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

Loulou Kosmala

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 en sciences du langage
/ Doctoral dissertation in English Linguistics

Loulou KOSMALA

A MULTIMODAL CONTRASTIVE STUDY OF (DIS)FLUENCY ACROSS LANGUAGES AND SETTINGS

TOWARDS A MULTIDIMENSIONAL SCALE OF INTER-(DIS)FLUENCY

Thèse co-dirigée par / *Dissertation co-supervized by*
Madame la Professeure Aliyah MORGENSTERN
et / *and* Madame la Professeure Maria CANDEA
Soutenue le 3 décembre 2021 / *Defended on December 3rd 2021*

Jury/ *Committee* :
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

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 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, positionelles, fonctionelles, 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 recurrentes et pertinentes aux activités langagières des locuteurs.

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

Table of Contents

List of figures

List of tables

List of abbreviated terms

AM: American
CA: Conversation Analysis
CL: Cognitive Linguistics
EDT: Explicit Editing Phrase
EFL: English as a Foreign Language
FIG: Figure
FP: Filled Pause
FR: French
ICM: Interactive Communication Management
IL: Interactional Linguistics
IR: Identical Repetition
L : Line
L1: First Language
L2: Second Language
LH: Left Hand
LRH: Lexical Retrieval Hypothesis
MLU: Mean Length Of Utterance
MS: Morpho-syntactic Marker
NL: Non-Lexical Sound
NNS: Non-Native Speaker
NS: Native Speaker
OCM: Own Communication Management
EN: English
POH: Palm Open Hand
PUOH: Palm-Up Open Hand
PHW: Per Hundred Words
RH: Right Hand
PR: Prolongation
SLA: Second Language Acquisition
SI: Self-Interruption
SR: Self-repair
SRB: Scope of Relevant Behavior
TCU: Turn-Constructional Unit
TR: Truncated Word
UP: Unfilled Pause
VOC: Vocal Marker

Acknowledgments

This thesis could not have seen the light of day without the tremendous support of my two advisors Aliyah Morgenstern and Maria Candea, who have been guiding me for the past few years, even before my Ph.D when I started my master's degree. Their continuous guidance, enthusiasm, insight, and sympathy, whether it was online or in person, has truly helped me complete this journey. To them, I would like to express my deep and sincere gratitude. Thank you both for giving me the opportunity to do research and work under your guidance, and for giving me the theoretical and methodological tools that I needed to carry out this thesis. I would also like to thank Camille Debras, Gaëlle Ferré, Gaëtanelle Gilquin, Anne Salazar and Eve Sweetser for accepting to be part of my thesis committee.

This work would also not have been possible without the financial support of my university and my Doctoral School *MAGIIE*, as well as the research grant provided by the LABEX EFL, thanks to whom I successfully pursued my Ph.D and traveled to U.C Berkeley for my research project. I am especially indebted to Florence Baillet, the director of English, German, Indian, and European Doctoral studies for being involved in my project and for continuously showing her support. I also wish to thank all the participants of the DisReg Corpus who accepted to be filmed for my study.

My sincere thanks also go to members of the *SeSyLiA* workgroup from Sorbonne Nouvelle University, for offering me opportunities to work on diverse and exciting group projects, and participating in inspiring seminars. Special thanks to Céline Horgues and Sylwia Scheuer for giving me access to their wonderful corpus, but also for their valuable insight on tandem interactions. Many thanks to Charlotte Danino who has always supported me in my work, and who has given me opportunities to take part in truly interesting research projects, such as *Birth(ing) stories*. I am also forever indebted to Christelle Exare for her everlasting enthusiasm, and her thorough proofreading of this thesis. I would also like to express my gratitude towards Pauline Beaupoil Hourdel, who has been a sort of mentor since my research internship in 2016. She has helped me with many aspects of my work, and has always given me the best of advice throughout my study period. I also wish to thank other Ph.D students and young doctors from the department, mainly Alice, Eric, Manon, and Rachel. I also extend my thanks to another colleague from Nanterre University, Christophe Parisse,

for all his tips and techniques to help me with ELAN and statistical tests, and who always replied to my late-night emails.

I am also deeply grateful to members of the (dis)fluency community whom I have met over the years, and who have all helped me complete my study in their own way. Special thanks to my other mentor, Ludivine Crible, for her insightful comments, suggestions, and assistance at different stages of my research project. I would also like to thank other members of the Balibel group at Louvain-La-Neuve University, especially Lisebeth Degand for inviting me to her seminar in 2018, and again Gaëtanelle Gilquin for her continuous support since the (Dis)fluency Conference in 2017. Big thanks to friends and colleagues from Bielefeld (a place which actually does exist), Simon, Jana, and Loredana. Looking forward to our next "DisfluentThree" project. I would also like to thank Petra Wagner for inviting me to her workgroup meeting in 2018, and for her valuable suggestions at the early stages of my project. I also extend my sincere thanks to Ralph Rose, Gunnel Tottie, and Kerstin Fischer, whom I was always eager to meet at conferences, and who have also offered their guidance on various aspects of my work. I am also deeply thankful to Christelle Dodane, Ivana Didirkovà and Fabrice Hirsch for giving me the opportunity to work with them on (dis)fluency in typical and atypical speech as part of their *BENEPHIDIRE* program.

My thanks also go to the lovely people I have met during my stay at UC Berkeley and more specifically at the *Gesture and Multimodality Group*. I am again deeply grateful to Eve Sweetser for supervising and supporting me, even after I returned home following the health crisis of Covid-19. I wish I had had the chance to collect the data I wanted, but I am still forever grateful to all the opportunities she provided me. I would also like to offer special thanks to Laura Sterponi, Associate Professor at the School Education, for also taking me under her wing and offering her guidance. My thanks also go to Marjorie Harness (Candie) Goodwin, Professor of Anthropology at UCLA, for inviting me to her Co-Op Lab, and for her insightful comments and suggestions during the data session. I would also like to thank my dear friends Anne, "Snacks", and "Mini" Geovane, and wish we had had more time on campus, at Raleigh's, Cafe Leila's, or simply, together.

Big thanks to my London friends, Jessica, Angelica, James, and Matt, who were the very first participants who accepted to take part in my experiment as part of my

Master's project in 2015. Thank you for still being a part of my life, and for joining me at the pub, the theater, the coffee shop, or at home, at times when I needed you.

I cannot thank enough the wonderful friends I have made at the *Maison de La Recherche*, our second home (not to mention *La Montagne*), who have made this adventure truly memorable. Big thanks to my *comrades* Julien and Hugo, Bastien, Alex, Cindy, Jerem, Clémence, Mélaine, Didier, Elise-Anne and of course Étienne! I also thank Corrado and Camille who have been by my side since our fun adventure at ISMBS in Chania ("relax, ok?"). I will always cherish these memories.

I am also grateful to my friends and family who have been following me in the past ten years, even before I started my Ph.D, and who will always support me no matter what. To my dad, who gave me *Carnet de Thèse* as a present after I finished my bachelor's as he knew how much I wanted to pursue a Ph.D. To my mom, who always asks me about what I do, but who has always supported me nonetheless. To my sisters Mimi, Violette, and Mirabelle (*Master*), and my little brother, Félix. To my cousin, Sandra *(Bouyou)*, who will always be *a giant in our tiny, tiny love*. To my best friends from high school, Alex, Léa, and Eileen – thank you for sticking around and believing in me. To Elisa, a.k.a *Ili Anthropos,* who has been my partner in crime since our very first year at Censier, and Guillaume (*Guigui*), my role model, who has taught me so much – I am so glad you are part of my *multimodal adventure*.

My sincere thanks also go to the Rouaud family and their wonderful home in Quinsac where I finished writing the last words of this thesis. And of course, I wish to thank *my* Hugo, who despite his constant teasing, has always genuinely shown his support.

Lastly, I would like to thank you all, speakers of this world, for speaking, gesturing, laughing, saying *uhm*, and interacting in your first, and second, or third – who knows how many- language; without you I could have never found such a fascinating topic of research.

Singing for me is -s -s sweet relief (…) it is the only -s -s time when I (…) fff feel -f -f fluent.
(Megan Washington, *Ted Talk Radio Hour with NPR*, 2014)[1]

-Tout ça enfin de ce (…) masque euh non pas grec mais ce masque euh…
-Saganesque.
(Françoise Sagan, *l'élégance de vivre*, Arte documentary directed by Marie Brunet-Debaines, 2015)[2]

The laughter died out, and only gestures of arms, movements of bodies, could be seen shaping something in the room. Was it an argument? A bet on the boat races? Was it nothing of the sort? What was shaped by the arms and bodies moving in the twilight room?
(Virginia Woolf, *Jacob's Room*, 2008 Penguins edition, p. 56.)

[1] Retrieved from https://www.npr.org/2014/11/21/364151177/how-does-singing-help-achieve-stillness (August 16th 2021)
[2] Retrieved from https://www.arte.tv/fr/videos/073617-000-A/francoise-sagan-l-elegance-de-vivre/ (August 16th 2021)

Introduction

I. Introducing inter-(dis)fluency: beyond cognitive oriented models of speech production

What happens when we speak? To put it simply, according to previous models of word production (e.g. Levelt, 1983, 1989, 1999), our brains go through a series of mental operations, as we first select and activate a lexical item stored in our minds, assess its morphological and phonological code, then formulate and articulate the intended expression in the speech channel. While spoken languages typically follow a linear order, as speakers utter one word after the other in the acoustic channel, they are in fact governed by a number of nonlinear processes, since speakers constantly work on their production to re-shape the course of their delivery. They may pause to think about what to say next, re-start a previously uttered constituent, add a new item, or abandon a current utterance. Spoken languages are thus typically governed by cycles of what have commonly been labeled "fluent" (grammatically correct syntactic structures with lexical content) versus "disfluent" speech (non-lexical items which add no propositional content and disturb the fluidity of the surface structures). However, spoken speech production covers a multiplicity of genres, from theatrical performance, talk show, prayer, to spontaneous face-to-face encounters, film interview, political speech, etc., which inevitably affects how we speak. In early psycholinguistic work on disfluency phenomena, a common distinction was often drawn between "spontaneous" and "prepared" speech, with a clear focus on the recognition of disfluency phenomena in spontaneous productions , i.e., productions elicited spontaneously with no preparation beforehand, as opposed to read or laboratory speech. As many researchers (e.g. Bailey & Feirrera; Shriberg, 1994, among others) have claimed, unprepared spontaneous speech necessarily gives rise to many disfluencies, such as "uh" and "um", self-repairs, repetitions, truncations, and the like. Indeed, disfluencies are a very common part of human speech, and are said to occur at the rate of six to ten per hundred words (Bortfeld et al., 2001; Dollaghan & Campbell, 1992; Fox Tree, 1995; Shriberg, 1994). Disfluencies are virtually everywhere, not only in our spontaneous exchanges, but in TV shows, video games, or

animated films. These markers of spontaneity, which help distinguish between careful read speech and spontaneous speech events, are also what makes us human. In fact, a number of new technologies have emerged recently and aimed to model disfluencies for speech synthesis, such as the *Google Duplex*, an Artificial Intelligence system using conversational data and its many disfluencies to model natural sounding speech, to carry out "real world" tasks over the phone[3]. Speech disfluency has also been a popular topic of research in speech modelling and human-machine dialogue since the late 1990s, with the rise of computer technology (to name but a few, Betz et al., 2018; Eklund, 2004; Eklund & Shriberg, 1998). In sum, the study of disfluency is gaining more and more attention in various fields, such as psycholinguistics, cognitive science, or computational linguistics, and is now recognized as a legitimate topic of research, with the recurrent *DiSS* workshop (Disfluency in Spontaneous Speech)[4] first held in 1999, which brings together researchers from various academic disciplines interested in topics and issues surrounding disfluency. This common interest is mainly due to the large role disfluencies are said to play in speech production and comprehension, as Shriberg (2001, p. 53) reported: "disfluencies provide a window onto underlying processes affecting human speech and language production". But what exactly is disfluency? Where does the term come from? What are its implications for current linguistic research?

The notion of "dysfluency", which initially emerged in the 1950s in clinical linguistics to refer to stuttering phenomena, is now a common term in psycholinguistics to describe these processes which form an integral part of speech production. The term "disfluency" initially stems from a departure of the notion of *ideal delivery*, an expression used to describe the seemingly continuous *fluent* flow of speech, marked by an absence of "noise" in the signal or the presence of ungrammatical structures (Clark, 1996). Similarly, in Second Language Acquisition, the notion of "fluency" refers to ideally effortless native-like speech, which contains very few errors, as opposed to non-native speech, which is not yet "fluent" in the acquisition process (Lennon, 1990). In addition, "fluency" can also refer to the ability to produce wellformed expressions in a persuasive and stylistic manner, using rich vocabulary and eloquent utterances (Fillmore, 1976). As the present thesis will show,

[3] Information retrieved from https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html (August 16th, 2021)

[4] Information concerning the workshop series can be found in Ralph Rose's website https://filledpause.org/diss/ (last retrieved on August 16th, 2021)

the constructs of "fluency" versus "disfluency" have different definitions, and they have constantly been subject to a binary opposition in the literature across several theoretical fields, based on the monolithic and mythical assumption that a speaker is either fluent or disfluent, that a structure either reflects fluency or disfluency, or that certain phenomena are either deemed lexical (e.g. verbs, nouns, adjectives) or non-lexical (e.g., breathing, laughter, clicking sounds). The main goal of the present thesis is to go beyond this common opposition, and further explore the complexity of these phenomena by offering an integrated and multi-level approach, combining different frameworks and methodologies. The main issue with previous studies, as will be highlighted throughout this thesis, is that so-called disfluency markers (e.g., "uh", "um", pauses, repairs, repetitions, truncations, and the like) have too often been exclusively restricted to the level of *speech analysis*. But language is so much more than a series of spoken words in decontextualized utterances. Language is an embodied experience, grounded in our overall environment comprised of our own bodies, our movement in space, the people and objects around us, and our social background. Following the frameworks of interactional linguistics, linguistic anthropology, cognitive linguistics, and gesture studies, it will be maintained throughout this thesis that the complexity of human communication can only be fully understood in richly contextualized situations, where all semiotic features (voice, gaze, face, gesture, icons) can be deployed together at a specific moment in time to build meaning that is relevant to the task at hand. The present thesis is thus deeply influenced by the work of researchers who do not only consider language as a verbal or vocal phenomenon, but a multimodal one, deeply embedded within social structures (Candea, 2000, Cienki, 2015a, Goodwin, 1980, Morgenstern, 2014, Morgenstern et al., 2021, Streeck, 2009, Sweetser, 1998, among others).

These assumptions have a number of consequences for the study of so-called disfluency phenomena. The issue with the term "disfluency" is that it presupposes a problem to be fixed, or a disruption, which is too restrictive, and does not truly capture the complexity of these phenomena. A novel term will thus be introduced in this thesis to describe these processes, following previous authors' initiatives (Candea, 2000, Crible et al., 2019, Allwood, 2017, McCarthy, 2009) labeled *inter-(dis)fluency*. Following Crible et al's (2019) functionally ambivalent approach to (dis)fluency, the "dis" in brackets captures the ambivalence of fluency and disfluency phenomena, without systematically opposing them, hence regarding them as dynamic systems,

with the potential for the same a priori "disfluent" forms to serve both "fluent" and/or "disfluent" functions. The prefix "inter", as will be further explained in the first Chapter, captures the interactive process of *doing fluency*, or *confluence* (McCarthy, 2009) where several levels of analysis (speech, gesture, and interaction) are combined to build the overall flow of communication. The core term "fluency" is to be understood here not as second language proficiency, but as a metaphor embodying the notions of continuity, flow, progressivity, and fluidity. This metaphor will be found throughout our analyses. Let us illustrate this idea with a short example, which can be analyzed in many different ways, depending on the approach taken:

*SPK1: **&uh (..)** well **you'd have [//] you wouldn't &hav [/] you wouldn't [//] you'd have** to ask **&uh** other people on that.
*SPK1: **&uh** I always seem you know growing up in **&b +//.**
*SPK1: **I think [/] I think in fact [/] I think** that's my house over there.

These three utterances are taken from an interview recorded in 1992 with American film director Tim Burton, transcribed with CHAT transcription conventions (described in Chapter 2). An expert in disfluency research would commonly make the following observations regarding: the number of occurrences found in this passage (10 in total in bold, one pause, three filled pauses, two truncations, one repair, one self-interruption, two repetitions), where they are located in the utterance (utterance-initial, medial, and final), and whether they co-occur with one another or appear isolated, forming more complex sequences (e.g. a truncation clustered with a repair). Drawing from this type of analysis, we can make the preliminary observation that this particular speaker is highly *disfluent*, given the number of disfluencies found in his speech. We may thus wonder, how come this American native speaker, who was already quite famous at that time, and who was used to doing interviews on TV, produces so many "unwanted" "non-lexical" items in his speech? Does it reflect a lack of confidence? Stress and anxiety? Difficulties in grammatical encoding? While there is no straightforward answer to this question, we soon realize that this type of analysis is too restrictive, as it completely disregards other fundamental aspects of face-to-face communication, mainly visible bodily behavior and interactional dynamics. Let us now reconsider this example by taking into account these two additional layers of analysis, using a multimodal transcription system:

Tim Burton – 1992 Interview[5]

1 *INT: How strange were you as a child?
((gazes at TIM))

2 *TIM: uh (...) well you'd have you wouldn't h(ave) you wouldn't
((aborted gesture))
hhh. you'd have to ask uh other people on that
((brings down his right palm facing sideways in a rapid motion))


3 *TIM: uh I always seem you know growing up in B(urbank)
((bent fingers brought forward in his central gesture space with a series of repetitive beat motions))


4 *TIM: I think I think in fact I think that's my house over there.
((points to the transistor next to him with his index finger))

5 *INT: ((laughs))

From what this transcription shows us, Tim Burton was first asked a rather blunt question by the interviewer about his childhood (l.1), projecting a question-answer sequence. Tim Burton's utterance in line 2 is first delayed with a turn-initial filled pause clustered with an unfilled pause, indexing an initiation to take the turn and provide an answer to the interviewer's prior question. What is relevant to note here is that the director is not asked a question about his films or his career as a filmmaker,

[5] Tim Burton on Bob Costas, 1992 late night interview. The full video can be found at https://www.youtube.com/watch?v=krc8NVKtA0E&t=71s (last retrieved on August 16th, 2021)

but a personal one regarding his private life, accounting for a change of participant's status or discourse identity (from "film director" to perhaps "intimate friend"). All these features (turn-taking, participation framework, sequence formation, etc.,) which are highly common characteristics of talk-in-interaction in *Conversation Analysis* (Sacks et al., 1974), are altogether essential to further our understanding of this excerpt. In this view, Tim Burton is thus not solely a highly "disfluent" speaker, but the co-participant of a conversation, more specifically a film interview, who is recorded live on television, and who is expected to provide answers to a number of questions asked by the interviewer in a timely manner. He may decide, however, to display dispreferred actions by rejecting the current topic of conversation, which is what he does here, as he refrains from providing a straightforward answer to the interviewer's question. In this view, the notion of *fluency* may thus also be applied to the *flow* or *progressivity* of the interaction, going beyond the level of speech analysis.

We can further note that the speaker displayed different forms of participation towards the task at hand, reshaping the course of his emerging talk and redirecting his attention to external objects. These manifestations are conveyed in his visual-gestural behavior, as depicted in the illustrations within the transcript. We can see him moving his hands in space in an orchestrated manner with his speech: in line 2, Tim dismissed the previously asked question with a negative statement ("I wouldn't") then shifted the topic ("you'd have to ask other people on that") while producing a sort of "brushing away" gesture (Bressem & Müller, 2014), synchronized with an audible inbreath ("hhh.") moving his hand away from his body in a rapid motion, as to negatively assess the current topic of conversation. He further makes use of his surrounding space to build meaning with his hands: he places his right hand with bent fingers opposite him in the gesture space to refer to a specific location (probably his hometown Burbank, in line 3, but the word was truncated), and then points towards an external object (the transistor in the studio) to direct the interviewer's attention towards an imaginary location (his house). In sum, the speaker does so much more than producing a series of "disfluent" utterances in this excerpt, he in fact *languages* his experience (Morgenstern, 2020) with his hands, making use of his body and his surrounding material environment to build meaning. As Jürgen Streeck beautifully said, "the world of artifacts was not built by brains, but hands" (Streeck, 2020, p. 2). Hands, and arms, among other parts of the body (e.g. trunk and shoulders) thus play a fundamental role in the interactive process of building language. In this sense, spoken words are not

only generated by a series of mental processes working in the brain, they are captured in situated discourse, experienced by living embodied human beings in the world. While the study of embodiment and gesture is now being more and more recognized as a legitimate and full-fledged topic of research, with for instance the *International Society for Gesture Studies*[6], the journal *Gesture*[7] and the book series *Gesture Studies*[8], co-edited by Sotaro Kita and Adam Kendon, their relationship to (dis)fluency remains still quite underexplored in the literature.

To address this gap, it will be argued in this thesis that the constructs of fluency and disfluency should *not* be restricted to the level of speech production, reflecting a mental cognitive process associated with difficulty or uncertainty, but should include other fundamental features as well, illustrated in this short example. The speaker may have sounded "disfluent" from a verbal or vocal perspective, but he also performed a series of *fluid* actions embodied in his *gestural flow*. So where exactly do we draw the line between "fluency" and "disfluency"? Is it in fact even relevant to oppose these notions? The present thesis calls for further investigation and stresses the need to consider all aspects of multimodal communication to explore the complexity of inter-(dis)fluency phenomena.

II. Data under study

The aim of this thesis is to explore the ways the different dimensions outlined above (speech, gesture, interaction) may interact with one another to build the fluency, fluidity, or flow, of multimodal discourse, by targeting different languages and types of situations in a dataset of videorecorded productions. Starting with the assumption that inter-(dis)fluency is a multimodal, dynamic and ambivalent system, we expect a high degree of variability and dispersion in the distribution of (dis)fluency markers (which will be labelled *fluencemes*, further explained in Chapters 1 and 2), as well as gestures, according to task type, or language. To that aim, this study will focus on a specialized video dataset of university students engaged in different activities in their first and second language.

The first dataset under study is a selected sample of the SITAF Corpus (Horgues & Scheuer, 2015) which includes video recordings of 21 French and American students

[6] https://www.gesturestudies.com (las consulted on August 17th 2021)
[7] https://benjamins.com/catalog/gest (last consulted on August 17th 2021)
[8] https://benjamins.com/catalog/gs (last consulted on August 17th 2021)

from Sorbonne Nouvelle University engaged in an argumentative task. The students interacted in pairs in L1-L2 settings, alternating between their first and second language in French and English. The students were asked to debate on a given topic and decide on their level of agreement. The pair knew each other fairly well, since they met once a month as part of a tandem exchange program to practice their second language. During the exchanges, they were thus invited to share and co-construct ideas on a given topic (i.e. do prisoners have the right to vote, are teenage years the best years of your life, etc.) leading to joint multimodal productions.

The second dataset under study is a selected sample of the *DisReg Corpus,* collected by myself for this dissertation (Kosmala 2020a), which comprises video recordings of 12 French students from Sorbonne Nouvelle University engaged in two different tasks in different settings. The students were first recorded during their presentation of a graded oral assignment in class, performed in front of the teacher and the whole classroom. They were then recorded face-to-face in pairs, and asked to discuss everyday topics (last film seen on TV, funny anecdote at university, etc.). Just like the participants from the SITAF Corpus, the pairs knew each other from university, and could hence discuss their common experience and display several tokens of understanding, leading to the *co-fluency* of discourse.

This dataset was compiled for its multimodal quality, as well as for the number of similarities found between the two data samples (similar speaker profiles, discourse identities, university setting, etc., see Chapter 2) which enables us to triangulate evidence from language proficiency, setting, task type, genre, and setting, as to capture the multifunctionality and multimodality of inter-(dis)fluency across different contexts of use. In addition, a *mixed-method* methodology will be used, relying on quantitative annotations of fluencemes and gestures and several variables (form, position, co-occurrence, duration, gesture type etc.,), coupled with rich multimodal qualitative analyses of a selection of excerpts. This type of methodology allows us to capture general tendencies and patterns of behavior of (dis)fluency and gesture found across different groups of speakers on average, with the aim to capture the complexity of inter-(dis)fluency in situated interactive sequences.

Overall, the present study is based on the quantitative annotation of 3172 fluencemes and 2381 hand gestures, as well as a total of 40 qualitative analyses. Since the aim of this thesis is to shed light on the different dimensions of fluency (speech, interaction, and gesture), some excerpts will purposefully be analyzed multiple times

across Chapters 3, 4 and 5, in order to zoom in and out on specific features that will be relevant to the topics explored in the chapters.

III. Preview of the thesis

The present thesis comprises five different chapters, which all target both theoretical and methodological issues regarding the multimodal and multidimensional status of fluency, as well as its place in linguistic research. Each chapter thus contributes to the development of our integrated and multi-level approach to inter-(dis)fluency phenomena, and aims to go beyond narrow definitions of "fluency" and "disfluency" found in previous research, integrating a number of relevant theoretical frameworks introduced in Chapter 1.

The first chapter is thus mainly theoretical, reviewing a number of interdisciplinary research fields relevant to the study of (dis)fluency, mainly psycholinguistics, usage-based linguistics, gesture studies and interactional linguistics. This chapter also contributes to the current terminological debate regarding the use of the term "disfluency" in the literature, further justifying our choice of terminology, leading us to the construction of our integrated framework.

The second chapter is methodological, and presents our mixed-methods methodology which relies on quantitative treatments performed with several tools and softwares (ELAN, CLAN, Excel, and statistical tests) as well as micro analyses of the data, using conversation-analytic tools borrowed from Conversational Analysis (Mondada, 2007, Sacks et al,, 1974). Based on previous (dis)fluency coding schemes and gesture classification systems, we present our annotation model, targeting three different levels of analysis (individual fluenceme, fluenceme sequence, and visual-gestural level). This chapter also explains our motivations for working on a specialized dataset, and discusses the different transcription methods and units of transcriptions used for the purposes of multimodal speech annotation.

Our last three chapters are empirical, and present the results of our two corpus-based studies conducted on the *SITAF* and *DisReg Corpus*. Some of the analyses presented in these chapters already feature in previously published work (Kosmala, 2019, 2020a, 2020b, 2021; Kosmala et al., 2019), and they have been re-examined more deeply and adapted to the present thesis.

In Chapter 3, we focus on the *SITAF Corpus*, targeting aspects of native versus non-native language use. Based on a literature review of the L2 Fluency and gesture

literature, we formulate a number of research questions and hypotheses regarding the distribution of fluencemes and gestures in L1 and L2 in French and English, and apply the annotation model presented in Chapter 2 to generate our quantitative findings. These findings are then further exploited with fine-grained qualitative analyses of the data, drawing potential relations between the notions of fluency, language proficiency, pedagogical intention, and interactional competence.

Chapter 4, which focuses on the DisReg Corpus, follows the same structure as Chapter 3. Our research questions and hypotheses stem from a brief review of the literature regarding the different variables affecting (dis)fluency and gesture in discourse, mainly type of delivery, mode of speech, language style, and social setting, painting a complex picture of the various features characterizing different speech situations. The aim of this chapter is hence to describe potential differences in fluency and visual-gestural behavior across two distinct styles and settings (individual graded class presentations versus face-to-face casual conversations). Just like Chapter 3, we rely on quantitative and qualitative analyses, drawing relations between fluency, gesture, audience design, and setting.

Lastly, in Chapter 5, we present a number of analyses from the two data samples, focusing this time exclusively on micro qualitative analyses to further explore the multimodality of inter-(dis)fluency phenomena, and shed light on the multiple features affecting their use in discourse, thus going beyond our previous quantitative annotations. We document the different forms and functions of gestures co-occurring with fluencemes or within their vicinity, and establish a typology of the gestural variants found in the data in relation to inter-(dis)fluency across several recurrent social practices. This chapter further questions the notion of "language", and our understanding of it in linguistic research, giving more support to the multimodal status of (dis)fluency and its place in gesture research.

All our findings are then discussed in the *General Conclusion*, where we summarize the main results obtained in Chapters 3 and 4, and compare the two datasets. A number of recurrent characteristics differentiating fluency and gesture behavior across languages and settings is further presented, based on the different variables used in our annotation model. Specific attention is also paid to individual differences, which play a fundamental role in corpus-based research, and which will further reveal that fluency and gesture are in part language- and speaker- specific, not restricted to general tendencies or average scores. Finally, we conclude this thesis with

the presentation of our multidimensional scale of inter-(dis)fluency, evaluating different degrees of fluency and disfluency in a tridimensional continuum based on the different dimensions explored throughout the chapters (speech, gesture, interaction) hence going beyond our common, but restricted, understanding of so-called disfluency phenomena in linguistic research.

Chapter 1. Theoretical Background

Introduction to the chapter

The notions of fluency and disfluency have been largely associated with three different disciplines, mainly (1) clinical psycholinguistics, (2) second language acquisition, and (3) computational linguistics (Grosman, 2018, p. 7). Consequently, several distinct approaches to disfluency have emerged, resulting in radically different views and theoretical implications of the same phenomena. For instance, the term "disfluency" can refer to a performance error on the one hand, or as a strategic signaling device on the other. These opposite views reflect the functional ambivalence of these phenomena (Crible et al., 2019; Götz, 2013), which cannot easily be categorized under one simple label. The aim of this chapter is to discuss the different theoretical and methodological frameworks grounding the concepts of fluency and disfluency in the study of spoken language. It will review very different theoretical backgrounds, such as psycholinguistics, cognitive usage-based linguistics, and second language acquisition; but other academic fields will also be reviewed, some of which have not as often been in the scope of "traditional" disfluency research, mainly interactional linguistics and gesture studies. The overview of these various theoretical approaches will highlight the need to view inter-(dis)fluency phenomena as complex, dynamic, multi-level, and multi-modal processes.

The choice of the terms (*dis)fluency and inter-(dis)fluency* (which will be adopted throughout this dissertation) will reflect our mixed theoretical and methodological approach, and will be compared to the various terms that have been used in the literature. This chapter thus also aims to contribute to the terminological debate regarding the use of the term "disfluency". We may ask ourselves whether it is really accurate to use such a "negative" (but common) term, which originally referred to stuttering and verbal blundering, when dealing with spontaneous face-to-face spoken interaction. Isn't it time, as Allwood (2017) or Tottie (2014) suggested, to make a change in the terminology? While there is no straightforward answer to this question, this chapter will still attempt to paint a consistent picture of the diverse and complex phenomena related to the construct of fluency and disfluency. This multi-approach

construct, as we shall see, is not restricted to the verbal and vocal dimensions of speech, but will incorporate other relevant visible features of face-to-face communication as well. This chapter is structured as follows: first, we begin with the traditional production-based approach to disfluency phenomena, related to speech planning and cognitive processes; secondly, we sketch out the different terms associated with these phenomena and their different approaches, and we discuss our choice of terminology; thirdly, we review three other theoretical frameworks (Cognitive Grammar, Interactional Linguistics, and Gesture Studies) relevant to the present study of inter-(dis)fluency. Finally, we conclude this chapter with an introduction to our integrated framework, and address our main theoretical assumptions.

I. What is Disfluency? A Psycholinguistic Production Model

The study of disfluency has been analyzed thoroughly over the past sixty years by a number of researchers in different academic fields, such as speech pathologists, speech scientists, phoneticians, and psycholinguists (Goldman-Eisler, 1958; Johnson, 1961; Lickley, 2015; Maclay & Osgood, 1959; Shriberg, 1994; to name but a few). The present section will focus on the theoretical framework mainly adopted by (but not restricted to[9]) psycholinguists, and will present a brief review of their theoretical and methodological approaches to disfluency[10].

1.1. Disfluency as a deviation in speech from the ideal delivery

In the field of pycholinguistics, the study of disfluency, or *speech errors* (see Levelt, 1983, 1989; Menn & Dronkers, 2016) has been of particular interest with regard to how speakers' brains create and understand meaningful language, based on different speech production models (Menn & Dronkers, 2016). These models incorporate several levels of processing, such as the conceptual level, the functional level, the positional level, or the phonological level (Men & Dronkers, 2016). As briefly noted in the *Introduction*, before speaking, speakers need to choose what kind of information

[9] This section will also mention the work of computational linguists (e.g. Shriberg, 1994).
[10] The field of pathological "dysfluency" (typically concerned with pathological speech such as stuttering or aphasia) goes beyond the scope of this dissertation so it will not be reviewed in this chapter.

they are going to put into words, and to do so, they must go through several stages of production and execution. Psycholinguist Levelt (1983, 1989, 1999) dedicated most of his work to the study of speech production, and took into account several aspects of speech, such as the constraints on conversational appropriateness, the processes of articulation, and the use of self-monitoring. His introduction to a paper discussing word production models (1999, p. 223) begins as follows:

> How do we generate spoken words? This issue is a fascinating one. In normal fluent conversation we produce two to three words per second, which amounts to about four syllables and ten or twelve phonemes per second. These words are continuously selected from a huge repository, the mental lexicon, which contains at least 50-100 thousand words in a normal, literate adult person. Even so, the high speed and complexity of word production does not seem to make it particularly error-prone. We err, on average, no more than once or twice in 1000 words.

The study of word production thus lies at the forefront of the analysis of human spoken speech. Levelt (1983, p. 305) discusses, for example, the *speakers' linearization problem*, mainly that the channel of speech prohibits the "simultaneous expression of multiple propositions"; consequently, a linear order has to be determined, which compels speakers to constantly work on their production. In order to do so, speakers can monitor their own speech (Levelt 1983) by following several steps (*message construction*, *formulating*, *articulating*, *parsing*, and *monitoring*). Self-monitoring can take two forms, (1) overt repairs, and (2) covert repairs. Overt repairs involve morphological changes (e.g. a truncated morpheme) while covert repairs only constitute an interruption point without change (e.g. a pause). Crible (2017) further discusses this linearization issue and argues that, even though spoken speech constitutes a linear stretch of phonemes, the act of speaking itself involves many non-linear processes. The distinction between "overt" and "covert" repairs made by Levelt (1983) actually reflects this notion of non-linearity (Crible, 2017, p. 18):

> The process of monitoring either one's own or someone else's speech involves playing with and moving along the linear articulatory channel, either backwards for retracing and reformulating, or forwards by announcing upcoming material.

Another author, Clark (1996, p. 253), also pointed out the nonlinearity of spoken speech:

> Utterances are nonlinear and have more than one track. Connie may produce an expression, change her mind, and start over. She may make a mistake and repair it.

Thus arises the notion of "ideal delivery": every use of a word, phrase or sentence has an ideal delivery, defined as "a single action with no suspensions – no silent pauses, no fillers, no repeats, no self-corrections, no delays except for those required by the syntax of the sentence" (Clark, 1996 p.253). When conversing, the primary goal of speakers is to produce maximally acceptable speech in both content and form (Hieke, 1981, p. 150). However, despite speakers' efforts to be in control of the communication channel and remain intelligible, they are often obliged to stop, backtrack, or interrupt their current planning, which often leads to utterances which turn out to be very different from their initial ideal delivery (Hieke, 1980). Given the fluctuating nature of speech production (moving backwards and forwards from *conceptualization* to *articulation*) it first appears that the construct of disfluency is originally based on a reflection of the non-linearity of speech, but also on a departure from the representation of this ideal delivery. In fact, Ferreira & Bailey (2004, p. 234) defined disfluency as "any deviation in speech from the ideal delivery". They also claimed that some disfluencies (such as repeats, abandonments and repairs) create *ungrammatical utterances*, which further implies that disfluency constitutes a deviation from an ideal, grammatical, and well-formed utterance.

1.2. The role of disfluencies in speech production

Disfluency, "when speech breaks down" (Lickley 2015, p. 12), thus very often describes an interruption in the speech flow (Fox Tree, 1995; Merlo & Mansur, 2004). As a matter of fact, Levelt (1989) mentions, in his work on self-repairs, *The Main Interruption Rule* (discussed by Nooteboom, 1980) whose principle is to "stop the flow of speech immediately upon detecting the occasion of repair". In other words, whenever (phonological or lexical) "trouble" in speech is detected, speakers must immediately interrupt it, which will result in a "disfluent" utterance. Such utterances tend to follow the same surface structure (Shriberg 1994) which can be divided into different parts, or regions: (1) the original delivery, (2) the reparandum, (3) the edit term, (4) repair, and (5) resumption (Fig. 1). Despite terminological differences, there seems to be a general consensus on the surface structure underlying disfluency phenomena, which is illustrated in the following figures. Figure 2 is taken from the

work of Shriberg (1994) which was originally based on Levelt (1989). This model was later taken up by Pallaud et al., (2019) who worked specifically on self-interruptions, and who distinguished between "disfluent" and "suspensive" self-breaks (Fig. 3).

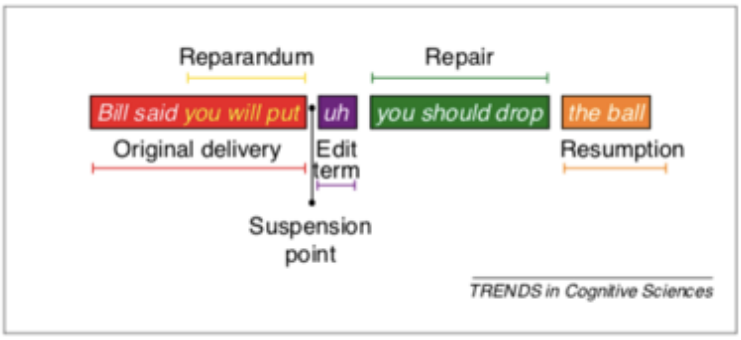


Figure 1. *A disfluent utterance (Ferreira & Bailey, 2004, p. 232)*

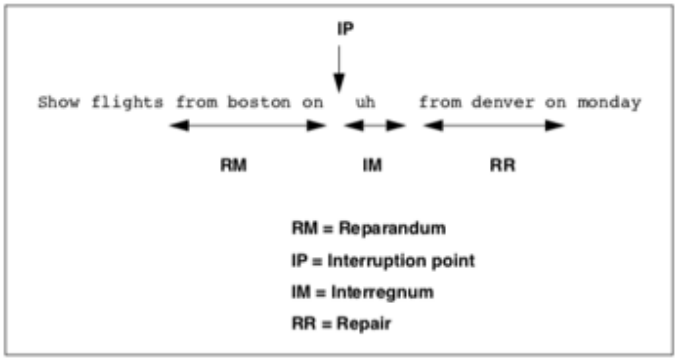


Figure 2*. Disfluency Regions (Shriberg 1994, p. 8)*

Interruption		Reparandum	Interregnum	Reparans	
disfluent	tu perds	un peu	*comment dire euh*	un peu	des repères
suspensive	tu perds	un peu	*comment dire euh*	des repères	

Figure 3. *Structure of disfluent and suspensive self-break (Pallaud et al., 2019, p. 2)*

First, the *reparandum* region shows the item that needs to be repaired. This region ends at the *interruption point* (or *suspension point*), the point in which the speech flow breaks down. It is then followed by the *editing phase* (also called *interregnum*, or *hiatus*) defined as "the time interval between the point of suspension of fluent speech and the point of its resumption" (Clark, 2006, p. 245). This time interval can be empty, or contain a silent or filled pause. When the interregnum is filled, the utterance does not necessarily have to be followed by a disruption, but can be resumed with no repair (*suspensive interruption*). However, in some cases, a repair (or *reparans*) does occur (*disfluent interruption*), which will then lead to the *resumption* of fluency (i.e. the fluent delivery). We thus find several variations of verbal disfluency; the flow of speech

can be interrupted with morpho-syntactic disruptions (e.g. self-repairs, non-lexical repetitions, false starts), but sometimes it can also be momentarily suspended by vocal markers (e.g. pauses and lengthening). This has led researchers to develop different classifications on disfluency markers, which will be discussed in the following section.

