionnement du corps etc.).
 - Par conséquent, les processus d'inter-(dis)fluence ne relèvent pas uniquement du verbal, mais également du posturo-mimo-gestuel, où les notions d'interruption et suspension, qui sont inhérentes à la fluence verbale, se retrouvent également dans la fluence gestuelle, où une synchronisation s'opère entre la suspension de la parole et du geste.

II. Apport méthodologique : méthode mixte et multi-niveaux

Notre construction théorique interdisciplinaire se retrouve également dans notre méthodologie qui emprunte ses outils à plusieurs domaines de recherche et qui réunit analyses quantitatives et qualitatives. Cette méthode est appliquée à nos données qui

sont tirées de deux corpus vidéo, le corpus SITAF et le corpus DisReg, décrits dans la section suivante.

2.1. Données étudiées : corpus SITAF et DisReg

Le Corpus SITAF (*Spécificité des Interactions Verbales dans le Cadre de Tandems Linguistiques Anglais-Français*) est un corpus vidéo collecté à l'Université Sorbonne Nouvelle entre 2012 et 2014 (Horgues & Scheuer, 2015) qui contient des interactions entre 21 paires d'étudiant·es niveau licence, anglophones et francophones, âgé·es de 17 à 22 ans. Notre étude repose sur une sélection de ce corpus comprenant 22 participants et participantes, 11 francophones et 11 américain·es lors d'échanges en tandem L1-L2 en anglais et en français (1h21, 14121 mots). Le corpus contient des vidéos où chaque tandem effectue une tâche argumentative (« Like Minds ») qui consiste à débattre d'un sujet donné (« est-ce que les prisonniers devraient avoir le droit de vote » etc.) et à donner leur niveau d'accord. Chaque échange se déroule en anglais et en français, ce qui donne la possibilité aux participants et participantes de parler à la fois dans leur première et deuxième langue. Avant les enregistrements, les étudiant·es ont également estimé leur niveau de compétence écrite, orale, et de compréhension en se donnant une note de 0 à 10. Ce corpus a été choisi pour sa qualité vidéo, dans la continuité d'études précédentes qui ont exploité sa dimension multimodale (e.g., Debras et al., 2015, 2020; Debras & Beaupoil-Hourdel, 2019) ; mais il a également été choisi pour notre étude car il permet de comparer les mêmes locuteurs et locutrices dans deux langues différentes, ce qui est moins courant dans les études en EFL (*English as a Foreign Language*) ou SLA (*Second Language Acquisition*) de manière générale (voir par exemple De Jong, 2016; Gilquin, 2008; Stam, 2001).

Notre deuxième corpus, le corpus DisReg (*DISfluency across REGisters*) a été collecté en 2018 dans le cadre de ce travail de thèse (Kosmala, 2020b). Il comporte des enregistrements vidéo de 12 étudiants et étudiantes francophones, de la Sorbonne Nouvelle, âgé·es de 18 à 22 ans, filmé·es dans deux situations différentes. Le corpus sélectionné contient d'une part des enregistrements faits en classe lors de présentations orales notées d'étudiant·es dans un cadre formel et institutionnel, et d'autre part des conversations semi-dirigées entre ces mêmes étudiant·es dans un cadre plus informel (durée de l'échantillon : 1h08, 12590 mots). L'intérêt de ce corpus repose à nouveau sur la comparaison des mêmes locuteurs et locutrices dans deux

situations de communication différentes. Il permet de faire un parallèle avec le corpus SITAF en incluant un autre ensemble de variables (style, registre, degré de préparation etc.) sur un échantillon et une population similaire à celle de SITAF (étudiant-es à l'université).

Notre échantillon total est de 2H30 et 26.000 mots, ce qui est relativement « petit » comparé aux larges corpus écrits que l'on trouve en linguistique de corpus (voir Vaughan & Clancy, 2013). Dans la section I.1.4 du Chapitre 2 de la thèse, nous défendons plus en détail notre choix de petit corpus et son intérêt pour notre étude multimodale, dans la lignée de Danino (2018) et Debras (2018). Il convient de noter que malgré la petite taille de notre échantillon, nous avons obtenu un nombre d'occurrences assez conséquent au total, soit 6042 fluencemes (3172 dans SITAF et 2870 dans DisReg) et 2381 gestes manuels (1362 dans SITAF et 1019 dans DisReg). De plus, les traitements quantitatifs effectués sur nos données sont complétés d'analyses qualitatives (40 au total) ce qui permet d'une part d'avoir une impression globale de ces phénomènes (i.e., distribution, fréquence, durée en moyenne etc.) et d'autre part de rendre compte de leur richesse et complexité en contexte. Ce type d'analyse, employé tout le long de ce travail de thèse, est connu sous le nom de « méthode mixte » (Candea, 2017; Morgenstern et al., 2021; Stivers, 2015; Tashakkori & Creswell, 2007) et rassemble approches quantitatives et qualitatives. Cette méthode permet de faire le pont entre les études quantitatives et qualitatives en (dis)fluence : les premières présentent des résultats quantitatifs sur un échantillon sous la forme de tests ou modèles statistiques, mais sans les illustrer dans des analyses qualitatives ; et les secondes offrent des outils pertinents à l'analyse séquentielle de l'interaction en contexte, mais ne se concentrent que sur une fraction d'occurrences présentes dans toutes les données.

2.2. Notre système de transcription multimodale

Dans ce travail de thèse, nous avons mis en place un système de transcription multimodale, en nous inspirant des conventions employées chez les études précédentes regroupant le domaine de la (dis)fluence avec les conventions du logiciel CLAN (MacWhinney, 2000), l'Analyse Conversationnelle (Jefferson, 2004), et les études gestuelles (Kendon, 2004; Mondada, 2018, voir section I.1.5. du Chapitre 2 de la thèse). Le travail de transcription est toujours sélectif, et reflète de ce fait le choix de notre approche théorique (Ochs, 1979). Bien qu'une transcription ne représente

jamais l'exhaustivité des phénomènes présents en interaction orale, elle permet tout de même de mettre en lumière certains aspects répondant à des questions de recherche précises. Dans notre cas, nous nous appuyons en particulier sur la temporalité des gestes manuels avec la production des fluencemes sur le flux gestuel, afin d'accentuer la dimension multimodale des phénomènes d'inter-(dis)fluence. Nos conventions de transcriptions, tirée du tableau 8 de la thèse (Chap. 2, section I. 1.5.3.) sont décrites ci-dessous :

Tableau 1. Conventions de transcription utilisées dans la thèse

Conventions CHAT (MacWhinney, 2000, pp. 48-74)	
+/	interruption par un.e autre participant.e
+//	auto-interruption
[/]	répétition
[//]	auto-réparation
+...	énoncé inachevé
(0.250)	pause silencieuse (durée indiquée en millisecondes)
wo:rd	allongement de voyelle ou consonne
+< >	chevauchement de parole
(a)bout	formes réduites
+/ +"/.	citations
xxx	Mots inintelligibles
Conventions d'Analyse Conversationnelle (Jefferson, 2004)	
[!]	clic de la langue
.hhh	inspiration
hhh	expiration
creaky	crissement de la voix
(())	description d'actions
Annotation gestuelle (Kendon, 2004)	
~ ~ ~	Préparation de l'apex
***	apex
***	geste suspendu
-.-.-	retraction du geste

2.3. Système d'annotation et outils utilisés

Dans la lignée du travail de Crible et al. (2019) et de notre étude pilote (Kosmala & Morgenstern, 2017) nous avons mis en place une grille d'annotation pour nos traitements quantitatifs, qui repose essentiellement sur trois niveaux : (1) le niveau du

fluenceme individuel, (2) le niveau de la séquence de fluencemes (i.e. fluencemes combinés), et (3) le niveau visuo-gestuel (cf Figure 26, Chap. 2, section II. 2.1. de la thèse). Nous empruntons le terme « séquence » à Crible et al. (2019) pour parler de la structure interne des fluencemes combinés.