As explained earlier, in psycholinguistic research on disfluency, the main goal is to comprehend how speakers understand and produce language in spoken utterances; unlike representations of an ideal and wellformed delivery of speech, spoken utterances uttered in actual conversations contain several disfluencies. Indeed, the latter are a key feature of unplanned, spontaneous speech, as speakers typically do not know in advance what they are going to say and how they are going to say it[11]. They plan their utterances as they produce them. In fact, the course of human language can never be in a continuous flow, as O'Connell & Kowal (2005, p. 457) pointed out:

> (1) Every speaker must breathe, and breathing inevitably disrupts the flow of speech. (2) The capacity of listeners to understand is limited by the density of speech per time unit; intelligibility is diminished by failure to interrupt speech. (3) Language is reductively dialogical; listeners turn into speakers and speakers in turn into listeners. Turn taking disallows continuity[12].

Since speakers are constantly confronted with several cognitive demands when producing speech (facing multiple semantic or syntactic possibilities), they may need *extra time* to decide on what they are going to say next, or find the right phrase, which ultimately leads to *delays* (Clark, 2006). Disfluency is thus not only associated with speech *disruption or interruption*, but also speech *suspension*. The notions of speech suspension and interruption will be of particular interest to our study, and will be further discussed throughout this dissertation.

When speaking, time becomes a crucial feature of conversation, as speakers need to synchronize some of their internal processes with those of their addressees (Clark, 2002). When speakers face a problem in speech, they may wish to signal their intention to initiate the delay of the upcoming message (Clark & Fox Tree, 2002). In this sense, disfluencies can be seen as the by-product of problems with planning utterances (Clark, 2002), leaving overt traces of speech processing (Hieke, 1981). Clark

[11] However, the notion of disfluency can also be applied to written text, cf Eitel & Kühl (2016)
[12] It should be noted that we do not exactly agree with this last point. Turn-taking can also lead to joint, continuous productions, which does not necessarily disrupt the flow of the conversation. This will be further discussed throughout this thesis.

(2002) in fact argued that they should be seen as *solutions* to speaking rather than *problems*. They serve a fundamental role in the structuring of spontaneous speech, as they can be used to announce delays related to planning load or upcoming topic changes, or even mark discourse structure (Clark, 2002; Swerts, 1998). Therefore, instead of being simply categorized as speech errors or slips of the tongue, they can be viewed as major speech signals that can shed light on the speech production process (Goldman-Eisler, 1968; Swerts, 1998). For this reason, a number of researchers have investigated the role of disfluencies in speech comprehension. Evidence suggests that words following filled pauses (such as "uh" and "um") will be better remembered by listeners (Corley et al., 2007), which ultimately affects the comprehension process (Brennan & Schober, 2001; Ferreira & Bailey, 2004). Disfluencies have also shown to help listeners with syntactic parsing (Bailey & Ferreira, 2003). However, not all disfluencies are "equal"; filled pauses, for example, which are the most frequent types of disfluencies (Rose, 1998; Shriberg, 1994) may have a larger effect on speech comprehension than pauses, and may increase listeners' attention to what they will hear next (Fraundorf & Watson, 2011). It is thus important to distinguish between the different types of disfluencies, which leads us to the next section.

1.3. Major disfluency types and classifications

One of the first major studies conducted on disfluency phenomena dates back to the 1950s with Maclay & Osgood's (1959) seminal paper entitled *Hesitation Phenomena in Spontaneous Speech*. In this paper, the authors provided a formal categorization of disfluency (or "hesitation") phenomena, based on Mahl's (1956) categorization of "disturbances[13]", mainly: (1) filled pause, (2) unfilled pause, (4) false starts and (4) repeat. In their taxonomy, filled pauses were labelled as "English hesitation devices" as they were very commonly associated with hesitation and uncertainty. Unfilled pauses included "silences of unusual length" and "non-phonemic lengthening of phonemes" (p.24). Both were considered "abnormal" hesitation markers. Repeats were judged "non-significant semantically", and false starts refer to incomplete or self-interrupted utterances. It would appear that the annotation of disfluency markers was first based on judgments on the part of the authors, who focused on the "abnormal" aspects of speech variation, associated with negative terms, such as "NON-significant"

[13] These terminological differences (*hesitation* and *disturbances*) will be discussed in Section II.2.1.

or "UNusual length", but the definition of the categories is in fact quite equivocal, with for example the "unfilled pause" category which contains both silences and prolongations.

A few decades later, with the rise of speech technology and computerized applications, disfluencies were annotated at a much larger scale and more systematically by a number of speech technologists who were interested in the annotation of natural speech data. They worked on large digitally recorded corpora, and modeled well defined annotation schemes in order to improve the automatic processing of spontaneous speech (e.g., Meteer et al., 1995). One of the first major and highly influential annotation schemes in this area was introduced by Shriberg (1994). She worked on a large dataset comprising different types of corpora (human-human and human-computer dialogues), and her aim was to find regularities in disfluencies across corpora in order to improve the automatic processing of disfluencies in speech application. She considered disfluencies as removable material which "must be deleted to arrive at the sequence the speaker "intended", likely the one that would be uttered upon a request for repetition" (Shriberg, 1994, p. 1). Once again, this view of disfluencies echoes the notion of ideal delivery introduced earlier (cf section 1.1.) which underlies the "traditional" definition of disfluency.[14] Her annotation model, which initially targeted disruptive features of spoken speech (i.e. repetitions, substitutions, insertions, misarticulations, word fragments etc.[15]), relied on a complex and efficient classification system, which paved the way for later annotation models of disfluency (e.g., Christodoulides et al., 2018; Eklund, 2004; Moniz, 2013; Pallaud et al., 2013)[16]. More recently, the DUEL (*Disfluency, Exclamation, and Laughter in Dialogue*) project (Hough et al., 2016) introduced a crosslinguistic annotation model of disfluency and laughter for practical and computational dialogue modelling, which included the annotation of several disfluencies (e.g. filled pauses, repairs, and restarts, among others[17]). The main commonality between the different Disfluency annotation

[14] This is also similar to the annotation models used by speech therapists known as *the Systematic Disfluency Analysis* (SDA) which identifies a range of disfluent behaviors going from typical patterns to behaviors reflective of stuttering problems (cf Campbell et al., (1991). It includes the annotation of repetitions, duration of prolongations or hesitations, increases in tension, rate, loudness and speech.
[15] Shriberg's detailed model can be found in Appendix 1.
[16] More recently, researchers introduced novel annotation schemes based on this previous formal disfluency taxonomy, but which contain additional (para)linguistic markers. Such markers include smallwords, discourse markers, and even gestures and body language (Götz, 2013; Crible et al. 2019). These multi-level schemes are further described in section III.3.1.3.
[17] This is a reference to the 3-year collaborative project between Université Paris Diderot and Bielefeld University (information found on this website https://www.dsg-bielefeld.de/DUEL/ , last retrieved on

models is the systematic reference to Levelt (1983)'s Repair Model which introduced the different Disfluency regions in their surface structure (i.e. *reparandum*, *interruption point*, *interregnum*, *repair*; cf section I.1.2.).

In sum, despite theoretical and terminological differences, there seems to be a general consensus regarding the formal categorization of disfluency overall (based on Levelt's and Shriberg's model), which includes the following disfluency types: (summarized by Lickley, (2015, p. 16):

- Filled pauses (e.g., "um","uh")
- Repetitions (of part-words, whole words, phrases)
- Substitutions (where a part-word, word, or string of words is replaced by another word or string)
- Insertions (where a speaker repeats a string, but adds a word or more)
- Deletions (where a speaker abandons the utterance mid-stream)

However, in this list are not included other typical "hesitation markers" such as vocal lengthening and silences which can also be considered a subpart of disfluency phenomena (Betz, 2020; Candea, 2000; Gilquin, 2008; Vasilescu & Adda-Decker, 2007). In fact, there seems to be a common distinction between "disfluency" and "hesitation", where "disfluency" is attributed morpho-syntactic features, and hesitation vocal features. But as we shall see, this distinction is far from being consistent and reliable across all disfluency classifications, which is most certainly due to the overlapping terminologies (discussed in Section II.2.2.2).

Another distinction can be made between *forward-* and *backward-looking disfluencies* (Ginzburg et al., 2014) which are again in tune with the different disfluency regions. *Backward looking* disfluencies refer to "an alteration that refers back to an already uttered reparandum" (p. 95), mainly repetitions, false starts, insertions, deletions etc. *Forward-looking* disfluencies, on the other hand, refer to "the completion of the utterance which is delayed by a filled or unfilled pause (hesitation) or a repetition of a previously uttered part of the utterance". This distinction, which is similar to Levelt's (1983) distinction between *overt* and *covert* repairs, draws attention to the different roles played by disfluencies in speech; while backward-looking disfluencies rectify breaks in production which were previously

August 26th 2021). Its aim was to model disfluencies in different spoken languages (German, French, and Chinese) and create formal models and computational systems on speech processing.

uttered in the utterance, forward-looking disfluencies stall speech and may thus serve time-buying functions. This is also similar to Hieke (1981)'s earlier taxonomy which contains two main categories, (1) stalls, and (2) repair. The former includes silent pauses, filled pauses, prolongations, and "prospective repeats". The latter includes anything which involves an alteration in the utterance, i.e. false starts, repairs (in the phonology, syntax, and rhetoric), and "prospective" repeats". A distinction is thus made between "retrospective" and "prospective" repeats. Retrospective repeats rectify breaks in a previous segment, while prospective repeats are followed by a pause, which serves a stalling function. Once again, this distinction is based on the temporal nature of speech (i.e. the process of delaying and stalling speech) as well as on the non-linear processes of speech production (i.e. handling past, current, and future speech segments simultaneously).

Other authors, rather focus on specific acoustic and/or morphosyntactic features of disfluency markers. As Pallaud et al., (2019) noted, disfluencies are concerned with phonetic, acoustic, and prosodic levels, as well as morpho-syntactic ones. For instance, Guaïtella (1993) emphasized the acoustic and perceptible features of vocal "hesitations". She argued that vocal hesitations (e.g. filled pauses, prolongations etc.) have specific acoustic features (Guaïtella, 1993, p. 131):

> During vocal hesitations voice sounds like a sustained note. However, while this sustained note shows, at acoustic level, a decreasing of pitch, it corresponds to a diminution of subglottal pressure, i.e. the physiological dimension of declination.

Conversely, Pallaud et al., (2019, p. 1) rather focused on *morpho-syntactic disruptions*, which have specific morpho-syntactic and syntagmatic effects (with regard to their position within the different disfluency regions, cf Shriberg, 1994). The main object of their study was thus to identify the points of interruption which occurred when the flow of speech broke down (e.g. with self-breaks). In the present dissertation, in line with Guaïtella (1993) and Pallaud et al., (2019) a distinction will be made between *markers of suspension*, i.e. *vocal markers* (e.g. "uh"/"um", silences, lengthening, and other non-lexical sounds) and *markers of interruption*, i.e. *morpho-syntactic markers* (self-repairs, repeats, truncated words, and self-interruptions). Because we focus on the notion of fluency as an embodiment of *flow* (cf sections II.2.3. and IV), we believe that the distinction between vocal (suspending the acoustic flow)

and morphosyntactic markers (interrupting the speech signal) is more adequate than the backward/forward-looking or overt-covert distinction. This will be further explained in Chapter 2 (cf Chap. 2, section II. 2.2.1.).

So far, we have given a brief overview of the processes underlying the construct of "disfluency" from the perspective of Psycholinguistics (and computational linguistics). Disfluencies, which are very frequent in natural language production, include several markers (e.g. filled and unfilled pauses, substitutions, deletions, interruptions), and they can be seen as the hallmarks of speech production processes (i.e. planning, processing, articulating, monitoring etc.). They are inherent to spontaneous spoken speech, and consequently, as many psycholinguists have argued (e.g. Clark 2002; Pallaud et al., 2013), there is nothing dysfunctional about them, despite what the prefix "dis" suggests. There are, however, several issues regarding the term disfluency, and as we shall see in the next section, there are a number of overlapping terms used in the literature to describe the same phenomena.

II. Fluency, disfluency, and hesitation: a terminological debate beyond terminological issues

The confusion and lack of consensus over the terminology have been pointed out by several researchers (to name but a few, Allwood, 2017; De Jong, 2018; Eklund, 2004; Kormos & Dénes, 2004; Lickley, 2015). These discrepancies come from the multiple points of view taken (*speaker* versus *listener's perspective*) but also from the different disciplinary approaches (psycholinguistics, applied linguistics, interactional linguistics etc.). In the previous section, we mainly sketched out the field of psycholinguistics and computational linguistics, and in the present section, we introduce other approaches related to the construct of fluency and disfluency, and discuss several common terms used in the literature (mainly *fluency*, *disfluency* and *hesitation*) as well as more novel terms (such as *(dis)fluency*, *confluence*, or *communication management*) introduced more recently. The choice of terminology, as we shall see, will reflect the different theoretical views of these phenomena. First, we take stock of the different definitions of the term "fluency", followed by a review of the two main approaches to disfluency and their underlying terminological issues. Finally, we discuss our choice of terminology.

2.1. Definitions of fluency

2.1.1. Smoothness of speech versus language competence

As pointed out earlier, the notion of disfluency has often been associated with a deviation in speech from the "ideal delivery", in other words, a *fluent* delivery. But what does the term *fluency* suggest precisely? If we look at a dictionary entry, "fluency" is defined as "a smooth and easy flow; readiness, smoothness of speech" (Oxford English Dictionary). This concept of "smoothness" was also taken up by Lickley (2015, p. 2) who regarded speech fluency as multidimensional[18]. He listed three different dimensions of fluency, where the notion of smoothness is recurrent:

(1) Planning fluency: smoothness of the speech flow.
(2) Surface fluency: a smooth flow from one sound to the next.
(3) Perceived fluency: the listener's impression that the speech they are listening to has been produced smoothly.

As Lickley observed, an "intuitive" definition of fluency is based on the listener's perception of the speech flow. From the listener's perspective, fluency can be viewed as "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (Lennon, 1990, p. 291). Once again, the idea of speech being produced "easily" and "efficiently" seems to some degree related to the notion of "smoothness", and perhaps again the construct of an ideal and uninterrupted delivery of speech. This view focuses on the dynamic mechanisms underlying speech production which we described in the previous section. Additionally, Fillmore (1976, p. 93) gave four different definitions of fluency:

(1) The ability to talk at length with few pauses and fill time with talk.
(2) The ability to express a message in a coherent manner with "semantically dense" sentences.
(3) The ability to talk in a wide range of contexts.
(4) The ability to be creative and imaginative in language use.

These four aspects of fluency seem to strongly rely on the speakers' "abilities" to perform a series of actions in different contexts of language use, but not as much of on the flow of speech (except, perhaps, in 1). As Kormos & Denès (2004) pointed out, there are two main approaches to fluency, one which regards it as a *temporal*

[18] The multidimensional view of fluency will also be discussed in section III.3.1.3.

phenomenon, and another one as a *spoken language competence*. The latter suggests a speaker's level of proficiency (e.g. "I speak English fluently"), which is in sharp contrast with the first definitions given by Lickley (2015). We will further discuss this distinction in the following subsection.

2.1.2. Fluency in Second Language Acquisition

We will now turn to the study of fluency in native (L1) and non-native (L2) speech in the field of Second Language Acquisition (SLA)[19]. One of the first perceptible features of non-native speech is often attributed to the learners' "foreign accents", defined by Rasier & Hiligsmann (2007, p. 43) as: "the perception of a language that are reminiscent of another language". The authors further acknowledged the three aspects of prosody which L2 learners often have difficulty with (p.44):

(1) The manipulation of the components of the L2's prosodic phonology. This type of error has to do with the inappropriate use of e.g. intonation, stress, accent, rhythm, pauses.
(2) The way phonological entities are implemented in the speech signal. This category encompasses errors relative to the phonetic realization of e.g. intonation, tone, stress, accent.
(3) The expression and/or perception of linguistic and paralinguistic meaning using prosodic cues.

Except perhaps in (1) with the inappropriate use of pauses, these aspects have little to do with the notion of speech fluency put forward by Lickley (2015). In fact, Lennon (1990, p. 291) argued that speech fluency is considered as a "purely performance phenomenon" and differs from elements of oral proficiency such as idiomaticness, appropriateness, lexical range, and syntactic complexity. Within an EFL (*English as a Foreign Language*) environment, Lennon (1990, p. 389) further explained that the term "fluency" has a narrow and a broad sense. The broad sense is a general cover term for oral proficiency, that is foreign language ability (cf previous example "I speak English fluently", in 2.1.1). It is usually used for academic reference with entries such as "fair", "good", "fluent"'. The narrow sense, on the other hand, is one component of oral proficiency. It can be found in procedures for grading oral examinations used by

[19] This subsection is not meant to be exhaustive. The study of fluency in SLA will be further analyzed in section 3.1.3 and in Chapter 3.

teachers. He gives the following example: a learner may be "fluent, but grammatically inaccurate", or "fluent but lacks a wide and varied vocabulary".

In SLA, many studies on fluency share essentially the same goal, which consists in finding objective measures of a speaker's speech fluency in order to evaluate their speaking proficiency (De Jong, 2018). These measures include speech rate (the number of syllables articulated per minute), mean length of utterance (average number of syllables produced in utterances), and number of filled pauses, silent pauses, repetitions, and repairs per minute, (for an extensive review see De Jong, 2018, p. 4). The rate of disfluency markers is thus a key component of speech fluency, and is often used as a measure to subjective ratings of perceived fluency. Riggenbach (1991) investigated this issue by comparing three L2 speakers who were judged "highly fluent" with three L2 speakers who were judged "highly nonfluent". The highly fluent speakers were found to speak faster and with fewer pauses than those judged highly nonfluent (according to the attested judges). Similarly, Watanabe & Rose (2012) pointed out on their review of pausology in SLA that fast speech tended to be associated with perceptions of fluency in L2 speech. They noted:

> While there are differing views of what constitutes fluency in a second language, one common theme in all of these views is speed: That is, fluent second language speech is rapid, comparable to native speech (Watanabe & Rose, 2012, p. 3).

The notion of fluency in L2 speech is thus systematically compared to several temporal variables found in L1.

In short, disfluencies have often been used by researchers in the field of SLA to help measure the perceived fluency of a learner's speaking performance. Here the goal for L2 learners is to "produce speech at the tempo of native speakers, unimpeded by silent pauses and hesitations, filled pauses ("ers" and "erms"), self-corrections, repetitions, false starts, and the like" (Lennon, 1990, p. 390). However, Lennon also pointed out the mystical monolithic view of fluency as a target to "native-like levels". He argued: (p.292)

> It is often assumed that the fluency target of the language learner is "native-like levels". However, a moment's reflection shows that the idea of monolithic and unitary fluency for native speakers is mythical. Native speakers clearly differ among themselves in fluency, and, more particularly, any individual native

> speaker may be more or less fluent according to the topic, interlocutor, situation, "noise", stress, and other factors.

Consequently, the concept of fluency is difficult to grasp because it is originally based on the realization of an ideal and "native-like" fixed language. But how can we measure the proficiency of a non-native speaker based on measures that are also inherent to native speech? As Lickley (2015, p. 2) observed, there can be a "mismatch" between "the flow of the processes underlying speech production" and "the listener's perception of fluency". He added:

> An utterance that is perceived as fluent may still have contained hitches during the production processes. Minor disturbances in the flow of overt speech are easily missed by the listener, and may be detectable only on close inspection of the acoustic signal.

This further justifies Lickley's argument in favor of a multidimensional view of fluency (see 2.1.), which includes the planning level, the surface level, and the perception level. We will return to this issue in Section III.3.1.3. As we shall see throughout this thesis, the notions of fluency and disfluency should not only be restricted to levels of production or perception but should include other dimensions as well, such as interaction. For instance, McCarthy (2009), in his paper entitled *Rethinking Spoken Fluency*, re-examined the notion of fluency by putting forward its interactive dimension. While fluency has typically been conceived as a monologic achievement, essentially judged with temporal measures (e.g. speed of delivery, number of pauses etc.), McCarthy focused on the *co-creation* of fluency in a conversation, rather than the fluency of an individual speaker. Therefore, he offered the metaphor of "confluence" to replace the term "fluency". Similarly, Peltonen (2019) considered fluency within a *Fluency Resources Framework* which views spoken L2 fluency as a collaborative, problem-solving activity, linked to strategic language use. These frameworks are further described in Chapter 3 (Chap. 3, section I. 1.2.2.)

The interactional and collaborative aspect of speech fluency has in fact received little attention in the field of Second Language Acquisition or in psycholinguistics which have only provided a partial picture of the phenomena under study. This motivates our need to study fluency and disfluency within an additional interactional framework, described in section III, 3.2.

2.2. Definitions and approaches to Disfluency

In the previous section, we provided several definitions of fluency based on different dimensions of spoken communication, mainly production, perception, and interaction. We also emphasized the fact that the term *fluency* can be problematic to define, because it is originally based on the notion of *ideal delivery*. If fluency embodies the ideal, efficient, and smooth delivery of speech, then disfluency presents a "failure" to maintain the smoothness of speech. We are now faced with a challenging question, raised by Lickley (2015, p. 13):

> Is fluent speech the norm? If a speaker can produce a stream of spontaneous speech without having second thoughts about whether they are conveying the correct message at the right time, without spotting and reacting to an inaccuracy in the message or an error in its production, without struggling to find the right words and getting the sounds right, and without being interrupted by another speaker or some other distraction, then that stream of speech is likely to be completed smoothly, without interruption or revision: Fluently, in other words. However, both casual observation and corpus studies of unrehearsed speech suggest that such fluency is the exception, rather than the rule.

The issue with the term disfluency[20] is that it entails a pathological problem, or something dysfunctional (which was originally the case with the term "dysfluency" used to refer to speech pathologies and stuttering). But as emphasized earlier (cf section 1.2) disfluencies are inherent to spontaneous speech, and play a large role in speech production. In fact, disfluencies can be used to restore *continuity* in speech. Clark & Wasow (1998), who worked on word repetitions, offered a repetition repair model known as "Commit-and-Restore Model", which is similar to Levelt's repair model (1983, 1989) as it consists in several stages (initial commitment, suspension of speech, hiatus, restart of constituent). However, in their model, word repetitions were not viewed as speech errors which reflected an interruption in speech. On the contrary, they argued that word repetitions could help to retore continuity in speech. They put forward the "continuity hypothesis", which, unlike the *Main Interruption Rule*, claims that speakers repeat a previous constituent in order to restore continuity in the delivery. Similarly, Hieke's (1981, p. 152) study on retrospective and prospective

[20] We also find the term "non-fluency" in the literature, which is explained in more detail in Eklund's dissertation (Eklund, 2004, p. 158).

repeats accounted for a view of continuity, whereby repeats were viewed as "bridging devices" which promoted continuity and reestablished fluency. This view of "continuity" is in sharp contrast with the view of disfluency as a negative disruption or interruption. In fact, two different views on disfluency phenomena have emerged in the literature, resulting in radically different theoretical implications, and thus further questioning the use of the term *disfluency*. These views are presented below.

2.2.1. The two main views of Disfluency

In her doctoral dissertation, Nicholson (2007, p. 94) pointed out two main views of disfluency: the first one, called the *Strategic Modelling* view, suggests that disfluencies are used for strategic and communicative purposes in order to signal their commitment to the listener. The second view, called the *Cognitive Burden view*, considers the speech production process as a highly complex one which can be cognitively overburdened, thus leading to the production of disfluencies. Disfluencies are thus viewed as a manifestation of cognitive load. Similarly, Clark & Fox Tree (2002) shed light on the different conceptions of one specific class of disfluency markers, mainly filled pauses. They presented three views, labeled (1)" filler-as-symptom", (2) "filler-as-signal", and (3) "filler-as-word". In the first view, "fillers[21]" are seen as symptoms of problems in speaking. In the second view, they are regarded as nonlinguistic signals which initiate a delay in speech. The third view, which has led to numerous debates (see Corley & Stewart, 2008; Kowal et al., 1983, Tottie, 2016) states that "uh" and "um" should be seen as word interjections commenting on a speaker's performance. While the third view could only be applied to filled pauses (and hardly to other disfluency markers such as repetitions, self-repairs or silences), we will examine the two views proposed by Nicholson (2007), which are similar to (1) and (2).

The *Cognitive Burden* (or *disfluency-as-symptom*) view is essentially reflected in the work of psycholinguists who explored and investigated the contextual, lexical, and cognitive determinants of disfluencies. Beattie (1979), who worked on *hesitation phenomena* (we will discuss the choice of the term "hesitation" in the following subsection), has shown that long clauses (containing 6-10 words) were more likely to contain more pauses than short ones (2-5 words). This is supported by the hypothesis that hesitation pauses occur at high points of uncertainty: unpredictable or infrequent lexical items are more likely to be preceded by pauses than frequent ones (Beattie &

[21] Note that the authors used the term "fillers" in their paper to refer to filled pauses.

Butterworth, 1979). Consequently, hesitation disfluencies are said to reflect "an act of choice" (Beattie & Butterworth, 1979, p. 202) between different lexical items. More evidence suggests that lexical access problems and planning difficulties often lead to disfluencies (Brennan & Schober, 2001; Hartsuiker & Notebaert, 2009; Schnadt & Corley, 2006). For instance, Hartsuiker & Notebaert (2009) conducted an experimental study in which participants were asked to describe networks of lines drawings and paths connecting these drawings. Results indicated that pictures which had a low name agreement led to more disfluencies than those with high name agreement. They concluded that difficulties at certain stages of language production (i.e., lexical access) resulted in distinct patterns of disfluencies (i.e. self-corrections and repetitions). The relationship between disfluency and discourse domain, cognitive effort, and utterance complexity is further described in Chapter 4, section I.1.2.

In sum, a number of studies have insisted on the fact that disfluencies reflect signs of "trouble", "problems", and "difficulties" in speech. They occur when speakers detect *trouble* in processing: as Levelt (1983) suggests, "uh" is a symptom of recency of *trouble* indicating that the trouble is still present at the moment of interruption. Similarly, filled pauses are said to provide a consistent picture of *difficulties* encountered in sentence planning (Holmes, 1988), as they are consistently affected by message-levels *difficulties* (Fraundorf & Watson, 2014). Furthermore, disfluencies indicate the depth of speakers' retrieval *problems* (Smith & Clark, 1993), reflect cognitive *difficulty* (Finlayson & Corley, 2012) and production *difficulty* (Fraundorf & Watson, 2011). Merlo & Mansur (2010, p. 491) define them as "*verbalized difficulties* in which the speaker notices a problem after or during the speech".

However, advocates of the *Strategic Modelling* view (or disfluency-as-signal) suggest that disfluencies have little to do with trouble. Disfluencies can serve a wide array of pragmatic functions, such as initiating a turn (Schegloff, 2010), keeping the floor (Kjellmer, 2003; Maclay & Osgood, 1959; Tottie, 2014), or managing interpersonal relations (Fischer, 2000). Tottie (2011, 2014, 2015, 2016, 2019) conducted several studies dedicated to the *pragmatic* uses of filled pauses in British and American English. In her work, which focuses primarily on the distribution of filled pauses in naturally occurring conversation, she put forward the idea that "uh" and "um" should be considered as a class of pragmatic markers, which are strongly

determined by setting and register[22]. Indeed, her corpus-based studies revealed effects in age, gender, socio-economic class, context, and register (Tottie, 2011, 2014). In line with Clark & Fox Tree's (2002) hypothesis that "uh" and "um" signal a delay in speech in order to keep or cede the floor, or to attract attention (Kjellmer 2013), Tottie (2011) stressed the fact that "uh" and "um" functioned as a *planning* device, and thus suggested the term *planner*. She argued that filled pauses could serve several overlapping functions: they can help structure upcoming discourse (following Swerts 1998), but they can also be used intentionally with a stylistic purpose. She further stated (Tottie, 2014, p. 21):

> Uhm can be much more than a filler of pauses, a sign of hesitation or disfluency: as a stance marker, it can be used to initiate discourse paragraphs, to clarify meanings, to correct an utterance, to achieve precision, and to mark stance, among other functions.

This view of filled pauses, which is not restrictive to situations of cognitive load or production difficulties (put forward by the *Cognitive-Burden* view), takes into account the pragmatic dimension of speech, and thus defends a more positive approach to these phenomena. We will return to this dimension in section III.

In light of this approach, the terms *disfluency* or *hesitation* (cf next subsection) are often found to be inadequate. In fact, Tottie (2014, p. 26) suggested that "uhms" should deserve to be called markers of *fluency* rather than *disfluency*. While her suggestion only concerns filled pauses and does not cover the rest of the disfluency markers (e.g. prolongations, silent pauses, repairs, repetitions etc.) Moniz et al., (2009) classified all disfluency phenomena as *fluent* communicative devices. Similarly, Hieke (1981) argued that disfluencies formed an integral part of speech production in a *positive* sense, and should thus be viewed as "a normal component of fluency" and "wellformedness phenomena rather than disfluencies, at least as far as they serve as devices by the speaker to produce more error-free, high-quality speech." (Hieke 1981, p. 150). As further discussed in section 2.2.3 and III.3.1, the constant overlap and confusion of the terms *fluency* and *disfluency* result in the fact that these constructs are very often restricted to a binary opposition between two abstract notions, while in reality they embody so much more.

[22] The effect of setting and register on disfluency will be examined in Chapter 4.

Further in line with the view of disfluency as a signal, Clark (1996, 2002) put forward the idea that disfluencies should be considered as *collateral signals*. Clark distinguished between primary signals (i.e. the linguistic devices by which speakers accomplish discourse) and collateral signals, which are lexical, syntactic, prosodic, and gestural devices which help coordinate speakers' primary signals. They are used by speakers to manage their on-going performance, and fall into four main categories (Clark & Fox Tree, 2002, p. 78):

(1) *Inserts*: parenthetical asides such as editing expressions (I mean, you know, sorry) certain discourse markers (well, now, oh, like), and laughter, sighs, and tongue clicks.
(2) *Juxtapositions*: replacements and repairs of one stretch of speech against another (e.g. *Mallet was / Mallet said*).
(3) *Modifications*: modifications of a syllable, word or phrase within a primary utterance.
(4) *Concomitants*: collateral signals produced in a different modality at the same time as speech (head nods, eye gaze, smiles, iconic gestures, pointing etc.)

This approach to disfluencies as collateral signals offers a much broader perspective as it incorporates the visual-gestural modality of speech. Similarly, Allwood et al., (2005) pointed out the pragmatic and multimodal dimensions of disfluency in their study of *communication management*. Communication management, or *speech management* (Allwood et al., 1990) are defined as a "linguistic and other behavior which gives evidence of an individual managing his own communication while taking his/her interlocutor into account". Allwood et al., (1990) argued that speakers can manage their own communication with the use of gaze aversion, pausing, repetitions, and the like. They further distinguished between own *communication management* (OCM) and *interactive communication management* (ICM) (Allwood et al., 2005). OCM is concerned with how speakers continuously manage the planning and production of their own communication, while ICM pertains to the management of the interaction through turn-taking, feedback, and sequencing. They also added: "both types of management serve to share the main messages[23] with other communicators and make communication more flexible and fluent by adapting face-to-face interaction

[23] The "main messages" can be understood here as the primary signals of communication introduced by Clark (1996).

demands on production and comprehension". (Allwood et al. 2005, p. 2). Once again, we can note the use of the term *fluent*, which suggests a more positive viewpoint: "OCM, contrary to what this term [disfluency] suggests, often contributes to the fluency and flexibility of speech" (Allwood et al., 2005).

Allwood et al., (1990, p. 10) further distinguished between two main functions within OCM: (1) *choice-related functions*, "to enable the speaker to gain time for processes having to do with the continuing choice of content and types of structured expressions", (2) *change-related functions*, "to enable the speaker, on the basis of various feedback processes (internal and external), to change already produced content, or expressions". These functions are similar to the two main types of disfluency presented in section I.1.3. (*forward-looking* versus *backward-looking*, *overt* versus *covert*, and *stalls* versus *repairs*). However, despite similarities in formal categorization, their approach to disfluency is innovating, because it offers central components of *multimodal interaction* (e.g. interactional dynamics, and the use of gestures) which were lacking in previous approaches. We will return to this point in section 2.2.3.

To conclude, the study of disfluency, which was originally investigated in the field of psycholinguistics (cf section I), is not only restricted to cognitive and internal speech processes, (in line with the Cognitive Burden view) but can include other dimensions of communication as well, such as pragmatics and gesture, which accounts for a more positive approach. We will return to this key aspect in Section III.

2.2.2. Disfluency or Hesitation?

In the previous subsections, we outlined the different uses of the terms *fluency* and *disfluency* in the literature, as well as their different theoretical implications. We emphasized the fact that these notions can be quite difficult to define, given the different views and approaches adopted by a number of researchers. Another traditional term found in the literature to describe spontaneous speech phenomena is *hesitation*. We already touched upon this term when we mentioned one of the first studies conducted by Maclay & Osgood (1959) on hesitation phenomena in spontaneous speech (section I.1.3). Their four hesitation types, (repeats, false starts, filled pauses, and lenghtenings) are very similar to the ones categorized in Shriberg's (1994) or Eklund's (2004) taxonomies of disfluency markers. In fact, it would appear

that hesitation and disfluency phenomena are closely related, and the terms could easily be interchangeable. Lickley (2014, p. 21) defined hesitation as the following:

> Hesitation usually involves the temporary suspension of flowing speech. It may be achieved by stopping altogether and remaining silent for a moment, by prolonging a syllable, by producing a filled pause or a lexical filler or by repeating the onset of the current phrase.

This definition seems to relate to previous definitions of disfluency with the notion of speech suspension, and it also includes the major disfluency types (mainly filled and unfilled pauses, prolongations, and repetitions). But as we will see, there are many inconsistencies regarding what types of markers should be included in the "disfluency" or "hesitation" category. For example, Merlo & Mansur (2004) listed the following disfluencies in their study on descriptive discourse:

- Fillers
- Interaction pauses (e.g. *you know, ok?)*
- Hesitation pauses with duration up to or equaling 250 ms,
- Hesitant prolongations (prolongation without prosodic intention)
- Lexical pauses (e.g *well, look, for example, that is*)
- Repetitions
- Retraced false starts (verbalizations that are corrected)
- Unretraced false starts (verbalizations that are abandoned)

Here the adjective "hesitant" and the noun "hesitation" are added to other disfluencies, forming collocations such as "hesitant prolongations" and "hesitation pauses", but note that these adjectives are not added before "repetitions". It is thus not exactly clear how "hesitation" differs from disfluency.

Moreover, some authors have included filled pauses and unfilled pauses in their taxonomy of hesitation/disfluency phenomena (e.g., Beattie, 1979; Ginzburg et al., 2014; Riggenbach, 1991; Vasilescu & Adda-Decker, 2007, among others) while others have studied "uh" and "um" specifically without labeling them as hesitation or disfluency markers, but "fillers" (e.g. Clark & Fox Tree, 2002; Merlo & Mansur, 2004) or "uhm" (e.g. Schegloff, 2010; Tottie, 2014). Conversely, other researchers have chosen not to include silent pauses in their taxonomy of disfluencies (e.g., Bortfeld et al., 2001; Boulis et al., 2005; Nicholson, 2007), while others distinguished between different types of pauses (e.g. silences, gaps, and lapses; cf Edlund et al., 2009 or

lexicalized versus non-lexicalized pauses, cf Schettino et al., 2020). Similarly, repetitions are sometimes included or excluded from hesitation phenomena. Lickley (2014, p. 14) argues that a repetition can be classified as a hesitation because when speakers pause in the middle of their utterance and start again, they "often restart by backtracking one or two words and repeating them with a fluent continuation". However, Ginzburg et al., (2014) distinguished between (1) hesitations – disfluencies that are followed by the completion of the utterance delayed by a filled or unfilled pause; and (2) repetitions – repetition of the previous constituent.

Another striking observation is the fact that in some cases the term "hesitation" is not only used to cover a (temporal) feature of disfluency, but also a function. Lickley (2001, p. 93) claimed that the two major functions of disfluency were "hesitation" and "self-repair", although he also gave a formal description of hesitation (characterized by filled and unfilled pauses, word prolongations, or a combination of all). Therefore, the function of "hesitation" is not very clear. In addition, Nicholson (2007, p. 127) investigated the use of "deletion disfluencies" in a multimodal map description task, and she explained that there were two main functions served by deletions. The first one, called "planning deletions" was when speakers abandoned an utterance because they needed to re-plan. The second one, called "hesitation deletions" was when speakers decided to restart the utterance or rephrase it in a different way. Once again, it is difficult to grasp the exact meaning of "hesitation" here. Shriberg (2001, p. 155) made an interesting comment on the notion of hesitation. She explained that one of the most commonly observed effects of disfluency is "a lengthening of rhymes or syllables preceding the interruption point" (p. 161). She also claimed that "lengthening found in the disfluency suggests a uniform probability of additional time in a hesitation", and that "lengthening, like uttering a filled pause, allows speakers to pause in the production of message content without ceasing phonation" (p. 161). Finally, she concluded:

> However, not all disfluencies are associated with hesitation: some disfluencies are associated with detection of error. In such cases, the reparandum is usually not lengthened, but rather *shortened*. (Shriberg, 1994, p. 161, our emphasis)

In sum, she makes a striking distinction between disfluency and hesitation on the basis of time and duration. When speakers detect a problem in speech, they can either repair or repeat the previous constituent, or they can insert a filled or unfilled pause.