Notre système d’annotation est largement inspiré des travaux de Crible et al. (2019) et se situe dans la continuité d’autres études issues du domaine de la (dis)fluence (e.g. Allwood et al., 1990 ; Lickley, 2015 ; Shriberg, 1994). Il intègre également une dimension visuo-gestuelle, jusqu’ici peu étudiée. Notre grille d’annotation est décrite en détail dans la section II du Chapitre 2 de la thèse, et cible un certain nombre de variables, synthétisées ci-dessous :

Tableau 2. *Modèle d’annotation quantitatif synthétisé*

Niveau du fluenceme
<ul style="list-style-type: none"> - Forme du marqueur (<i>pause remplie, pause silencieuse, allongements, répétitions, réparations, auto-interruptions, troncations, marqueurs explicites d’édition</i> etc.) - Type de marqueur employé (<i>marqueurs vocaux, morphosyntaxique, ou non-lexical</i>) - Durée des marqueurs vocaux (en ms)
Niveau de la séquence
<ul style="list-style-type: none"> - Type de séquence de fluencemes (<i>simple, complexe</i>) - Longueur de la séquence (i.e. nombre de marqueurs combinés au sein d’une séquence. - Modèle de combinaison (i.e. type de combinaison récurrente) - Position dans l’énoncé (<i>initial, médial, final, isolé</i>) - Gestion de la communication (<i>communication management</i>, d’après Allwood et al. 1990 ; fonction des fluencemes, soit intra- ou interpersonnelle)
Niveau visuo-gestuel
<ul style="list-style-type: none"> - Phase gestuelle (<i>gesture phase</i>, position de repos, préparation, apex...) - Type de geste (pragmatique, référentiel) - Sous-type (représentationnel, deictique-anaphorique, interactif, discursif, geste de réflexion/recherche) - Direction du regard (vers l’interlocuteur, sur le côté, vers les notes de papier)

Les données ont d’abord été transcrites avec le logiciel CLAN (MacWhinney, 2000), puis transférées sur ELAN (Sloetjes & Wittenburg, 2008) pour l’annotation des catégories décrites ci-dessus. Notre grille de codage a ensuite été convertie sur Excel pour effectuer les traitements statistiques (fréquence, proportion, moyenne...), et les

analyses ont ensuite été effectuées à l'aide de calculateurs statistiques (t test, Z test, log-likelihood, Wilcoxon, etc., voir section II. 2.3.1. du Chapitre 2).

Pour les analyses qualitatives, nous empruntons nos outils à l'analyse conversationnelle (Mazeland, 2006; Mondada, 2007; Pekarek Doehler, 2006; Sacks et al., 1974), avec des principes tels que le cadre participatif (Goodwin, 2007) les mécanismes de tours de parole, la notion de structure préférentielle, les paires adjacentes, et l'alignement. Nos analyses qualitatives permettent également de mettre en lumière la dimension multimodale de l'inter-(dis)fluence, en nous appuyant sur l'analyse des gestes et leur coordination avec les fluencemes verbaux. Le logiciel PRAAT a également été employé dans certaines analyses pour rendre compte de certains paramètres prosodiques, tels que la courbe intonative (voir Chap. 2, section II. 3.2.).

En bref, nos analyses reposent à la fois sur des méthodes quantitatives et qualitatives, et empruntent des outils méthodologiques à plusieurs disciplines (linguistique de corpus, analyse conversationnelle, études gestuelles, et l'étude de la prosodie). Ces méthodes sont appliquées à nos deux études de corpus, qui sont synthétisées dans les parties suivantes.

III. L'inter-(dis)fluence en langue première et seconde : étude du Corpus SITAF

Le chapitre 3 de la thèse est consacré à l'étude du corpus SITAF, et s'intéresse d'une part à la distribution des fluencemes et des gestes en L1 et en L2, et d'autre part à la relation entre fluence, compétence de langue, compétence interactionnelle, intention pédagogique, et stratégie communicationnelle. Nos questions de recherche et résultats sont synthétisés dans les deux sections suivantes.

3.1. Questions de recherche et hypothèses

Nos questions de recherche et hypothèses sont issues de notre revue de la littérature qui couvre les domaines de l'acquisition des langues secondes, la didactique des langues, l'analyse conversationnelle, et les études gestuelles (cf Chap. 3, section I). Dans cette revue, nous argumentons qu'un grand nombre d'études sur la fluence en acquisition des langues seconde repose essentiellement sur des variables temporelles et des indices de compétence langagière (Cucchiaroni et al., 2000; Götz, 2013; Hilton,

2009; Lennon, 1990). Nous avons associé ici la notion de fluence à la notion de *compétence interactionnelle* (Galaczi, 2014) qui se concentre sur la capacité à co-construire l'interaction, plutôt que la gestion individuelle de problèmes de production. Cette idée rejoint également le concept de « CA-SLA » (Pekarek Doehler, 2006), l'étude de l'acquisition des langues secondes sous le prisme de l'analyse conversationnelle. Cette position assume que les processus cognitifs liés à l'acquisition d'une langue sont également « médiatisés par les pratiques sociales » (Pekarek Doehler, 2006, p. 128). Dans le cas du corpus SITAF, ces pratiques sociales peuvent également être considérées comme discours « didactique » avec la notion de didacticité seconde explorée par Moirand (1993) qui fait référence aux discours qui ne sont pas didactiques en soi (e.g. contexte scolaire) mais qui relèvent d'une intention didactique (e.g. se faire comprendre, aider l'autre etc.). Nous mettons dès lors l'accent sur l'étude de la fluence et la gestuelle dans le cadre d'analyses contextualisées, imbriquées dans l'action, afin de dépasser la notion trop restrictive de la fluence comme compétence langagière, et rendre compte du caractère contextuel et interactionnel de ces phénomènes.

Dans le domaine des études gestuelles, on retrouve un nombre d'études en acquisition des langues secondes qui s'intéressent au taux des gestes en L1 et L2, avec des théories telles que la *Lexical Retrieval Hypothesis* (Krauss et al., 2000). Cette théorie stipule que la production de gestes référentiels faciliterait l'accès lexical ; cela expliquerait pourquoi les apprenants produisent plus de gestes dans leur L2 que leur L1 afin de « compenser » des déficits langagiers ou difficultés lexicales (e.g. Stam, 2001). D'autres études plus récentes (Graziano & Gullberg, 2013, 2018) réfutent cette théorie et s'intéressent plutôt à la manière dont les gestes pragmatiques se synchronisent avec la parole (dis)fluente en L2. Nos questions de recherche se situent dans la continuité de ces études, et notre objectif principal a été de mesurer les potentielles différences distributionnelles entre la L1 et la L2 via des variables temporelles, positionnelles, séquentielles, interactionnelles, et visuo-gestuelles. Nos principales hypothèses peuvent se résumer ainsi (voir Chap. 3, section I. 1.4):

- Plusieurs différences entre la fluence L1 et L2 sont attendues, soit un taux de fluencemes plus élevé L2 qu'en L1, mais aussi des différences positionnelles, temporelles et séquentielles. Des différences individuelles et translinguistiques sont également attendues entre les groupes francophones et américains.