The latter is more likely to be interpreted as a hesitation. Similarly, Betz (2020, p. 11) defined hesitation as "anything that temporally extends the delivery of the intended message". In this sense, hesitations are viewed as a *temporal extension* of the message, which may help speakers to buy time in order to solve problems when speaking. Therefore, the notion of hesitation is closely related to the concept of *buying time* (Brennan & Schober, 2001; Fehringer & Fry, 2007).

Candea (2000, p. 18) investigated so-called hesitation phenomena in French spontaneous speech and emphasized the differences found in the terminology. She argued that the term hesitation was more adequate than disfluency because it did not entail a problem or a speech pathology. From what Shriberg and others have noted, a hesitation can be defined by the notion of time suspension. If we look at a dictionary entry, we find the following definition: "The action of hesitating; a pausing or delaying in deciding or acting, due to irresolution; the condition of doubt in relation to action." (Oxford English Dictionary). Therefore, the concept of hesitation is attributed two core semantic features, mainly time suspension and uncertainty, as Candea (2000, p. 15) explained:

> Le trait sémantique principal reste le « temps d'arrêt », la suspension dans le temps du processus de production de parole, explicitement donné par le dictionnaire, mais on remarque qu'il est en fait clairement affirmé que la cause de ce temps d'arrêt doit être attribuée à une « incertitude » du locuteur : on est à nouveau renvoyé vers la difficulté liée à une prise de décision, et donc à l'existence implicite de plusieurs possibilités entre lesquelles le locuteur hésite.

What seems to distinguish a hesitation from a disfluency is the fact that hesitations are used to delay information and result in a speech suspension, while disfluencies are the result of a sudden interruption in the speech flow, usually related to problems encountered in speech. They are nonetheless, tightly related; a hesitation can cause an interruption, and vice versa. However, Candea (2000, p. 18) pointed out that the term "hesitation" was still too broad, and had several underlying problems: (1) it is an ambiguous term because it can be expressed vocally with the production of "hesitation markers" (such as "uh" and "um", syllable prolongations etc.), but it can also be expressed verbally with the production of lexical expressions such as: "I don't know what to do—I'm hesitating between X and Y". Moreover, it can also be expressed non-verbally (through facial expressions, gestures etc.). (2) not only is the term too

ambiguous, it is also too specific. It presupposes that a specific class of markers, which are very common in spoken spontaneous speech (such as "uh" and "um"), are systematically associated with the cognitive notion of "hesitation" which is, according to Candea, too simplistic and not often the case. In other words, saying "uhm" does not necessarily imply that speakers are currently "hesitating" in the strict sense (i.e. making a difficult choice). As we have seen earlier with the work of Tottie (2014, 2016, 2019) or Clark & Fox Tree (2002), "uh" and "um" can serve many pragmatic functions other than "hesitation".

Further grounded in French theories of *co-énonciation* and *colocution*, Candea (2000, 2017) took into account three essential dimensions of language when studying spontaneous speech phenomena, mainly (1) syntactic and informational structuring, (2) language processing and encoding, and (3) the construction of intersubjectivity. For this reason, Candea (2000) refrained from using the term "hesitation" or "disfluency" in her work, as the latter failed to truly embody these different dimensions of language. Therefore, she opted for the term "travail de marque de formulation" (*formulation marker*) instead, following Morel & Danon-Boileau (1998). This novel term further integrates the different cognitive, enunciative and interactional mechanisms associated with the production of so called "hesitation" phenomena, without being contingent upon error or indecision. We agree with this view, so the term *hesitation* will also be avoided in this thesis, except perhaps in some specific cases when speakers are overtly displaying their uncertainty, for example through verbal expressions (*I am not sure)* or visual displays (e.g. thinking face). This will be further examined in Chapter 5.

2.2.3. Beyond terminological issues: a functionally ambivalent approach to (Dis)fluency

Given the complexity and multiplicity of processes underlying the concept of the phenomena under study, as well as the range of perspectives and angles adopted by different researchers, the overlapping terms "fluency", "disfluency" or "hesitation" may be too restrictive, and at times even confusing. For this reason, a new body of research emerged (in line with the *Strategic Modelling View*) and offered new terms to define these phenomena. We have briefly sketched out some of these novel terms introduced in the literature more or less recently, such as "planner" (exclusively for filled pauses; Tottie, 2016; Jucker, 2015), "collateral signals" (Clark, 2013) "travail de

formulation" (Candea, 2000; Morel & Danon-Boileau, 1998), "wellformedness phenomena" (Hieke, 1981) "own communication management" (Allwood et al., 1990; 2005), or "confluence" (McCarthy, 2009). All of these terms are a blatant departure from the initial term "disfluency" or "hesitation", as they account for a more positive approach to these phenomena[24].

More recently, another body of research, partly based on the work of Götz (2013) on fluency enhancing strategies in SLA, put forward the term *(dis)fluency*, with the prefix "dis" in brackets (Crible, 2018; Crible et al., 2019; Dumont, 2018; Grosman, 2018; Notarrigo, 2017). This term captures both the notion of fluency *and* disfluency, and goes beyond the binary divide between the two concepts. Instead of opposing fluency with disfluency, or arguing in favor of a positive versus a negative view of disfluency (cf section 2.2.1), Crible et al., (2019) argued that this duality should be considered on a scale or a continuum of (dis)fluency. This implies that the same forms, called "fluencemes" (suggested by Götz, 2013), vary systematically according to language, context, and genre. Fluencemes are defined as "an abstract and idealized feature of speech that contributes to the production or perception of fluency, whatever its concrete realization may be" (Götz 2013, p. 8). The choice of the term *fluenceme* is central because, unlike *disfluency marker or hesitation marker,* it avoids the underlying notion of dysfunction, problem, or uncertainty.

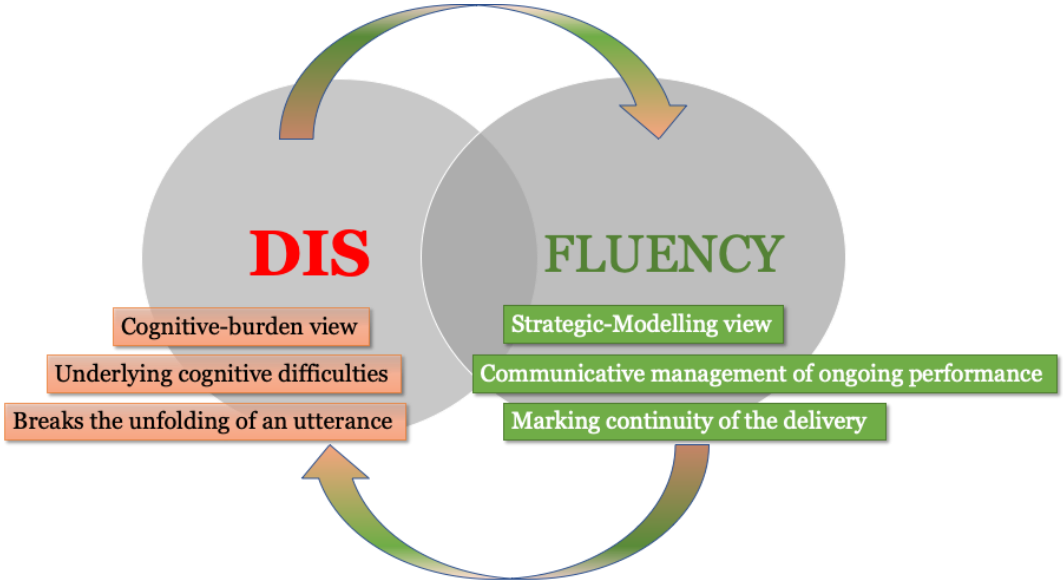


***Figure 4**. An ambivalent approach to (Dis)fluency: two sides of the same coin (following Crible et al., 2019).*

[24] We will also mention the term "repair" used by conversation analysts in section III.3.2.2.

Additionally, the term *(dis)fluency* implies that disfluency and fluency reflect *two sides of the same coin* (Crible et al. 2017, p. 71); in other words, it does not disregard the Cognitive-burden view nor the Strategic-Modelling view altogether, but considers them both: the same forms have the potential to perform both fluent and/or disfluent functions. This is exemplified in Figure 4.

Even though Allwood (2017) fervently argued in favor of a positive view of disfluency embodying efficient mechanisms to interactive communication, he raised a more nuanced point when discussing the fact that speakers were not necessarily fluent or disfluent in all types of communicative activities; he noted (p.3):

> It seems fairly clear that most of them would be "disfluent" in written language, if we are not trying to capture authentic speech in writing. It also seems clear that many of them might be disfluent in many types of public speaking. But this does not mean that they are disfluent in interactive (small) talk, where it is important that you are able to hesitate, change your mind, repeat for clarity, be flexible, and non-categorical and give continuous unobstrusive feedback.

Once more, the idea that speakers are deemed "fluent" or "disfluent" does not only rely on temporal measures of fluency (e.g. length and frequency of pauses), but on contextual features as well, such as the type of communicative activity or register. Additionally, it also relies on the social expectations and conversational constraints in a given situation. Fluencemes thus emerge from speakers' intentions and expectations in a specific context, which can potentially lead to (un)successful communication.

Similarly, Tottie (2016, p. 116) made an interesting comment on the duality and functional ambivalence of filled pauses, which captures the *disfluency-as-signal* and *disfluency-as-symptom* view:

> Whether UHM should be regarded as a symptom of ongoing planning or a signal to listeners has been discussed in the literature – for a good summary, see de Leeuw (2007). My view is that like pragmatic markers in general, UHM must have both functions simultaneously. The symptom view is well expressed by Goffman (1981:293): "... the speaker, momentarily unable or unwilling to produce the required word or phrase, gives audible evidence that he is engaged in speech-productive labor ..." – i.e. the speaker is planning what to say. Similarly, speakers do not consciously decide to say well, you know, I mean or like – but nevertheless, like UHM, these items signal something to listeners about speakers' attitudes and states of mind.

Therefore, the term *(dis)fluency* accounts for a *unified* approach to fluency and disfluency, instead of rejecting one view and supporting another. Moreover, this body of research suggests that Fluency and Disfluency phenomena should not only be restricted to a holistic view based on the global impression of efficiency or naturalness (see 2.1.1.), but should also adopt a *componential* approach which takes into account situational and contextual features of language[25]. This is in line with the framework of Cognitive Grammar and *usage-based linguistics*, which will be introduced in section III.3.

2.3. Summary of the overlapping terms and our choice of terminology

The previous section has outlined several overlapping terms used in the literature to refer to spontaneous speech phenomena. As we have seen, the choice of terminology is central because it reflects a certain theoretical standpoint. While some authors have chosen the term *hesitation* to refer to an act of choice (e.g. Beattie, 1979; Goldman-Eisler, 1968) or the acoustic and phonetic features of pauses (e.g.Duez, 2001), others focus on the surface structure of *disfluency* which embodies an interruption point (e.g. Pallaud et al., 2019; Shriberg, 1994). Conversely, several authors refused to use the term *disfluency* or *hesitation,* and coined novel terms, such as *communication management*, *confluence*, or *collateral signals*. Allwood (2017) even suggested to change the terminology and abandon the term *disfluency* altogether for a more positive and neutral one (*communication management*). However, the latter may be too large, and does truly not capture the notion of *flow* which will be put forward in this thesis. A list of the main existing terms is summarized in the following table, including a "tentative" definition of the terms, and a (non-exhaustive) list of the authors who have used them.

[25] For a more extensive review on holistic and componential approaches to fluency, read Crible (2017, p. 19) and Grosman (2018, p. 20)

***Table 1**. Summary of the overlapping terms used in the literature*

Term	Tentative definition	Authors
Disfluency	A deviation in speech from the ideal delivery. A temporary suspension or interruption of the speech flow.	Shriberg (1994) Eklund (2004) Lickley (2015) Bailey & Feirrera (2004)
Fluency	Ideal delivery of speech, global impression of smooth speech. One component of oral proficiency.	De Jong (2018) Lennon (1991) Fillmore (1979)
Hesitation	Temporal extension of the message (through pausing and delaying), often associated with uncertainty and an act of choice.	Gilquin (2008), Maclay & Osgood, (1959) Duez (1991) Betz (2020)
Confluence	Co-creation of fluency in a conversation.	McCarthy (2009)
(Dis)fluency	Functionally ambivalent phenomena made of fluencemes which can potentially serve fluent and/or disfluent functions.	Götz (2013) Crible et al., (2019)
Communication Management	Linguistic and other behavior which gives evidence of a speaker managing his communication while taking his interlocutor into account.	Allwood et al. (1990; 2005) Ginzburg & Poesio (2016)
Wellformedness	Devices used by the speaker to produce more error-free and high quality speech.	Hieke (1981)
Travail de Formulation	Markers used for planning and formulation.	Candea (2000) Morel & Danon-Boileau (1998)
Collateral signals	Communicative signals which comment and manage speakers' ongoing performance.	Clark (1996, 2003)

The terms *confluence*, *communication management*, *wellformedness*, *travail de formulation* and *collateral signals* are all a blatant departure from the initial term *disfluency*, as they all account for a more positive approach to these phenomena. These terms thus offer a fresh perspective on the phenomena under study, which does not linger on cognitive problems, feelings of uncertainty, or the production of speech errors. However, given the profusion of these novel terms, the issue still remains to find the appropriate and most relevant term for the present dissertation.

The reason for the term *disfluency* to be an overarching term in the literature is certainly related to the fact that it has been the most widely used since the late 1950s. As we have seen, despite what the negative prefix “dis” suggests, most researchers interested in disfluency phenomena do not view them as negative processes, but on the contrary, they consider them as an integral part of speech planning and processing. Moreover, from a strictly *formal* perspective, disfluencies do mark a disruption in the speech flow or in the acoustic signal, but that does not mean that they are necessarily

disruptive per se, as they can serve communicative and interpersonal functions. Therefore, the real issue is not only terminological, but theoretical as well. The "dis" in disfluency, from a psycholinguistic point of view, refers to a breakdown in the speech flow, thus focusing primarily on the linear verbal and vocal channel of communication. Moreover, this breakdown is often understood at the surface level of the verbal *utterance*, thus disregarding all other aspects of communication at the level of the interaction. The integration of interactional dynamics in the study of disfluency was put forward in the work of Clark (1996), McCarthy (2009), Candea (2000), or Allwood (2017) which enables us to regard these phenomena as communicative devices contributing to the (co)-construction of fluency. This contribution is highly relevant to our study as we believe that disfluency should be investigated at several levels of analysis, integrating the verbal, vocal, and visual-gestural communication channel, as well as the speech, visuogestural, and interactional level of fluency. This is further elaborated in Section IV.

This leads us to our choice of terminology: because disfluency is such a complex and multi-faceted phenomenon, resulting from several cognitive, interactional, and speech processes, it cannot easily be categorized under one label. While the terms *confluence*, *travail de formulation*, *communication management* and *collateral signals* are truly innovative, they do not quite grasp the functional ambivalence of these phenomena, put forward in the work of Götz (2013) and Crible et al., (2019). By keeping the core notion of *fluency*, understood here in broad terms (i.e. speech, discursive, interactional, and gestural fluency, or flow) and adding the "dis" (i.e. disruption, discontinuity) in brackets, we focus on the potential for the same fluencemes to serve fluent and/or disfluent functions, depending on contextual and situational features. Moreover, we are also in line with the various theoretical implications underlying the term *(dis)fluency*. This term reflects a view grounded in the framework of cognitive and usage-based grammar (e.g. Langacker, 1987) which regards language as a dynamic system containing fluid categories (i.e. fluencemes with a range of variation and different degrees of fluency and disfluency according to their context of use). This approach to the phenomena under study led to several usage-based models of (dis)fluency introduced by Götz (2013) and Crible et al., (2019), which will be further described in section III.3.1.3.

Moreover, our usage-based and dynamic approach to (dis)fluency is also grounded in an interactional and conversation analytic framework (e.g. Goodwin,

2017; Mondada, 2007; Sacks et al., 1974) which focuses on occurrences of an instance and its sequential unfolding in a specific interactional sequence captured in embodied interaction. Therefore, the notion of ambivalence reflecting two sides of (dis)fluency, will also be extended to notions of (dis)alignment, (dis)continuity, and (dis)engagement in the interactional flow. For this reason, we will also speak of *inter-(dis)fluency* in order to emphasize their potential fluent or disfluent contribution at the interactional level. This aspect was also addressed in the work of McCarthy (2009) and Allwood et al. (2013), but we will use additional conversation analytic methodological tools to support our analyses.

Lastly, our understanding of inter-(dis)fluency will also take into account the kinetic and visual-gestural features of communication, where gesture[26] and speech jointly regulate communicative interactions (Kendon 2004). Therefore, inter-(dis)fluency will no longer be seen as a strictly verbal or vocal phenomenon, but as an embodied and multimodal one as well. The notion of functional and interactional ambivalence will thus also be reflected in the gestural and bodily actions enacted by fluencemes.

To conclude, our study of inter-(dis)fluency phenomena, grounded in an integrated theoretical framework (further described in section III and IV) aims to go beyond the traditional and "pathologized" view of disfluency as a speech disruption, and offer a fresh interactional and multimodal perspective (in line with Allwood et al. 2005, McCarthy, 2009, and Clark, 1996, but with a larger theoretical framework). Our choice of terminology, which focuses on the duality of fluencemes and their degree of fluency and/or disfluency, is in tune with both the *disfluency-as-symptom* and *disfluency-as-signal* views addressed in section 2.2.1, as we believe that (dis)fluency should not be restricted to a single view or one particular label. This will also reflect our methodological choice to combine quantitative annotations of fluencemes and their positional, temporal and combinatory features at the utterance level, with multimodal fine-grained analyses of embodied fluencemes within interactional sequences. This will be further described in Chapter 2.

[26] "Gesture" is understood in the broad sense here, including all movements of the body (head, upper and lower body, arms, and hands).

III. Beyond the Production Model: An interdisciplinary approach to Inter-(Dis)fluency

It has been shown throughout this chapter that defining notions of fluency and disfluency can be particularly challenging, given the fact that they are rooted in terminological, methodological, and theoretical differences. In the previous sections, we mainly reviewed the psycholinguistic production-based approaches to Disfluency, centering on the simultaneous processing, planning, and production processes underlying the constructs of fluency and disfluency. This section will focus on different theoretical approaches relevant to the present study, which we briefly touched upon in the previous subsection, mainly Cognitive Grammar and usage-based linguistics (3.1.), Conversation Analysis and Interactional Linguistics (3.2.), and Multimodality and gesture studies (3.3.). This review will reflect our integrated and multidisciplinary approach to (Dis)fluency, embedded within our mixed-method framework, further described in Section VI and Chapter 2.

3.1. Cognitive Grammar and Usage-based linguistics

In the present section, we first present the core features of the Cognitive Grammar framework (3.1.1.), and explain how it is relevant to the study of (dis)fluency (section 3.1.2). We then conclude with an introduction to different cognitive and usage-based models of (dis)fluency (in section 3.1.3).

3.1.1. Key principles of Cognitive Grammar and usage-based linguistics

Cognitive Grammar, initially introduced by Langacker, (1987, 1995, 1999), provides a framework to language which considers grammar as not built up out of syntax and semantics respectively, but consisting of symbolic units, made of form-meaning pairings. Its central claim is that grammar is "per se a symbolic phenomenon, consisting of patterns for imposing and symbolizing particular schemes of conceptual structuring" (Langacker, 1998, p. 2 in Cienki, 2015b). Therefore, the main linguistic components of a language, such as grammar, lexicon, and phonology, are not independent, but inter-related. The understanding of these linguistic components relies essentially on speakers' general cognitive abilities, such as perception, attention, and categorization. The semantics of a language, for instance, can be related to some of these cognitive abilities, such as perception; i.e. the way speakers draw on their

perceptual experience to conceive and construe a situation. Geeraerts (2006, p. 4) reviewed four specific characteristics of Cognitive Linguistics in relation to linguistic meaning and semantic categories, of which we summarize the main points below. These four tenants are central to the understanding of the theory of cognitive grammar:

1) Linguistic meaning is perspectival
2) Linguistic meaning is dynamic and flexible
3) Linguistic meaning is encyclopedic and non-autonomous
4) Linguistic meaning is based on usage and experience

The first point implies that meaning is not simply an "objective reflection of the outside world" (Geeraerts, 2006, p. 4) but a way of shaping it through perception and conceptualization. Speakers thus construe an objective situation in the world in different ways, based on their viewpoint of the event (e.g. the particular spatial perspective of a speaker will affect his or her choice of a prepositional phrase, saying "in front of" versus "at the back of" when describing an object in space). The second point states that speakers' experiences of the world can change, which requires them to adapt their semantic categories accordingly. Therefore, a semantic category (e.g. *electronics*) is not fixed and static but can give room to more nuanced meanings (e.g. a circuit, or the branch of physics and technology concerned with circuits), which means that language is made of dynamic and flexible structures shaped by speakers' experiences. The third point further emphasizes the idea that meaning reflects speakers' whole experiences as it is not separated from other forms of knowledge of the world, but integrated with other cognitive abilities. In this sense, speakers are seen as embodied beings, and their organic nature influences their experience of the world, which is in turn reflected in their language use. Lastly, the last point claims that meaning is primarily based on usage and deeply rooted in experience. Speakers' experience of a language is thus based on its actual use in real life conversations, not on word entries in a vocabulary, or syntactic structures in a grammar book. In light of this approach emerges the framework *of usage-based linguistics* (e.g. Barlow, 2013; Bybee, 2008; Croft, 2000) which considers linguistic structures grounded in their observation of actual language use.

Speakers' knowledge of a language, in terms of language processing or language acquisition is influenced by their categorization and conceptualization of experience.

These assumptions stand in sharp opposition with the generativist and structural Saussurean tradition of *langue* and *parole* which distinguish between the level of the message structure (*langue*) and the level of language use (parole). Similarly, in the perspective of generative grammar (Chomsky, 1965), a distinction is made between the concept of *competence* versus *performance*; the latter being of little significance according to generativists. The usage-based model rejects that hierarchy, and views language as a dynamic system whereby linguistic units emerge from general cognitive processes. Bybee (2010) identified several cognitive processes influencing the use and development of linguistic structure (summarized by Ibbotson, 2013, p. 2): (1) categorization; identifying tokens as instances of a particular type, (2) chunking; the formation of sequential units through repetition or practice, (3) rich memory; the storage of detailed information from experience; (4) analogy; mapping of an existing structural pattern onto a novel instance, and (5) cross-modal association; cognitive ability to form link and meaning. The first two processes (categorization and chunking) are particularly relevant to the study of (dis)fluency, and will be further developed in the following subsection. The notion of frequency plays a crucial role in the integration of usage-based processes, as the more frequently items consistently co-occur together, the more likely they will become automatized.

This usage-based model of language is particularly relevant to the domain of language acquisition. For instance, Tomasello's (1995, 2003) groundbreaking work on language acquisition and social cognition has shown the way children acquire linguistic conventions through social and cognitive skills. For example, infants begin to understand utterances in a communicative context at around the age of nine to twelve months when they start acquiring joint attentional skills involving outside objects, such as following a parent's pointing or gaze (Tomasello et al., 2005). In other words, children come to understand and acquire language in interaction, captured in intersubjective contexts embedded within joint attentional formats. Joint attention, the process through which interlocutors rely and focus their attention on the same experience, is a prerequisite for social interaction, and plays a fundamental role in language development (Tomasello et al., 1995). In this sense, language use can be viewed as a form of social interaction, involving cognitive processes such as the ability to take other people's knowledge, intentions, and beliefs into account (Clark, 1996; Tomasello, 2003).

In sum, language is made of fluid categories and dynamic structures, altogether shaped by experience, usage, communication, and other cognitive abilities such as conceptualization or processing. Its acquisition and development rely on actual language use in intersubjective environments i.e., through interaction. When speakers interact with one another, they engage in two main activities, one which consists in conveying social intentions, and another establishing joint attention (Tomasello 2003). While this perspective was developed mainly in terms of its implications for first language acquisition, it also has an effect on second language acquisition (e.g. Bybee, 2008; Segalowitz, 2016; Wulff & Ellis, 2018). Wulff & Ellis (2018, p. 37) presented two assumptions characterizing usage-based approaches to language learning: (1) "the linguistic input learners receive is the primary source for their second language learning", and (2) "the cognitive mechanisms that learners employ in language learning are not exclusive to language learning, but are general cognitive mechanisms associated with learning of any kind". In other words, second language learners acquire conventionalized constructions (i.e. form-function mappings, or syntactic frames, see Goldberg, 2006) in their target language through repeated exposure, but also through processes of abstraction (e.g. deriving a general rule from the usage of a prototypical construction such as -s + 3rd person singular). Bybee (2008) looked at the effect on token frequency (i.e. the number of times a unit appears in speech) in the process of SLA, and its "conserving effects" (e.g. how repetition strengthens memory representations for linguistic forms and make them more accessible (Bybee, 2008, p. 218). This plays a key role in the process of acquisition, as the more exposed a learner is to a given construction (such as irregular forms), the more likely they will produce the constructions correctly. Bybee also points out a "reducing effect" of frequency, which is the observation that repeated constructions (such as greetings) tend to reduce phonetically. This poses a challenge for language learners who need to acquire the right phonological variants of a phrase (e.g. *gonna* for *going to*). Moreover, L2 learners may be faced with another challenge, which deals with the ability to take into account the social demands of communication (i.e. establishing joint attention and conveying social intentions) in order to communicate successfully, thus *fluently* (Segalowitz, 2016). We will examine this point more closely in the following section.

3.1.2. Why study (Dis)fluency in the framework of Cognitive Grammar?

We shall now turn to the relevant theoretical contribution of Cognitive Grammar and its usage-based framework to the study of (Dis)fluency. As mentioned earlier, one of the central tenants of usage-based grammar focuses on the embodied and social nature of interactional discourse, which is of interest to SLA, as the social demands of communication (i.e. conveying social intentions and establishing joint attention) may have an impact on the speakers' ability to communicate fluently. As Segalowitz (2016, p. 14) pointed out, the study of L2 fluency has often been decontextualized from the social and communicative situations in which the target language is acquired. On that account, he argued that L2 fluency should be regarded as "the outcome of a dynamical system where cognitive, social, sociolinguistic, pragmatic, and psycholinguistic considerations interact in complex ways" (Segalowitz, 2016, p. 18). This argument is in tune with the central assumptions of usage-based grammar. Based on these considerations, Segalowitz presented a framework situating fluency in a larger theoretical context, which offers three dimensions of fluency (utterance, cognitive, and perceived fluency)[27]. We believe that these assumptions regarding the dynamic and embodied nature of L2 fluency should also be applied to L1 fluency, which contributes to our multilevel and integrated approach to (Dis)fluency, briefly introduced in section 2.3, and which is further developed in section IV.

However, while the framework of Cognitive Linguistics takes into consideration several central claims of usage-based grammar *in theory*, it is not systematically done in practice. Cienki (2012, 2015b) discussed several implications regarding the analysis of "usage events" (i.e. contextualized linguistic units) within the framework of Cognitive Grammar, and pointed out one important limitation found in previous analyses. He gave several examples based on the analyses of Langacker (2008), such as "The students had collected a lot of money for the trip" (Langacker, 2008, p. 121, in Cienki, 2012), and argued that most analyses using Cognitive Grammar were based on constructed sentences, and not actual language use captured in interaction. The contribution of Conversation Analysis (CA) and Interactional Linguistics[28] is thus truly relevant to the analysis of usage events, as it considers turn units in a conversation which play a fundamental role in the structure of the talk. CA research also emphasizes

[27] This framework will be further described in section 3.1.3.
[28] The field of CA and Interactional Linguistics will be presented in Section 3.2.

the importance of repairs and their role in negotiating interaction, which further accounts for a positive view of fluencemes (cf section II. 2.2.1). As Cienki (2015b, p. 502) argued: "they are not mere dysfluencies". Talk and usage events of spoken language in general thus involves complex systems that change dynamically moment by moment". Cienki (2012, 2015b) further addressed this matter and discussed other kinds of recurrent behavior found in usage events, other than words and syntactic structures which have been the traditional subject of linguistic analysis. He focused on the three following behaviors: (1) non-lexical sounds, (2) intonation, and (3) gesture[29], and argued that despite being nonlinguistic per se, these recurrent structures could also gain symbolic status. This has a number of implications for the present study of (dis)fluency. First, it considers the variability and variation of fluencemes according to their context of use. As Cienki (2012,2015) pointed out, in certain contexts some non-lexical words (e.g. "uhm") can constitute a turn of talk, but less so in other settings. This implies that non-lexical sounds can gain a symbolic relation to certain meanings, with for example the association of "uhm" with uncertainty. But this mapping can be done to varying degrees; as we have seen (section 2.2.2.) filled pauses are not always related to feelings of uncertainty, and in some contexts, they may not have such a strong form-meaning correspondence. Moreover, filled pauses have a more fixed form-meaning mapping than other non-lexical sounds, such as inbreaths for example, as the latter are more likely to be associated with the marking of prosodic-syntactic boundaries (Trouvain et al., 2019). However, in some specific contexts, inbreaths can also gain symbolic status, and be used to mark a dispreferred answer (Hoey, 2014). While Cienki's (2012, 2015b) arguments are only made on non-lexical sounds (i.e. "uh", "mm", "uh uh"), we will show how these claims can be applied to other fluencemes as well. The idea that non-lexical sounds, and hereby fluencemes, can be considered lexical (hence "fluent") in different degrees and according to their context of use is one of the main theoretical claims of the present study, based on Crible et al., (2019)'s functionally ambivalent approach to (Dis)fluency. We will return to this idea in section 3.1.3. and in section IV.

In sum, in order to get a full understanding of fluencemes, one should consider their occurrence in naturally occurring speech captured in situated interaction, as well as their degree of symbolic meaning and conventionality according to their context of

[29] See section 3.3. for more details.

use. This leads us to another central research field closely related to usage-based theory, highly relevant to the present study of (dis)fluency, which is *corpus-based* linguistics. As Geeraerts (2006) pointed out, corpus-based studies have not yet gained a prominent status in Cognitive Linguistics, except in language acquisition research which is a "domain par excellence" to test usage-based models of language (Geeraerts, 2006, p. 17). Corpus-based studies are also of great interest to the study of (dis)fluency, as more of them have emerged recently, some of which will be presented in the following section.

In her corpus-based study of (dis)fluency and discourse markers, Crible (2018) described two of the major tenants of usage-based grammar (among others) which lie at the core of the study of (dis)fluency, mainly frequency and schematicity. Frequent combinations of fluencemes, such as filled pauses and unfilled pauses, can become conventional to a certain degree if they are exposed to highly repeated instances, and thus form recurrent patterns of combination, which can then be schematized (e.g. *filled pause+ unfilled pause*). This claim is based on the corpus-based and experimental evidence that fluencemes very often co-occur with one another rather than exclusively on their own (e.g. Benus et al., 2006; Betz & Lopez- Gambino, 2016; Duez, 1991; Grosjean & Deschamps, 1972; Shriberg, 1994). Therefore, Crible et al., (2017, p. 71) discussed the *sequential* aspects of (dis)fluency: the fact that fluencemes can form specific patterns of combination, known as "sequences". They are thus better understood as constructions, which can be either "simple" (isolated tokens) or "complex[30]" (combined tokens). The more frequent the combinations, the less cognitively demanding they will be perceived. Rare combinations on the other hand (e.g. *filled pause+repetition+unfilled pause+repair),* will appear less automatic, thus more disfluent. One of Crible's (2018) major contributions was to treat sequence frequency as one factor of fluency in order to examine the extent in which rare and frequent combinations could reveal different degrees of (dis)fluency. This was done through systematic quantitative analyses which combined different variables, such as co-occurrence, position, and register variation. This will be further described in the following subsection. Additionally, Crible (2018) largely investigated the clustering of (dis)fluencies with discourse markers (e.g. coordinating conjunctions, adverbs, or

[30] The term "complex" is borrowed from Shriberg (1994, p. 58) which refers to disfluencies that overlap with one another.

interjections, all included in the fluenceme category[31]) which was also conducted in previous corpus-based studies in line with the usage-based framework. For example, Schneider (2014, p. 9) studied the formation of chunks, defined as multi-word units, and the frequency effect of filled pauses in the process of chunking (i.e. when words in a sequence become gradually more connected). She looked at the placement of filled pauses, and found *sentence-initial hesitation chunks* (Schneider, 2014, p. 237) which comprised chunked combinations of sentence-initial markers and hesitations. Her results showed that frequent coordinating conjunctions (e.g. *and* and *but*) often merged with filled pauses, forming chunks such as "anduh" or "butuh". Similarly, Tottie (2016) further claimed that filled pauses and discourse markers belonged to the same category of *planners* (cf section 2.2.2), as their frequent cluster revealed time-buying strategies. More recently, she investigated the frequency of filled pauses in a large dataset of American English journalistic prose (the *TIME* corpus and the *COCA*), and presented evidence that "*ehm*" ("uh", "um", and "er"), the use of which has largely increased in magazines and newspapers in the past decades, was on the cline of lexicalization (Tottie, 2019). She argued that their high frequency in written speech (7.5 per million words in the *TIME,* and 6.4 in *COCA*) revealed conscious choices on the part of the writers to use *ehm* as they would use a word, to convey their attitude towards the text. Tottie gave the example of the following sentence taken from the COCA: "He showed his dancing, um, skills" (Tottie, 2019, p. 120), and classified this instance as serving a sort of "tongue-in-cheek" function, implying that the matter should not be taken seriously by the reader. This corpus-based study of filled pauses in written speech thus provides further evidence that their status is highly flexible, with the potential of serving specific pragmatic functions: they behave like stance adverbs in written text, or are used unconsciously by speakers to buy time in a spoken conversation. Tottie (2019, p. 127) concluded: "Their appearance in writing definitely qualifies them as words, whereas their status in spoken language is better described as a continuum with low-to-high degrees of wordhood". Once again, this idea reflects the different degrees of conventionality and symbolic meaning found within fluencemes, following assumptions from usage-based theory.

[31] Crible's (2017) annotation scheme can be found in Appendix 1. Her typology of fluencemes (of which the present thesis is partly based) also includes discourse markers, but it should be noted that the latter are excluded from our analysis. This will be justified in Chapter 2.

To conclude, a number of corpus-based studies have shown the relevant contribution of usage-based theory to the study of (dis)fluency, which can be redefined by Crible et al., (2017, p. 71) as being:

(1) Sequential; fluency is the result of specific patterns of combination or "sequences."
(2) Situational: these patterns are confronted with social and contextual norms.
(3) Ambivalent: a particular pattern can be either fluent or disfluent depending on its distribution in the micro and macro-context.

Three central assumptions thus emerge. First, fluencemes are better understood in terms of constructions, whose degree of entrenchment (i.e. the process whereby linguistic units become entrenched in speakers' memories) relies on the high frequency of specific patterns. Secondly, this degree of entrenchment and conventionality is determined by social and contextual factors. Thirdly, the status of fluencemes is highly flexible and dynamic, showing either sides of fluency and/or disfluency depending on their context of use. These assumptions lie at the core of the present study which is also based on different cognitive and usage-based models of (dis)fluency, presented below.

3.1.3. Cognitive and Usage-based models of (Dis)fluency: towards a multi-dimensional model

We briefly mentioned the work of Segalowitz (2016) in the previous section, and shall now turn to a more detailed account of his three-fold model of L2 fluency. Segalowitz put forward three central ideas based on his framework, summarized as the following:

(1) a speech act is a dynamic process,
(2) fluent speech is characterized by rapid, automatic, and efficient speech,
(3) a speech act relies on its communicative acceptability according to the expectations of the interlocutor.

These claims have already been introduced throughout this chapter by different researchers in different research fields (cf section 1.1.;2.1.1., 2.2.2., and 3.1.2.) In a similar vein, Segalowitz drew on the work of several authors (e.g., Goldman-Eisler, 1958; Meisel, 1987; Rehbein, 1987) to situate fluency in a larger theoretical framework, thus overreaching the field of SLA, and integrating cognitive, sociolinguistic, pragmatic, and psycholinguistic considerations. This is reflected in his multidimensional model of L2 fluency, which distinguishes between three different levels of analysis (Segalowitz, 2016, p. 5):

(1) *Utterance fluency*: Fluidity of observable speech as characterized by measurable temporal features (e.g. syllable rate, duration, and rate of hesitations).
(2) *Cognitive fluency*: Fluid operation (i.e. speed, efficiency) of the cognitive processes underlying L2 speech acts. It does not only include the articulatory act itself (i.e. utterance fluency) but the mobilization and temporal integration of mental processes that give rise to the utterance (Goldman-Eisler, 1968). These cognitive processes include all the demands inherent to utterance construction (e.g. semantic retrieval, planning).
(3) *Perceived fluency*: subjective judgments of L2 speakers' oral fluency.

In this framework, a distinction is made between the fluidity of observable speech at the level of the utterance, the cognitive processes underlying the production of utterances, and the perception of the final speech output. All of these dimensions are inter-related, as a disfluent execution of cognitive operations (i.e. semantic retrieval) can potentially result in a disfluent speech output, which will in turn be perceived and judged as disfluent.

Another relevant theoretical and methodological framework of L2 fluency was introduced by Götz (2013) in which she coined the term *fluenceme* (cf section 2.2.3). The latter were also categorized in three different types: fluencemes of production (similar to *utterance fluency)* which are related to temporal variables and fluency-enhancement strategies (e.g. filled and unfilled pauses, repeats, and repairs); perceptive fluency (similar to *perceived fluency*) based on the listener's attention; and nonverbal fluencemes which include aspects of non-verbal communication (e.g. hand gestures and facial expressions). The whole list of fluencemes is given in Table 2.

Götz's definition of fluency is truly valuable for the perspectives of the present multi-approach and multi-level investigation of inter-(dis)fluency, as it encompasses numerous aspects of spoken communication (e.g. prosody, lexicon, pragmatics, discourse, non-verbal communication). Her framework, on a par with Allwood et al., (2005)'s which integrates the visual-gestural modality of discourse, most certainly goes beyond the first definition of (dis)fluency we presented in Section I. Her three-fold typology of fluencemes is not quite the same as Segalowitz's model, as she mostly focused on a speaker-based approach which includes observable features of communication, while Segalowitz targeted a more abstract cognitive based approach (as pointed out by Crible, 2018, p. 21).