- Un taux de gestes plus élevé en L2 est également attendu, mais cela n'implique pas nécessairement les gestes référentiels, contrairement à la *Lexical Retrieval Hypothesis*. D'après Graziano & Gullberg (2018) nous nous attendons plutôt à un taux de gestes pragmatiques plus élevé pendant les fluencemes, quel que soit la langue.
- Pas de corrélation attendue entre le taux de fluencemes et le niveau de langue, mais plutôt un grand nombre de différences individuelles. Même si l'on s'attend à un taux de fluencemes plus élevé en L2 qu'en L1, cela ne reflète pas systématiquement une pauvre maîtrise de la L2 ou des difficultés lexicales. Nous illustrons lors de nos analyses qualitatives l'importance du contexte et des stratégies interactionnelles mobilisées par les apprenant·es par le biais de fluencemes et autres ressources sémiotiques.

3.2. Synthèse des résultats quantitatifs et qualitatifs de l'étude

Un total de 3172 fluencemes a été relevé dans le corpus, avec 1173 en L1 et 1999 en L2 (dans les deux groupes). Cela équivaut à un taux de 42,2 fluencemes tous les 100 mots chez les Américain·es dans leur L2, et 16,1 dans leur L1, ce qui représente une différence significative ($LL = 495.45$, $p < 0.0001$). Chez les Français·es, on retrouve un taux de 23 fluencemes tous les 100 mots dans leur L1, et 33,1 dans leur L2, ce qui est également une différence significative ($LL = 48.69$; $p < 0.0001$).

On remarque des différences similaires dans la distribution des gestes, avec chez les Américain·es un taux en moyenne de 16 gestes ($N = 376$) tous les 100 mots en L2, versus 11 ($N = 375$) en L1 ($LL = 27.91$; $p < 0.0001$), et chez les Français·es un taux de 10 gestes tous les 100 mots en L1 ($N = 375$), et 15 dans leur L2 ($N = 275$; $LL = 34.34$; $p < 0.0001$). Ces différences de fréquence entre L1 et L2 confirment de manière générale nos hypothèses issues de la littérature, mais nos résultats montrent également des différences plus nuancées selon le type de variable étudiée dans notre étude quantitative (position, combinaison, durée type de geste etc.). Elles sont résumées dans le tableau ci-dessous, tiré du tableau 45 de la thèse (cf General Conclusion, section II. 2.3.).

Une grande partie de notre travail est constituée de micro analyses d'extraits en situation tandem, décrites plus en détail dans la section II.2.2 du chapitre 3. En bref, nos analyses montrent qu'un grand nombre d'apprenant·es mobilisent un certain

nombre de stratégies vocales et gestuelles pour remédier à des difficultés de production, marquer leur engagement dans la tâche en cours, ou co-construire le discours. Ces analyses mettent en lumière l'idée que les différences que l'on trouve entre la L1 et la L2 ne peuvent se restreindre à des différences de fréquence, mais doivent également inclure la richesse du contexte et les différences individuelles. Ce type d'analyse nous a également permis de déconstruire la notion de compétence en L2 en tant que processus cognitif individualisant qui se limite à des difficultés de production (Pekarek Doehler, 2018).

Tableau 3. Synthèse des résultats du Chapitre 3 sur SITAF

CARACTÉRISTIQUES VERBALES/VOCALES (FLUENCEMES)		
	L1	L2
Taux de fluencemes	Taux plus élevé en L2 qu'en L1 (pour les deux groupes)	
Distribution	plus d'auto-interruptions et de pauses silencieuses (groupe américain) plus d'auto-interruptions et allongements (groupe français)	plus de répétitions et de pauses remplies (groupe américain) plus de sons non-lexicaux (groupe français)
Type de séquence	plus grande proportion de séquences complexes (seulement pour le groupe américain, pas de différences significatives dans le groupe français)	plus grande proportion de séquences simples (seulement pour le groupe américain)
Position dans l'énoncé	plus grande proportion en position finale (groupe français)	plus grande proportion en position médiale (groupe américain)
Gestion de la communication (inter vs intrapersonnel)	pas de différence significative entre L1 et L2 (pour les deux groupes)	
CARACTÉRISTIQUES VISUO-GESTUELLES (PENDANT OU HORS FLUENCEMES)		
	L1	L2
Taux de gestes	Taux plus élevé en L2 qu'en L1 (pour les deux groupes)	
	Taux de gestes suspendus plus élevé pendant les fluencemes en L2 qu'en L1	
Distribution des gestes	plus de gestes référentiels (pour les deux groupes)	plus de gestes de recherche lexicale (pour les deux groupes)
Direction du regard	aucune différence significative	

Bien que l'on trouve un taux de fluencemes plus élevé en L2 qu'en L1 dans les deux groupes, aucune corrélation a été trouvée entre ce taux et la note d'auto-évaluation des apprenant-es, puisqu'un grand nombre de différences individuelles ont été mises en évidence dans les données. Cela confirme que la fluence L2 n'est pas seulement propre

à la compétence langagière, mais également à l'individu dans un contexte donné (De Jong, 2018).

Pour conclure, l'objectif principal de ce chapitre était à la fois d'offrir une vision globale de la fluence L2 dans la continuité des études précédentes en Acquisition des Langues Secondes, mais également d'offrir une vision plus orientée vers des mécanismes interactionnels multimodaux, afin de rendre compte de la multidimensionalité de ces phénomènes, qui opèrent à différents niveaux d'analyse. Nous avons également adopté ce type de démarche dans le chapitre 4, synthétisé dans la section suivante.

IV. L'inter-(dis)fluence à travers les situations de communication : étude du corpus DisReg

Le chapitre 4 expose notre deuxième étude quantitative sur le Corpus DisReg, en nous concentrant cette fois-ci sur des différences situationnelles, avec d'une part les présentations orales en classe, et d'autre part les conversations spontanées.

4.1. Questions de recherche et hypothèses

Ce chapitre commence par proposer une courte revue de la littérature problématisée qui couvre plusieurs domaines théoriques (psycholinguistique, analyse conversationnelle, sociolinguistique, études gestuelles) afin de formuler nos hypothèses et questions de recherche. Notre première observation porte sur l'identification des traits ou caractéristiques qui différencient les deux types de situation de notre corpus (présentations orales versus conversations). On remarque que dans la littérature les auteur·es proposent de manière globale des caractéristiques limitées à une seule dimension ; par exemple, dans certaines études en psycholinguistique, on oppose souvent les données « spontanées » aux données « préparées » (O'Shaughnessy, 1992; Shriberg, 2001, 1999), ou des situations « dialogiques » à d'autres « monologiques » (Bavelas et al., 2008; Michel, 2011). Par conséquent, on retrouve au sein de la littérature sur la (dis)fluence des résultats contradictoires portant sur les situations de dialogue/discours spontané et monologue/discours préparé (voir Witton-Davies, 2014). Certaines études montrent un taux de fluencemes plus élevé en situation de dialogue qu'en monologue (e.g. Duez,

1982; Skehan, 2001) tandis que d'autres montrent le contraire (e.g. Kowal et al., 1983; Michel, 2011).