***Table 2**. Götz's (2013) model of L2 fluency*[32]

Productive fluency	Perceptive fluency	Nonverbal fluency
Temporal variables		
▪ Speech rate	▪ Accuracy	▪ Gestures
▪ Mean length of run	▪ Idiomaticity	▪ Facial expressions
▪ Unfilled pauses	▪ Intonation	▪ Body language
▪ Phonation/time ratio	▪ Accent	▪ Looks
Fluency-enhancement	▪ Pragmatic features	▪ Emblems
▪ Speech management strategies (repeats, filled pauses)	▪ Lexical diversity	
	▪ Register	
	▪ Sentence Structure	
▪ Discourse markers		
▪ Smallwords		

More in line with Cognitive Grammar theory, Grosman (2018) introduced a socio-cognitive framework of fluency, based on Schmid & Gunther's framework on salience (2016). In her model, three dimensions of fluency were also evaluated: the grammatical, the discourse-level, and socio-interpersonal dimension[33]. This multi-level approach accounted for an evaluation of speech as :

> being disfluent when the (grammatical, discursive, socio-interpersonal) discourse expectations of the hearer are over-confirmed or over-deceived. This evaluation depends on the degree of convergence of the discourse with the hearers' expectations (Grosman et al., 2019, p.23).

This tridimensional evaluation of fluency was used for experimental purposes (cf Grosman et al., 2019; Grosman, 2018), and was based on six assertions which corresponded to an evaluated dimension of fluency, presented below (Grosman et al., 2019, p. 24).

(1) Grammatical

- The sentence is well formed.
- The sentence includes hesitations.

[32] This table is adapted from her book, and is inspired by Crible's (2017, p. 29)

[33] These are translations made by Grosman and colleagues in their paper, but the original terms were "fluence componentielle linguistique-phrasique", "fluence socio-interpersonnelle", and "fluence situationnelle" (same order, cf Grosman, 2018, p. 296). Their translation caused slight terminological changes, as "socio-interpersonnelle" became "discourse-level", and "situationnelle" became "socio-interpersonal".

(2) Discourse-level

- The speech is fluid.
- The speech is nice to listen to.

(3) Socio-interpersonal

- The speech in this context appeals to me.
- The speech in this context is improper.

The first dimension, similar to the *utterance* and *productive* levels of fluency presented earlier, refers to decontextualized linguistic expectations, in relation to wellformedness and grammaticality judgments. The second dimension is based on the discourse flow with regards to social expectations. Lastly, the last dimension is related to interpersonal norms and expectations in a specific register.

The notable contribution of the fluency models presented above is their acknowledgement of the different dimensions of (dis)fluency (also proposed by Lickley, 2015 cf section 2.1.1.). While psycholinguistic and SLA research has mainly focused on the grammatical, utterance, or productive dimension of disfluency (cf section I)[34], others, who disagreed with the term "disfluency" have rather focused on its pragmatic and socio-interpersonal dimension (cf section 2.2.2). The understanding of these different dimensions may thus further justify the multiple terminological backlashes found in the literature. Being disfluent at the level of the utterance is not quite the same as being disfluent at the level of the interaction, where social expectations differ greatly from grammaticality judgments. For instance, Grosman et al., (2019) found that speech samples which contained repetitions were found to be judged more disfluent by listeners when evaluating the socio-interpersonal dimension of fluency than the discursive or grammatical dimensions. However, this perception of fluency also strongly relies on the type of speech produced, in line with the social expectations of the situation. The combination of these different dimensions will be further accounted for in section IV.

In sum, the three proposals presented by Segalowitz , Götz, and Grosman are a valuable contribution to the present study of inter-(dis)fluency for two main reasons: first, their typology of (dis)fluency goes beyond the traditional definition of disfluency presented in section I, as it integrates several central aspects of human communication beyond the level of speech production (e.g., social expectations, situational features,

[34] A lot of psycholinguistic work has also been done on the perception of disfluencies, which was briefly mentioned in Section 1.2.

cognitive processes, aspects of nonverbal communication, etc.); secondly, their tridimensional account of (dis)fluency vouches for a multi-approach perspective on these phenomena, which further justifies our need to situate our study within a larger integrated theoretical framework. However, it should be noted that these models were used for different purposes, mostly experimental in the case of Grosman et al., (2019) and Götz (2013), with a strong focus on the hearer's perspective, which goes beyond the scope of this thesis. Moreover, as Crible (2018) pointed out, Götz only extracted several features from the data semi-automatically, but without annotating them systematically with an annotation scheme, "a methodological choice which is time-saving but perhaps questionable from the point of view of replicability, exhaustivity and granularity" (Crible; 2018, p. 22[35]). Götz (2013)'s framework is thus mostly conceptual, and has not been formalized into an annotation scheme. This gave rise to a novel corpus-based annotation model of (dis)fluency, introduced by Crible et al., (2019)[36], which is described below.

Borrowing from Shriberg's (1994) formal *disfluency classification* (cf section 1.3.) and grounded in a functionally ambivalent view of (dis)fluency (cf section 2.2.3.), the collaborative work of Crible et al., (2019)[37] combined a fine-grained identification of (dis)fluencies with a thorough and reliable technical format. Their typology of *fluencemes* (following Götz, 2013) offers a systematic and detailed annotation of ambivalent devices, which includes "typical" non-lexical (dis)fluencies (e.g. filled pauses, repeats, deletions, repairs etc.) and more lexical ones, such as discourse markers (e.g. *well*, *I mean*, *but*). Their work is based on three central assumptions, which are in tune with the tenants of Cognitive and usage-based grammar (Crible et al., 2019, p. 21):

(1) Formally similar (dis)fluency markers can be functionally different (i.e. fluent or disfluent).

[35] For a more detailed review on the number of technical and theoretical drawbacks found in previous approaches to disfluencies, read Crible (2017, p. 21)

[36] This research team was part of a five-year research project entitled "Fluency and disfluency markers. A multimodal contrastive perspective" conducted at the University of Louvain and Namur. They were involved in a large scale usage-based study of (dis)fluency markers in spoken French, L1 and L2 English, and French Belgian Sign Language. For more information go to *https://uclouvain.be/fr/instituts-recherche/ilc/fluency-and-disfluency-markers-a-multimodal-contrastive-perspective.html (last retrieved on August 26th 2021)*

[37] This includes the work of Crible (2018), Dumont (2018), Grosman, (2018), and Notarrigo (2017).

(2) This difference is due to the variation in distributional factors (such as frequency, syntactic position or co-occurrence) and interactional factors (e.g. register expectations).

(3) (Dis)fluency markers are the result of general cognitive processes and may therefore be incorporated into typologies covering different spoken and signed languages.

Their annotation model was applied to spoken L1 and L2 speech (cf Dumont, 2018), as well as signed languages (with the cross-modal study of palm-up signs and filled pauses, cf Notarrigo, 2017). Their typology of fluencemes included simple (i.e. when their structure only involves one part, e.g. a filled pause or discourse marker) and compound markers (i.e; when they consisted in two main parts, e.g. a repetition or a substitution), as well as diacritics (e.g. misarticulations, vocal lengthening)[38]. The clustering of immediately adjacent fluencemes were called *sequences* (cf 3.1.2.), which is a term that will also be adopted in this thesis. The internal structure of the sequences was analyzed, by taking into account their content (e.g. the number of markers found within a sequence and their combination pattern, such as *FP+UP+TR*) and their cluster (i.e. whether the markers occurred on their own or as part of a sequence, see Crible, 2017, p. 119). Moreover, the application of this annotation scheme was adjusted to the respective research agendas of the research team members. For instance, Crible's work (2017; 2018) targeted specifically the combination of discourse markers and other fluencemes, but did not systematically annotate all the other types (e.g. pauses, repeats, repairs etc.). Conversely, Grosman (2018) targeted all (dis)fluencies, with a focus on pauses, prolongations, and repeats, but did not annotate discourse markers. Similarly, the present corpus-based study of (dis)fluency phenomena is adapted from Crible et al.,'s annotation scheme (2019), with a number of conceptual and technical adjustments made to address our research purposes (cf Chapter 2).

To conclude, this section has outlined different theoretical and methodological frameworks of (dis)fluency situated within the scope of Cognitive Grammar and usage-based linguistics. These typologies altogether provided relevant analytical and methodological tools for the present study, which represents the first step toward the construction of our integrated multi-level and multi-modal approach to (dis)fluency.

[38] A more detailed list of the fluencemes included in the typology can be found in Appendix 1.

We shall now turn to the review of another major theoretical framework which further situates our study of inter-(dis)fluency in a larger theoretical construction.

3.2. Interactional Linguistics and Conversation Analysis

In the previous section, we mentioned Cienki's (2012, 2015) observation that most examples in Cognitive Grammar were typically based on constructed sentences, isolated from a larger interactional context (cf section 3.1.2). Similarly, most studies conducted on (dis)fluency phenomena focus on *the utterance level* of fluencemes (cf Segalowitz, 2016 section 3.1.3.) and the simultaneous cognitive, processing, and planning processes associated with their production[39]. Even though some researchers have highlighted their contextual, situational and interpersonal dimension (cf section 3.1.3.), their qualitative examples are much too often based on isolated utterances which do not illustrate their sequential unfolding within the course of interaction. Let us draw on the following examples, taken from Kjellmer, 2003 p. 185):

(1) It does take a bit of time to get to know the Mexicans erm er and I suppose in that sense they're sort of like the British of Latin American.

(2) I came to nearly all the university open days and asked er questions and got to know extra people there er er and er I took some interest in the departmental activities.

(3) He's just horrible and erm <M01> Oh I'm so sorry FX. <F01> Yeah.

The three examples listed above show decontextualized material taken from the Cobuild-Direct corpus, which were used by the author to illustrate turn-holding and turn-yielding functions of filled pauses. However, this method seems somewhat at odds with the actual mechanisms of turn-taking which rely on the organization of turns in a conversation, and this information is entirely left out from his analyses. As a matter of fact, a large number of studies on (dis)fluency tend not to illustrate their findings with qualitative examples, but exclusively rely on quantitative results. While it is true that quantitative findings can give a robust, representative, and statistically valid overview of the data, they fail to illuminate particular and complex instances within their context of use. This issue is even more relevant to the study of (dis)fluency which has shown to be functionally and interactionally ambivalent. We believe that this ambivalence can be further examined in detailed qualitative analyses, through the

[39] Except for a few notable exceptions, such as Tottie (2011,2015,2016,2019), Clark & Fox Tree (2002), Bortfeld et al., (2001) among others.

medium of conversation analytic tools, further grounded in the framework of interactional linguistics.

3.2.1. Introduction to the interdisciplinary framework of Interactional Linguistics

The interdisciplinary framework of Interactional Linguistics (IL)[40] emerged in the early 21st century, altogether with a growing community of linguists who were interested in studying grammar and prosody from a specific interactional approach. This new community, which "takes an interdisciplinary and a cross-linguistic perspective on language" and whose goal is to understand "how languages are shaped by interaction" (Couper-Kuhlen & Selting, 2001, p. 3) includes a number of linguists from different theoretical backgrounds i.e., discourse analysis, sociolinguistics, anthropological linguistics, discourse functional linguistics, and usage-based grammar. With this in view, Couper-Kuhlen & Selting (2001) traced back three main theoretical influences (discourse-functional linguistics, Conversation Analysis, and anthropological linguistics) which represented major stepping stones towards establishing the Interactional Linguistics Framework. These theoretical contributions are summarized below.

First, *functional linguistics* (e.g. Halliday, 1973; Jakobson, 1960) focuses on the way social and cognitive language functions influence the organization of the linguistic system. It completely rejects the formalist paradigm (e.g. Chomsky, cf section 3.1.1) which regards grammar as an autonomous system and a mental phenomenon, independent from other social and cognitive functions. The discourse-functional tradition draws on the relation between particular linguistic units (e.g. utterances) and language functions. For instance, Jakobson (1960) offered six different language functions (referential, phatic, emotive, poetic, conative, and metalinguistic) which relate to different components of the speech situation (e.g. the context, the type of message produced, the addressor, or the addressee; see Schiffrin, 1994, p. 33). Functional linguistics were also interested in the integration of discourse and grammar by looking at the preference of certain syntactic configurations (e.g. the preference of a noun phrase over a pronoun) in discourse (cf Du Bois, 1985 on the *preferred argument structure*). As Couper-Kuhlen & Selting (2001, p. 2) pointed out: "functional linguistic research – although it did not focus on conversational

[40] There is another closely related research field dealing with aspects of social interaction known as *Language and Social Interaction* (LSI), see LeBaron et al., (2003) for review.

interaction – was instrumental in establishing a mindset for the study of language which saw linguistic form as something "to do things with" on situated occasions of use".

The second major influence of Interactional Linguistics is *Conversation Analysis* (CA; Sacks et al., 1974) which introduced major analytic tools for the study of social interaction, through qualitative micro-analyses of *talk-in-interaction* (i.e. naturally occurring speech in every day conversation, (cf Schegloff, 1991). CA regards interaction as "the home environment of language", (Sidnell, 2016, p. 2), an orderly, interactionally managed system, whereby norms and practices are shaped by speakers' *actions*. Actions refer to what the co-participants of a conversation are doing interactionally in relation to one another (Pomerantz & Fehr, 2011; Schegloff, 1996a). In other words, the act of speaking does not only involve the individual productions of one speaker, but its coordination and cooperation with other participants of a conversation within turns. Goodwin & Heritage (1990, pp. 292–293) further elaborated on this conversation-analytic approach to interaction:

> Within most traditional perspectives, analysis focuses exclusively on the speaker. The hearer is treated as a figment of the speaker's imagination. From the CA perspective, however, hearers are co-participants who can decline as well as accept the status offered them [...] hearers are active participants in the process of building a turn at talk.

This perspective is thus very different from the psycholinguistic assumptions that speech planning results from several cognitive operations that are mentally performed by an individual speaker (cf section I.1.1.) without taking into account the co-speaker's contribution in the interaction. We will return to this point in section 3.2.3.

Sacks et al., (1974)'s seminal paper, entitled *A simplest systematics for the organization of turn-taking for conversation* sketched out some of the fundamental aspects underlying the construction of talk-in-interaction, and demonstrated the way speakers, when engaged in ordinary, everyday practices, co-produce stretches of talk in orderly ways, which can be subject to detailed qualitative analyses. As Schegloff (1991) further argued, the expression of messages in specific linguistic forms (i.e. utterances) does not result from mental cognitive processes, but is shaped by the orderly structure of the interaction. Sidnell (2016, p. 2) reviewed some the

fundamental aspects regarding the orderliness of interaction, summarized in the following points[41]:

(1) *Distribution of turns*. Conversations are governed by turn-taking mechanisms which are organized in various ways by the participants (e.g. turn holding, turn yielding, turn allocation) through *turn constructional units* (i.e. lexical items, phrases or clauses which determine the shape of a speaker's possible turn).

(2) *Addressing problems of hearing, speaking or understanding*. During the course of their talk, participants may need to signal speaking, hearing, or understanding troubles through repairs[42] which can be initiated by the speaker or another participant (Schegloff et al., 1977).

(3) *The formation of actions*. Participants are able to project or recognize what type of action is being performed (e.g. responsive action, pre-sequence action, sequence-initial action) by adding stretches of talk. For example, disaffiliative actions can be prefaced by interjections such as "well" (McHoul, 1978 in Goodwin & Heritage, 1990).

(4) *The sequential dimension of actions*. Actions are organized into sequences (e.g. *adjacency pairs*, sequences made of two turns, such as question-answer or request-accept) as a way to construct "an architecture of intersubjectivity—a basis for mutual understanding" (Sidnell, 2016, p. 2).

These aspects are reflected in the following short extract, taken from Sidnell (2016). This excerpt comprises a conversation between two speakers, Shelley and Debbie. Shelley's first sequence-initial action, which is a declarative TCU ("you were at the halloween thing"), is followed by Debbie's lexical TCU ("huh") which signals her trouble understanding Shelley's initial statement, and further invites Shelley to reformulate.

01 Shelley:	you were at the halloween thing.
02 Debbie:	huh?
03 Shelley:	the halloween p[arty
04 Debbie:	[ri:ght.

Figure 5. *Shelley & Debbie (Sidnell, 2016, p. 3)*

[41] Some of these aspects will be further exemplified in section 3.2.2., and throughout our qualitative analyses in Chapters 3, 4, and 5.

[42] Note that the term "repair" does not have the same implications as Levelt's (1989) term (cf section 1.2.). This terminological difference will be briefly discussed in section 3.2.2.

Just before Shelley reaches the point of her TCU's completion (l.3), Debbie joins in with a lexical TCU ("right" l.4), as a way to signal her understanding and alignment with Shelley, following the reformulation. Each utterance, or TCU, are thus better understood within the overall structure and orderliness of the ongoing talk-in-interaction.

To sum up, CA is characterized by "the conversation-analytic understanding of speech exchange as social interaction and the conversation-analytic tools of micro-analysis and participant-oriented proof procedures" (Couper-Kuhlen & Selting, 2001, p. 2) which constitutes one of the main foundations of Interactional Linguistics.

Lastly, the field of *anthropological linguistics*, which draws on the study of "real people in real time and real space" (Duranti, 2011, p. 3) is the final stepping stone towards establishing the framework of Interactional Linguistics. Some anthropological linguists have conducted cross-cultural studies in relation to *language socialization* (e.g. Ochs, 1996; Ochs & Schieffelin, 2011; Schieffelin & Ochs, 1986), which refers to the process whereby children and novices are socialized through language (Ochs, 1996). Linguistic anthropology research integrates discourse and ethnographic methods to investigate the way language practices are shaped socially and culturally within different situations. The term *situation* encompasses a variety of social and communicative dimensions, such as the social identity of the speaker, the type of activity (e.g. storytelling, interviewing, giving advice), and affective and epistemic stance (Ochs, 1996). The integration of the cross-linguistic and cross-cultural perspective into the analysis of interaction is, according to Couper-Kuhlen & Selting (2001, p. 3) "crucial to the interactional linguistic enterprise".

In sum, the field of Interactional Linguistics brings together a variety of disciplines[43] and theoretical fields whose main goal is to investigate the different linguistic resources and practices that are incrementally accomplished over time in the course of the talk, altogether shaped by the social, cultural, and sequential aspects of conversation. The accomplishment of discourse is inherently interactional (Schegloff, 1982) as it draws on the collaborative achievement of the co-speakers, who coordinate their actions with one another. This area of research is thus in sharp contrast with the

[43] However, each discipline has their own research interests that do not necessarily overlap; cf Fox et al., (2013) for a review on the differences between CA and Interactional linguistics, and Duranti (2011) on the distinction between linguistic anthropology and other domains of linguistics. On the other hand, Goodwin & Heritage (1990) also discussed the contributions of CA to the field of anthropological linguistics.

formalist approach to discourse and grammar which views language primarily as a mental and autonomous phenomenon, independent from human social or cognitive abilities. In this thesis, we will especially draw on the analytic tools provided by CA, as well as the *participation framework* further described below.

A major contribution in the field of social interaction and linguistic anthropology is found in the work of C. Goodwin and M.H. Goodwin (Goodwin, 1981, 2003, 2007, 2010, 2017; Goodwin & Goodwin, 1996, 2004, 1986) who studied embodied *participation frameworks* (initially introduced by Goffman, 1981). Participation refers to "action demonstrating forms of involvement performed by parties within evolving structures of talk" (Goodwin & Goodwin, 2004, p. 222). Within this framework, the focus is essentially on two interactive practices, mainly (1) how participants orient themselves in ways relevant to the activities they are engaged in, and (2) how situated analysis of an emerging course of action shapes the further development of action (Goodwin & Heritage, 1990, p. 292). In this respect, participation is viewed as a "situated, multi-party accomplishment" (Goodwin & Goodwin, 2004, p. 231), in which the status of the participants (e.g. *speaker* or *hearer, addressee or recipient* etc. cf Goodwin & Heritage, 1990) can shift depending on the organization of particular situated activities (e.g. assessment, topic initiation, story preface). For instance, during a storytelling activity, a speaker may need to create a complex participation framework which will include multiple participants, such as the hearers of the story, the character in the story who is doing its retelling, as well as the characters within the story who are absent from the telling (Goodwin & Goodwin, 2004). Additionally, during a homework activity (cf Goodwin, 2007) the answer to a question can be provided by a participant through the simultaneous use of different semiotic resources (i.e. speech, gesture, and the actual homework paper) which mutually work with one another. The participation framework is thus also established through the alignment of the participant's *bodies,* who can make use of hand gestures to build an embodied action during the course of the talk (Kendon, 2004). Speakers thus have a multiplicity of semiotic resources at their disposal, co-deployed altogether to build actions oriented to the hearers, and which are all relevant to the ongoing situated activity. The speakers' deployment of multiple semiotic resources for building action is hence another central aspect of the interactionist approach to social interaction. The integration of different semiotic fields (i.e. the stream of speech, hand gestures, and body posture and orientation) as well as different levels of analysis

(syntax, lexical, prosody, pragmatics) are essential to the study of social interaction and their underlying situated practices. The study of embodied and multimodal interaction (e.g. Stivers & Sidnell, 2005) is further addressed in section 3.3.1.

To conclude, the interdisciplinary field of Interactional Linguistics, which includes multiple contributions from CA, anthropological linguistics, and discourse-functional linguistics, among other fields, altogether provides a relevant theoretical and methodological framework to the study of fluencemes, which leads us to the following subsection, focusing on the study of "repairs" in CA.

3.2.2. The study of conversational repairs in talk-in-interaction

Before addressing studies investigating (dis)fluency phenomena in the field of interactional linguistics, it should first be noted that the term *disfluency* is virtually excluded from all researchers' analyses. This is not surprising, given what the term *disfluency* entails (cf section I), which is in sharp opposition with the central assumptions of Conversation Analysis and Interactional Linguistics. The latter used the label *repair* (Schegloff et al., 1977) instead, as well as other terms, such as *non-lexical sounds*, *vocalizations, uhm*, etc. The term *repair*, although homonymous with the one used by Levelt (1983), which presupposes that an item needs to be repaired, has different implications. In this section, we provide a brief overview of the study of repair within the framework of Interactional Linguistics.

Repair is a key area of research within Conversation Analysis. It refers to an organized system which is addressed to deal with recurrent problems in speaking, hearing, and understanding. Repairs can be distinguished between self-corrections and other-corrections, which is similar to the distinction presented by Levelt (1983) [44], but this type of phenomena, unlike what Levelt addressed in his model, is "neither contingent upon error, nor limited to replacement" (Schegloff et al., 1977, p. 363). Repair mechanisms are in fact regarded as a sequential[45] phenomenon, which have a specific organization within the course of interaction. They include different segment parts, such as "initiation" and "outcome". Their initiation can be placed in three main positions, (1) within the same turn, (2) in a turn's transition space, and (3) in a subsequent turn. Self-initiations within the same turn are said to use "a variety of non-

[44] Cf section I.1.2.

[45] Note that the term "sequential" is different from the one mentioned in section 3.1.3. when discussing fluenceme sequences. In this case, it refers to sequences of turns or of talk found within the interaction, while in Crible et al's. (2019) case, it refers to the combinatory patterns of adjacent fluencemes.

lexical speech perturbations, e.g. cut-offs, sound stretches, "uh" etc. to signal the possibility of repair-initiation immediately following" (Schegloff et al. 1977, p. 367). It is interesting to note that this comment on the repair structure is similar to the description of the interregnum region of the *disfluency surface structure* (cf section I.1.2). This was also found in Goodwin & Goodwin (2004, p. 230) who illustrated the display of a repair structure, by taking into account the participants' gaze:

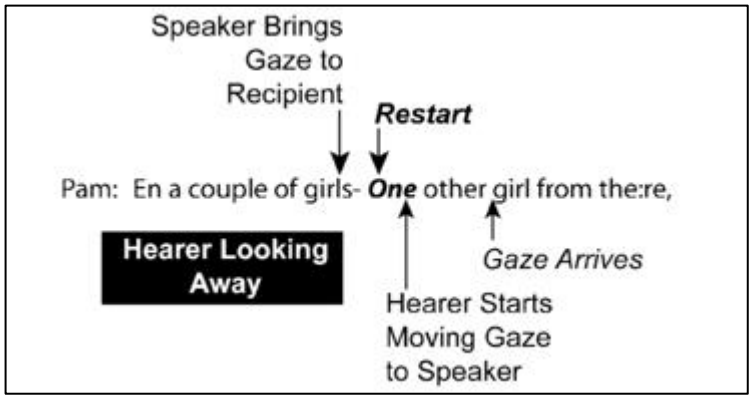


***Figure 6**. Example of Repair (Goodwin & Goodwin, 2004, p. 229)*

In this figure, Goodwin & Goodwin (2004) illustrate a case of restart initiated by a speaker during a conversational exchange. Within a participation framework, repairs are said to occur when speakers "lack the visible orientation of a hearer" (Goodwin & Goodwin, 2004, p. 229). In other words, when the positioning of a hearer is not oriented to the ongoing talk, a speaker can modify stretches of their speech to repair the participation format. During face-to-face encounters, hearers may resort to a range of visual or vocal backchannels (i.e. head nods or vocal continuers, cf Stivers, 2008) to display their status as recipient, or listener. Hearers can also display that they are attending to the speakers' talk by gazing towards them (Goodwin, 1981). However, in some cases (as in Fig. 6) hearers may not be fully oriented to the ongoing talk, which may invite speakers to interrupt and modify their utterance in order to secure the gaze of their addressee. Therefore, the grammatical structure of a sentence can be modified to adapt potential changes in the participants' status. During a word search for instance, repairs can be used to achieve a state of mutual orientation between speakers and hearers (cf section 3.3.1). Goodwin (1981, chapter 4) further explored similar types of repair phenomena, such as lengthening, additions, or phrasal breaks, which are said to achieve interactive tasks, still in relation to gaze. He showed how speakers needed to delay the final syllable of a word to coordinate its completion with the arrival of the

hearer's gaze. This kind of coordination was also achieved by adding an "uh" in the utterance. In this sense, instances of repairs are not limited to contexts of speech error, as they can also function as requests to secure mutual gaze, and help to achieve a state of mutual orientation in the talk. Similarly, producing "uhm" at the beginning of a turn can signal a speaker's commitment and understanding towards an ongoing activity. In a study on phone conversations, Schegloff (2010) showed how turn-initial "uhms" could be used as a reason for calling, or for launching a new topic in the upcoming sequence. Another study conducted by Morita & Takagi (2018) on *eeto*, the Japanese equivalent of "uhm", argued that its primary function was to indicate an "interactional concern" (Morita & Takagi, 2018, p. 32), as to mark attentiveness and commitment to the conversational task at hand. However, "uhm" can also mark a dispreferred answer, as Schegloff (2010) argued:

> Uhm [...] mark the "reason-for-initiating" an episode of interaction, that a dispreferred response is upcoming, that a dispreferred sequence is being launched, or that a sequence's ending has resisted consummation and is being tried again (Schegloff, 2010, p. 38).

Repair phenomena can thus also be associated to the *preference structure* of an interaction. The concept of preference refers to a "socially determined structural pattern" (Yule, 1996, p. 76) which is expected in a speaker's next action. For instance, during a greeting sequence, the expected answer to a first greeting (e.g. *hello*) would be another greeting (e.g. *good morning*), which conforms to social norms. This illustrates a case of preferred action. A dispreferred action, on the other hand, would mark an unexpected structure (Yule, 1996, p. 79), which does not meet the social requirements of the situation. Yule (1996, p. 81) gave a list of common patterns associated with dispreferred responses[46], one of which includes the entry "delay/hesitate" with pauses and "uhm".

In sum, repair phenomena (which includes instances of restarts, "uhms", pauses, repeats, and self-breaks) have been subject to thorough investigation within CA and the framework of Interactional linguistics. These analyses were conducted through micro qualitative examples of the data, drawing on CA analytic tools such as turn-taking or preference structure. Despite a few similarities in the analysis of the repairs' underlying structure, the conversation-analytic and interactionist approach is

[46] This table can be found in Appendix 1.

in sharp opposition with the psycholinguistic ones presented in section I. In the following section, we discuss the ways these two conflicting methodological and theoretical frameworks could be combined for the present multi-approach and multi-level study of inter-(dis)fluency.

3.2.3. Contribution of the field to the study of Inter-(dis)fluency

We shall now discuss the relevant contribution of the field of Interactional Linguistics to the present study of inter-(dis)fluency. It would seem, at first, that the terms *disfluency* and *talk-in-interaction* are incompatible, since the former focuses on the mental planning processes of an individual, while the latter deals with the social and cultural mechanisms shaping social interaction. However, the phenomena under study are analyzed in both fields (i.e. psycholinguistics and interactional linguistics), but with clearly different methodological tools and research purposes. As we claimed earlier (section 2.2.3. and 2.3.) we believe that (dis)fluency phenomena should not be restricted to a single label nor a single model, but should instead integrate a variety of perspectives. We suggest that the study of fluency and disfluency could gain insight from CA and Interactional linguistics for a number of reasons, elaborated below:

First, the study of (dis)fluency has much too often been excluded from traditional formal linguistics (cf *performance* vs *competence* distinction section 3.1.1.) while they can actually represent meaningful aspects of speech production (cf section I.1.2.). However, their role is not only restricted to the level of verbal utterance production, but to the level of interaction as well. As we have seen, fluencemes can exhibit essential features of talk-in-interaction, shaping the course of the talk which is constantly adapted to the exigencies of the interaction. Fluencemes thus result from both cognitive and socio-interactional processes. This view is also in line with the perspective of usage-based grammar described earlier (cf section 3.1.2.).

Second, within an Interactional Linguistics framework, fluencemes are seen as the byproducts of a dynamic and flexible grammar, shaped by conversational, social and discourse constraints (cf Mondada, 2001). This emphasizes the flexible and dynamic nature of fluencemes, whose status is determined by its contexts of production. This notion of flexibility was also presented within the scope of cognitive grammar (cf section 3.1.2.).

Third, the notion of fluency, and its negative counterpart DISfluency, which have mainly been restricted to the flow of speech, can include other types of flow, such

as the interactional flow. This could be applied to the principle of *progressivity* in CA (e.g., Schegloff, 2007; Sterponi & Fasulo, 2010; Stivers & Robinson, 2006) which refers to the smooth unfolding of interaction, based on the temporal development of the talk. As Schegloff (2007, p. 14) put it: "moving from some element to a hearably-next-one with nothing intervening is the embodiment of, and the measure of, progressivity". In some contexts, however, the interaction does not always go as smoothly e.g., Stivers (2001) who spoke of "interactional complications" (Stivers, 2001, p. 252) during pediatric encounters. Such interactional difficulties, which interfere with the immediate contiguity of the talk, can take the form of silence, or non-answer responses. Similarly, Sterponi & Fasulo (2010) showed that when speakers try to achieve intersubjectivity (i.e. mutual understanding) it can potentially lead to a breakdown in the interaction if they fail to reach mutual understanding, which does not respect the progressivity of the talk.

This third point is particularly relevant to the present study of ambivalent fluencemes. While in some contexts they can help speakers display their engagement to the ongoing activity (e.g. indicate mutual cooperation, display understanding, or yield a turn), they may also display a form of disengagement. For instance, during silences, speakers may choose not to cooperate with turn-allocation techniques (Hoey, 2015), or they can display their lack of involvement in the ongoing activity (Szymanski, 1999). In addition, when speakers make assessments (i.e. claim knowledge of what they are asserting, cf Pomerantz, 1984), they can be declined by second assessments (i.e. responses to the previous first assessment through agreement, or disagreement[47]) in the form of "delay devices" (Pomerantz, 1984, p. 70) such as "uh" or "well", as a way to preface disagreement. This can also relate to notions of *alignment* and *affiliation*. In a storytelling context, for example, *alignment* indicates a speaker's aligning with the turn-taking principles of the storytelling activity (i.e. that the teller of the story has the floor until its completion), and *affiliation* refers to a hearer's display of support to the speaker's stance (cf Stivers, 2008). Examples of *disaligned actions* (during storytelling activities) can thus illustrate the hearer's failure to treat the story told by the speaker as either in progress or over.

[47] Pomerantz (1984) in fact distinguishes between upgrade agreement, same evaluation, and downgrade agreements.

To conclude, the overall construct of (dis)fluency can further be extended to general principles addressed in CA such as (de)salignment[48], (dis)engagement, (dis)affiliation, (dis)preference and (dis)agreement. In this view, the process of *fluency,* understood as *flow,* or *continuity,* can refer to different types of flow at different levels of analysis (i.e. utterance, cognitive, interpersonal, cf section 3.1.3). This type of analysis is made possible with the combination of different methodological and theoretical approaches, which vouches for a multi-level and multi-approach study of inter-(dis)fluency. This is further elaborated in section IV.

3.3. Gesture studies and multimodal interaction

Though briefly, we mentioned several times throughout this chapter that the presence of visual-gestural behavior in discourse (i.e. gaze and hand gestures) was central to the analysis of face-to-face communication. In this respect, the goal of the present thesis is to investigate inter-(dis)fluency phenomena from a multimodal perspective i.e., by integrating multiple visual-gestural features of communication to the study of verbal fluencemes within embodied interaction. In the present section, we outline some of the central aspects related to the study of multimodal interaction (3.3.1), followed by the review of a few papers from the field of *gesture studies* which investigated the relationship between speech and gesture (3.3.2), and introduced different gesture classifications (3.3.3). The last subsection (3.3.4) is dedicated to an overview of the studies which worked specifically on gesture in relation to (dis)fluency.

3.3.1. Multimodality in the study of embodied interaction

In the past few decades, the study of what has commonly been labeled "nonverbal" or "non-linguistic" communication has increasingly become a central interest of research among scholars in various disciplines (e.g. cognitive linguistics, psycholinguistics, linguistics anthropology, interactional linguistics). With the rise of interactionist approaches to social interaction (cf section 3.2.1.) who started working on video recordings of everyday interactions, new perspectives emerged for studying language practices as *embodied* within their social, material, and spatial environment. This includes the study of gesture, gaze, head movements, facial expressions, body movements, as well as the manipulation of external objects in the environment (cf Boutet, 2018; Goodwin, 2003; Morgenstern & Boutet, forth.; Streeck et al., 2011). In

[48] Mondada & Traverso (2005) spoke of *(dés)alignements.*

this respect, this perspective on *embodied interaction* regards cognition as not a separate mental action working in the brain, but as an "embodied action" (De Jaegher & Di Paolo, 2007, p. 486 in Streeck, 2015). This "embodied turn" in social sciences (cf Mondada, 2016, p. 338; Streeck, 2015), marked the conception of what has been termed *multimodal* communication. The term *multimodality,* which has become an overarching term in the field of interaction studies and gesture studies, refers to the plurality of communication channels and modalities deployed in interaction. It is defined as "the various resources mobilized by participants for organizing their action – such as gesture, gaze, facial expressions, body postures, body movements, and also prosody, lexis and grammar" (Mondada, 2016, p. 337). Similarly, Stivers & Sidnell, (2005, p. 1) claimed that:

> Face-to-face interaction is, by definition, multimodal interaction in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours[49].

This multimodal and integrative approach is also applied to child language, which views the process of language acquisition as inherently multimodal, through the lens of a multimodal construction grammar (cf Morgenstern, 2014, 2019). In addition, Cienki (2017, 2015a) introduced the notion of *dynamic scope of relevant behaviors*, (SRB) which takes into account the kinds of symbolic status gestures can have. He suggested that in a communicative context, speakers may be invited to choose among a scope of behaviors (i.e. gesture, speech, gaze, etc.), by focusing on the *vocal* modality of discourse (which is the "default focus", Cienki, 2015, p. 628), or by including a set of behaviors (i.e. the combination of the vocal and visual-gestural features) in a specific context. Further in line with CG (cf section 3.1.2.) his claim is that repeated instances of gestures paired with certain functions are more likely to become entrenched linguistic signs. This theory thus further reflects the multimodality of discourse which includes a range of vocal and visual-gestural behaviors.

Stivers & Sidnell (2005), following Enfield (2005), further distinguished between the vocal-aural and the visual-spatial modalities of face-to-face multimodal

[49] However, it should be noted that linguistic anthropologists take on a slightly different approach which does not dwell on the distinction between different modalities, but rather insist on the "abstraction of the interacting body from the material world", see Streeck et al., (2011, p. 9).

communication, which will be further described below. This distinction does not only dwell on differences in modality, but also with respect to "semiotic ground" (cf Enfield, 2005, p. 52). In sum, the combination of different modalities and semiotic resources can further our understanding of the multiple processes underlying the organization of talk-in-interaction. As Stivers & Sidnell (2005, p. 2) argued:

> Much can be gained from examining a turn-at-talk for where it is situated vocally (e.g., sequentially, prosodically, syntactically) as well as visuospatially (e.g., body orientation, facial expression, accompanying gestures).

We shall now turn to a brief review of the two modalities of discourse, based on Enfield (2005) and Stivers & Sidnell (2005). The vocal modality of talk-in-interaction was investigated in the work of conversation analysts who focused on the organizations of shared practices in the course of social interaction (cf section 3.2.1.). It includes the lexico-syntactic channel (e.g. work on lexical items such as "okay"), as well as the prosodic channel (e.g. upward or downward intonation, prosodic contour, see Selting et al., 2010 for review). For instance, Ogden (2013; 2018) studied the uses of peripheral linguistic objects known as *non-verbal vocalizations*, such as tongue clicks. He combined phonetic, kinetic, and conversation-analytic methods to study speakers' stance-taking displays. With regard to the vocal modality of interaction, he looked specifically at the timing of clicks within the speech signal, and claimed that clicks could be used as temporal markers and thus function as "metronomes" (Ogden, 2013, p. 314). Such evidence can be provided by looking at a spectrogram and waveform window, illustrated in the following figure:

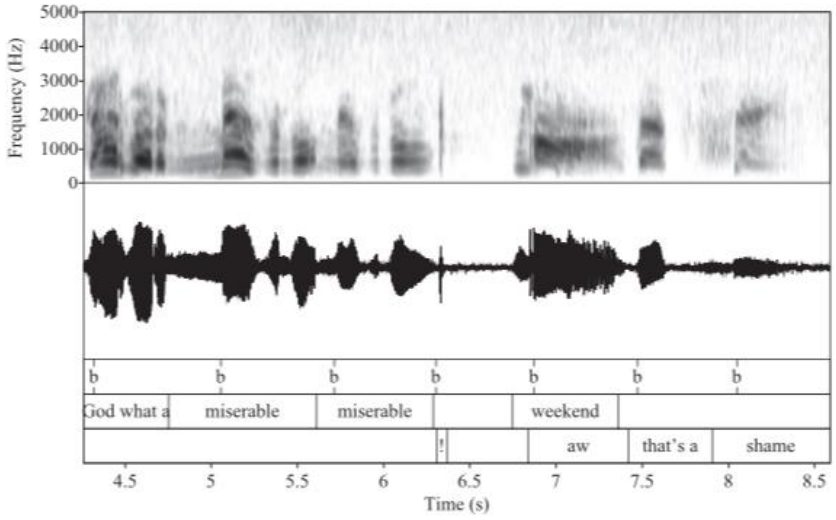


Figure 7. *Spectrogram and waveform window (Ogden, 2013, p. 315)*

In Ogden's (2013) example, the click is indicated with the *!* symbol within the transcription line below the waveform. As Ogden noted, the click is produced at the fourth beat (indicated with the *b* above the transcription line)[50], which is also synchronized with the two turns at talk, as it occurs near the end of A, the first speaker's turn completion, and prefaces B, the second speaker's turn initiation. As Ogden further noted:

> The temporal placement of B's click thus demonstrates an orientation to the rhythm established by A, and projects an upcoming, on-beat turn. B's turn is indeed aligned in time and action with A's (Ogden, 2013, pp. 315-316).

The co-construction of actions within an interactional sequence can thus be analyzed not only syntactically but prosodically with regard to several phonetic properties such as rhythm, voicing, loudness, voice quality etc. While these aspects are not the main focus of this dissertation, some of them will be explored to support our multimodal qualitative analyses. This is further elaborated in Chapter 2.

The visuo-spatial modality of discourse has been examined by a number of researchers who were interested in the study of hand gestures specifically (cf section 3.3.2.), but it also includes the study of other types of behavior, such as gaze, and body orientation within the spatial environment. For instance, Kendon (1976, 1990) and Ciolek & Kendon (1980) analyzed the way speakers oriented their bodies during social encounters, and defined the concept of *F-formation* (Kendon, 1976) which refers to a specific configuration of speakers in a jointly interactional space. As Mondada (2013) pointed out, the analysis of embodied interaction cannot solely focus on the speaker's body posture or orientation alone, but also with regard to the arrangement of other bodies within the spatial environment. She further suggested the term *interactional space* (Mondada, 2013, p. 248) which refers to the following:

> the situated, mutually adjusted changing arrangements of the participant's bodies within space, as they are made relevant by the activity they are engaged in, their mutual attention and their common focus on attention, the objects they manipulate and the way in which they coordinate in joint action.

[50] The beats were produced automatically using a specific algorithm (retrieved from http://cspeech.ucd.ie/fred/beatExtraction.php, last consulted on August 26th 2021).
See Ogden (2013, p. 314) for more details.
More information regarding the spectrogram and the waveform are given in Chapter 2.

Similarly, Sweetser & Sizemore (2008) distinguished between three types of space: (1) *the personal gesture space*; space in front of the speaker's trunk and head (McNeill, 1992); (2) *the inter-speaker space*; space between the two personal gesture space which can be reached by speakers to mark common ground; (3) *the unclaimed surrounding space*; an adjacent space away from the personal and interpersonal gesture space. When speakers are co-oriented towards each other around a central interpersonal space, they may rely on their body movement and gaze to display their engagement to one another or to the ongoing activity. This is illustrated in Figure 8, taken from Goodwin (1981, p.96) who studied the different "engagement displays" participants convey through their body orientation and gaze (also see section III. 3.2.1.)

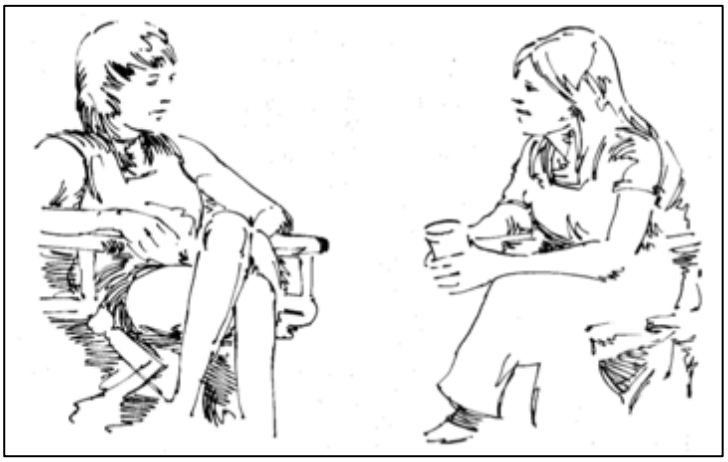


Figure 8. *Engagement display (Goodwin, 1981, p. 96)*

In a similar vein, another body of research (e.g., Debras, 2017; Debras & Cienki, 2012; Streeck, 2009) studied the uses of head tilts and shrugs in relation to postures of disengagement, as well as stance (i.e. the expression of a speaker's feelings, attitudes and judgement, see Kärkkäinen, 2006). Shrugs, which are defined as a "compound enactment" (Streeck, 2009, p. 189) involve a manifestation of the hands, combined with a movement of the shoulders, as well as a particular facial expression and a head movement, depicted below:

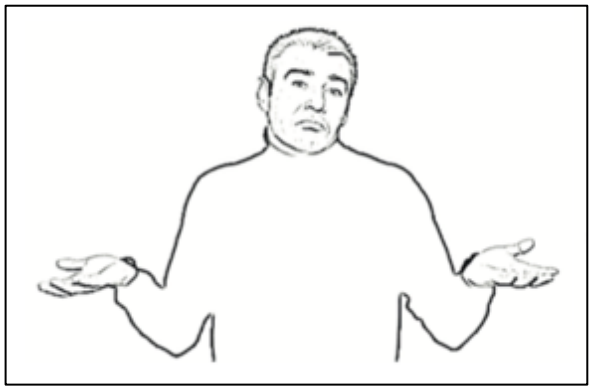


Figure 9. *Example of a prototypical shrug (Debras, 2017, p. 2)*

The illustration above shows an example of a "prototypical shrug" (Streeck, 2009, p. 189) which includes a combination of distinct features (i.e. an upward rotation of the forearm, one or two lifted shoulders, a head tilt, raised eyebrows, and a mouth shrug, see Debras, 2017, p. 2) but as Debras (2017) pointed out, not all instances of shrugging include all of these formal features illustrated above. For instance, a shrug can also be only performed with the face, or by lifting one's shoulder. In their study on face-to-face interactions between British university students, Debras & Cienki (2012) showed several examples of speakers who displayed mitigated affiliation via the combination of the vocal and visuo-spatial modalities: (1) they produced utterances such as "yeah but", accompanied by a lateral head tilt; (2) they remained silent while lowering their eyebrows and pulling their lips downwards (3) they acknowledged a "stance differential" (Du Bois, 2007) between another stance and their own positioning, with verbal expressions such as "oh yeah.. that makes sense", combined with shrugging. Conversely, when speakers displayed their orientation towards their interlocutor's utterance, they tended to gaze at them, lean forward, and tilt their head. In summary, the combination of gaze, head, and shoulder movement with or without speech can convey pragmatic and interpersonal functions related to various aspects of stance.

Another way to convey disengagement from a situated language activity is through the display of a distinct *thinking face* (Goodwin & Goodwin, 1986). This very iconic thinking face, which depicts an individual in "deep thinking" (i.e. eyebrow raised looking upwards or frowning) is rather stereotypical and highly recognizable across situations. Figure 10 illustrates three types of thinking faces which were displayed by three different speakers in different situations. The speakers' language and cultural backgrounds were different (two of them are French, and one of them is American), as

well as the type of activity they were engaged in[51]. However, the thinking faces displayed in the figure share similar features: (1) a frown, or a wince, (2) gaze looking upwards, or eyes closed (3) hands alongside face (in a and d), (4) thumb and index finger resting on the chin (in c). As we shall see (cf Chapter 2,3, and 4) a majority of the instances of thinking faces annotated in the data across corpora occurred during the production of verbal fluencemes.

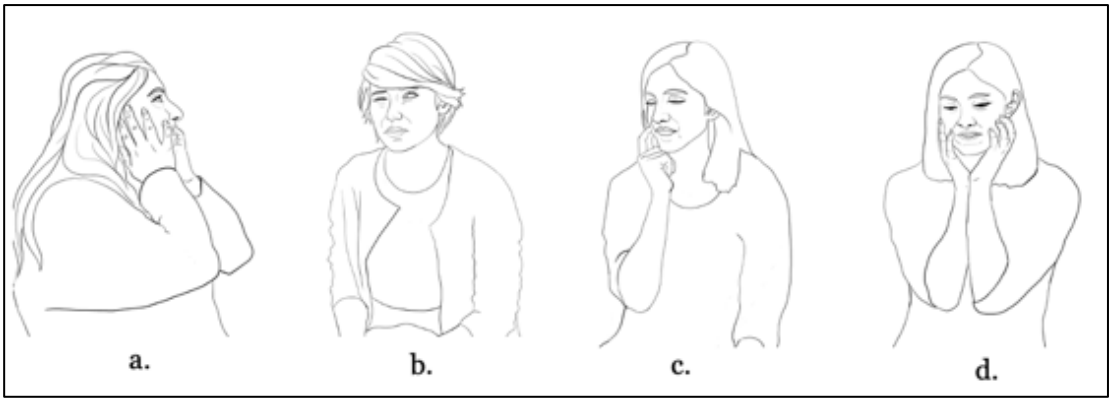


Figure 10. *Embodied displays of thinking face (SITAF and DisReg)*

In recordings of face-to-face conversations in natural settings, Goodwin & Goodwin (1986) explored the display of thinking faces and instances of gaze withdrawal during word searches. In their paper, they showed that the activity of searching for a word did not only involve vocal phenomena, but visual ones as well. They drew on several examples from their data to illustrate cases of gaze withdrawal and thinking face near "perturbations in the talk displaying initiation of a word search" (Goodwin & Goodwin, 1986, p. 57). Such instances of visual display during word search can provide relevant information to the hearers, who are hereby informed that a change in the current activity has occurred. A change in participation status can also take place (i.e. from recipient to active participant) when the hearer's coparticipation in the search is appropriate. This is illustrated in the following example:

[51] These examples are taken from the data under scrutiny in the present thesis (the SITAF Corpus and the DisReg Corpus, cf Chap. 2). More examples of thinking faces will be provided across Chapters 3, 4 and 5. Special thanks to Violette Kosmala for the illustrations.

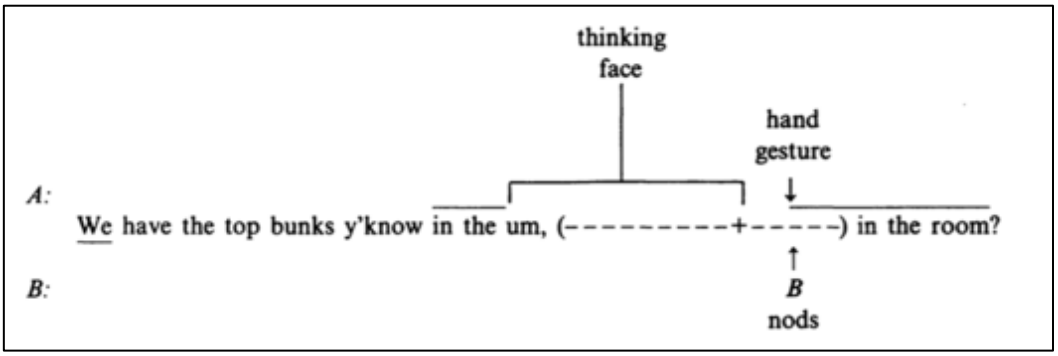


Figure 11. *Example of a multimodal word search (Goodwin & Goodwin, 1986: 71)*

In this example, Speaker A begins her word search and displays a thinking face, followed by a hand gesture oriented towards Speaker B to invite him to take part in the search. B's visual response, a nod, conveys his understanding towards what A is trying to say. Therefore, the hand gesture produced by Speaker A suggested a shift in the participant's status who solicited her interlocutor's coparticipation in the search.

To conclude, face-to-face interaction can be regarded as a joint and embodied manifestation of the vocal-aural and visual-spatial modalities of discourse, which continuously supplement one another through regulatory work. The present thesis mainly focuses on the visual-gestural actions deployed by gesturers, who make use of them respectively or conjointly to regulate interaction within their spatial environment. This allows them to regularly display their state of engagement or disengagement to the different activities they participate in. These central aspects of multimodal social interaction are highly relevant to the present study of embodied inter-(dis)fluency, as they serve as a basis for further development on the relationship between (dis)fluency and gesture (cf 3.3.4.).

3.3.2. The different approaches to gesture

As Harrison (2009, p. 27) pointed out in his thesis work, gesture research can broadly be distinguished between two dominant approaches, the (1) *cognitive-psychological* approach (how gestures are related to the expression of thought), and (2) *the functional-communicative approach*[52] (how gestures function and are used by speakers to structure speech acts in interaction). These two different approaches have led to different gesture classification systems, which is addressed in the following section. In this section, we present a brief overview of these two dominant traditions

[52] There are, of course, different existing approaches to gesture which will not be described in this thesis; see Beattie & Shovelton (1999), De Ruiter (2007) or Iverson & Thelen (1999) for review.

in the field of gesture studies, as well as their implications for the study of gestures and their relation to speech. We conclude with our choice of approach adopted for the present thesis.

The cognitive-psychological approach was first reflected in the influential work of McNeill (1985, 1992) who viewed gesture as "a window onto the mind" (also see Goldin-Meadow, 1999; Goldin-Meadow et al., 1993; Kita, 1993; Kita & Özyürek, 2003). In this view, gestures are said to share the same cognitive processes as speech, with the two being part of the same psychological structure. The combination of speech and gestures thus allows researchers to observe two simultaneous views of the same production process. He defines gestures as the following (McNeil, 1985, p. 351):

> These are movements that (with a class of exceptions to be described) occur only during speech, are synchronized with linguistic units, are parallel in semantic and pragmatic function to the synchronized linguistic units, perform text functions like speech, dissolve like speech in aphasia, and develop together with speech in children.

The idea that speech and gesture are connected "internally" (McNeill, 1985, p. 353) is supported by the evidence given above. In light of this approach, Butterworth & Hadar (1989) explored the production of gestures during lexical retrieval, and related it to the model of speech production, which involves several computational stages (e.g. message construction, selection of a lexical item in the mental lexicon, retrieval of a phonological word, instruction to articulators (Butterworth & Hadar, 1989, p. 172). They argued that, during lexical retrieval, speakers make use of iconic gestures (cf section 3.3.3.) to assist word finding by "exploiting another route to the phonological lexicon" (Butterworth & Hadar, 1989, p. 175). Additionally, Krauss (1998, p. 11) claimed that gestures could help speakers work on their production at different stages of speech processing. At the conceptualizing stage, they may help to formulate a concept that will be expressed in speech; at the stage of grammatical encoding, the information found in the gesture can help speakers to map the concept onto their mental lexicon; and at the phonological encoding stage, gestures can facilitate the retrieval of a word form. In this sense, gestures can serve a compensatory role in the speech production system. We will return to this question in Chapter 3 when we discuss the *Lexical Retrieval Hypothesis* (Krauss et al., 2000) and the potential role gestures play during L2 lexical access. In the same vein, Kita et al., (2017) supported

the idea that many gestures served self-oriented functions, facilitating conceptualization and activating spatio-motoric information for the purposes of thinking and speaking. Gestures can thus help express spatial information and take on different viewpoints within the spatial environment (Alibali, 2005; McNeill, 1992). Goldin-Meadow (1999) further suggested that gestures accompanying speech served two main functions: (1) they provide speakers with a representational format that can reduce cognitive effort in speech production and thus serve as a tool for thinking; (2) they provide hearers with this very representational format which gives them access to the speaker's unspoken thoughts.

In sum, the cognitive-psychological approach to gesture mainly revolves around the idea that the process of *thinking* can be made visible through gesture. Hand gestures thus have the potential to display thoughts that are not overtly expressed in speech, which can in turn facilitate the production process. We shall now turn to a different approach to gesture, which focuses on their practical use and significance within social interaction.

The functional-communicative approach to gesture is reflected in the pioneering work of Kendon (1980, 2004, 2014, 2017), Müller (1998, 2017) and Streeck (2009b, 2010, 2015), among others, who have documented their different forms and functions across languages and cultures within the study of social interaction.[53] Kendon's groundbreaking work on gesture, defined as the "utterance uses of visible bodily actions" (Kendon, 2004, pp. 1-2) regards it as an integral part of utterance construction. Utterances, which are not restricted to verbal productions, refer to the following (Kendon, 2004, p. 7):

> an "utterance" is any unit of activity that is treated by those co-present as a communicative "move", "turn" or "contribution". Such units of activity may be constructed from speech or from visible bodily action or from combination of these two modalities.

[53] However, it should be noted that Kendon, Müller and Streeck's approaches have a few distinctions of their own. Kendon's work essentially focuses on gestures as an integral part of utterance construction, further shaped by their context of use; but Müller argues that the context-of-use alone is not enough to explain the meaning of certain specific gesture forms. She suggests that gestures are motivated by cognitive-semiotic techniques and different gestural modes of representation (cf section 3.3.3.). Finally, Streeck's approach offers a view of gesture as "craft" (Streeck, 2009) building on sensemaking practices and practical actions of the hands.

In this view, the term "gesture" (not restricted to manual gestures) refers to the kind of visible bodily action that contributes to the construction of the utterance. Kendon's approach to the gesture-speech relationship is thus different from McNeill's in the sense that, while Kendon views gesture and speech as two different kinds of "expressive resources" available to speakers (Kendon, 2004, p. 111), McNeill views them as two separate channels of "observation of the psychological activities that take place during speech production" (McNeill, 1985, p. 350). Conversely, Kendon (2004, p. 111) does not believe that gesture and speech are part of the same production processes, but are rather "an integral part of what a speaker does in fashioning an object, the utterance, that is shaped to meet the expressive and communicative aims and requirements of a given interactional moment." In sum, the difference between McNeill and Kendon's approaches to gesture lies essentially in their contrasting views on language production: while McNeill considers language as the result of internal computations, Kendon regards it as a series of actions contributing to the utterance construction in interaction.

Kendon's approach to gesture thus reflects a view of language as inherently multimodal (cf section 3.3.1), and constituting a mode of action (i.e. speakers engage in different kinds of visible bodily actions that are integrated to the ongoing speaking activity). According to Kendon (2014), utterances are always constructed for others (i.e. whether for the speakers themselves, their interlocutor, or a virtual interlocutor). They are thus always produced with respect to bodily posture, head orientation, and gaze orientation. In the field of interaction research, which focuses on the interactional and social role served by bodily actions within talk-in-interaction, manual actions (i.e. hand gestures) have shown to be systematically deployed to manage turn-taking, for instance by displaying a participant's request for a turn (Mondada, 2007), or by citing the other participant's previous contribution (Bavelas et al., 1992). Streeck (2009a) further speaks of *speech-handling* gestures, which refers to a specific class of gestures which enact metapragmatic and communicative functions. We will return to this point in the following subsection.

The discourse-functional approach to gesture is thus further grounded in an "embodied, cultural sense-making praxis that draws on all the capacities of the human hand" (Streeck, 2009, p. 30). Manual actions can achieve specific semantic and pragmatic functions based on their different modes of gestural representation (Müller, 2014, 1998) mainly *acting* and *representing* (cf section III. 3.3.3.) These modes arise

from the hands' practical actions, i.e., their capacity to grasp or manipulate virtual objects in a virtual world. Their symbolic and pragmatic use in interaction, which is originally derived from manipulatory actions, is further shaped by the context in which they occur. For instance, a hand with fingers extended and palm facing upwards directed towards the interactional space can be seen as an action of offering or holding the hand out to receive a virtual object (Kendon, 2014); in one context this same gesture form can be seen as a turn-taking action designed to yield a turn, while in another it can be used to deliver a new piece of information in the speaker's own discourse. Kendon's (2004) work on the *POH (Palm Open Hand) gesture family*, or Müller (2017)'s study of *recurrent gestures (*i.e. conventionalized and culturally shared gestures) further explores the different pragmatic contribution of gestures in different contexts of interaction.

To conclude, while the two contrasting views addressed above show some similarities with regard to the significant role gesture plays in speech, they are chiefly rooted in theoretical differences[54], thus focusing on different types of gestures (e.g. Kendon focuses on pragmatic co-speech gestures, while McNeill focuses on iconic ones). As Morgenstern & Goldin-Meadow (in press) put it:

> Gesture theories vary with respect to their view of their relation between language and gesture, and this variability may go hand-in-hand with the type of gesture that is the focus of the theory.

While the psychological-cognitive view focuses on gestures working alongside speech to manage human thought processes, the functional-communicative view draws on speakers' abilities to build emerging utterances in different contexts of use through visible bodily action. Since the latter is more in line with some of the central topics addressed in interactional linguistics (cf section 3.2.) we will adopt a functional-communicative approach to gesture for the present study of inter-(dis)fluency[55], thus vouching for a *functional classification* of gesture (cf section 3.3.3. and Chapter 2); this leads us to a review of the different existing gesture classifications.

---

[54] However, this does not imply that the two approaches presented here are strictly incompatible. They can also be complementary, see for instance *Metaphor and Gesture*, a multi-disciplinary volume on the use of metaphor in gesture, edited by Alan Cienki and Cornelia Müller.
(for more information, visit https://benjamins.com/catalog/gs.3) last retrieved on August 26th 2021.
Also see Rohrer et al., (2020) who offered a multimodal labeling manual which aimed to bridge the gap between the approaches used by Kendon and McNeill.

[55] We further argue in favor of a functional-communicative approach in chapters 3 and 5.

3.3.3. Gesture classifications

In the 6th chapter of his book on visible bodily action entitled *classifying gestures*, Kendon (2004) reviewed some of the major classifications on gesture in the course of the 20th and 21st century. As we shall see, researchers very often classified gestures on the basis of form, function, or on their relation to speech. The present section is partly based on a selection of some of the major classifications reviewed by Kendon, but it will also include other schemes not present in his book. Kendon's chapter first introduced Austin (1753-1837)'s early classification which distinguished between "significant" and "non-significant" gestures. Significant gestures are said to convey the expression of substantive meaning (e.g. depiction of objects or action, expression of attitudes or feelings), while non-significant gestures mark certain discourse structures (cf Kendon, 2004, p. 90). This distinction, as Kendon noted, would later be used to distinguish between gestures that relate to the expression of speech and those that relate to the structure of discourse. Additionally, while some researchers opted for a classification that was mainly semiotic (cf William Wundt who divided gestures according to how their form was related to their meaning in Kendon, 2004, p. 91), others, such as Ekman & Friesen (1969) analyzed the different ways "nonverbal behavior", defined as "any movement or position of the face and/or the body" (Ekman & Friesen, 1969, p. 49) could be used in relation to speech. They offered five different categories which later became well known in the field of gesture studies, summarized as follows:

1) *Emblems*: highly conventionalized forms of manual gestures with a stable and shared semantic meaning which can be used as alternatives to speech (e.g. the thumbs up gesture, or waving goodbye)
2) *Illustrators*: manual gestures mainly used by speakers when speaking. This includes batons (used to emphasize a word or a phrase), and deictic gestures (used to point to an object).
3) *Affect displays*: displays of facial expressions and emotions
4) *Regulators*: gestures used to regulate speech
5) *Adaptors*: movements performed on the speaker's own body (self-adaptors), the body of others (alter-adaptors), or objects (object-adaptors).

As Kendon (2004) pointed out, Ekman & Friesen's classification, despite being highly influential, is difficult to apply into a gesture annotation scheme because it has not been established with a common set of criteria. Some categories, such as emblems, are

distinguished with regard to their social and conventional status, illustrators on the other hand, are distinguished on the basis of their relation to speech.

Another major well-known gesture classification is McNeill's which, as Kendon (2017) noted, was partly semiotic and semantic. His widely used categories include the following:

1) *Iconic gestures*: gestures which have a "formal relation to the semantic content of the linguistic unit" (McNeill, 1985, p. 354).
2) *Metaphoric gestures*: gestures that are related to abstract meanings.
3) *Beats*: "bi-phasic small low energy rapid flicks of the fingers or hand" (McNeill, 1992, p. 80) which serve punctual and discourse marking functions.
4) *Deictic gestures*: "pointing movements which are prototypically performed with the pointing finger" (McNeill, 1992; p. 80)

As Kendon (2004) noted, several researchers have distinguished between different kinds of manual gestures based on whether they relate to the content of discourse or to its structure. For instance, McNeill's *beats* and Efron & Ekman's *batons* categories, as well as Austin's *non-significant gestures* refer to gestures that are related to the structuring of discourse, as opposed to *significant gestures*, *iconic* and *metaphoric gestures* which refer to the propositional content of discourse. However, the categories offered in these classifications are not exactly consistent with this distinction, given the confusion over formal, functional and semantic criteria. This was pointed out by Müller (1998) who criticized these classification systems (in Cienki, 2005, p. 425). For instance, in McNeill's classification, beat gestures are defined both on the basis of their form (i.e. rapid flicks of the finger) and their function (i.e. used to mark emphasis). Furthermore, iconic gestures and metaphoric gestures are distinguished on the grounds of different semiotic and conceptual criteria, which can be conflating (see Sweetser 1998 who reviewed a variety of gestures which *metaphorically* make use of the gesture space to mark discourse structure). The idea of metaphoricity is thus not restricted to referential gestures only, but can be applied to pragmatic gestures as well (cf Streeck, 2008), so the distinction between "iconic" (concrete) and "metaphoric" (abstract) within the same referential gesture category is not entirely valid. Therefore, Müller (1998) vouched for a *functional classification* of gestures, which was later adapted by Cienki (2004, 2005)[56]. The different categories of gestures were

[56] This gesture classification model can be found in Appendix 1.

determined based on their function in different speech situations. It distinguished between *referential gestures* (with concrete or abstract reference) and other types of *pragmatic gestures* which includes (1) *performative gestures* (gestures with an expressive function e.g hand clapping, or an appeal function such as requesting, or dismissing) and (2) *discursive gestures* (used for emphasis and discourse structuring).

In sum, several authors, such as Müller (1998), Kendon (1995, 2004, 2017) and others (Bavelas et al. 1995; Cienki, 2005; Lopez-Ozieblo, 2020 with differences in terminology) distinguished between two main classes of gestures: *referential gestures* and *pragmatic gestures*. This distinction can be made on the basis of which aspect of the communicative situation they are about: whether they pertain to a specific thematic object in the content of discourse; or whether they relate to the interaction itself i.e., the expression of stance, an illocutionary force (the speaker's intention when producing the utterance), or the structuring of discourse.

According to Kendon (2004, 2017) referential gestures contribute to the propositional or referential meaning of utterances in two ways, (1) through pointing[57], and (2) by performing actions which make a physical object or an action visible. The latter is the equivalent of the iconic and metaphoric gesture described by McNeill. In the same vein, Müller (1998,204) spoke of *representational* gestures which had two main modes of representation (acting and representing, cf section 3.3.2), and Streeck (2008a) used the term *depictive gestures*. Bavelas et al., (1992, 1995) further distinguished between *topic gestures* and *interactional gestures,* which is another term for "pragmatic gesture". Topic gestures were defined as the depiction of "some aspect of the topical content of the conversation, such as the size of an object or (metaphorically) of a problem" (Bavelas et al., 1995, p. 397). Referential gestures can further fall into two different categories (see Cienki, 2004,2005 based on Müller, 1998), those with concrete reference (objects, properties, actions, location) and those with abstract reference (entities, properties, events).

As Kendon (2017) pointed out, McNeill's classification mainly revolves around gestures that are used to represent objects or spatial relations, but lacks in the description of gestures' *pragmatic contribution to discourse* (except for beats). Pragmatic gestures, also known as *interactive gestures* (Bavelas et al., 1992, 1995), *speech-handling gestures* (Streeck, 2009), *recurrent gestures* (Müller 2017) or

[57] Deictic gestures can also be categorized in a different category, see for example Navretta, (2001) and Gullberg's (1995) category of deictic-anaphoric gestures. This will be further developed in Chapter 2.

performative gestures (Cienki, 2004) draw attention to the pragmatic role of gestures in discourse and interaction, and how they may structure speech acts, indicate the relationship between different discourse segments, or "refer to some aspect of conversing with another person" (Bavelas et al., 1992, p. 473). Streeck (2009) regarded *speech-handling gestures* as a variety of open-handed unilateral or bilateral gestures used by the speaker as manipulative actions to manipulate virtual objects (e.g. offer or receive an object of discourse and organize relationships between them). These pragmatic, *recurrent gestures* (Müller, 2017) comprise a number of highly common and recurrent gesture forms such as the palm-up-open hand, the palm-away-open-hand, the cyclic gesture (Fig. 10), the finger bunch (Fig. 11), or the shrug (cf section 3.3.1.) (also see Kendon, 1995) some of which are illustrated below.

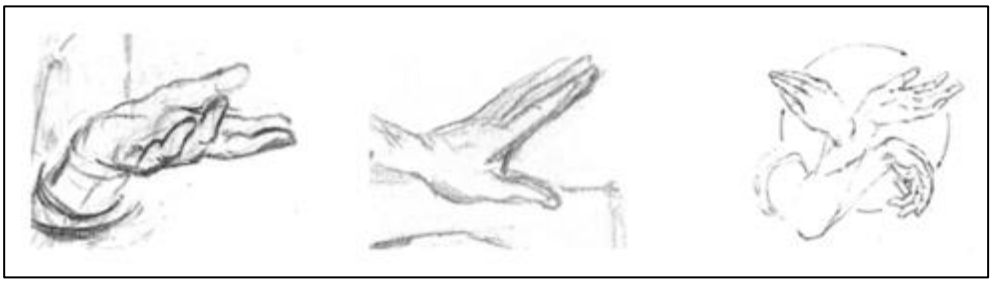


Figure 12. *Example of recurrent gestures (Müller, 2017, p. 3)*

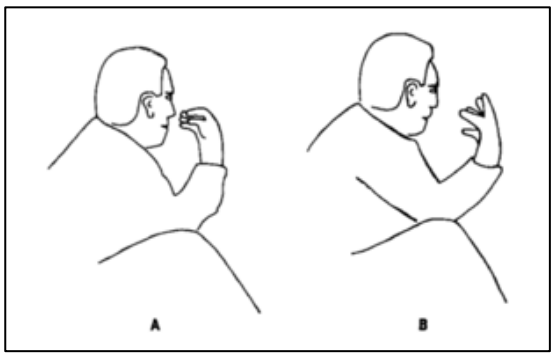


Figure 13. *Example of a Finger Bunch gesture (A) (Kendon, 1995, p. 265)*

The study of this specific class of gestures is thus grounded in "embodied motivations and embedded within a complex dynamic network of variable usage contexts" (Müller, 2017, p. 1). In other words, the same gesture forms, which have different degrees of conventionalization, form-meaning relation, and idiosyncrasy (cf Müller, 2017) are shaped by the different communicative practices of the speakers' everyday actions across cultures and languages (e.g. Italian, Kendon, (1995); German, Müller (2017); Ilocano Streeck & Hartge, (1992). In addition, pragmatic gestures can further be

distinguished between different subcategories, suggested by Kendon (2017, pp. 170-172):

- *Operational*: gestures that operate in relation to what is being expressed verbally (e.g. use of a headshake to express negation).
- *Modal*: gestures that give an interpretative frame for what the speaker is expressing (e.g. use of quotative mark gesture or finger bunch gesture).
- *Performative*: gestures that express the illocutionary force of an utterance (e.g. palm up used to give an example of something).
- *Parsing or punctual*: gestures used to make distinct segments of discourse, marking emphasis or contrast (e.g. beat gestures, or precision grip gesture to emphasize a stretch of speech).
- *Interactional regulation*: gestures used for waving, greeting, requesting, inviting someone to do something etc.

The terms used in these categories are also overlapping in other classifications (cf *performative gestures* in Cienki, 2004, 2005). For example, Bavelas et al.'s category of *interactive gestures* (1992, 1995) does not only include gestures used for interactional regulation, but all the gestures that are used to help maintain conversation as a social system. This includes four different aspects, such as (1) citing the interlocutor's previous contribution, (2) seeking agreement, understanding, help, (3) delivery of new versus shared information, (4) managing events around the turn. More recently, Lopez-Ozieblo (2020) offered a revised functional classification of pragmatic gestures, adapted from Kendon (2017), and further in line with a functional linguistic-based framework and a functional classification of discourse markers. It includes the following categories: (1) *cognitive gestures* (equivalent to modal gestures), (2) *metadiscursive* (equivalent to parsing), and (3) *interactive* (which includes performative, operational, and interactional regulators).

In sum, there are many different existing and overlapping gesture classifications which were introduced in the field of gesture studies, and these categories can be distinguished using a different set of criteria (semiotic, semantic, functional, conceptual etc.). In this section, we mostly focused on two dominant trends in the field of gesture studies originally introduced by McNeill and Kendon, but as we have seen, gestures can be analyzed in different ways, on the basis of their form,

their function, their iconicity, their relation to speech, to the material world (Streeck, 2009), or to the affordances of the context. Indeed, another body of research largely focused on form-based approaches to gesture, with for instance the ToGoG group (*Towards a Grammar of Gesture*, cf Müller et al., (2013) and Boutet's kiniesological approach (Boutet, 2008, 2010, 2018). Müller (2014) suggested that gestures operated on a limited set of gestural modes of presentation, mainly *acting* and *representing*. The representing mode focuses on the way speakers use hands to depict actions, objects, properties, and temporal relations; and the mode of acting, which includes specific forms of manual actions such as molding and drawing can be used to re-enact everyday actions, mold the shape of an object, or outline an object in the air (cf Chap. 5). These different gestural modes of presentation, as Müller (2014) suggests, may lead to the creation of *referential/iconic gestures* described earlier, but while Kendon (2004) focused on the interactive and context-of-use account of such gestures, Müller (2014) drew on their formational features, which can play a central role with regards to their expressivity and function. She added (Müller, 2004, p. 134):

> The articulatory effort to form a given hand shape to move the hands in a cyclic, rectangular, or straight way, to place it in the center, at the side of the gesture space, or move it towards an interlocutor are manifestations of deliberate expressivity.

In this view, Müller and colleagues started with the assumption that the articulation of hand shapes, movements, positions, and orientations of bodily behavior played a central role in the formation of "potentially meaningful units of body motion" (Müller, 2014: 138). In a similar vein, Boutet (2018) focused on the physiological role of the body and its dynamics, which are altogether grounded in constraints and potential articulation of people's body movements: "the meanings that emerge out of our gestural productions are the product of the natural articulation of our body" (Morgenstern & Boutet, forth., p. 6). Boutet's (2008, 2010) kinesiological approach, which is based on the movements of the body from a biomechanical view, adopts an exclusively form-based approach and describes gestural units on the basis of their formal characteristics, movement, and flow. The body's flow, which links one segment to the next, can spread from the shoulder to the fingertips, known as "proximal-distal", or from the hands to the shoulder, known as "distal-proximal" (cf Morgenstern, 2020

for review). This notion of flow is of particular interest to the present thesis as it will be related to other types of flow at the level of speech and interaction (cf. section IV).

Table 3. Overview of gesture terms

Gestures that relate to the lexical content of discourse					
Term	Referential **iconic** and **metaphoric** gestures	Referential **pointing gestures**	**Topic gestures**	**Depictive gestures**	**Representational gestures** (with **abstract** and **concrete** reference)
Author	McNeill (1985)	Kendon (2004)	Bavelas et al. (1992,1995)	Streeck (2008)	Müller (1998; 2004)

Gestures that relate to discourse and the interaction itself						
Term	**Beats** or **batons (illustrators)**	**Pragmatic gestures**	**Interactive gestures**	**Performative, discursive** gestures	**Speech-handling** gestures	**Recurrent gestures**
Author	McNeill (1985), Ekman & Friesen (1969)	Kendon (2004)	Bavelas et al., (1992) Lopez-Ozibelo (2020)	Cienki (2004, adapted from Müller 1998)	Streeck (2009)	Müller (2017)

Other gesture classifications, as we have seen, such as McNeill (1985)'s and Ekman & Friesen (1969) were criticized for mixing functional and formal criteria. Kendon (2017) and Cienki (2005; based on Müller (1998) thus offered a functional classification of gestures consistent with functional criteria, which can then be applied to larger annotation schemes (cf Graziano & Gullberg, 2018). The present thesis opts for a similar function-based approach, in line with the functional-communicative tradition. However, it does not disregard the form-based approaches introduced by Müller et al. (2013) and Boutet (2018) either, as we believe that the formational and dynamic aspects of gestural units play a central role in the building of their *deliberate expressiveness* (Kendon, 2004, pp. 13-14) altogether shaped by their context of use. Our identification and classification of gestures, which is partly based on Kendon (2004), Cienki (2004, 2005) and Bavelas et al.(1992,1995) will thus rely on both formal and functional characteristics; but it will adopt a consistent function-based labeling system, avoiding terms such as "beats" or "metaphoric" for the reasons outlined above. This will be further developed in Chapter 2.

With this in view, gestures can thus fall into two main categories: (1) those that relate to the content of discourse, and depict "actual, imaginary, and abstract worlds"

(Streeck, 2010, p. 27); and (2) gestures that relate to the interaction itself, and embody communicative acts and regulate interaction[58]. A summary of the overlapping terms is found in Table 3 above. We will discuss our choice of terminology in Chapter 2.

3.3.4. (Dis)fluency and gesture

In the previous sections, we emphasized the need to view social interaction as inherently multimodal (3.2.1) further situating this study within a functional-communicative approach to gesture (3.3.2.), and adopting a functional classification system (3.3.3.) We shall now conclude this section on gesture and multimodality by relating it to the present study of embodied inter-(dis)fluency. While gestures are said to exclusively occur during speech (cf McNeill, 1985) and thus very rarely during disfluencies (observed by Akhavan et al., 2016; Christenfeld et al., 1991; Esposito et al., 2001; Graziano & Gullberg, 2013; Yasinnik et al., 2005), their relationship is still of interest for two reasons. First, it can shed light on the temporal coordination between gesture and speech production, co-orchestrated and deployed together to lead to the building of utterances; second, it further supports the view of (dis)fluency as an *embodied* phenomenon, made of a vocal and *gestural flow*, altogether shaped by the contingencies of social interaction.