Ce manque d'homogénéité dans la littérature souligne le besoin de considérer plusieurs dimensions du discours afin de caractériser ses potentielles différences situationnelles. En sociolinguistique, un des éléments essentiels à prendre en compte dans l'étude de la variation est le *style* (e.g. Labov, 1972), qui, selon Eskénazi (1993) comporte plusieurs axes ou dimensions. La notion de style incarne à la fois le registre de la situation, le niveau de familiarité avec l'interlocuteur ou l'interlocutrice, et le degré d'intelligibilité requis dans une tâche. De ce fait, cette notion ne se limite pas à un mode de communication, mais prend en compte plusieurs aspects sociaux propres à une situation donnée. De même, en Analyse Conversationnelle, on distingue les situations « ordinaires » des situations « institutionnelles » (Heritage, 1997), et ces dernières apportent un nombre de restrictions sur certaines pratiques sociales (l'organisation du tour de parole, choix lexical, cadre participatif etc.). En résumé, il existe un grand nombre de caractéristiques qui différencient une situation sociale d'une autre, et le but du Chapitre 4 est de prendre en compte ces caractéristiques afin de voir en quoi elles peuvent avoir une influence sur la distribution des fluencemes et des gestes. Peu de travaux en études gestuelles se sont intéressés à la notion de style, ou aux différences situationnelles (à part quelques exceptions, voir Bavelas et al., 2008), donc ce chapitre a également pour but de faire dialoguer plusieurs disciplines, en accord avec notre modèle théorique interdisciplinaire (voir section I de ce résumé, ou Chap. 1 de la thèse). Nos hypothèses, qui découlent de ces différentes études, sont résumées dans le tableau suivant, tiré du tableau 29 de la thèse (Chap. 4, section 1.4.). Elles prennent en compte les traits caractérisant les deux situations (via la distribution des fluencemes et du posturo-mimo-gestuel) ainsi que les potentiels facteurs sous-jacents.

Tableau 4. Synthèse des hypothèses formulées dans le Chapitre 4

Traits caractérisant les deux situations	Potentiel facteurs sous-jacents
Taux plus élevé de fluencemes lors des présentations que les conversations Regard plus souvent porté sur l'autre lors des conversations ; à l'inverse, regard plus souvent porté sur les notes de papier en classe	Manque d'interactivité et contraintes institutionnelles lors des présentations (temps limité, public non familial, format de présentation spécifique etc.)
Fluencemes produits en classe associés à des mécanismes cognitifs auto-centrés Longueur moyenne de l'énoncé plus élevée en classe qu'en conversation	Registre formel, productions intelligibles à moitié lues en classe
Taux de gestes interactifs plus élevé en conversation qu'en classe Fluencemes produits en conversation associés à des processus interactionnels	Haut niveau d'interactivité lors des conversations, échange entre des participantes et participants familiaux

4.2. Synthèse des résultats quantitatifs et qualitatifs de l'étude

La partie qui présente les résultats (section II du Chapitre 4) est structurée de la même manière que le Chapitre 3, et utilise les mêmes traitements statistiques, décrits plus en détail dans section II de ce résumé. En bref, un total de 2870 fluencemes a été relevé dans notre sélection du corpus, 1472 lors des présentations, et 1398 lors des conversations. Cela représente un taux de 28,4 fluencemes tous les 100 mots en classe, versus 20,4 en conversation ($LL = 325.38$, $p < 0.0001$). On retrouve également un taux de gestes plus élevé lors des présentations orales (10,2 tous les 100 mots, 528 occurrences) que lors des conversations (7,6 tous les 100 mots, 491 occurrences ; $LL = 20.46$, $p < 0.0001$). Ces résultats suggèrent des différences de fréquence en moyenne, mais le tableau ci-dessous présente les autres variables étudiées dans notre étude, qui illustrent d'autres caractéristiques (voir Tableau 45 de la thèse, General Conclusion, section II. 2.3.).

Il convient de noter que nos résultats portant sur les gestes sont en contradiction avec l'étude de Bavelas et al. (2008) qui ont montré un taux de gestes plus élevé en situation de dialogue que de monologue. À nouveau, cela souligne l'intérêt d'étudier d'autres caractéristiques propres à certaines situations de communication (e.g. le cadre, le rôle des participant-es les mécanismes de tour de parole etc.), qui ne peuvent se limiter à des différences de mode (e.g. dialogue vs monologue).

Tableau 5. Synthèse des résultats du Chapitre 4 sur DisReg

CARACTÉRISTIQUES VERBALES/VOCALES (FLUENCEMES)		
	classe	conversation
Taux de fluencemes	Taux plus élevé en classe qu'en conversation	
Distribution	plus de sons non-lexicaux, pauses remplies et pauses silencieuses (corrélées à la longueur moyenne d'énoncé)	plus de répétitions, interruptions et allongements
Type de séquence	plus grande proportion de séquences complexes	plus grande proportion de séquences simples
Position dans l'énoncé	plus grande proportion en position initiale	plus grande proportion en position finale
Gestion de la communication (inter vs intrapersonnel)	majorité de fluencemes intrapersonnels	plus grande proportion de fluencemes interpersonnelles
CARACTÉRISTIQUES VISUO-GESTUELLES (PENDANT OU HORS FLUENCEMES)		
	classe	conversation
Taux de gestes	Taux plus élevé en classe qu'en conversation	
Distribution des gestes	plus de gestes pragmatiques discursifs	plus de gestes interactifs
Direction du regard	majoritairement vers les notes de papier	plus souvent vers l'interlocuteur/interlocutrice

Certaines différences quantitatives sont également illustrées dans nos micro analyses d'extraits. Celles-ci montrent plus particulièrement la manière dont les étudiants et étudiantes en contexte de classe sont très peu orienté-es vers leur public, car ils ou elles focalisent toute leur attention sur le contenu de leur présentation. En effet, les objets qu'ils ou elles ont à leur disposition (leurs notes, ordinateur, livre etc.) deviennent un réel outil de communication qu'ils ou elles emploient pour la construction de leur discours. Tandis qu'ils ou elles produisent plus de gestes qu'en conversation, la majorité de ces gestes sont employés à des fins discursives, notamment pour segmenter le discours, ou mettre l'emphase sur certains mots. En bref, une majorité des fluencemes et gestes produits lors des présentations orales sont auto-centrés, intrapersonnels, et cela peut s'expliquer par un nombre de contraintes situationnelles (présentation notée en temps limité, pas de participation de la part du public). Cela souligne également que ce type de situation ne peut se réduire à l'étiquette de discours « préparé », car même si le contenu des présentations était en effet préparé à la maison, la performance orale elle, était bien « spontanée ». Par

ailleurs, même si ce type de discours est « monologique », cela ne devrait pas empêcher les étudiants et étudiantes de s'adresser à leur public qui est présent dans la salle. Inversement, en situation conversationnelle, un réel alignement s'opère entre les participants et participantes qui mobilisent gestes et parole pour co-construire le discours, en s'appuyant sur leur connaissance partagée. Ces différences situationnelles reflètent à nouveau la dualité des phénomènes d'inter-(dis)fluence qui peuvent à la fois être fluent, communicationnels, interactionnels, ou disfluents, marquant un désengagement dans l'échange. Ces analyses soulignent à nouveau la multidimensionalité de ces phénomènes, en accord avec notre modèle théorique.