Before reviewing some of the main studies that investigated the relationship between gesture and (dis)fluency[59] specifically, we shall first briefly introduce the concept of *gesture phrasing* and the *phases of gestural action*, offered by Kendon (2004, p. 111)[60]. When a speaker engages in a gesturing activity, he or she will undertake a series of phases; first a phase of *preparation*, which leads to the core "expression" of the gesture, known as the *stroke*; which is then followed by a phase of relaxation, known as *recovery*. The stroke can also be followed by a phase in which the gestural movement is sustained, known as the *post-stroke hold* (Kita, 2003). Seyfeddinipur (2006), and Seyfeddinipur & Kita (2001) later used this model to investigate the coordination of disfluencies and gestures on a corpus of German semi spontaneous speech. They investigated the production of overt and covert repairs

[58] Note that Streeck (2010; 2015) offered a more complex classification based on the ecologies of gesture i.e., their relation to their communicative intention within the environment (see Streeck, 2015, p. 426 for review).
[59] It should be noted that all the studies described below used the term "disfluency" in their paper.
[60] This is also found in McNeill (1992, p. 83) who spoke of "three phases of gesticulation", mainly "preparation", "stroke" and "retraction".

based on Levelt's (1983) *Repair Model* (cf section I.1.1.) in relation to the different phases of gesture (Seyfeddinipur, 2006, pp. 107-109), listed and ilustrated as follows:

- *Preparation*; a movement of the hands to a location where a stroke is deployed.
- *Hold*; when hands are in a static position, other than the rest position.
- *Stroke*; a phase which displays the core meaning of the gesture.
- *Retraction;* when hands move back into rest position (on the lap, arm rests, arms folded in front of the chest).
- *Interrupted preparation/stroke;* when a phase was abruptly ended.

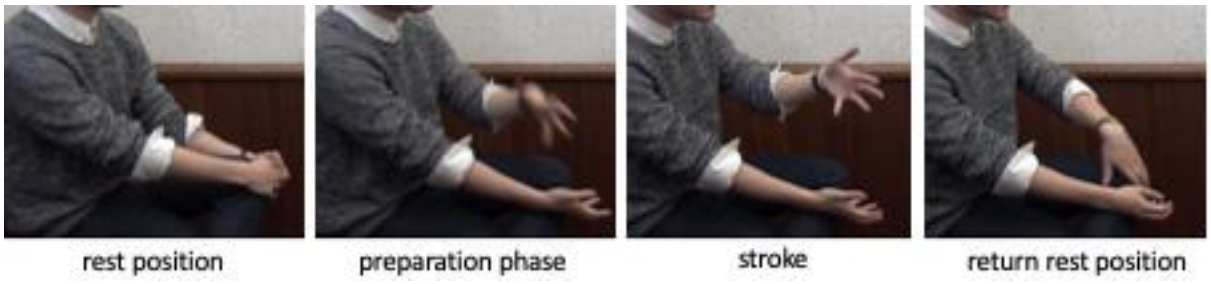


Figure 14. *Phases of gestural movement (SITAF Corpus)*

Their results showed that many gestures tended to be suspended prior to the production of disfluencies: out of 432 speech suspensions, 306 were accompanied by gestures. Seyfeddinipur (2006) gave the example of a speaker who executed a deictic gesture, interrupted it midway, and returned to the starting position at the same time as he produced the repair. She also illustrated cases of gestural suspension (i.e. hands dropping back into rest position) temporally coordinated with a vocal speech suspension. These findings gave support to their *Delayed-Interruption-For-Planning* Hypothesis (adapted from *the Main Interruption Rule*, cf section 1.1.) a process whereby a speaker, as soon as he or she encounters an error in speech, will not interrupt his speech right away, but will start replanning first (through delaying). This hypothesis further accounted for the *Interruption-Upon-Detection Hypothesis* which predicts that when a speech error is detected, a "stop signal" (Seyfedinnipur & Kita, 2001, p. 30) is sent to both modalities simultaneously. The fact that gestures were suspended prior to speech suspension further suggests that they could be seen as early indicators of upcoming interruption, which supports the view of speech and gesture as being part of the same planning process. In fact, in an earlier study conducted on hesitation and gesture, Butterworth & Beattie (1978) claimed that gestures could be seen as the product of *lexical preplanning processes*, indicating the onset of a lexical item currently unavailable in speech.

Similar results were reported in other studies. Esposito & Marinaro (2007) conducted a study on pauses and "gesture pauses" (i.e. holds) among adult and child speakers during an elicitation experiment, and showed a high frequency of overlaps between holds and speech pauses in both groups. This was also found in Yasinnik et al. (2005) who observed a high number of gestures which were temporarily held during disfluent speech in recordings of academic lectures. Similarly, Chui (2005) investigated the coordination of speech with the different phases of gesturing, and found that several gesture onsets occurred during *disfluent stretches of speech* (i.e. a self-repair accompanied by a hesitation pause). However, very few gestures were produced during speech, but the ones which occurred during *disfluent speech* were related to lexical problems or problems of planning. This last finding is also reflected in a different body of research who examined the role of gestures during lexical retrieval (e.g. Krauss & Hadar, 1999) in native and non-native speech, following the assumptions that gestures could help speakers compensate for their speech difficulties. We will address this issue more specifically in Chapter 3 when we present our study of inter-(dis)fluency in native and non-native speech. In the same vein, Graziano & Gullberg (2013, 2018) investigated the temporal coordination of disfluencies and gesture, based on a corpus of retellings done by different groups of speakers (competent L1 speakers, adult and child L2 learners). They looked at the distribution of disfluencies in relation to the gestural phases, but they also coded the functions of gestures, i.e. *referential* and *pragmatic*, following Kendon (2004). Their findings yielded similar results as the ones reported above, mainly that gestures occurred significantly more during fluent stretches of speech and that gestures tended to be held during disfluent speech. However, their results also indicated that speakers (of all groups) produced a majority of pragmatic gestures during disfluencies, suggesting that the latter did not necessarily occur when speakers were trying to compensate for their expressive difficulties through referential gestures, but can also occur when speakers comment on their own utterance. Similarly, Akhavan et al. (2016) provided further evidence that the main role of gestures was *not* to remedy speech problems, with again the striking observation that a majority of gestures occurred when speech was fluent. They also found that several disfluencies were accompanied by iconic gestures related to the lexical-semantic system, but a number of disfluencies was also found to occur with beat gestures. The latter, the authors suggested, may have a more communicative role in communication.

Taken together, the findings summarized above provide evidence that processes of speech suspension can to some extent be synchronized with gesture suspension, with the observation that very few gestures accompany fluent speech, and a great deal of them tend to be suspended during disfluent speech. This was found to be true across speakers, languages, age groups, and task types, which further gives support to the notion that the visual-gestural and the verbal/vocal modalities reflect a unified planning process. In the same vein, the present study of inter-(dis)fluency will also investigate the temporal coordination of (dis)fluency and gesture production across languages, speech genres, and communicative tasks (see Chapter 5).

However, while the results found in the studies reviewed above acknowledge the importance of disfluency phenomena in relation to gesture, we believe that it still presents a number of limitations. First, these studies focus mostly on the lexical and planning processes associated with disfluency and gesture production, without much taking into account their potential pragmatic role in the interaction (except for Graziano & Gullberg, 2018 and Akhavan et al., 2016 who observed a co-occurrence between disfluencies and pragmatic gestures, but did not quite investigate it further). Their view of disfluency and gesture thus exclusively relies on the speaker's perspective and his or her own planning processes, without much taking into account the contribution of his or her interlocutor within the interaction (cf section. 3.2.). Conversely, Stam & Tellier (2017) and Tellier et al. (2013) investigated the functions of gestures during pauses in native and non-native interactions, and hypothesized that gestures produced during pauses could also be partner-oriented, and thus be used as a teaching strategy for the learners. They distinguished between gestures that were *production oriented* (e.g. word searching), *comprehension oriented* (e.g. introducing a concept, marking the word), and *interaction-oriented* (eliciting an answer, helping the interlocutor). Their results showed that, unlike previous work, a considerable number of gestures co-occurred with pauses, especially in non-native speech. But this result was not interpreted as a sign that gestures were used to reflect speakers' lexical difficulties, but were rather treated as strategies used by speakers to facilitate comprehension. Similarly, *fluencemes* can also positively contribute to the co-construction of meaning, for example through *embodied completions*. This practice can be defined as the completion of an action, previously initiated, through *gesture or embodied display* (Mori & Hayashi, 2006). In a study of interactions between non-native learners of Japanese, Mori & Hayashi (2006) demonstrated the way native and

non-native speakers coordinate their talk through gesture and embodied completions in the context of L2 use. Although there is no overt mention of fluencemes in their paper, they are nonetheless present, specifically during *embodied word searching sequences* (Rydell, 2019). As Rydell (2019) argued, searching for a word is not only an internal process resulting from language difficulties, it is also an embodied visible activity (cf Goodwin & Goodwin, 1986) which can be collaboratively negotiated by two or more speakers through the mobilization of gaze and gesture.

Therefore, a second major limitation found in the studies addressed above is the fact that they have often been too restricted to a view of "disfluency" phenomena as speech error, affecting the speech (and hence gesture) production apparatus, while we have shown them to be a dynamic, functionally, and interactionally ambivalent system (cf 2.2.3., 3.1., 3.2.), further shaped by the social and interactional contingencies of talk-in-interaction. We thus believe in a view of inter-(dis)fluency as not only contingent upon speech errors or difficulties, but *embodied* and situated within a relevant set of language practices involving multiple bodily actions within the course of interaction.

To conclude, the present study of embodied inter-(dis)fluency draws on the different theoretical and methodological perspectives adopted by previous work on gesture and speech in different research fields (i.e. psycholinguistics, cognitive linguistics, interactional linguistics, linguistics anthropology etc.), which reflects our interdisciplinary and integrated framework (addressed in the following section), as well as our mixed-method approach (discussed in Chapter 2). We first build on the work of the different researchers reviewed in this section, who examined the synchronicity between speech suspension and gestural suspension based on the analysis of gestural phrases and gesture functional types. In addition, we expand on these findings by closely examining the emergence of embodied fluencemes in detailed qualitative micro-analyses of multimodal and embodied interactions. These qualitative analyses will further shed light on the different multimodal interactional practices embodied by fluencemes, by taking into account their position with regard to the *participation framework*, the *turn-at-talk*, and their relation to principles such as *progressivity or preference,* which further situates our study within a conversation-analytic and interactional approach.

IV. Towards an integrated framework of inter-(dis)fluency

In this chapter, we have sought to provide a complex and reliable picture of fluency and disfluency phenomena, by reviewing a number of different theoretical and methodological frameworks, which will be summarized below. Fluency and disfluency have shown to be highly complex constructs, revolving around multiple processes at the same time; these processes combine the verbal and vocal channel, the visuo-gestural channel, and the different practices shaping social interaction. This motivated our choice of the term *inter-(dis)fluency*[61], which provides a unified view of fluency and disfluency. The aim of this chapter was to introduce different theoretical backgrounds in order to stress the need to situate inter-(dis)fluency in a larger integrated framework, and thus bridge the gap between production-based psycholinguistic studies conducted on disfluency and usage-based, interactional, multimodal approaches to social interaction. This also invites us to consider (dis)fluency from multiple dimensions, based on Lickley (2015), Segalowitz (2016), Götz (2013), Candea (2000, 2017) and Grosman (2018). The present section is structured as follows: first we summarize the main approaches adopted in this thesis, and explain how they can be combined and integrated to our study of inter-(dis)fluency (4.1); then, we provide a definition of inter-(dis)fluency by taking into account its different dimensions (4.2). Lastly, we address our main theoretical assumptions.

4.1. Summary of the approaches adopted in this thesis

In this chapter, we reviewed multiple theoretical frameworks which all had different (but sometimes interrelated) perspectives on (dis)fluency phenomena. We started with **psycholinguistics,** which was one of the first major field of research in the late 1950s to systematically investigate (dis)fluency, based on the earlier work conducted by clinical linguists on stuttering and verbal "dysfluency". Their approach to (dis)fluency was thus closely related to **speech production models** (e.g. Levelt, 1983, 1989) which investigated the different stages of speech production, from conceptualization of the main message to its articulation. This line of research

[61] Our definition of inter-(dis)fluency is developed in section 4.2.

provided several analyses of the *disfluency regions* and listed the major *disfluency types*, which altogether served as a basis for later work on (dis)fluency in other research fields such as **corpus-based linguistics** and **second language acquisition**. Psycholinguistics thus introduced the fundamental premises of *disfluency* as a **legitimate topic of research**, which had traditionally been excluded from "traditional" formal and generativist approaches to language.

However, as we have seen, there are a number of issues underlying the term "disfluency", as the latter presupposes a problem, or a deviation from **ideally fluent speech**. The term **fluency,** on the other hand**,** is traditionally found in second language acquisition research, and refers to aspects of L2 speech performance based on several temporal variables, which includes different "disfluency markers" such as pauses, repairs, etc. The notions of fluency versus disfluency have thus been consistently distinguished from one another, despite the **constant overlap of terms** found across theoretical research fields. Some researchers argued that disfluencies should be called **markers of fluency instead of disfluency**, as they have been shown to serve many positive structuring functions in discourse other than just signaling an interruption in the speech signal. This constant opposition between fluency and disfluency is reflected in two contrasting views of these phenomena, one that regards them as a **cognitive burden**, and another one which considers them as a **communicative signal**. It has come to our understanding that the real conflicting issue regarding the notions of fluency versus disfluency is not only terminological, but theoretical as well. The fact that these phenomena have been systematically analyzed from a strictly verbal and formal perspective (except for a few notable exceptions, e.g. Tottie, 2014; Allwood et al., 2015; McCarthy, 2009, Clark & Fox Tree, 2002 etc.) has hindered their evaluation on the basis of discourse, interaction, or even gesture. This led us to a review of other theoretical fields which took on a new perspective to these phenomena, which are altogether relevant to our study.

First, the framework of **cognitive grammar and usage-based linguistics** accounted for a dynamic approach to these phenomena which considers them as fluid categories whose degree of symbolic meaning, conventionalization, and entrenchment are shaped by repeated instances of specific patterns in different contexts of use. In light of this approach emerged different cognitive and usage-based frameworks on (dis)fluency (e.g. Crible et al. 2019; Grosman, 2018; Götz, 2013; Segalowitz, 2016) which serve as a basis for the present definition of (dis)fluency (cf section 4.2)

Secondly, the framework of **interactional linguistics** provided essential conversation-analytic tools to the study of inter-(dis)fluency based on their **sequential development within talk-in-interaction**, which considers their position within conversational turns, their relation to **stance, intersubjectivity**, or speakers' positioning in a *participation framework*. Interactionist approaches also rely on **data-driven methods** such as the *single case analysis*, which focuses on the analysis of a single episode with respect to a specific aspect, or a *collection study,* which generalizes the results of a cumulative series of single case analyses with regards to a relevant aspect (cf Mazeland, 2006). These types of analyses are rarely addressed in usage-based frameworks which rather focus on isolated utterances, or quantitative findings alone (cf Cienki, 2016). Thirdly, the field of **gesture studies and multimodal interaction**, which is further grounded in an interactional framework, accounted for a view of (dis)fluency as an **embodied and multimodal** phenomenon, tightly related to the deployment of visible bodily actions (e.g. manual gestures, gaze direction, body movement, facial expressions etc.) which altogether form an integral part of the utterance construction. In this respect, we view language as a mode of action, reflecting the notion of **"languaging"** (Linell, 2009; Kendon, 2014; Morgenstern, 2020). The latter being defined as "linguistic actions and activities in actual communication" (Linell, 2009, p. 274). This term was later mainly adopted in the work of Morgenstern (2020) and Morgenstern & Boutet (forth.) and is deeply rooted in multimodal analyses of motivated and conventionalized language forms (including sounds, words, tones, gestures). Language is thus inherently embodied within its ecological and multimodal environment, which is comprised of our material, interactional, and cultural space. Additionally, the field of gesture studies provided relevant **gesture classifications,** grounded in a functional-communicative approach, which can be applied to the study of embodied inter-(dis)fluency through the analysis of its temporal coordination with gesture.

To conclude, the present study offers an **integrated theoretical construction of inter-(dis)fluency**, combining multiple approaches and perspectives, thus going beyond the first traditional psycholinguistic approach to disfluency. However, even though the present study is not exactly grounded in a psycholinguistic approach to (dis)fluency, it still borrows from its **formal classification** (i.e. the fluenceme classification, in line with Crible et al., 2019) for the purposes of corpus annotation. This has not been extensively done in other fields such

as interactional linguistics which only focused on a selection of markers in specific interactional sequences. The integration of these different theoretical frameworks thus further reflects our mixed-method approach which relies on quantitative and qualitative methods (cf Chapter 2). In sum, the combination of the usage-based, interactional, and multimodal frameworks allows us to bridge the gap between large corpus-based quantitative studies and data-driven single case analyses or collection studies. While quantitative methods give a robust and statistically valid overview of the data, they fail to illuminate particular instances in a specific interactional sequence, whose complex information can never be truly conveyed in quantitative findings. On the other hand, single case analyses, although truly illuminating of all the ongoing relevant interactional and social processes shaping the course of the talk, only rely on a small selection of instances, thus disregarding all other instances of the same phenomena in the whole dataset. This is highly relevant to the present study of fluencemes which are highly frequent in speech, and which consequently do not systematically exhibit essential features of talk-in-interaction. Reciprocally, their use is not systematically restricted to contexts of speech error. This further emphasizes their dynamic and fluid nature, whose degree of fluency, understood here as communicativeness, flow, or stream, depends on a number of contextual features. We shall now turn to our definition of inter-(dis)fluency, which, as we shall see, considers different dimensions of fluency.

4.2. Our definition of inter-(dis)fluency

Our understanding of (dis)fluency follows the innovative and valuable contributions of several authors in fluency and disfluency research, such as McCarthy (2009), Allwood et al., (2015), Götz, (2003), Segalowitz (2016), Candea (2000, 2017), Crible et al., (2019), and Clark (2003) (among others) who integrated essential aspects of face-to-face interaction, e.g. gesture, to the study of (dis)fluency, and who considered its different dimensions (e.g. utterance, cognitive, interpersonal etc.). Moreover, our understanding of (dis)fluency is also shaped by the different theoretical frameworks addressed in this chapter which altogether provided a rich picture of these phenomena. The present study thus vouches for a definition of inter-(dis)fluency as inherently **multimodal, multidimensional, and multilevel**. We believe that it is essential to consider (dis)fluency by taking into account different levels of analysis, considering the verbal, interactional, and visual-gestural channels of communication,

which motivated us to adopt terms such as *inter-(dis)fluency* and *inter-fluency*. The notion of *fluency* is thus broadly understood in general terms such as "communicativeness" "smoothness" "fluidity" and "flow" which can all be applied respectively to: (1) speech production (i.e. **flow of speech**), (2) interaction (i.e. **fluidity and progressivity of the exchanges**), and (3) gestures (i.e. **gestural and body flow**). The *dis* in brackets is equally important, as it symbolizes the **potential** interruption or discontinuity of all these flows, e.g. the cut off of a speech segment, a communication breakdown, or the interruption of a gestural movement. This interruption further reflects the potential for the same forms to be either fluent and/or disfluent, self-oriented or other-directed, lexical or non-lexical, thus placing (dis)fluency phenomena in a **multilevel continuum** which considers a series of markers which can potentially gain symbolic status (in line with Cienki, 2015b). Lastly, *inter* views **this multidimensional flow** not as one way, but a *two-way-flow* influencing one another, in line with McCarthy's notion of *confluence*. We preferred to use the prefix "inter" here, as it also draws a parallel to notions of *intersubjectivity*, *interpersonal relations*, and *interaction*. Our tripartite model of inter-(dis)fluency, whose dimensions are partly based on Grosman (2008), Segalowitz (2016), and Götz, (2003) is illustrated in the following figure:

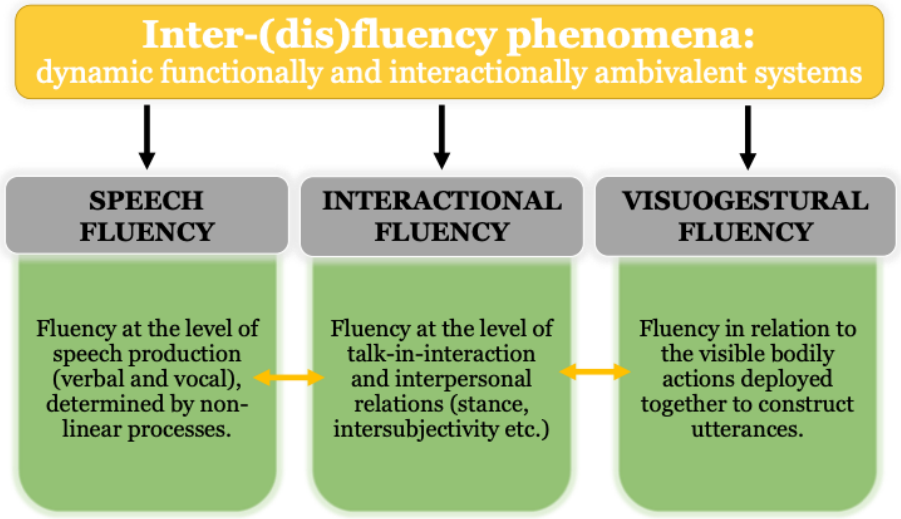


Figure 15. *Multidimensional model of inter-(dis)fluency*

This model comprises three different dimensions. First, the *speech dimension* is the equivalent of Segalowitz's (2016) *utterance fluency* (cf section 3.1.3.). We refrained from using the term "utterance" here because it has different implications, e.g. Kendon who defines it as: "the ensemble of actions, whether composed of speech alone, of

visible action alone, or of a combination of the two." (Kendon, 2004, p. 111). This dimension is only restricted to the level of speech (verbal and vocal) production which relies on multiple non-linear processes introduced in section I.1.1. It also takes into account the combination level of fluencemes, whether they occur clustered with other markers, or isolated (cf Crible et al., 2019, section 3.1.3.). The temporal features of fluencemes are also investigated within this dimension, which includes the duration of the vocal markers (e.g. filled and unfilled pauses, and prolongations). In sum, this dimension essentially reflects the work of psycholinguists as well as researchers in SLA and corpus-based linguistics. The second dimension, the *interactional dimension (inter-fluency)*, is similar to Grosman (2008)'s socio-interpersonal dimension (cf section 3.1.3.), but it further includes the situated conversational language practices shaping social interaction introduced in section 3.2.1, which is also in line with French theories of *co-énonciation* reflected in the work of Candea (2000, 2017). Finally, the *visuogestural dimension*, which echoes Götz's (2013) *nonverbal fluency*, reflects the influential work of different gesture researchers, such as Cienki (2005), Kendon (2004), Morgenstern (2014, 2020), Müller (2017), and Streeck (2009). In this dimension, the term "utterance" is to be understood according to Kendon's definition provided above. We excluded the *cognitive* and *perceived* dimensions of fluency in our model (cf Segalowitz, 2016; Götz, 2013), as we believe that they cannot be as easily "measurable"; *cognitive fluency* is rather an abstract construct, and *perception fluency* would require perception experiments, which goes beyond the scope of our study.

In sum, our definition of inter-(dis)fluency involves multiple dimensions which are not mutually exclusive, but interactively complementing one another in the course of the interaction (hence the term *inter-(dis)fluency*). In some contexts, a verbal utterance that is considered highly "disfluent" in the speech flow will not necessarily impede the interactional flow of the multimodal interaction; in other contexts, however, the presence of a single fluenceme could disrupt the progressivity of an interactional sequence. Once more, this view vouches for a dynamic approach to (dis)fluency which includes different degrees of fluency and or/ disfluency.

4.3. Main theoretical assumptions

In light of the multiple approaches addressed above, we will now address our main theoretical assumptions underlying our multimodal and contrastive study of inter-

(dis)fluency, grounded in an integrated interdisciplinary theoretical framework. These assumptions, listed below, are applied to both native and non-native (dis)fluency, as opposed to Segalowitz (2016) and Götz (2013) who offered fluency models which specifically targeted L2 speech. Starting with the analysis of typical "disfluency markers" (cf section 1.3.) which have traditionally been viewed as an interruption of the speech flow with no propositional content (Fox Tree, 2007), the present study aims to uncover the different discursive and interactional roles the same *a priori disfluent forms* can serve in different situations by taking into account different levels of analysis (speech, gesture, and interaction). Following Götz (2013) and Crible et al., (2019) we will speak of *fluencemes*, which are better understood in terms of constructions, whose degree of conventionality and entrenchment may rely on their frequency, position, and combinatory patterns. Highly frequent clusters (e.g. *filled pause + unfilled pause*) can show a high degree of automaticity, while highly complex combinations show greater disruption in the speech signal. This is one way to reflect the two sides of (dis)fluency, which is the starting point of this thesis. But as we have seen, the ambivalence found within (dis)fluency can cover other aspects of language such as pragmatics and multimodal communication.

- In this sense, inter-(dis)fluency phenomena should not be regarded in terms of binary opposition between fluency and disfluency, but rather as a **multi-level embodiment of the notion of fluidity and flow** (fluency) with its **potential interruption** (disfluency).
- The notion of interruption is to be understood on **the basis of different levels**: (1) *the disruption of the verbal flow and the acoustic signal*; or an interruption in the speaking activity; (2) *the interruption of the interactional flow* through postures of disengagement, disalignment, or disagreement; and (3) *the suspension or interruption of a gestural activity*.
- In line with Allwood (2017) the present study considers fluency as resulting from two systems of communication, mainly **interactive communication management** (ICM) and **own communication management** (OCM). In this sense, inter-(dis)fluency phenomena do not only deal with internal cognitive processes (OCM), but also exhibit essential features of talk-in-interaction (ICM).
- Fluencemes are thus **highly flexible and dynamic categories** which display different degrees of convention and different sets of meanings which are altogether shaped by their **context of use**. Context is understood here in terms of (1) the

immediate neighboring environment of the fluencemes at the combination level (e.g. fluenceme sequence) (2) the syntactic position of fluencemes within the verbal utterance, (3) their sequential position within a turn (e.g. turn transitional place); (4) their co-occurrence with bodily actions; and (5) the situated language activity speakers are currently engaged in (e.g. storytelling activity) (6) their overall material environment (e.g. the objects they are manipulating).

To conclude, many of the assumptions addressed above follow assumptions from the different theoretical frameworks discussed in this chapter, mainly *cognitive grammar*, *interactional linguistics*, and *gesture studies*. The notion of **fluidity and dynamicity**, for instance, are found both in interactional linguistics (cf Mondada, 2007) and cognitive linguistics (cf Cienki, 2005), and the notions of embodied cognition, embodied experience and **embodied interaction** resulting in the situatedness of gesture and language are found across the three frameworks. A number of hypotheses regarding the distribution, combination, frequency, and visuogestural manifestation of fluencemes, which emerge from these different assumptions, will be further addressed in Chapters 3 and 4.

Highlights of Chapter 1:

- The term "Dysfluency" emerged in the 1950s in clinical linguistics to study stuttering phenomena, and is now a common term in psycholinguistics to refer to spoken spontaneous speech processes (pauses, repairs, repetitions, etc.,)
- However, the constructs of fluency and disfluency have very often been opposed to one another in the literature, leading to a restricted definition of disfluency based on speech error or speech disruption.
- Other terms are also found in the literature, such as "hesitation", but it is often inadequate as it is too contingent upon cognitive difficulty or indecision.
- The present thesis offers an integrated approach to inter-(dis)fluency, going beyond the traditional production- and cognitive-oriented view of disfluency and hesitation, and integrating different frameworks, mainly usage-based linguistics, interactional linguistics, and gesture studies.
- The term (dis)fluency follows a functionally ambivalent and dynamic approach, whereby the same a priori disfluent forms all have the potential to serve more or less fluent and/or disfluent functions, thus going beyond the binary opposition between "fluency" and "disfluency".
- The core term *fluency* is understood here not only as language proficiency, but as a multidimensional flow embodying the flow of speech, the gestural flow, and the interactional flow, thus vouching for a tridimensional and multilevel continuum which will be used throughout our analyses in the following chapters.

Chapter 2. Corpus and Method

Introduction to the chapter

The present interdisciplinary and integrated approach to inter-(dis)fluency, which draws on a multilevel fluency-disfluency continuum, aims to analyze the distribution and behavior of ambivalent fluencemes in multimodal discourse. As we have seen in the previous chapter, their categories are highly flexible and dynamic, and their ambivalence can be evaluated by looking at several variables, such as utterance position, language proficiency, register variation, and fluenceme sequence complexity. These findings can be yielded using a corpus-based methodology, which relies on quantitative treatments (e.g. frequency measures, percentages, average values etc.), in line with corpus-based approaches to cognitive linguistics and pragmatics (e.g. Crible, 2018; Crible et al., 2019; Schneider, 2014; Tottie, 2014, 2015). Furthermore, the degree of fluency and/or disfluency of fluencemes can be evaluated qualitatively at the interactional level, by integrating the social, sequential, and bodily actions participants may turn to when engaged in specific interactional practices (e.g. Kendon, 2004; Mondada, 2013; Sacks et al., 1974). Therefore, the present study relies on a *mixed-method* approach (cf Morgenstern et al., 2021; Stivers, 2015; Tashakkori & Creswell, 2007) which includes quantitative and qualitative analyses of the data. Quantitative analyses rely on the treatment of dependent and independent variables in the whole dataset using statistical tools, while qualitative analyses rely on a close observation of specific occurrences in the data, examined within their ecological environment. Therefore, our empirical and usage-based study of inter-(dis)fluency requires natural speech data to address our hypotheses (cf Chap. 1, section IV. 4.4), and conduct qualitative micro-analyses.

The present chapter is structured as follows: First, we explain our motivations for working on a videotaped dataset which comprises two corpora, the SITAF Corpus and the DisReg Corpus. Secondly, we discuss the different transcription methods and units of transcription used for the purposes of multimodal speech annotation. We then describe the annotation protocol used for quantitative annotation analyses, using a specific annotation scheme, which went through a number of methodological and technical changes to obtain a finalized version applied to the two corpora. Lastly, we

end this chapter with the description of the methods used to conduct our qualitative analyses.

I. Data

As Morgenstern & Goldin-Meadow (in press) have noted, different methods in linguistic research have been offered to study and document language and gesture, using either naturalistic or experimental data, or a combination of the two. The naturalistic approach, which we briefly described in Chapter 1, strongly relies on video recording tools that have led researchers to conduct detailed analyses of spontaneous interaction data within talk-in-interaction (e.g. Goodwin, 2007; Mondada, 2019; Sacks et al., 1974). These recordings capture the habitual and daily social activities people often engage in (e.g. arguing about politics, retelling an anecdote, helping with homework etc.) which further enables researchers to study the deployment of linguistic and gestural units in context, captured in situated discourse (cf section 1.1).

The experimental approach, on the other hand, often relies on laboratory-based controlled experiments in which participants are asked to perform a series of tasks with the help of an external stimulus. For instance, Cochet & Vauclair (2014) conducted a study in which French university students (male and female, aged 17-27) took part in an experiment which consisted in eliciting pointing gestures. The participants were seated at a table with the experimenter sitting across them, and the latter read out loud 7 specific communicative scenarios (e.g. asking for salt at the dinner table; showing your friend where the keys are; giving directions to a stranger, etc.). The participants were then asked to produce a pointing gesture that was specifically related to the situation. This study enabled the investigators to analyze the speaker's gestural preference patterns (i.e. variation in hand shape and function) in relation to their contextual features (i.e. the type of communicative scenario). Conversely, a different study on pointing gestures conducted by Kendon & Versante (2003) examined the hand shapes and configurations of pointing gestures (e.g. index finger pointing, index palm down, thumb pointing, open hand pointing etc.) in different video recordings of naturalistic conversations between middle-aged speakers in the south of Italy. The researchers first reviewed these recordings to identify all instances of manual pointing, and then analyzed them according to their shape, configuration, and usage in context. They illustrated their instances with a series of detailed examples which showed the different contexts in which the gestures appeared,

the transcription of the interactional exchange, and a pictorial illustration of the gesture.

While the two studies presented above share similar research goals (i.e. examine the different forms of pointing gestures according to their use in context) they reflect radically different methods. Cochet & Vauclair (2014)'s study relies on the elicitation of a gesture in a highly controlled environment (i.e. an elicitation experiment), while Kendon & Versante (2003)'s work illustrates the different deployments of a gesture within situated activities (i.e. friends playing cards together, conversations between members of a club at a council meeting). Consequently, the results of the respective studies are presented very differently in the two papers. While Cochet & Vauclair (2014) present quantitative findings with statistical evidence (e.g. percentages, p values, z scores etc.), Kendon & Versante (2003) illustrate their findings with qualitative detailed analyses, in line with interactionist approaches.

The present study of inter-(dis)fluency, which relies on both qualitative and quantitative methods (cf section II and III) is based on a video-recorded dataset of English and French, which comprises semi-directed interactional data (cf section 1.2 and 1.3). As we shall see, our data is neither essentially naturalistic nor experimental per se, as it presents instances of naturally occurring situations (i.e. students interacting with one another, talking about school assignments), but within a relatively controlled setting. This will be further developed in sections 1.2, 1.3 and 1.4. In the following subsection, we address the importance of video recorded data in natural settings for the study of inter-(dis)fluency.

1.1. The importance of video collected data in naturally occurring situations

In the field of Interactional Linguistics, Conversation Analysis, and ethnomethodology (e.g. Garfinkel, 1967; Sacks, 1992), a significant interest has been given to the collection of oral data in spontaneous, naturalistic, and ecological settings within situated and social activities, known as *talk-in-interaction* (cf Chapter 1, section 3.2.1). Conversational talk is, as Thornbury & Slade (2006, p. 1) put it:

> the primary location for the enactment of social values and relationships. Through talk we establish, maintain and modify our social identities. The role that conversation plays in our formation as social beings starts at an early age.

Three essential analytic orientations emerge from this conversation-analytic approach to interaction (Atkinson et al., 2002, p. 104): first, talk and bodily behavior are the primary "vehicles through which people accomplish social activities and events"; secondly, the significance of the participants' social activities is contingent on their immediate context, as they progressively shape it moment-by-moment; thirdly, participants rely on social practices to make sense of their actions and of others'. As we have seen earlier (see Chap. 1, section III.3.3), in face-to-face interaction, social activities are accomplished through the deployment of multiple semiotic modalities: spoken, visual, and tactile. The development of the talk is entangled within the material environment and the bodily actions performed by participants. In the same vein, gestural actions also pertain to the ecologies of their neighboring environment: they can project a turn or an action, and provide co-participants with a "forward understanding"; an anticipation of what will come next (Streeck, 2010, p. 228). In addition, the participants' gestural actions are made within tangible, physical settings, enabling them to rely on the relevant objects found in the local environment (Goodwin, 2007), which altogether offers a multimodal and interactive frame for the situated activity. This was also put forward by Cienki (2016, 2015b), who stressed the need to study videos of face-to-face spontaneous conversations in order to capture relevant aspects of multimodal communication that are not otherwise observable in decontextualized speech only: (Cienki, 2016, p. 606):

> Communicative usage events based on the canonical face-to-face encounter, even if they are digitally mediated audio-visually, are different in nature and in substance from those when the interlocutor is not co-present and cannot be seen or heard.

Similarly, in sociocultural theories of learning, derived from Vygotsky's (1934) research in children development, and further situated within a usage-based perspective (Tomasello, 2003), multimodal conversation is regarded as an essential medium for language learning and human socialization. Through routinely social activities, children learn to progressively *language* their experience, in other words "produce motivated, conventionalized language forms (sounds, words, tones, gestures)" (Morgenstern, 2021, p. 67). These everyday practices play a fundamental role in the child's language development, and this has led a number of researchers to develop video recording tools to capture interactional talk deployed in different

ecologies, whether "in time", during the sequential unfolding of the child's interactional environment, or "over time" during longitudinal studies of children's development within their family environment in the course of several years (Morgenstern, 2021, p. 68). This is done through fieldwork – when the researcher collects data directly *in the field*, in other words within a specific social, institutional, or cultural environment (e.g. a daily encounter between close friends, a medical consultation, or a business meeting) where the researcher observes, records and potentially interacts with the participants of the conversational exchange. Given the close proximity the observers may have with the participants of the interaction (i.e. their physical body present in the field, or the use of a microphone), their presence can also potentially interfere with the speakers' ongoing activities (see Labov's 1972 observer's paradox in Beaupoil-Hourdel, 2015 and Morgenstern, 2021, for review). These issues further invite researchers to consider themselves as actual social actors, who can, in turn, shape the contingencies of the ongoing exchange (Mondada, 2001).

In sum, speakers' individual and social processes are embodied within ordinary practices of language use, and these everyday situations have been documented and recorded with the help of video recording tools for the purposes of data collection. However, the idea of *truly* authentic, ordinary, and casual speech has also been criticized by some researchers for being too idealized and not easily accessible (cf Candea, 2017, p. 11). Another approach which can be applied to the study of language in use, as adopted by Candea (2000, 2017) Crible et al. (2019), or Tottie (2011, 2014, 2015) is *corpus linguistics* (cf Chap. 1, section 3.1.1). As Candea puts it (2017, p. 11):

> La position que j'ai adoptée depuis, a été celle de la linguistique de corpus : il ne s'agissait pas de forger des analyses sur des exemples fabriqués par l'imagination du chercheur, mais il ne s'agissait pas non plus d'écarter des masses d'enregistrements variés sous prétexte que les locuteurs et locutrices n'étaient pas en conversation amicale avec des pairs. Le cadre d'analyse grammatical de Morel & Danon-Boileau allait dans le même sens, en favorisant la diversité des corpus, des situations, des styles, des profils de gens.

Similarly, Crible (2018) conducted a contrastive corpus study based on a compilation of different corpora in French and English with a comparable corpus design. Her dataset included a large collection of corpora which comprises eight different registers (interviews, conversations, political speeches, classroom lessons, interviews, sports

commentaries, news broadcasts, and phone calls.). These settings show different degrees of preparation (spontaneous, semi-prepared, and prepared) as well as different degrees of interactivity (e.g. interactive, semi-interactive, non-interactive). This was also done in the work of Tottie (2011, 2014) who looked at different types of corpora (e.g. London-Lund Corpus, the British National Corpus) in order to show differences in age, gender, and register. This is further elaborated in Chapter 4.

Therefore, it is also important to distinguish between different types of settings when studying aspects of spoken language use. For instance, Clark (1996) distinguished between *personal settings* (face-to-face or telephone casual conversations), *nonpersonal settings* (monologues; e.g., an academic lecture, or a preacher's sermon), *institutional settings* (speech exchanges limited by institutional rules, such as a politician holding a news conference or a lawyer interrogating a witness in court), *prescriptive settings* (e.g. members of a church reciting readings from a prayer book), *fictional settings* (e.g. a theatrical performance), *mediated settings* (e.g. a lawyer reading a testament at a hearing, or a letter dictated to a secretary), and *private settings* (e.g. when people speak to themselves without addressing others).