Pour conclure, l'objectif principal de ce chapitre était de souligner notre vision multidimensionnelle et ambivalente des phénomènes d'inter-(dis)fluence en explorant leur distribution et fonctionnement à travers deux contextes de communication distincts. On retrouve en moyenne un taux de fluencemes beaucoup plus élevé en classe qu'en conversation, ce qui peut être dû à de nombreuses contraintes situationnelles, et qui confirme d'une part le lien étroit entre (dis)fluence et production de la parole. Nous soulignons néanmoins le besoin d'articuler analyses quantitatives et qualitatives qui permettent de mettre en lumière certains processus qui ne représentent qu'une minorité dans les traitements quantitatifs. Même si une grande majorité des fluencemes tendent à être auto-orientés, leur analyse contextualisée au sein de séquences interactionnelles ne devrait être ignorée, car elle souligne les différents rôles que ces marqueurs peuvent jouer en interaction multimodale. Cette dimension est d'autant plus exploitée dans le Chapitre 5 de la thèse, synthétisé dans la section suivante.

V. Inter-(dis)fluence et multimodalité

Dans les chapitres précédents, nous avons souligné l'importance des analyses qualitatives qui exploitent pleinement la richesse des ressources sémiotiques mobilisées dans une séquence précise ; le Chapitre 5 de la thèse se consacre entièrement à ce type d'analyse et propose des micro-analyses de 20 extraits tirés de nos deux corpus. L'objectif de ce chapitre est de souligner la dimension multimodale des phénomènes d'inter-(dis)fluence en explorant l'interrelation entre gestes, mouvements du corps, manipulation d'objets, et emploi de fluencemes verbaux en contexte. Notre première partie vise plus particulièrement la relation temporelle entre les fluencemes et les gestes, en nous intéressant aux cas de suspension du geste

synchronisé avec la suspension de la parole (par le biais de fluencemes), ou aux cas d'interruption de la parole synchronisé avec l'interruption du geste. Nous illustrons également d'autres cas où s'opère une coordination entre la préparation du geste et du discours (section I. 1.2., Chap. 5 de la thèse). En bref, ces exemples soulignent le lien étroit entre geste et parole, qui ne fonctionnent pas indépendamment l'un de l'autre, mais qui représentent un processus conjoint, permettant la construction du discours multimodal (Kendon, 2004). Nous soulignons à nouveau, en accord avec les théories de Cienki (2012), Goodwin (2000) et Müller & Tag (2010) (voir section I.1.3. de ce résumé), que les gestes ne sont pas employés pendant les fluencemes pour « compenser » une absence de la parole, mais qu'il s'agit plutôt d'un processus de sélection parmi les différentes ressources sémiotiques qui sont pertinentes à l'activité en cours.

Dans ce chapitre, nous avons également pris appui sur des méthodes plus formelles de l'analyse gestuelle, en suivant le travail de Müller et al. (2013), ce qui nous a permis d'identifier des variants gestuels propre à l'inter-(dis)fluence. Notre deuxième partie est consacrée à l'étude de différentes pratiques, soit celle du « doing thinking » (Heller, 2021) et des gestes de recherche lexicale (Gullberg, 2011, Tellier & Stam, 2012), mais aussi des pratiques intersubjectives propre à la co-construction du discours. Enfin, nous nous intéressons également aux différents modes de représentation gestuelle (Müller, 2014) employées en discours pour la construction référentielle. Ces différentes pratiques sont synthétisées dans le tableau de la section suivante.

5.1. Synthèse des principales pratiques multimodales récurrentes de l'inter-(dis)fluence

Notre synthèse, tirée du tableau 44 de la thèse (Chap. 5, section III. 3.2.) est présentée à nouveau ci-dessous.

Tableau 5. Synthèse des pratiques multimodales de l'inter-(dis)fluence

Pratique sociale récurrente incarnant l'inter-(dis)fluence	différentes ressources posturo-mimo-gestuelles
"Doing thinking", processus de réflexion, d'hésitation ou de recherche lexicale	"thinking posture" ; "thinking face" (Goodwin & Goodwin 1986) ; posture inflexible, froncement de sourcils
	contact physique sur soi (mains sur le visage, poing sur le menton)
	corps prostré, orienté sur le côté
	tête effectuant des mouvements vers la droite et la gauche
	mouvement du poignet cyclique continu
	claquement de doigt répétitif
Démonstration d'intersubjectivité, prise de positionnement	battement de doigt répétitif
	regard mutuel
	corps orienté vers l'autre
	haussement d'épaules
	paumes de la main levées vers le haut
(Co)-construction du sens	index relevé et orienté vers l'autre
	les mains rejouent une scène ("enact")
	les mains dessinent ou tracent une ligne dans l'espace ("trace")
	les mains représentent un objet dans son ensemble ("represent")
	les mains moulent la forme d'un objet ("mold")

L'objectif principal du Chapitre 5 était d'explorer la dimension multimodale de l'inter-(dis)fluence afin de dépasser la vision trop restrictive des phénomènes de « disfluence » telle qu'elle est décrite de manière traditionnelle dans la littérature. Cela nous a également invité à reconsidérer la notion de « langage » qui ne s'apparente pas uniquement à des processus verbaux, mais qui doit prendre en compte la richesse des ressources sémiotiques présentes lors de pratiques interactionnelles en contexte, imbriquées dans l'action. Nous soulignons dès lors que l'analyse multimodale mérite de trouver sa place au sein du domaine d'étude de la (dis)fluence, qui jusqu'ici a été très peu étudiée.

Conclusion

Pour conclure, cette thèse présente deux apports majeurs dans le domaine de la (dis)fluence, et des sciences du langage de manière plus générale, concernant notre modèle théorique et la méthodologie employée. D'une part, ce travail de thèse propose une vision multidimensionnelle de l'inter-(dis)fluence, qui considère la parole, l'interaction, et la posturo-mimo-gestuelle ; ces trois niveaux d'analyse sont en perpétuelle interaction, et s'ajustent les uns aux autres pour construire la fluence du discours.

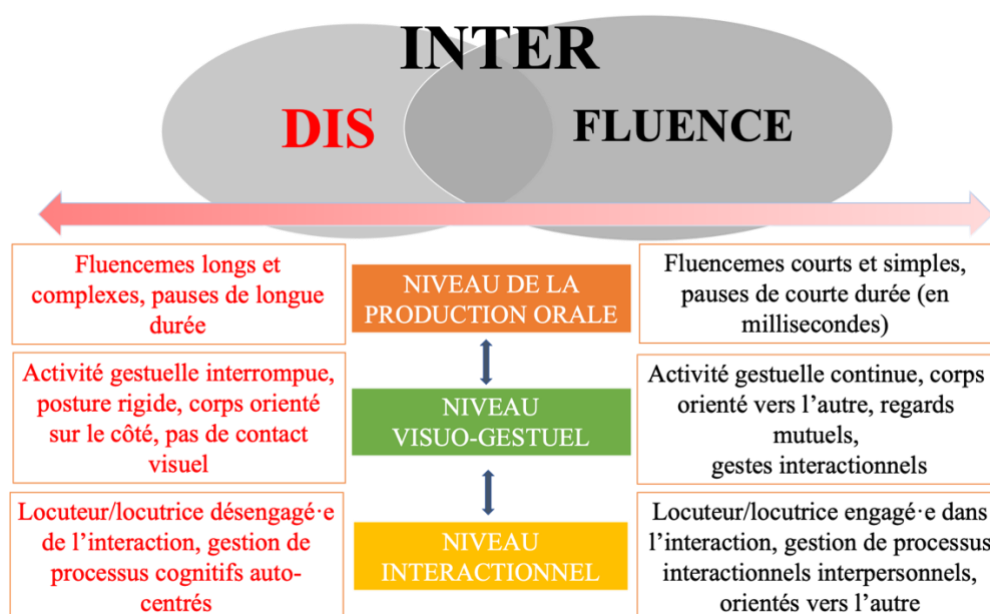


Fig. 2. Analyse de l'inter-(dis)fluence sur un continuum multi-niveaux

Par fluence, nous entendons à la fois le flux verbal de la parole, mais également la séquentialité de l'interaction et le flux gestuel. Notre vision conçoit ces phénomènes sur un continuum, allant de la fluence à la disfluence (dans la lignée de Crible et al. 2019) mais qui prend en compte nos trois dimensions, illustrées dans le schéma ci-dessus (tiré de la Fig. 71 de la thèse, *General Conclusion*, section III). Ce schéma met également en avant notre contribution méthodologique, qui prenait comme point de départ l'annotation des fluencemes verbaux, et de leurs variables temporelles, positionnelles et séquentielles, suivant la tradition des études en (dis)fluence (cf section I du résumé, ou Chap. 1 de la thèse). Notre analyse propose d'aller plus loin en incluant la dimension interactionnelle qui rend compte de l'organisation des tours de

parole, et leur structure progressive selon les contingences de l'interaction, en suivant les outils de l'Analyse Conversationnelle (cf section 1.3. du résumé). Enfin, nous prenons également en compte la dimension visuo-gestuelle qui considère la place des gestes manuels, des expressions faciales, de la direction du regard, et des mouvements du corps dans la construction de la fluence.

Toutes ces caractéristiques ont joué un rôle majeur dans le cadre de notre étude de corpus qui explore l'influence de la langue et du contexte dans la distribution des fluencemes et des gestes. Nos résultats prouvent l'intérêt d'une vision dynamique, fluide et adaptative des fluenceme en interaction multimodale.

Bibliographie

- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4), 307–331.
- Allwood, J. (2017). Fluency or disfluency? *Proceedings of DiSS 2017, the 8th Workshop on Disfluency in Spontaneous Speech*, 1.
- Allwood, J., Ahlsén, E., Lund, J., & Sundqvist, J. (2005). Multimodality in own communication management. *Proceedings from the Second Nordic Conference on Multimodal Communication*.
- Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58(2), 495–520.
- Beattie, G. W. (1979). *Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation*.
- Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22(3), 201–211.
- Betz, S. (2020). *Hesitations in Spoken Dialogue Systems* [Thèse de doctorat]. Bielefeld University.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, 44(2), 123–147.
- Buyse, L. (2019). ‘The (er) justice (er) the dep= you know the what’s it called’. The (dis) fluency of ‘you know’ in learner and native English. *Fluency and Disfluency Across Languages and Language Varieties*, 4, 243.
- Candea, M. (2000). *Contribution à l’étude des pauses silencieuses et des phénomènes dits « d’hésitation » en français oral spontané. Étude sur un corpus de récit en classe de français*. [Thèse de doctorat, Université Paris III- Sorbonne Nouvelle].
- Candea, M. (2017). *Pratiques de prononciation et enjeux sociaux. Approches post-variationnistes en sociophonétique du français de France* [Habilitation à diriger des recherches]. Université Grenoble Alpes.
- Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, 37(6), 871–887.
- Cienki, A. (2017). Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard*, 3(s1).
- Cienki, A. (2015). The dynamic scope of relevant behaviors in talk: A perspective from cognitive linguistics. *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication, August 6-8, 2014, Tartu, Estonia*, 110, 5–7.
- Cienki, A. J. (2012). Usage events of spoken language and the symbolic units we (may) abstract from them. *Cognitive Processes in Language*, 149–158.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201–242.
- Couper-Kuhlen, E., & Selting, M. (2001). Introducing interactional linguistics. *Studies in Interactional Linguistics*, 122.
- Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. John Benjamins Publishing.
- Crible, L., Dumont, A., Grosman, I., & Notarrigo, I. (2019). (Dis)fluency across spoken and signed languages: Application of an interoperable annotation scheme. In L. Degand, G. Gilquin, & A. C. Simon (Eds.), *Fluency and Disfluency across Languages and Language Varieties* Presses universitaires de Louvain.
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners’ fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, 107(2), 989–999.
- Danino, C. (2018). Les petits corpus—Introduction du numéro thématique. *Corpus*, 18.

- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, 54(2), 113–132.
- De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, 15(3), 237–254.
- Debras, C. (2017). The shrug: Forms and meanings of a compound enactment. *Gesture*, 16(1), 1–34.
- Debras, C. (2018). Petits et grands corpus en analyse linguistique des gestes. *Corpus*, 18.
- Debras, C., & Beaupoil-Hourdel, P. (2019). Gestualité et construction des chaînes de référence dans un corpus d'interactions tandem. *Cahiers de Praxématique*, 72.
- Debras, C., Beaupoil-Hourdel, P., Morgenstern, A., Horgues, C., & Scheuer, S. (2020). Corrective Feedback Sequences in Tandem Interactions: Multimodal Cues and Speakers' Positionings. In S. Raineri, M. Sekali, & A. Leroux (Eds.), *Linguistic correction/correctness*. Presse de Paris Nanterre.
- Debras, C., Horgues, C., & Scheuer, S. (2015). The multimodality of corrective feedback in tandem interactions. *Procedia-Social and Behavioral Sciences*, 212, 16–22.
- Deschamps, A. (1980). The syntactical distribution of pauses in English spoken as a second language by French students. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 255–262). Mouton.
- Disfluency: Interrupting speech and gesture*. (2006). [Thèse de doctorat]. Radboud University.
- Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A Psycholinguistic Perspective. *Studies in Second Language Acquisition*, 20(3), 349–385.
- Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, 25(1), 11–28.
- Duez, D. (2001). Signification des hésitations dans la parole spontanée. *Revue Parole*, 17–18.
- Eskénazi, M. (1993). Trends in speaking styles research. *Third European Conference on Speech Communication and Technology*.
- Fehrer, C., & Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica: Acta Societatis Linguisticae Europaeae*, 41(1–2), 37–72.
- Ferreira, F., & Bailey, K. G. D. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences*, 8(5), 231–237.
- Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, 19(5), 921–928.
- Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, 29(9), 1083–1096.
- Galaczi, E. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35(5), 553–574.
- Geeraerts, D. (2006). A rough guide to cognitive linguistics. In *Cognitive linguistics: Basic readings* (pp. 1–28). De Gruyter Mouton.
- Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference. In J. Romero-Trillo (Ed.), *Pragmatics and Corpus Linguistics: A Mutualistic Entente* (pp. 119–149). De Gruyter Mouton.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10), 1489–1522.
- Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse & Society*, 18(1), 53–73.
- Goodwin, C. (2010). Multimodality in human interaction. *Calidoscopio*, 8(2).
- Götz, S. (2013). *Fluency in native and nonnative English speech* (John Benjamins Publishing, Vol. 53). John Benjamins Publishing.
- Graziano, M., & Gullberg, M. (2013). Gesture production and speech fluency in competent speakers and language learners. *Presentado En TIGER, Tilburg University, Holanda*.

- Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology, 9*, 879.
- Grosman, I. (2018). *Évaluation contextuelle de la (dis) fluence en production et perception: Pratiques communicatives et formes prosodico-syntaxiques en français* [Thèse de Doctorat]. UCL-Université Catholique de Louvain.
- Guaïtella, I. (1993). Functional, acoustical and perceptual analysis of vocal hesitations in spontaneous speech. *ESCA Workshop on Prosody*.
- Gullberg, M. (2008). Gestures and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition* (pp. 276–305). Routledge.
- Gullberg, M. (2011). Multilingual multimodality: Communicative difficulties and their solutions in second-language use. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 137–151). Cambridge University Press.
- Hartsuiker, R. J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*.
- Heller, V. (2021). Embodied Displays of “Doing Thinking.” Epistemic and Interactive Functions of Thinking Displays in Children’s Argumentative Activities. *Frontiers in Psychology, 12*, 369.
- Heritage, J. (1997). Conversation analysis and institutional talk. *Handbook of Language and Social Interaction, 103*, 47.
- Hilton, H. (2009). Annotation and analyses of temporal aspects of spoken fluency. *Calico Journal, 26*(3), 644–661.
- Holmes, V. M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes, 3*(4), 323–361.
- Horgues, C., & Scheuer, S. (2015). Why some things are better done in tandem. In *Investigating English Pronunciation* (pp. 47–82). Springer.
- Jefferson, G. (2004). Glossary of transcript symbols. *Conversation Analysis: Studies from the First Generation. Amsterdam: John Benjamins*, 13–31.
- Jokinen, K., & Allwood, J. (2010). Hesitation in intercultural communication: Some observations and analyses on interpreting shoulder shrugging. In *Culture and computing* (pp. 55–70). Springer.
- Jucker, A. H. (2014). Uh and um as Planners in the Corpus of Historical American English. In I. Taavitsainen, M. Kytö, C. Claridge, & J. Smith (Eds.), *Developments in English: Expanding Electronic Evidence* (pp. 162–177). CUP.
- Kaushik, M., Trinkle, M., & Hashemi-Sakhtsari, A. (2010). Automatic detection and removal of disfluencies from spontaneous speech. *Proceedings of the Australasian International Conference on Speech Science and Technology (SST)*, 70.
- Kendon, A. (2004). *Gesture: Visible action as utterance* (Cambridge University Press). Cambridge University Press.
- Kjellmer, G. (2003). Hesitation. In defence of er and erm. *English Studies, 84*(2), 170–198.
- Kosmala, L. (2019). On the Multifunctionality and Multimodality of Silent Pauses in Native and Non-native Interactions. *Proceedings of the 1st International Seminar on the Foundations of Speech: Breathing, Pausing, and Voice*.
- Kosmala, L. (2020a). (Dis)fluencies and their contribution to the co-construction of meaning in native and non-native tandem interactions of French and English. *TIPA. Travaux Interdisciplinaires Sur La Parole et Le Langage, 36*, Article 36.
- Kosmala, L. (2020b). Euh le savez-vous ? Le rôle des (dis)fluences en contexte interactionnel : étude exploratoire et qualitative. *SHS Web of Conferences, 78*, 01018.
- Kosmala, L. (2021). On the Specificities of L1 and L2 (Dis)fluencies and the Interactional Multimodal Strategies of L2 Speakers in Tandem Interactions. *Journal of Monolingual and Bilingual Speech*.
- Kosmala, L., Candea, M., & Morgenstern, A. (2019). Synchronization of (Dis)fluent Speech and Gesture: A Multimodal Approach to (Dis)fluency. *Gesture and Speech in Interaction 6th Edition*.
- Kosmala, L., & Morgenstern, A. (2017). A preliminary study of hesitation phenomena in L1 and L2 productions: A multimodal approach. *Proceedings of DiSS 2017, the 8th Workshop on Disfluency in Spontaneous Speech*, 37.

- Kowal, S., Wiese, R., & O'Connell, D. C. (1983). The use of time in storytelling. *Language and Speech*, 26(4), 377–392.
- Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process. *Language and Gesture*, 2, 261.
- Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.
- Langacker, R. W. (1987). *Foundations of Cognitive grammar, vol. 1 theoretical prerequisites*. Stanford University press.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387–417.
- Lickley, R. J. (2015). Fluency and Disfluency. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 445–474). John Wiley.
- Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15(1), 19–44.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. transcription format and programs* (Vol. 1). Psychology Press.
- Mahl, G. F. (1956). Disturbances and silences in the patient's speech in psychotherapy. *The Journal of Abnormal and Social Psychology*, 53(1), 1.
- Mazeland, H. (2006). Conversation analysis. *Encyclopedia of Language and Linguistics*, 3, 153–162.
- McCarthy, M. (2009). Rethinking spoken fluency. *ELIA*, 9, 11–29.
- McNeill, D. (2000). *Language and gesture* (Vol. 2). Cambridge University Press.
- Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: Topic familiarity and disfluencies. *Journal of Communication Disorders*, 37(6), 489–503.
- Meteor, M. W., Taylor, A. A., MacIntyre, R., & Iyer, R. (1995). *Dysfluency annotation stylebook for the switchboard corpus*. University of Pennsylvania Philadelphia, PA.
- Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, 2, 141–173.
- Moirand, S. (1993). Autour de la notion de didacticité. *Les Carnets Du Cediscor. Publication Du Centre de Recherches Sur La Didacticité Des Discours Ordinaires*, 1, 9–20.
- Mondada, L. (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2), 194–225.
- Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, 51(1), 85–106.
- Morel, M.-A., & Danon-Boileau, L. (1998). *Grammaire de l'intonation l'exemple du français*. Editions OPHRYS.
- Morgenstern, A. (2014). Children's multimodal language development. *Manual of Language Acquisition*, 123–142.
- Morgenstern, A. (2019). Le développement multimodal du langage de l'enfant: Des premiers bourgeons aux constructions multimodales. *Multimodalité Du Langage Dans Les Interactions et l'acquisition*, 27.
- Morgenstern, A., Caët, S., Debras, C., Beaupoil-Hourdel, P., & Le Mené, M. (2021). Children's socialization to multi-party interactive practices. In L. Caronia (Ed.), *Language and Social Interaction at Home and School* (pp. 45–86). John Benjamins Publishing Company.
- Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family. *The Semantics and Pragmatics of Everyday Gestures*, 9, 233–256.
- Müller, C. (2014). Gesture as 'deliberate expressive movement.' *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, 127–151.
- Müller, C., Bressemer, J., & Ladewig, S. H. (2013). Towards a grammar of gestures: A form-based view. In C. Müller, A. Cienki, S. H. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body- Language—Communication. An International Handbook on Multimodality in Human Interaction* (Vol. 1, pp. 707–733). De Gruyter Mouton.
- Müller, C., & Tag, S. (2010). The dynamics of metaphor. *Foregrounding and Activation of Metaphoricity in Conversational Interaction. Cognitive Semiotics, Berlin et Al*, 10(6), 85–120.