In line with some of the approaches outlined above, the present study of inter-(dis)fluency is conducted on a video-recorded dataset comprised of two different corpora in semi-spontaneous, semi-naturalistic settings, reflecting different degrees of language proficiency (L1 and L2) and speech preparation (prepared versus spontaneous). This dataset was deliberately compiled to obtain different discourse characteristics and genres. The data therefore exhibits differences in language, setting, genre, and register, which will further account for contextual and situational differences found in the distribution of ambivalent fluencemes across data types (cf Chap. 3 and 4). In addition, the use of semi-guided interactions has also been adopted in previous work on multimodality (e.g. Boutet & Cienki, 2016; Debras, 2013; Cienki & Irishkhanova, 2018), which further justifies our choice of data. The motivations for working on our two different videotaped corpora are provided in section 1.4. We first begin with the description of the first corpus under scrutiny in the following section.

1.2. The SITAF Corpus

The SITAF corpus (*Spécificités des Interactions Verbales dans le Cadre de Tandems Linguistique)* was collected at Sorbonne Nouvelle University between 2012 and 2014 (Horgues & Scheuer, 2015). It was collected within the framework of the SITAF

project, a research project funded by Sorbonne Nouvelle University, which aimed to gather various semi-spontaneous tandem exchanges in English and in French. The main motivations for collecting this corpus were: (1) to see whether the learning environment of the tandem interactions was reflected in the pronunciation features observed in the data, and in what ways; (2) how and which L2 phonetic features were acquired by the participants in tandem exchanges (Horgues & Scheuer, 2015, p. 1).

The data consists of a 25-hour video recorded corpus, comprising 21 pairs of undergraduate students. It includes 21 native French participants, all female, and 21 native English participants, 16 female and 5 male, all aged from 17 to 22. The participants were recruited on a voluntary basis, and the members of the SITAF team selected the pairs using an online questionnaire which had previously been answered by the participants. The questionnaire addressed: (1) their linguistic background, (2) their level of proficiency in their second language, (i.e. English for the French speakers and French for the English speakers), by rating it on a scale from 1 to 10; and (3) their personal interests, e.g. favorite conversation topics.

A majority of the French students were enrolled in an English major (i.e. English as a foreign language) as part of their undergraduate program. Their average score in L2 oral expression was 6.8 out of 10. The average score for the English native speakers in their L2 (French) was 6.6 (see Table 5 in section 1.2.3). The English native speakers came from a variety of language backgrounds: American English, Canadian English, British English, Irish, and Australian. English-French bilinguals were excluded from the corpus. The latter include speakers whose parent is a native speaker of the target language, speakers who started learning their L2 before the age of 5, and speakers who attended L2-medium school for a long period of time.

1.2.1. Methods and data collection procedure

All the video recordings were made in a sound studio at Sorbonne Nouvelle university. The paired participants (who were tandem partners as part of the tandem project at university) were video recorded twice using three cameras – two filming each participant individually, and one recording the whole interaction, as shown in Fig. 16. The participants were recorded twice; the first time in February 2013, several days following their first encounter, and the second time in May 2013. They were encouraged to meet regularly between the recording sessions (e.g. once a week). By

the end of the experiment, they had met about twelve times on average during the three-month interval.

The experiment was conducted in two different settings: the L1-L1 "control" settings (English-English or French-French) and L1-L2 "tandem" settings (English-French, French-English). The participants first interacted in L1-L1 settings during the first recording session in February, then in L1-L2 settings during the second one in May. Before the experiment, the instructors emphasized on the notions of "solidarity" and "mutual assistance", as well as the need to separate English and French when performing the tasks. During the recording sessions, each pair was asked to perform three communicative tasks, first in English, and then in French.

Figure 16. *Camera configurations (SITAF Corpus)*

The first task, entitled "Liar, Liar" is a storytelling task in which one participant had to talk about his/her last vacation and insert three lies in the story, which was later identified by the tandem partner. The participants who told the story were allowed to prepare their narrative before the recording session, but they were not allowed to write anything down, except for a few key words to help them as they went along. The instructions given to the partners (who were not doing the retelling) were to carefully listen to the story, and only take part in the interaction when they needed to ask for clarification, or to assist their interlocutor with language difficulties. At the end of the

narrative, the partner had to guess the three lies, and if they failed to identify them, the participant who had initially told the story would reveal them.

The second task, entitled "Like Mind" required no preparation beforehand, and is a more collaborative argumentative task. It consists in discussing a relatively controversial topic, such as "the best years of your life are teenage years", "prisoners should have the right to vote", or "a good friend should always take your side, whatever happens." For this task, the participants were asked to respectively give their opinion on the topic, and later justify their position. At the end of the debate, they would both decide on their level of agreement on the topic, rating it on a scale from 0 (complete agreement) to 10 (complete agreement).

The third task was a reading task; the participants were instructed to read a small text written in their second language (i.e. in French for the English speakers, and in English for the French speakers), first with the help of their tandem partners, and a second time on their own. At the end of the two sessions, the participants filled in two questionnaires, and talked about their learning experience within the tandem setting to a member of the SITAF team and an expert in L2 pedagogy.

All the participants had to read and sign a consent form in order to take part in the study (cf Appendix 2). All their faces are blurred or hidden in this thesis in order to preserve their anonymity (Horgues & Scheuer, 2015). They were also assigned labels (e.g. F07, A07, A03, F03 etc.) as well as pseudonyms to protect their identity.

1.2.2. Why the SITAF Corpus?

As pointed out throughout this dissertation (cf section I.1.1. of this chapter, and Chap. 1 section 3.3.,) the approach adopted for the present study of inter-(dis)fluency is by essence multimodal, thus relying on the multiple semiotic resources deployed by speakers in the course of their social practices. Therefore, the first motivation for selecting the SITAF corpus was for its sound and video quality; the use of three cameras is particularly helpful for analyzing each participant's facial expressions, as well as their body movement and hand gestures. While more and more videorecorded corpora are being documented and shared on open resources in English and in French (e.g. the Talkbank database, the ORTOLANG repository, or the ColaJE corpus, among others[62]), it is still difficult to find multimodal data in two languages within the same

[62] For more information, visit https://www.ortolang.fr , https://www.talkbank.org , and http://colaje.scicog.fr/index.php/corpus (last retrieved on August 26th 2021)

groups of speakers. Several researchers have already explored the multimodal quality of the SITAF Corpus and worked on various topics, such as corrective feedback (Debras et al., 2015, 2020; Scheuer & Horgues, 2020) miscommunication (Horgues & Scheuer, 2017) and chains of reference (Debras & Beaupoil-Hourdel, 2019). The present study, which is based on an earlier preliminary study conducted on the same data (Kosmala, 2021, Kosmala & Morgenstern, 2017) offers another contribution to this corpus, following the multimodal approach adopted from past studies.

Another motivation for choosing the SITAF corpus is that it provides a scope for comparative analysis of L1 and L2 productions of the same speakers both in French and in English, and in similar contexts. In this respect, the aim of the present study is to compare the production and distribution of fluencemes in L1 and L2, in line with previous work on fluency in Second Language Acquisition (see Chapter 1 section 2.1.1, and Chapter 3). While a considerable amount of research has been conducted on EFL learners (English as a Foreign Language, or English as lingua franca), such as Spanish, Turkish, or French learners of English (e.g. de Jong, 2016a; Gilquin, 2008; Stam, 2001, among others, cf Chapter 3), these groups of speakers have systematically been compared to different groups of native English speakers. To our knowledge, less work has been done on fluency rates in L1 and L2 productions within the same speaker groups. This is the case with the SITAF Corpus, where each speaker alternated between their native and their non-native language, which allows for the analysis of intra-speaker variation.

In addition, tandem settings provide a relevant model for language learning through practices of cooperation and socialization. The model of tandem learning was originally developed in a German-French youth-exchange program in the 1960s in an informal learning context (Bechtel, 2003 in Elo & Pörn, 2018) and has been implemented in different contexts and languages ever since (e.g. in Swedish-medium schools, read Elo & Pörn, 2018 for review). Contrary to more formal teacher-student settings, tandem partners are not expected to assess or correct each other's L2 oral performance, but rather maintain a friendly relationship (Horgues & Scheuer, 2015, p.2). As Elo & Pörn (2018: 1-2) put it:

> Tandem learning embraces a socio-interactional perspective, emphasizing that learning and instruction are social processes situated in social contexts, in which participants are engaged in mutual social actions.

This type of peer interaction in informal settings (i.e. outside the classroom environment) thus relies on reciprocity (Calvert & Brammerts, 2003), and is a two-way language learning process, where both parties cooperate and can benefit from each other's expertise. This is particularly relevant for the present study of inter-(dis)fluency which is grounded in the framework of interactional linguistics, and more specifically in a view of *CA for SLA* (Pekarek-Doehler, 2006; cf Chap. 3); as we shall see in Chapter 3, our analyses will further explore the different ways fluencemes can contribute to the co-construction of meaning, the different positionings of the co-participants, and the multimodal communication strategies deployed by L2 learners to deal with their language difficulties.

1.2.3. Selected sample under scrutiny

As briefly mentioned in the previous section, the present study is based on an earlier pilot study (Kosmala, 2021, Kosmala & Morgenstern, 2017) which was conducted on a small sample of the SITAF Corpus. The sample initially included eight recordings from the data in L1-L1 and L1-L2 settings, from Task 1 and Task 2, and selected two pairs of speakers (2 French speakers and 2 American speakers). The aim of the pilot study was to test the first version of our annotation scheme, which was later subjected to several changes in order to be developed into a more reliable format (cf section II.2.1.).

Following the DiSS 2017 workshop, where the pilot study was initially introduced, a few adjustments were also made regarding the choice of the data. In the pilot study, we initially compared (dis)fluency rates in L1 and L2, in English and in French, as well as in Task 1 ("Liar Liar") and Task 2 ("Like Minds"). While results showed a higher rate of fluencemes in Task 1 compared to Task 2, no differences were found in the distribution of fluencemes in L1 and L2 (for Task 1 only). However, this may be explained by the fact that L1 French and L1 English were compared to L2 English only, and in Task 1 specifically, while Task 1 and Task 2 were compared regardless of language proficiency; as a result, too many different variables were combined (i.e. speaker, language proficiency, task type), which questions the validity of the results. Moreover, the issue with Task 1 is that it contains deceptive speech, so the high rates of fluencemes found in Task 1 may lead to different interpretations (i.e. increased cognitive load, planning processes, or deception behavior, e.g. Arciuli et al., 2010). Therefore, for the present study, we chose to work exclusively on Task 2 instead of Task 1 to avoid combining multiple variables such as truth-telling versus deception,

which goes beyond the scope of this thesis. Additionally, we only selected L1-L2 settings (and not L1-L1) in our novel data sample in order to compare the same participants' L1 and L2 productions in the same tandem settings. We wanted to treat task type and communication setting separately in a different study (in the DisReg Corpus, cf section 1.3) in order to avoid crossing too many different variables within the same investigation. Lastly, the present study does not target aspects of sociolinguistic variation such as age, gender, or language variety. In this respect, we only selected American speakers from the corpus and excluded other varieties of English. In addition, all the participants of the study were roughly the same age, i.e., in their early twenties.

11 pairs from Task 2 in L1-L2 settings were selected from the data. This includes 22 speakers in 22 video recordings. The interactions lasted on average 3:40 minutes. The total duration of the selected data is approximately 1h21. More information regarding sample size is provided in the table below.

Table 4. *Selected sample size of the SITAF Corpus*

	L1-L2 English		**L1-L2 French**	
	Duration (min)	Number of words	Duration (min)	Number of words
Pair 02	02:36	394	03:24	527
Pair 03	04:40	888	05:07	820
Pair 07	04:44	796	04:58	999
Pair 09	05:57	919	05:44	1153
Pair 10	02:00	267	01:31	239
Pair 11	05:30	1112	03:06	571
Pair 13	04:40	673	03:26	506
Pair 15	03:35	617	02:50	487
Pair 16	03:04	560	01:56	560
Pair 17	02:32	432	05:03	866
Pair 18	02:15	358	02:25	377
Total	41:33 min	7016 words	39:30 min	7105 words

Table 5 includes the self-assessment scores made by the participants when they evaluated their level of L2 oral proficiency (cf section 1.2); while these numbers do not officially assess the students' proficiency levels within a framework of reference such as the CEFR (*Common European Framework of Reference for Languages: Learning,*

Teaching, Assessment[63]*),* they still give an approximate idea of the L2 learner's oral skills with some accuracy (cf Ma & Winke, 2019). Unlike previous work on Fluency in SLA, the primary focus of the present thesis is *not* to assess proficiency through (dis)fluency rates; however, the relationship between proficiency and fluency will still be discussed in Chapter 3.

Table 5. *Students' self-evaluation scores*

	Listening comp.	**Oral prod.**		**Listening comp.**	**Oral prod.**
French participants			**American participants**		
F02 (Maria)	8/10	7/10	A02 (Haley)	7/10	6/10
F03 (Marina)	7/10	7/10	A03 (Julia)	7/10	6/10
F07 (Julie)	7/10	7/10	A07 (Amber)	8/10	6/10
F09 (Emilie)	7/10	7/10	A09 (Arthur)	8/10	7/10
F10 (Juliette)	5/10	5/10	A10 (Betty)	6/10	6/10
F11 (Sally)	6/10	6/10	A11 (Harry)	8/10	6/10
F13 (Elena)	7/10	8/10	A13 (Francis)	7/10	7/10
F15 (Melissa)	7/10	7/10	A15 (Simon)	8/10	7/10
F16 (Elisa)	7/10	8/10	A16 (Beth)	8/10	4/10
F17 (Lola)	7/10	8/10	A17 (Ruth)	7/10	6/10
F18 (Sophie)	7/10	6/10	A18 (Rosie)	7/10	5/10

1.3. The DisReg Corpus

The DisReg Corpus (DISfluency across REGisters) was collected as part of the present 3-year PhD project supervised by Professors Aliyah Morgenstern and Maria Candea at Sorbonne Nouvelle University. Following the exploratory work conducted on the SITAF Corpus, our objective was to apply our finalized annotation scheme (cf section II.2.2.) to a similar videotaped corpus in order to compare a new set of variables (i.e. register variation, level of interactivity etc.). The initial project was to videorecord 12 French students at a French university in two different communication settings (during oral presentations and face-to-face interactions), and later use the same research protocol to replicate it on 12 other American students at an American university. During the Fall semester of 2018 at Sorbonne Nouvelle University, 8.15 hours of videotaped data were collected. Unfortunately, due to the outbreak of

[63] For more information visit https://www.coe.int/en/web/portal/home (last retrieved on August 25th 2021)

coronavirus disease in 2020[64], our data collection project in the United States was severely impacted and could not reach its completion. Our research was thus only conducted on the existing French data, and we later decided to add more L1-L2 pairings to our SITAF sample (see Table 4) in order to have more data in English[65].

The corpus comprises 18 video recordings of 12 undergraduate French students enrolled in a French literature class held at Sorbonne Nouvelle University. The corpus is twofold: the first part includes recordings of the students giving an oral presentation in front of the whole class and their teacher. Their presentation was part of an evaluation which counted for approximately 50% of their overall grade. The presentation consisted in analyzing a sonnet or an excerpt taken from a novel, essay, or play, using French dissertation methods (i.e. introduction, three-part presentation and analysis, and conclusion). The presentations lasted 29.5 minutes on average. The second part includes video recordings of the same students who were filmed in pairs when engaged in semi-guided conversations. The interactions lasted 22.6 minutes on average. In the following subsections, we explain the methods and procedure used for the data collection, explain our motivations for collecting this corpus, and describe the size of our selected sample.

1.3.1. Methods and data collection procedure

All students were recruited on a voluntary basis. At the beginning of the Fall Semester in 2018, we (i.e., the investigator, Loulou Kosmala) first contacted several French literature teachers by email, described our research project, inquired about their students' oral presentations, and asked whether we would be allowed to come to the classroom and introduce ourselves, so we could present our data collection project in person. We then came to class and briefly presented our research project to the students, but withheld specific aspects of our research goals (i.e. the study of (dis)fluency phenomena and body behavior) which could have potentially led the students to be self-conscious about their way of speaking[66]. This information was later

[64] I stayed at UC Berkeley in the linguistics department as a Visiting Scholar during the Spring Semester of 2020 to collect the data, with the help of my sponsor Eve Sweetser. Unfortunately, due to the outbreak of Covid-19, my stay had to be shortened; two months and a half after my arrival, I was obliged to return to France. Even though my research protocol had already been approved by the CPHS (The Committee for Protection of Human Subjects), I was not able to find the participants in time to complete the data collection project.

[65] We had initially only selected 6 pairs from the SITAF Corpus before collecting the DisReg Corpus.

[66] This is known as "incomplete disclosure"; when subjects are not fully informed about specific aspects of the study. Visit https://cphs.berkeley.edu/deception.pdf for more information (last retrieved on August 26th 2021).

revealed to the participants at the end of the two recording sessions. We asked the students whether they would agree to be filmed in class during an oral presentation, and then again in pairs during a conversation. We insisted that the research was strictly conducted to achieve academic purposes, and that the recordings would only be shared with members of the scientific community if the students gave their consent. We added that we were not evaluating the content nor the quality of their presentation, and if they wanted to watch their oral performance, they were welcome to have a copy of the recording. A sheet of paper was then passed around to the students, and whoever agreed to participate in the study had to write down their name, email address, and date of their presentation.

Following this first encounter, we then contacted the students who had agreed to take part in the study, thanked them for participating, and reminded them of our project and its scientific implications. We further informed them that they were going to sign a consent form (cf Appendix 2), and that they were free to go back on their decision anytime they wanted. We then inquired about friends or classmates who would also accept to be filmed for the second recording session (the conversational one) so we could film them in pairs. We explained that the point of the study was to record the same students in the two conditions. When the two students accepted to be filmed, we either came to class for the first recording session, and then filmed them in pairs for the second session, or vice versa, depending on the students' availabilities and the dates of their presentations. The date for the second recording session depended entirely on the students. When we had not previously introduced ourselves in class, we contacted the students directly by email; this was the case when we had found friends or classmates of the participants who also accepted to be filmed for the project. We also made sure to contact the instructors before coming to class to inform them of our project, and ask them to grant us permission to film the students. When we came to the classroom, we first gave the consent form to the student, asked him or her to read it carefully, and sign it. The student gave us back the form, and we kept a copy. The student then came to the teacher's desk, sat or stood, and started his or her presentation in front of the class. Sometimes the presentations were also made in pairs (cf Fig. 17). The instructor usually sat in the corner of the classroom to take notes, and was not seen on camera. We sat in the front row of the class, held the camera in our hand, and tried not to sit directly in front of the student to avoid distracting him/her. We also made sure to film the participant only, and not the other students in the class.

The camera filmed the upper part of the participant's body (face, shoulders, and arms, see Fig. 17). We recorded the entire presentation, and did not interrupt the student. When the presentation was over, we quietly left the classroom without interrupting the teacher's feedback and the rest of the lesson. When the two students and the investigator reached an agreement for the date and time for the second recording session, we booked a classroom on campus, and asked the students to meet us there. Before the students arrived in the room, we arranged two chairs so that they would sit face to face during the exchange (cf Fig. 18).

Figure 17. *Participants in class during their oral presentation*

Figure 18*. Participants in pairs during the conversation-session*

We made the participants read and sign another consent form, and spent a few minutes to explain the point of this specific part of the study. Unlike the oral presentation in class, which was prepared at home in advance, and involved a certain degree of stress, this part was much more relaxed and spontaneous. We invited the participants to view it as a casual exchange. We gave them a sheet of paper which included a few topics that could help start the conversation; the topics included: (1) last film/ TV show you have seen; (2) last novel/article you've really liked; (3) last trip, and (4) funny anecdote at university. The participants did not have to go through all the topics, and they were free to talk about anything else if they wanted. They were also asked not to interact with us. We sat in front of them with the camera, and made sure that they both fitted in the frame, so their facial expressions, body movements, and hand gestures were visible. Overall, the students managed to talk freely about various topics, despite feeling a little nervous in the beginning.

Unlike the participants from SITAF, all the subjects accepted to have their faces shown, so they did not have to be blurred (cf Appendix 2). They were also assigned code names (A1, A2, B1, B2, etc.) as well as pseudonyms to protect their identity.

1.3.2. Why the DisReg Corpus?

The DisReg Corpus was first collected as an addition to our corpus-based study, and our primary objective was to collect data that had a comparable corpus design with the SITAF Corpus (cf section 1.4.2). In addition, the DisReg Corpus presents a number of qualities for the present study of inter-(dis)fluency. First, it includes video recordings, which, as emphasized earlier, are fundamental to the study of multimodal face-to-face spoken communication. Most importantly, the video recordings include productions of the same students engaged in different practices and in different communication settings. As mentioned earlier, (cf section I.1.1.) social situations can take place in a variety of spoken settings, with different degrees of spontaneity and interactivity, which ultimately result in differences in socio-interactional conventions and expectations.

This last point is a key aspect to consider when taking into account the situatedness of language: speakers constantly adapt their vocal and bodily behavior to align with their interlocutors within a specific participation framework (Goodwin, 2007). This type of behavior may vary depending on the relationship they have with the co-participants, the level of familiarity they may share, or the material

environment in which they are interacting. In class presentations, for instance, the participants' language productions are bounded by a number of institutional constraints. First, their oral performance almost exclusively relies on what is written on their notes prepared at home, which leaves very little room for spontaneity. As a result, a majority of the students keep their gaze fixed on the piece of paper or laptop they have at their disposal, and content themselves with reading. They also often become very self-aware of their own productions, and may thus wish to signal to the audience that they are currently searching for the right page in their notes, or apologize for misreading a word, in order to save face (see Goffman, 1967 on face-work[67]). Second, even though they have many interlocutors (the students in the class and the teacher), they rarely address them directly, because they are expected to give a lengthy formal presentation, which does not require the audience's coparticipation. This may further justify the students' needs to constantly rely on material objects they have within reach (e.g. a piece of paper, a book, a laptop, or a pen) instead of relying on their partner, which they would do naturally in conversational settings (i.e. the conversation-session in pairs). Third, their motivations for carrying out their task in class are entirely different from the ones expected in a social conversation. The quality of their oral presentation is most certainly driven by their wish to obtain a good grade, and perhaps make a good impression on the instructor, while the fluidity of their productions during a conversation is rather determined by the contingencies of the exchange, and their wish to express and co-construct ideas. These points are further developed in Chapter 4.

Following the usage-based assumption that fluencemes are dynamic and fluid categories showing different degrees of fluency and or disfluency (cf Chapter 1 section IV), and further in line with Cienki's (2017b; 2012) *scope of relevant behavior* theory (cf Chapter 1, section III. 3.3.1.) the DisReg Corpus allows for a multilevel contextual analysis of inter-(dis)fluency: their interactional and functional ambivalence can further be evaluated by observing their pattern of embodied behavior in two different registers and settings (i.e. *personal* versus *nonpersonal*, cf section 1.1.). This will be addressed more specifically in Chapter 4.

[67] These aspects of social interaction (e.g. face work and self-consciousness) will further be elaborated in Chapter 4.

1.3.3. Selected sample under scrutiny

As specified earlier, the initial objective of the present empirical and corpus-based study is to triangulate evidence from two different corpora with a comparable corpus design (cf section 1.4.1), and this also includes sample size.

Table 6. *DisReg Corpus sample size duration (number of words)*

	Conversation	**Class presentation**	
Pair A	04:56 (1048)	A1 (David)	02:08 (324)
		A2 (Jessica)	04:11 (590)
Pair B	03:53 (766)	B1 (Paul)	03:00 (431)
		B2 (Paula)	02:54 (402)
Pair C	05:53 (1295)	C1 (Dan)	02:44 (389)
		C2 (Laura)	03:06 (513)
Pair D	05:56 (1140)	D1 (Alex)	02:33 (559)
		D2 (Jenny)	03:36 (489)
Pair E	06:26 (1385)	E1 (Lea)	02:30 (306)
		E2 (Tina)	02:43 (391)
Pair F	06:20 (1347)	F1 (Linda)	02:42 (352)
		F2 (Matt)	03:23 (863)
Total	33:24 min (5609 words)	35:30 min (5609 words)	

The video recordings from the SITAF Corpus last on average 3:40 minutes (cf section 1.2.3.), while the ones from DisReg are significantly longer (over 20 minutes, cf section 1.3.). Therefore, since the latter comprises video recordings of considerably longer duration than SITAF, we randomly extracted 2-6 minutes from each video file from the DisReg Corpus (equally representing all participants) to approximately match the size of SITAF. The total duration of our selected sample is 1h08. The exact duration of the recordings is found in Table 6 above.

1.4. Motivations for working on a "small" corpus

We are aware that our selected sample (2h30) is relatively small compared to the actual size of the whole dataset (25 hours for SITAF and 8 hours for DisReg), but also compared to most corpus-based studies in linguistics, which are very often associated with large-scale collections of spoken or written corpora. In this section, we argue in favor of "small" and context-specific corpora, and emphasize their benefits for both quantitative and qualitative treatments, in line with Vaughan & Clancy (2013), Danino (2018) and Debras (2018). Table 7 summarizes our total corpus size in number of words (26,000) and total duration (2h30), broken down by speaker group and setting in the two corpora under scrutiny.

***Table 7**. Total corpus size*

	SITAF Corpus	**DisReg Corpus**
Number of words	Tandem interactions (Task 2 EN): 7016 Tandem interactions (Task 2 FR): 7105	Class presentations: 5609 Conversations: 6981
Duration (min)	Tandem interactions (Task 2 EN): 41:33 Tandem interactions (Task 2 FR): 39:30	Class presentations: 34:30 Conversations: 31:30
Participants	22 participants 11 American speakers 11 French speakers	12 participants French speakers

Despite the relatively "small" size of the corpus, it should be noted that the data still yielded a rather high number of tokens overall: 6042 fluencemes (3172 in SITAF and 2870 in DisReg) and 2381 hand gestures (1362 in SITAF and 1019 in DisReg) in total, which can still be used efficiently for quantitative treatments[68]. Moreover, our findings do not solely rely on quantitative treatments, but also draw on qualitative analyses, which focus on several case studies of local pragmatic patterns. This further reflects our **mixed-methods** approach to corpus linguistics and conversation analysis (cf Hashemi & Babaii, 2013; Johnson et al., 2007; Stivers, 2015; Tashakkori & Creswell, 2007), which can be defined as the following (Tashakkori & Creswell, 2007, p. 4)[69]:

[68] See Chapters 3 and 4.
[69] Read Candea (2017), Hashemi & Babaii (2013), and Johnson et al . (2007) for a more detailed review and definition.

> Research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches in a single study or program of inquiry.

As Stivers (2015) noted, there has recently been an increase in the use of a mixed methods approach combining CA methods with quantitative treatments. She argued that this combination of methods enabled CA research to target a broader audience. This kind of approach is not only beneficial to CA research, but to (dis)fluency research too. For instance, Peltonen (2020) further argued in favor of mixed-methods in L2 fluency research, and criticized studies for being mostly quantitatively oriented, involving frequency-based analyses of fluency, without paying attention to their functions or the contexts in which they occurred. She thus promoted the use of a qualitative approach to provide "a more comprehensive picture of fluency, enabling detailed analyses of fluency-related features in their immediate contexts" (Peltonen, 2020, p. 23).

In light of these approaches, we further defend our choice to work on a relatively small sample in order to combine quantitative and qualitative analyses, by looking into our data in depth, taking into account the local and global context of use of the fluencemes, their overall frequency across languages and settings, and their multimodal use in embodied interaction. We further argue in favor of small corpora in the following subsection.

1.4.1. "Size doesn't matter": the benefits of using "small" corpora

As Vaughan & Clancy (2013) pointed out, significant value has been given to the study of considerably large corpora, with the emergence of modern corpus-based linguistics. In their paper on small corpora and pragmatic research, they provided several examples of the largest English corpora available, such as the British National Corpus (BNC), which contains 100 million words of written and spoken English, and the COCA Corpus (Corpus of Contemporary American English) which is made up of over 450 million words. Used by a multiplicity of beneficiaries, such as linguists, teachers, or translators, these corpora provide a diversity of samples in British and American English in different genres and contexts. Large corpora have thus successfully been exploited in many corpus-based studies to explore different aspects of language, such as language variation, historical linguistics, or language pedagogy. The larger the data sample, the more reliable it becomes for efficient quantitative treatments. However,

as Vaughan & Clancy argued, there are equally many benefits to studying smaller corpora which do not only rely on generalized findings of frequency measures. Their arguments are presented below.

First, relevance to corpus size is relative, as it depends on the language modality. Spoken corpora often tend to be smaller than written corpora, as the data needs to be collected, transcribed as adequately as possible using transcription conventions (cf section 1.5.2), and then manually annotated by researchers, which can be a long and difficult enterprise. When it comes to videotaped corpora, which are often collected by researchers in CA, language acquisition, and ethnomethodology (cf section 1.1.), their analysis can be even more challenging and time consuming, as they rely on careful manual annotations of the observed phenomena (e.g. gestural actions and contextual features) at several levels of analysis (e.g. prosody, phonology, syntax and gesture, see Debras, 2018, p. 9). Therefore, what is traditionally considered a "small" corpus, in fact comprises a multiplicity of richly annotated multimodal features, carried out manually by one or several members of a research team. When the annotation is carried out alone, which is often the case with researchers who are limited by time constraints within a Ph.D project (e.g. Debras, 2013, who collected a 2-hour videotaped corpus during her Ph.D), it is often virtually impossible to build a very large corpus, for practical reasons.

There are, however, other advantages to working on smaller corpora that are not necessarily motivated by time limitations. In the field of pragmatics, one of the central benefits of working on a small corpus is that it enables researchers to "access authentic, naturally occurring language and to maintain a close connection between language and context." (Vaughan & Clancy, 2013, p. 6). Smaller corpora thus give easier access to contextualized findings which further illustrate specific instances of a given phenomenon. As Vaughan & Clancy (2013, p. 6) put it:

> While it is certainly possible to investigate phenomena such as hedging using large corpora, this can be a major challenge due to the variety of (para)linguistic selections available for use as hedges. Using a small, context-specific corpus offers significant advantages. These phenomena can not only be investigated in their original context of use, it is also usually possible to investigate virtually all occurrences and essay a refined analysis which takes the polysemous nature of many pragmatic features into account. Therefore, we can move from

> quantitative observations regarding frequency of items with pragmatic potential, which only tell a part of the story.

These remarks are also truly relevant to the study of fluencemes. Their overall frequency rates in the data can be interesting to get a general idea of their distribution across languages and settings, but being more "disfluent" (i.e. producing a higher number of fluencemes) in one particular language or setting does not necessarily mean that speakers are experiencing more language or cognitive difficulties (cf Chapter 1 section II. 2.2.1). This type of finding has to be completed with in-depth contextual analysis of the phenomena, which takes into account their different dimensions (i.e. speech, interactional, and visuo-gestural, cf Chapter 1, Section IV.4.2). In fact, the construction of our integrated framework (cf Chapter 1, Section IV) is not only grounded within the multiple theoretical research fields relevant to our study (i.e. usage-based linguistics, interactional linguistics, and gesture studies, cf Chap. 1), it is also based on a careful observation of our data. Moreover, as Vaughan & Clancy (2013) emphasized, a majority of small corpora were compiled by the researchers themselves, which reflects a close relationship between corpus and researcher (cf the mention of investigators as social actors, section 1.1.). As pointed out by Koester (2010) researchers often have a close proximity and a high degree of familiarity with the data they compiled, as they are more aware of the contexts in which it was collected. This better ensures that the generated quantitative findings are also complemented with qualitative contextual analyses. For instance, Cutting (2001, 2002) deliberately chose to work on a small corpus (26,000 words) which includes casual conversations of six students who took part in a Master's course in Applied Linguistics. She stated (Cutting, 2002, p. 62):

> Each day's recording lasted from 10 to 30 minutes; the total number of hours was seven. It was decided to focus on a small group of English native speakers to permit the researcher to become familiar enough with each of them to detect any tendencies caused by speakers' idiosyncrasies.

This type of approach, as Vaughan & Clancy (2013) pointed out, is hardly attainable with a larger corpus. The deliberate choice of working on a "small" sample can thus be motivated by the wish to produce a specialized corpus delimited by register, setting, speaker idiosyncrasy and discourse domain. One key consideration regarding corpus design is that it should be suitable for specific research purposes; while larger corpora

tend to be built for "general" linguistic phenomena, specialized corpora often target more specific research questions (Koester, 2010). This is also the case with the two corpora under study. The SITAF Corpus was originally collected to address research questions related to pronunciation and phonetic features, linguistic transfer, and L2 acquisition processes (Horgues & Scheuer, 2015). The present study targets more specifically multimodal inter-(dis)fluency phenomena in semi-spontaneous tandem settings, which motivated our choice to work on a smaller specialized sample (cf section 1.2.3.). Similarly, the DisReg Corpus was collected for the purposes of studying inter-(dis)fluency across communication settings and language styles, and the selected sample was motivated by our wish to approximately match the size of our existing SITAF sample (cf section 1.3.3.). In the following section, we further elaborate on our objective to work on two corpora with a comparable corpus design.

1.4.2. Comparable corpus design

Even though (dis)fluency rates will not be compared statistically in the two different datasets, it was still deemed important to work on similar corpora, as to ensure continuity between the two investigations. In this section, we describe the commonalities between the SITAF Corpus and the DisReg Corpus.

First, the two corpora comprise similar speaker profiles. All of them are students studying at the same university, or who at least had some experience when staying at this university (i.e. the American students who only stayed for a semester or a year). All of them are also undergraduate students, studying social sciences (i.e. with an English or French major), and they all know each other from university. Their relationship is, in fact, bound to the university, to a certain extent: In SITAF, the participants are tandem partners who met through the tandem exchange program, and in DisReg, the participants are classmates, who spent a considerable amount of time together at the university. Some of them developed close relationships, and the different degrees of proximity are reflected in their conversational exchanges. As we shall see in Chapters 4, and 5, some participants in DisReg referred to common past experiences that they shared outside university, while others mostly talked about assignments they prepared for school. During the conversation recording sessions, all the students thus had the opportunity to establish mutual understanding and common ground, which are central topics of research in the study of social interaction and language in use. These topics will be further explored in the two corpora, as well as

with other essential features of talk-in-interaction that are altogether relevant to the study of inter-(dis)fluency, such as turn-taking, intersubjectivity, co-construction of meaning, collaborative word searches, etc.

Secondly, the students were also recorded in a relatively familiar institutional setting, i.e. on campus: in a sound studio for SITAF, and in an actual classroom for DisReg. As pointed out before, (cf section 1.1., and Chapter 1, section 3.2 and 3.3.) social conversational exchanges are intrinsically grounded and embodied within their surrounding spatial and material environment. In both corpora, all the participants had access to some kind of material object, most oftentimes a piece of paper; whether it was the instructions for the argumentative task (SITAF), the lists of different topics of conversation, or the students' notes (DisReg). As we shall see (across Chapters 3, 4, and 5), the participants interacted differently with the piece(s) of paper they had within reach; some used it to deal with their language difficulties, some fiddled with it, and others pointed towards it to establish a state of mutual understanding. This diversity of behaviors will also reflect a degree of variability found within fluencemes.

One last point to consider is the nature of the data. Our two studies are neither fundamentally experimental nor naturalistic (cf section I.), but rather rely on semi-structured elicitation techniques (Eisenbeiss, 2010). The latter refer to techniques that "keep the communicative situation as natural as possible, but use interviewing techniques, videos or games to encourage the production of rich and comparable speech samples" (Eisenbess, 2010, p. 1). This is particularly relevant to SITAF, which involved speaking tasks, or games ("Liar Liar" and "Mind Games", cf section I.1.2.) to encourage participants to interact in their first and second language. The nature of the data in DisReg is slightly different, since the participants did not have to perform a specific task, and were simply "guided" by a list of topics to start the conversation. It still involved semi-structured elicitation techniques on the part of the investigator (i.e. eliciting productions on a given topic), but to a lesser extent than in SITAF. As to the recordings of students in class, the latter are perhaps closer to naturalistic techniques, since the recording situations are very close to a real-life situation (i.e. a student giving an oral presentation at the university), despite being in a highly institutional and nonpersonal setting.

To conclude, despite some differences regarding the nature of the data, the two corpora under study share a set of similar features. They both involve a certain degree of researcher control over the data, but still offer ecological validity in the sense that

they include semi-naturalistic real-life situations of students interacting within a shared institutional and social environment, the university. Additionally, given the semi-controlled and semi-structured design of the data, our contextual and multimodal analysis of inter-(dis)fluency phenomena can be conducted on relatively comparable speakers who were subjected to the same semi-structured elicitation tasks. This allows for an efficient and reliable quantitative treatment of our corpus sample, despite its relatively "small" size. As argued earlier, the present study defends the use of a "small", but specialized audio-visual dataset, in order to explore the linguistic, contextual, and interactional variables influencing the uses of ambivalent fluencemes across languages and communication settings. We shall now turn to the description of the transcription techniques used on our data.

1.5. Data transcription

As Thompson (2010, p. 98) noted, one of the first decisions the researcher needs to make when preparing transcriptions and annotations of audio-visual corpora is which transcription and spelling conventions to use. This choice is generally determined by the nature of the research and the potential uses of the corpus. The transcriptions need to be consistent with a common set of transcription conventions, which differ according to the type of data, the research discipline (e.g. linguistics, sociology, anthropology etc.), and the approach taken. But transcription is also an act of representing dynamic oral speech into fixed written words, which is ultimately interpretive and political, as Lapadat (2000, p. 204) put it:

> Verbatim transcription serves the purpose of taking speech, which is fleeting, aural, performative, and heavily contextualized within its situational and social context of use, and freezing it into a static, permanent, and manipulable form. The researcher chooses what talk to write down, and how to represent it – a choice that is both interpretive and political (Green et al. 1997).

Transcription is thus essentially about choice: what kind of unit of talk to choose, how to represent paralinguistic events, which symbols to use, etc. These choices are embedded within the transcription itself, and they leave traces of the researcher's authorship, in other words, his or her position towards the text (Bucholtz, 2000). Green et al., (1997) distinguished between the interpretive level and the representational level of transcription. The former deals with what should be

transcribed in the transcript, and the latter with how it is transcribed (e.g. how to represent nonstandard English). These two levels show further evidence of the transcriber's act of choice and selection.