- Nicholson, H. B. M. (2007). *Disfluency in dialogue: Attention, structure and function* [Thèse de Doctorat] University of Edinburgh.
- Ochs, E. (1979). Transcription as theory. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43–72). Academic Press.
- O’Shaughnessy, D. (1992). *Analysis of False Starts in Spontaneous Speech*. The International Conference on Spoken Language Processing, Banff, Alberta, Canada.
- Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, 9(1), 19–36.
- Pallaud, B., Bertrand, R., Prevot, L., Blache, P., & Rauzy, S. (2019). *Suspensive and Disfluent Self Interruptions in French Language Interactions*.
- Pallaud, B., Rauzy, S., & Blache, P. (2013). Auto-interruptions et disfluences en français parlé dans quatre corpus du CID. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, 29.
- Pekarek Doehler, S. (2006). «CA for SLA»: Analyse conversationnelle et recherche sur l’acquisition des langues. *Revue Française de Linguistique Appliquée*, 11(2), 123–137.
- Pekarek Doehler, S. (2018). Elaborations on L2 interactional competence: The development of L2 grammar-for-interaction. *Classroom Discourse*, 9(1), 3–24.
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14(4), 423–441.
- Sacks, H., Jefferson, G., & Schegloff, E. A. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Schachter, S., Christenfeld, N., & Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, 60(3), 362–367. <https://doi.org/10.1037/0022-3514.60.3.362>
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I* (Vol. 1). Cambridge university press.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54(2), 79–95.
- Shriberg, E. (2001). To ‘errrr’is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1), 153–169.
- Shriberg, E. E. (1994). *Preliminaries to a Theory of Speech Disfluencies* [Thèse de Doctorat]. University of California.
- Shriberg, E. E. (1999). *Phonetic consequences of speech disfluency*. SRI INTERNATIONAL MENLO PARK CA.
- Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second language learning, teaching, and testing* (pp. 167–185). Harlow: Longman.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. *6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32, 25–38.
- Stam, G. (2001). Lexical failure and gesture in second language development. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), *Oralité et gestualité: Interactions et comportements multimodaux dans la communication* (pp. 271–275). L’Harmattan.
- Stam, G., & Tellier, M. (2017). The sound of silence. *Why Gesture?: How the Hands Function in Speaking, Thinking and Communicating*, 7, 353.
- Sterponi, L., & Fasulo, A. (2010). “How to go on”: Intersubjectivity and progressivity in the communication of a child with autism. *Ethos*, 38(1), 116–142.
- Stivers, T. (2015). Coding social interaction: A heretical approach in conversation analysis? *Research on Language and Social Interaction*, 48(1), 1–19.
- Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning* (Vol. 2). John Benjamins Publishing.
- Streeck, J., Goodwin, C., & LeBaron, C. (2011). *Embodied interaction: Language and body in the material world*. Cambridge University Press.

- Tarone, E. (1980). Communication strategies, foreigner talk, and repair in interlanguage 1. *Language Learning*, 30(2), 417–428.
- Tashakkori, A., & Creswell, J. W. (2007). *The new era of mixed methods*. Sage Publications.
- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65(1), 71–79.
- Tellier, M., & Stam, G., (2012) Stratégies verbales et gestuelles dans l'explication lexicale d'un verbe d'action. In Rivière, V. *Spécificités et diversité des interactions didactiques*. Paris : Riveneuve éditions. 357 – 374.
- Tottie, G. (2011). Uh and Um as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, 16(2), 173–197.
- Tottie, G. (2014). On the use of uh and um in American English. *Functions of Language*, 21(1), 6–29.
- Tottie, G. (2016). Planning what to say: Uh and um among the pragmatic markers. In G. Kaltenböck, E. Keizer, & A. Lohmann (Eds.), *Outside the Clause. Form and function of extra-clausal constituents*. (pp. 97–122). John Benjamins.
- Vaughan, E., & Clancy, B. (2013). Small corpora and pragmatics. In *Yearbook of Corpus Linguistics and Pragmatics 2013* (pp. 53–73). Springer.
- Witton-Davies, G. (2014). *The study of fluency and its development in monologue and dialogue* [PhD Thesis]. Lancaster University.
- Yule, G. (1996). *Pragmatics*. Oxford University Press.

A multimodal contrastive study of (dis)fluency across languages and settings : Towards a multidimensional scale of inter-(dis)fluency

Abstract – The research presented in this thesis deals with so-called “disfluency” phenomena, a topic of study traditionally concerned with the annotation of a priori “disfluent” forms, such as “uh” and “um”, silences, repairs, repetitions, and the like, marking an interruption or a suspension in the verbal channel. More recently, a number of researchers have vouched for an ambivalent approach to these markers, also known as “fluencemes”, to uncover the potential for the same forms to serve both fluent and disfluent functions depending on the context. The present study is situated within this field of research, and offers an additional multimodal and interactional approach, taking into account the multiple modalities available to speakers in situated interactional practices, such as hand gestures, gaze, facial displays, or artefacts, used to build meaning in discourse. The purpose of this thesis is to go beyond production-oriented models of disfluency, and evaluate the degrees of fluency, fluidity, or flow, of face-to-face communication with a tridimensional scale, considering the levels of speech, gesture, and interaction. Our analysis targets more specifically the durational, positional, functional, sequential, and visual-gestural properties of fluencemes, and combines quantitative annotations with micro-analyses of the data. Based on a video dataset in French and English of university students engaged in different tasks across different settings and languages, this research shows that the construct of disfluency should not be restricted to the level of speech production, as it also exhibits recurrent interactive multimodal practices which are relevant to speakers’ language activities.

Keywords: *disfluency, fluency, gesture, interaction, second language acquisition, multimodality, register*

Une étude multimodale et contrastive de la (dis)fluence à travers les langues et contextes : Vers évaluation multidimensionnelle de l’inter-(dis)fluence

Résumé – Ce travail de thèse porte sur les phénomènes dits de « disfluence », un domaine de recherche qui s’appuie traditionnellement sur l’annotation de formes a priori « disfluentes », telles que « uh » et « um », les silences, les réparations, les répétitions, etc., qui marquent une interruption ou une suspension de la chaîne parlée. Plus récemment, des chercheurs ont mis en avant une approche ambivalente de ces marqueurs, aussi connus sous le nom de « fluencemes » afin de dévoiler le potentiel qu’ont ces mêmes formes à avoir des emplois à la fois fluents et disfluents selon les contextes de production. La présente étude se situe dans la continuité de cette démarche, et intègre une approche multimodale et interactionnelle, en prenant en compte les différentes modalités qui participent à la construction du discours, tels que les gestes, le regard, les expressions faciales, ou l’utilisation d’objets. L’objectif de cette thèse est d’évaluer les degrés de fluence dans la séquentialité de l’interaction multimodale, via une échelle tridimensionnelle qui considère la parole, la gestualité, et l’interaction. Notre analyse porte plus particulièrement sur les caractéristiques temporelles, positionnelles, fonctionnelles, et visuo-gestuelles des fluencemes, en combinant des annotations quantitatives et micro analyses des données. A partir d’un corpus vidéo en français et en anglais comprenant des échanges entre étudiants universitaires dans différentes langues et contextes, cette étude montre que la notion de disfluence ne saurait se réduire à une difficulté cognitive sur le plan verbal, puisqu’elle incarne également des pratiques interactives multimodales récurrentes et pertinentes aux activités langagières des locuteurs.

Mots-clés: *disfluence, fluence, gestualité, interaction, acquisition langue seconde, multimodalité, registre*

UNIVERSITÉ SORBONNE NOUVELLE

ED 625 – MAGIIE (Mondes Anglophones, Germanophones, Indiens, Iraniens et Études Européennes)
EA 4398 PRISMES (Langues, Textes, Arts et Cultures du Monde Anglophone)

Maison de La Recherche

4, rue des Irlandais, 75005 Paris