With this in view, the researcher's transcription needs to be selective (Duranti, 2006; Ochs, 1979), but the selection process should not be "random and implicit" (Ochs, 1979, p. 44); it should reflect the researcher's conscious choice of his or her theoretical goals. In addition, one important feature of a transcript is that it should not contain too much information, in order to facilitate its readability for a larger scientific community. For instance, a transcript which is too detailed will be difficult to follow and assess by readers (Ochs, 1979). In sum, transcription is, to a larger extent, *theory* (Ochs, 1979), in the sense that it relies on the researcher's theoretical interpretation of the event being transcribed. When transcribing video recordings, the researcher must find ways to transcribe multimodal events in different temporalities (Kendon, 2004; Mondada, 2018), which presents a number of challenges. In the following sections, we introduce the different theoretical choices we made regarding our transcription methods for the purposes of representing fluencemes in multimodal talk. We first discuss our choice of unit (1.5.1), followed by a brief review of existing transcription conventions (section 1.5.2), and we conclude with a presentation our of transcription system (section 1.5.3).

1.5.1. Units of transcription

As briefly mentioned above, the transcription of a spoken event ultimately requires the selection of a consistent linguistic unit of segmentation. Even though talk is carried out through a continuous stream of speech and gestures, it can be segmented and broken down into smaller units at different levels, such as the level of interaction (e.g., the turn-at-talk), syntax (e.g. clause[70]) or prosody (e.g. intonation unit). For instance, Du Bois et al., (1993, p. 47) divided discourse into recognizable *intonation units*, defined as:

> a stretch of speech uttered under a single coherent intonation contour. It tends to be marked by cues such as a pause and a shift upward in overall pitch level at its beginning, and a lengthening of its final syllable.

[70] see Foster et al., (2000, p. 365) who transcribed data into speech units, consisting of "an independent clause, or a subclausal unit, together with any subordinate clause(s) associated with either".

This type of segmentation relies on supra-segmental aspects of speech, recognizable by its overall pitch movement, and it has been largely used by discourse analysts who were interested in intonation from a functional perspective. In Du Bois et al., (1993)'s paper, intonation contours are defined on the basis of their function; for instance, the final pitch movement of an intonation unit can either mark the "projection" or "continuation" of a unit of discourse, or of the turn. Du Bois et al., (1993) further spoke of "transitional continuity", which refers to "the marking of the degree of continuity that occurs at the transition point between one intonation unit and the next" (De Bois et al., 1993, p. 47). Other researchers, who further maintained this functional perspective, preferred the term "informational phrases" to refer to these units of talk. They offered a similar definition, but also further emphasized the semantic and discursive criteria used to characterize these phrases (Gumperz & Berenz, 1993, p. 4):

> The best way to characterize an informational phrase is as a rhythmically bounded, prosodically defined chunk, a lexical string that falls under a single intonation contour. Prototypically, these are set off from surrounding phrasal units by pausing and constitute semantically interpretable syntactic entities [...] in less prototypical cases, determination of phrase boundaries depends on what divisions make sense in terms of the rhythmic and thematic organization of the surrounding discourse.

Prosodic and phonological cues have thus largely been used as relevant criteria for the identification of linguistic units in spoken discourse, altogether coupled with other cues at different linguistic levels such as semantics and discourse. However, as Reed (2009, p. 353) pointed out, it is not clear how relevant these cues are with respect to the orientation and participation of co-speakers in a conversation. In fact, interactants may wish to draw on several interactional cues, such as turn-taking, when finishing off a phrase and starting another, and this is not based on intonational cues alone. Reed (2009) thus suggested the term "chunk" instead of intonational phrase, to avoid classifying these units solely on prosodic grounds. The idea of transition from one unit to the next, interpretable by taking into account the sequential context of the surrounding interaction, is central in the field of CA. The latter refer to these fundamental units of talk as *turn-constructional units* (TCU) (Sacks et al., 1974; cf Chapter 1, section III. 3.2.2). They are defined as interactional units which constitute the different building blocks of a turn-at-talk. They further provide cues regarding the

potential completion of a speaker's turn, through a "transition relevant place" (TRP): a place near the end of a turn where a turn transition is made relevant. TCUs include sentential, clausal, and lexical constructions (Sacks et al. 1974), and are altogether delimited on the basis of syntax, prosody, and pragmatics. These different cues (i.e. intonation, syntax, and pragmatics) are said to "work together and interact in complex ways" (Ford & Thompson, 1996, p. 137).

In sum, the delimitation and recognition of linguistic units in face-to-face interaction are fundamental for the purposes of discourse transcription, as well as for in-depth, sequential, and moment-by-moment analyses of talk-in-interaction. This identification relies on different linguistic cues (prosodic, discursive, interactional, pragmatic etc.), which altogether uncover complex aspects of turn construction. Another highly common unit of segmentation found across studies in linguistic research is the *utterance*, which has different definitions. For instance, Goodwin (1981, p. 7) defined it as *the actual stream of speech*, which includes a wide range of vocal phenomena such as inbreaths, laughter, crying, pauses etc. However, for clarity's sake, his identification of utterances is restricted to vocal phenomena, which includes the recognition of smaller intonation phrases within the utterance, but which excludes other pragmatic and syntactic cues. Conversely, other researchers from different research fields, such as SLA, opted for a definition that includes several other criteria, such as Crookes & Rulon (1985, p. 9)[71]:

> An utterance is defined as a stream of speech with at least one of the following characteristics:
> (1) Under one intonation contour,
> (2) Bounded by pauses, and
> (3) Constituting a single semantic unit.

Similarly, Parisse & Le Normand (2007, pp. 5–6), who worked on spontaneous language productions of two to four-year-old children, applied three main criteria to their identification of utterances, listed as follows:

(1) coherent syntactic unit,
(2) single intonation contour;
(3) bounded by a silence of at least 400 ms, or by a speaker's turn.

[71] Read Crookes (1990) for a detailed review of the utterance in second language discourse analysis.

The first two criteria are similar to the ones found in Crookes & Rulon (1985), as well as those offered by Du Bois et al. (1993) and Gumperz & Berenz (1993), which mainly revolve around syntactic, semantic, and prosodic cues. However, the issue with the last criterion is that many silent pauses in spontaneous speech last longer than 400 ms and occur in medial position, but they do not necessarily signal the end of an utterance[72], as in:

*B1: hhh. eum (0.461) qui:i est en fait assez eum **(0.737)** euh étrange[73].

This example is taken from the DisReg Corpus, during B1's oral presentation. Here the silence (in bold) is 737 milliseconds long, but it does not appear to signal the end of the speaker's utterance, as it ensures syntactic continuity between the intensifier ("assez") and its predicate adjective ("étrange"). It is also surrounded by two filled pauses ("euh" and "eum"). Moreover, if we look at the pitch pattern of the utterance (Fig 19), we can see that the adjective "étrange" is produced with a falling intonation, which further suggests that it signals the end of the intonation unit. Therefore, despite its relatively long duration, the pause in this example could hardly be used as a criterion for utterance boundary, if we take into account the other criteria considered earlier (i.e. syntactic and intonational).

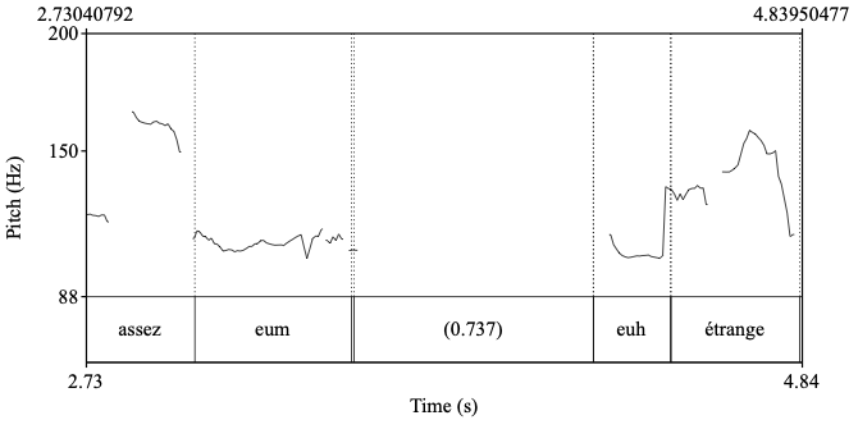


Figure 19. *Intonation contour of an utterance (PRAAT window)*

In addition, while all the criteria reviewed above focus on verbal, vocal, or pragmatic aspects of language use to identify boundaries between different units of talk, none of them pay attention to visible bodily conduct. In fact, if we look closely at the video

[72] The authors (Parisse & Le Normand, 2007, p. 6) did point out that their criteria were not absolute and could actually contradict one another.
[73] Our transcription conventions are described in section 1.5.3.

recording, we can see that when B1 remained silent for 737 milliseconds, he produced a palm-up open hand gesture oriented towards his audience (the classroom) at the exact same time (Fig. 20).

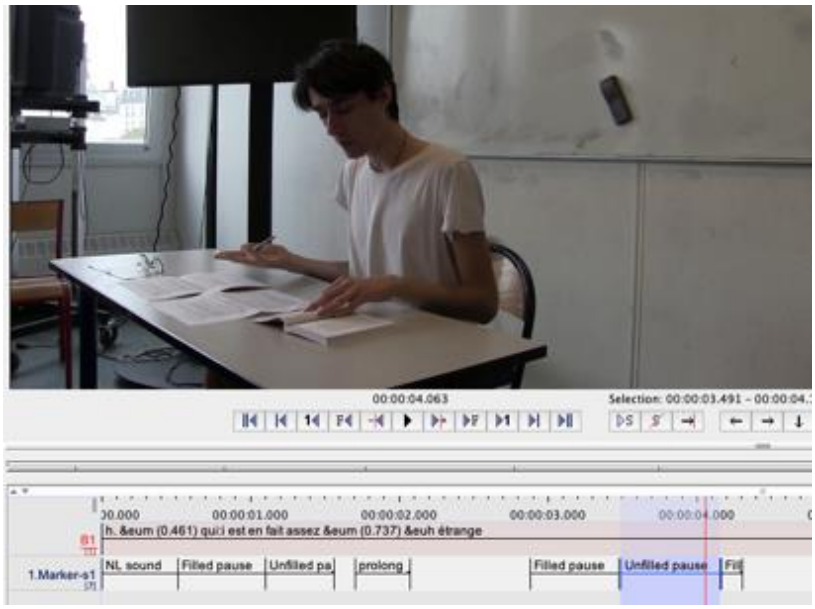


***Figure 20**. Gestural activity during a pause (ELAN window)*

This hand gesture, which can be interpreted as a pragmatic discursive gesture here (cf section II) does not signal the end of the speaker's ongoing utterance, but rather metaphorically projects an upcoming piece of information, which is later verbalized with an adjectival phrase (*étrange*); this is what Streeck calls "forward gesturing" (Streeck, 2009a, p. 161) defined as the following:

> An adaptive mechanism or method that draws on the multimodality of the communicating body to enable others to anticipate the trajectory of an action, and, thus, to facilitate interpersonal coordination.

The notions of projection and continuation, used by Du Bois et al. (1993) and Sacks et al. (1974) (among others) to refer to the continuity of a linguistic unit within discourse or a turn-at-talk, thus also needs to be applied to the multimodality of the exchange in order to successfully determine the delimitation of multimodal utterances. This was also pointed out by Debras (2013, p. 134), who chose to work on a unit called "proposition multimodale" (multimodal clause); the latter was identified using syntactic, semantic, and sequential criteria, and took into account intonational, as well as visual-gestural cues. This can be done with the help of annotation tools such as PRAAT and ELAN, which are further described in section II. In this view, a multimodal utterance is not only made of a prosodically, syntactically and semantically unified

stream or string of words, but is altogether combined with visible action, further in line with Kendon's definition (2004).[74]

To conclude, while there is no ideal unit of transcription to choose from when transcribing the complexity of multimodal discourse, it is essential to raise awareness of the issues related to transcription and segmentation methods. While intonation cues provide a fundamental prosodic unit of speech, it has been pointed out that the definition of an intonation unit was not entirely defined on prosodic grounds alone (Reed, 2009), as it can also be combined with syntactic and semantic criteria. Moreover, the recognition of these units, or chunks (Reed, 2009) is intricately bound to the exigencies of the ongoing interactional exchange (Sacks et al., 1974), which also plays a major role in their segmentation. In line with Goodwin (1981) and Kendon (2004), and further grounded within our integrated interdisciplinary framework (cf Chapter 1, section IV) we will speak of multimodal *utterances* (Cienki, 2017), and adopt intonational, syntactic, interactional, as well as visual-gestural criteria to identify them, summarized below:

(1) *Under a single coherent intonation contour* (Du Bois et al. 1992 ; Crookes & Rulon, 1985; Parisse & Le Normand, 2007)
(2) *Constitute an interpretable syntactic and semantic entity* (Gumperz & Berenz, 1993).
(3) *Include a point of possible completion (TRP) which potentially leads to speaker change* (Sacks et al. 1974).
(4) *This point of potential continuation or completion is further determined by the multimodality of the exchange, e.g. forward gesturing, gesture hold* (Kendon, 2004; Streeck, 2009).

While it is important to identify each of these different types of criteria and resources, they are not in the least mutually exclusive, as they continuously work together in complex ways (Ford & Thompson, 1993). Therefore, anything that constitutes a multimodal communicative move, whether it is a head nod, a backchannel, or an interjection, is considered as part of a multimodal utterance. It was also decided not to establish an a priori duration threshold for the pauses to identify utterances' boundaries (e.g. Debras, 2013; Parisse & Le Normand, 2007), as we have seen that silences of relatively long duration do not necessarily signal the completion of an

[74] See Chapter 1, section III. 3.3.2 and section IV. 4.2

utterance (cf Fig. 20). However, a self-break or a self-interruption on the other hand (i.e. when an utterance is interrupted by the speaker) does signal the breaking off of an utterance, as we shall see in Section II. Our choice of unit is particularly important for the present study of inter-(dis)fluency, as to uncover the different ways ambivalent fluencemes may signal the continuation (fluency), or interruption (DISfluency) of the multimodal communication flow.

1.5.2. Transcription conventions and multimodality

In the previous section, we discussed the different theoretical and methodological orientations underlying the researcher's choice of segmentation unit when transcribing multimodal talk. As emphasized earlier (section 1.5), transcribing is *a situated act* (Green et al. 1997) which involves decisions about the significance of a specific language event that needs to be foregrounded, or on the contrary that is not necessarily worth mentioning. This act of transcribing further requires the selection of a relevant transcription convention system, depending on the type of data under study, and the investigator's specific research goals. Conventions are important, because they "improve consistency within a transcript and within and across databases" (Lapadat, 2000,p. 205). In this section, we will briefly introduce some of the major and most widely used transcription convention systems[75], mainly *the Jeffersonian Transcription System* (Jefferson, 1996, 2004), as well as others that were designed specifically for multimodal transcription (e.g. Goodwin, 2010; Kendon, 2004; Mondada, 2018). In section 1.5.3., we will discuss our choice of convention, and introduce another widely used transcription format from the CHILDES and TalkBank Project (MacWhinney, 2000; MacWhinney & Wagner, 2010).

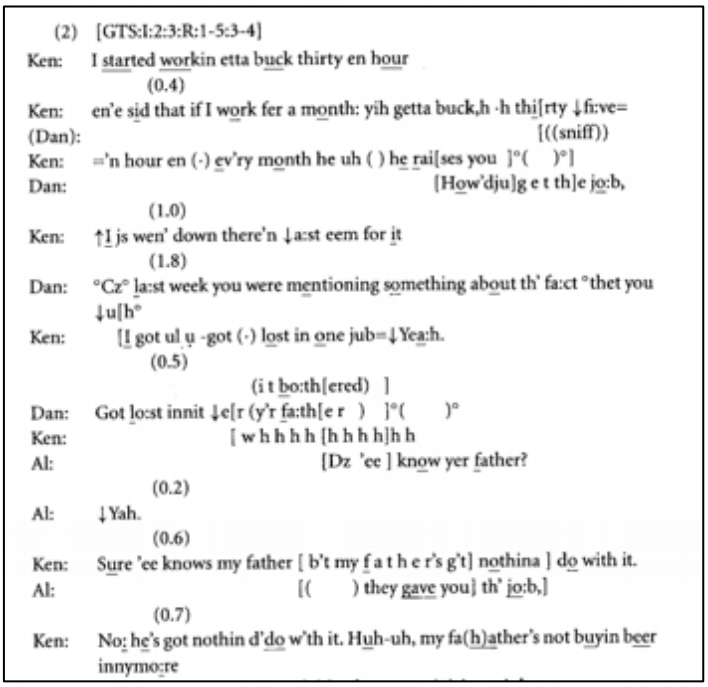
In the field of CA, the process of transcription relies on a careful observation of the sequential development of participants' actions and the deployment of other visible conduct in the course of the talk. While a detailed transcript, no matter how richly annotated, will never truly replace the actual data (Hepburn & Bolden, 2013), the goal of conversation analytic research is to develop ways of representing talk that truly capture the significant details of interactants' orderly social practices. This was first carried out by Gail Jefferson, who was a student of Harvey Sacks, one of the three founders of CA[76]. Jefferson was a pioneer of the conversation-analytic transcription

[75] There are, of course, many other transcription systems available, see for example the GAT2 system developed by Selting et al. (2009) used by Sikveland & Ogden (2012) among others.
[76] For more information, read Heritage (2009) and Ten Have (2007).

system, and she provided a systematic method for transcribing spontaneous talk with respect to overlapping talk, intonation patterns, speaking rate, voice intensity, etc. Jefferson (and to a larger extent, CA) transcription conventions mainly involve the five following aspects (Hepburn & Bolden, 2013, p. 58):

(1) Transcript layout – where speakers are identified on the page, line numbers.
(2) Temporal and sequential relationships – overlapping talk; gaps, pauses, etc.
(3) Speech delivery and intonation – unit-final intonation, volume, pitch variations, voice quality etc.
(4) Transcriptionist's comments and uncertain hearings—description of extralinguistic events.
(5) Features accompanying talk – aspiration, laughter, cries.

Specific attention is thus paid to a scope of phenomena, whether it is the sequential aspect of the talk (1), some of the extralinguistic features accompanying it (5), or information regarding intonational features of the speech delivery (3).



(2) [GTS:I:2:3:R:1-5:3-4]
Ken: I started workin etta buck thirty en hour
(0.4)
Ken: en'e sid that if I work fer a month: yih getta buck,h ·h thi[rty ↓fi:ve=
(Dan): [((sniff))
Ken: ='n hour en (·) ev'ry month he uh ( ) he rai[ses you]°( )°]
Dan: [How'dju]g e t th]e jo:b,
(1.0)
Ken: ↑I js wen' down there'n ↓a:st eem for it
(1.8)
Dan: °Cz° la:st week you were mentioning something about th' fa:ct °thet you ↓u[h°
Ken: [I got ul ų -got (·) lost in one jub=↓Yea:h.
(0.5)
(i t bo:th[ered)]
Dan: Got lo:st innit ↓e[r (y'r fa:th[e r)]°( )°
Ken: [w h h h h [h h h h]h h
Al: [Dz 'ee] know yer father?
(0.2)
Al: ↓Yah.
(0.6)
Ken: Sure 'ee knows my father [b't my f a t h e r's g't] nothina] do with it.
Al: [( ) they gave you] th' jo:b,]
(0.7)
Ken: No: he's got nothin d'do w'th it. Huh-uh, my fa(h)ather's not buyin beer innymo:re

Figure 21. *Example of a transcript made by Jefferson (Jefferson, 2004, p. 15)*

Figure 21 further illustrates an example of an actual transcript made by Gail Jefferson, following specific transcription conventions, to name but a few: underlined syllables show emphasis; arrows indicate a rise or a drop in intonation; square brackets show where speech overlaps; colons indicate a stretched sound; the equal signs illustrate an

instance of "latching" (e.g. Schegloff, 2000) when two speaker's units are latched together; number in parentheses indicate pauses represented in tenths of a second.

While this transcription is extremely detailed and pays attention to a variety of features (as opposed to orthographic transcriptions that are more concise and succinct) it is still, as Jefferson (2004, p. 15) said so herself, "a nightmare". Indeed, this transcript is not easily readable at first glance for the common reader, as it contains a great deal of special characters and symbols. In fact, Jefferson (2004, p. 15) raised an interesting question on the matter:

> Why put all that stuff in? Well, as they say, because it's there. Of course there's a whole lot of stuff "there", i.e., in the tapes, and it doesn't all show up in my transcripts; so because it's there, plus I think it's interesting. Things like overlap, laughter, and "pronunciational particulars" (what others call "comic book" and/or stereotyped renderings), for example. My transcripts pay a lot of attention to those sorts of features.
>
> What good are they? I suppose that could be argued in principle, but it seems to me that one cannot know what one will find until one finds it, so what I'll do is show some places where attention to such features turned out to be fruitful.

Her last point is particularly interesting. The goal of discourse transcription is to make specific phenomena that are of interest to the researcher noticeable on a written transcript, so it becomes readily accessible to the reader. However, it is virtually impossible to give an account of all (extra)linguistic phenomena, so the researcher needs to select what specific aspects of the talk are deemed relevant, depending on their research goals. Here Jefferson chose to focus on pronunciational and sequential features (i.e. laughter, alternative pronunciations, and overlaps), but she completely omitted the description of visible bodily conduct, such as hand gestures, body movement, or direction of gaze[77]. This issue was raised by several authors, such as Mondada (2018), Bezemer & Mavers (2011), or Ayaß (2015), among others. Ayaß (2015) discussed the methodological status of transcription in CA, and compared different multimodal transcription systems, with for instance Goodwin (2010), who inserted images as well as visual representations within his transcripts (Fig. 22).

[77] This is probably because her transcript is taken from an audio recording. Schegloff and Sacks started to work on filmed data in the 1980s (cf Mondada, 2018, p. 86), and this transcription was initially done in 1964 (Jefferson, 2004, p. 14). The status of transcription is thus inevitably determined by the nature of the data.

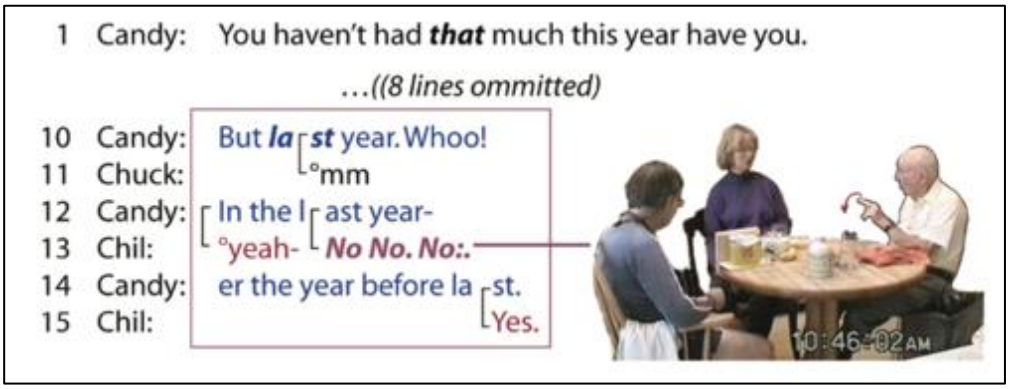


***Figure 22**. Excerpt from Goodwin's "Multimodality in human interaction" (2010, p. 89)*

As shown in Figure 22, the author focused on a specific language practice, where interactants are negotiating and co-constructing meaning with one another. This excerpt includes an actual picture of the exchange, illustrating the body, gaze, and gestural behavior of the participants. Several lines of the transcription are omitted, but the events are retold by the author in his analysis (Goodwin, 2010, p. 88). The author thus chose to focus specifically on the sequential development and the multimodality of this particular exchange, rather than on prosodic features like Jefferson in Fig. 21. In a similar vein, Mondada (2018) developed a specific system of conventions for the transcription of multimodal practices in CA. In her paper, she discussed several technical and practical issues surrounding transcription choices, and provided several solutions as to how the richness of video data could be further exploited through multimodal transcription. As we have seen in the previous chapter (cf Chap. 1, section III.3.3), the notion of multimodality revolves around the mobilization of multiple semiotic resources in different modalities and temporalities within the ecology of a situated activity. Therefore, it can be challenging to come up with a standard set of conventions for multimodal transcripts, as they cannot solely be based on orthographic conventions which characterize the linearity of speech. Multimodal transcripts must be able to illustrate two fundamental aspects of visible conduct, mainly: (1) the temporality of a gesture, that is, its unfolding in different gesture phrases (Kendon, 2004) from preparation to recovery; and (2) their shape and movement (Mondada, 2018, p. 90). This is illustrated in the two figures below:

Figure 23. *Multimodal transcription taken from Mondada (2018, p. 90)*

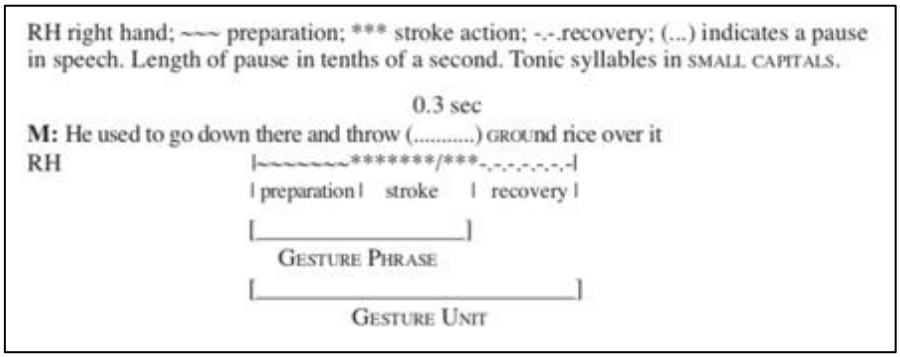


Figure 24. *Multimodal transcription taken from Kendon (2004, p. 114)*

In Fig. 23, specialized symbols are used to delimit the beginning and end of the gesture (the + sign), its preparation (indicated by …), and its withdrawal (indicated by „„). It also includes a concise description of the gesture (e.g. "points eggs"), accompanied by a picture. This is very similar to Kendon's (2004) transcription techniques (Fig 24) who focused specifically on the timing of gesture phrases within gesture units (cf Chap. 1, section IIII, 3.3.). As we can see, these transcripts are much shorter than the one introduced earlier by Jefferson (Fig. 21). Instead of drawing from longer transcripts in order to illustrate the overall sequential development of the talk and several pronunciation features (as in Jefferson's Fig. 21), the multimodal transcripts shown above (Fig. 22, 23 and 24) zoom in on selected pictures of particular instances, thus further reflecting the authors' own specific research interests (Ochs, 1979). Similarly, we will adopt a comparable approach in our qualitative analyses, and focus more specifically on the temporal development of fluencemes and co-occurring gestures (cf Chapter 3, 4, and 5).

In sum, several authors have used specialized notational systems to accurately illustrate the temporal development of relevant bodily actions performed in the course of interaction (see Goodwin, 2010; Kendon, 2004; Selting et al., 2009; Sikveland &

Ogden, 2012, among others). These fundamental features of multimodal talk reflect the researcher's authorship and act of transcription, which is primarily a selective and dynamic process. However, a highly annotated transcript which contains too much information can be difficult to assess (Ochs, 1979), so the researcher also needs to be conscious about the readability of his or her transcription. In the next section, we further discuss our theoretical and methodological choices regarding our transcription format.

1.5.3. Transcribing fluencemes in multimodal talk: a dynamic process

As Lapadat (2000, p. 205) rightly pointed out: "researchers' transcription systems need to reflect their data and their purposes, hence different approaches to doing transcription have arisen." She further commented on her own work, and described how her transcription methods evolved progressively, depending on the type of research she was conducting. Transcribing is thus a dynamic process; even if the ultimate representation of a transcript remains fixed on a page, transcriptions can change over time, as Duranti (2006, p. 307) noted in his beautiful paper entitled *Transcripts, Like Shadows on a Wall*:

> More generally, transcripts have a life or rather, we give them a life. Transcripts are born, get longer and fatter, and change in character, sometimes through our revisions, other times by simply sitting in a drawer for a few years. When we pick them up, they read differently. We should be aware of these changes and thematize them. The different versions or interpretations of the same transcript provide us with a record of our epistemological and theoretical changes. In some cases, we might even call these changes "progress".

Transcripts change and transcription methods change too; this happens because the transcriber's research interests shift, or evolve over time, depending on his or her theoretical orientations. This was also our case, as our theoretical and methodological orientations towards inter-(dis)fluency phenomena have constantly been reconsidered in the past 4-5 years. At the early stages of our preliminary work on (dis)fluency (cf Kosmala & Morgenstern, 2017), we targeted a transcription software that would accurately transcribe and annotate fluencemes in spontaneous speech: the CLAN (Child Language Analysis) program (MacWhinney, 2000; MacWhinney & Wagner, 2010). Originally applied to child language, CLAN was designed by Catherine Snow and Brian MacWhinney who developed a system for sharing language-learning

data, known as the CHILDES (Child Language Data Exchange System) in 1984 (MacWhinney, 2001). The CLAN software has now been used from the past two decades by a number of researchers in a growing scientific community, following the creation of the Talkbank system in 2002[78]. Talkbank was a follow-up of CHILDES, and offered a data archiving system for transcribed video and audio data. The goal of Talkbank was to provide a common framework for data sharing, transcription methods, and analysis in four major disciplines (classroom discourse, animal communication, field linguistics, and computational analysis; see MacWhinney, 2001, p. 7). They also added more specialized research groups, including CA, Gesture, L2 learning, Corpus Linguistics, and disfluency production (see MacWhinney, 2001, pp. 9-10). MacWhinney further developed a specific manual, called CHAT (*Codes for the Human Analysis of Transcripts*), which introduced different computational tools and conventions required for the CHAT transcriptions. In the first volume of the manual, several sections are dedicated to the transcription of retracing phenomena, self-interruptions, and prosody within words (MacWhinney, 2000, pp. 61-75), which are all part of (dis)fluency phenomena. These conventions have been used by several researchers in (dis)fluency research, especially in child language acquisition (cf Didirkovà et al., 2019; Dodane et al., 2016) and stuttering (Ratner et al., 1996). In sum, the data-sharing initiatives of the CHILDES and Talkbank systems prompted several researchers in different fields to use the CLAN software with CHAT transcription conventions to transcribe their video recordings (e.g. Beaupoil-Hourdel, 2015, p. 20; Debras, 2013; Morgenstern & Parisse, 2007, 2012; among others). This further motivated our choice to work with CLAN, using a similar set of conventions (cf Table 8). One of the main benefits of using unified transcription conventions is, as explained earlier (cf section 1.5.2), that it facilitates its readability and interpretability for the scientific community. This was particularly important for the transcription of fluencemes, which have been assigned a specific set of codes and symbols in CHAT[79]. Additionally, another benefit of using CHAT conventions in CLAN is that it allowed us to run several commands, such as frequency and MLU, which was very practical for our quantitative treatments (see section II.2.3.2.).

[78] For more information, read https://www.talkbank.org (last retrieved on August 28th 2021)
[79] It should be noted that a number of annotation systems have also been created specifically for the transcription and annotation of disfluency, see for example the *DisMo* Project (Christodoulides et al., 2018), the *DUEL* project (Hough et al., 2016),or *The Dysfluency Annotation Stylebook for the Switchboard Corpus* (Meteer et al., 1995).

The first recordings of the SITAF Corpus were thus transcribed with CLAN, following CHAT transcription conventions. Some video recordings had already been transcribed by the annotators of the SITAF project, using a different software called *Transcriber* (Barras et al., 2001). The existing transcriptions were thus exported to a CLAN format[80], and adapted to CHAT transcription conventions, in order to establish a standardized transcription system for all the recordings. In addition, all our recordings were analyzed with an annotation software called ELAN (Sloetjes & Wittenburg, 2008), which is further described in section II. Our recordings have thus been transcribed and annotated with different annotation and transcription tools, which further invited us to reconsider the methodological status of our transcription. While CHAT transcription conventions target a variety of sequential, syntactic, vocal, as well as gestural phenomena (i.e. overlaps, utterance terminators, trailing offs, creaks, pauses, gestures etc.), they include a lot of special symbols and codes, which are necessary within the software to run the commands, but which are not easily readable in a finalized transcript. For instance, the transcription of a simple event or a gesture requires the use of an amperstand followed by an equal sign, such as *&=reads*, or *&=head:shake, &=eyes:open*, which, in CA, is plainly transcribed with double parentheses, e.g., *((smiles))*. Since fluencemes are already transcribed with specific symbols (cf Table 8), we did not want to impede the readability of the finalized transcripts[81] by adding more codes. Additionally, the specialized notation systems for multimodal transcripts introduced earlier (cf section 1.5.2.) cannot easily be inserted within a CHAT transcription in CLAN, so a simplified version of their system was also added to our transcription conventions. Lastly, some vocalizations, such as audible inbreaths and tongue clicks, which are very frequently transcribed in CA (e.g. Hoey, 2014; Ogden, 2018; Wright, 2011) do not appear on the CLAN manual, so they were also included in our conventions.

To conclude, our transcripts went through various technical, theoretical, and methodological changes, which ultimately reflects the development of our interdisciplinary multi-level and multimodal approach to inter-(dis)fluency. While a transcript will always require selectivity, and never truly replaces the actual data, we

[80] This was done thanks to the TEICONVERT program: https://ct3.ortolang.fr/teiconvert/index-en.html (last retrieved on August 25th 2021)

[81] By finalized transcript, we mean the final version of the transcript shown in our qualitative analyses, which is different from the initial transcript we used to transcribe with CLAN or ELAN. For instance, filled pauses are typically transcribed with an amperstand (&uh &um &euh) in CLAN, because it allows them to be ignored as words when running the MLU and freq commands (MacWhinney, 2000, p. 65).

sought to provide a set of conventions which maximally reflected our theoretical orientations. Our conventions, summarized in Table 8, are mostly based on the CHAT manual, but they also borrow a few symbols from CA. Information about voice intensity, intonation, pitch movement, stress, and the like (cf Fig. 21) was not transcribed, as the main focus of our work is on inter-(dis)fluency phenomena and multimodal events, so we wished to keep our transcripts as specific as possible. In addition, we chose Kendon's (2004) gesture annotation system, because it has been widely used (e.g. Debras, 2017; Sikveland & Ogden, 2012) so it may be more easily recognized by members of the gesture community.

Table 8. *Transcription conventions used in this thesis*

CHAT conventions (MacWhinney, 2000, pp. 48-74)	
+/	interruption by other participant
+//	self-interruption
[/]	word repetition
[//]	self-repair
+...	trailing off
(0.250)	unfilled pause (number in milliseconds)
wo:rd	prolonged vowel or consonant
+< <>	overlapping talk
(a)bout	shortenings
+/ +"/.	quoted utterance
xxx	unintelligible words
CA conventions (Jefferson, 2004; Ogden, 2018, 2013)	
[!]	tongue click
.hhh	inbreath
hhh	outbreath
creaky	creaks
(( ))	description of events, or analyst's comment
Gesture annotation (Kendon, 2004)	
~ ~ ~	preparation of gesture stroke
***	gesture stroke
<u>***</u>	hold
-.-.-	return to rest position

However, information concerning the deployment of gestures within gesture phrases is not systematically included in our transcripts, depending on the kind of examples we are presenting. Several of them also include pictorial illustrations and a succinct

written description of gestures, to give a sense of their composition and the ecology in which they happen (cf Mondada, 2018, p. 90).

II. Annotation protocol for the quantitative analyses

In the previous section, we described how the data was acquired (for SITAF) and produced (for DisReg) and discussed the challenges surrounding transcription techniques. We shall now turn to the description of the management of the data, how it was analyzed, treated, and converted to work on different interfaces (data interoperability). Our data management plan is summarized in the figure below[82]. As mentioned throughout this dissertation, the present study of inter-(dis)fluency phenomena combines quantitative and qualitative methods, further reflecting our mixed-methods approach and integrated theoretical framework. Our goal is twofold: (1) to explore the overall distribution of fluencemes across languages and settings, in line with previous corpus-based studies and psycholinguistic work on (dis)fluency, and (2) to zoom in on selected social practices in interaction to highlight their pragmatic and multimodal dimension, in line with conversation-analytic approaches.

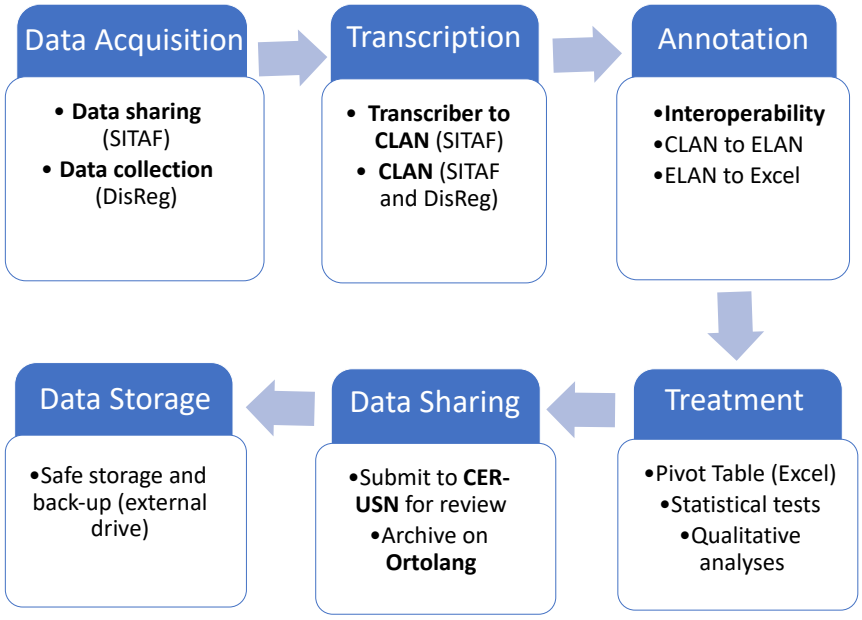


Figure 25. *Data Management Plan*

[82] CER-USN refers to *Le Comité Ethique de Recherche de l'Université Sorbonne Nouvelle*. It should be noted that the description of the study procedures (cf 1.3.1 and 1.2.1.) should have been submitted to CER-USN prior to the video recordings, but it had not been created at the time. The CER was only established in 2021.

In this section, we describe the methods and tools used for our quantitative treatments more specifically.

2.1. Early versions of the annotation protocol

It should be noted that our annotation protocol underwent several important technical and methodological changes[83], some of which are concisely described in this section. Several categories used in the initial version of our annotation scheme (Kosmala & Morgenstern, 2017) were questioned by ourselves and by members of the scientific community, which invited us to reconsider their categorization. Consequently, several categories had to be adjusted, while new ones were added, and others entirely removed. These changes are addressed in the list below.

- *Fluenceme sequence*: The notion of sequence (i.e. combination of markers), borrowed from Crible et al., (2019) was already used as a category in the early version of our protocol, but the different types of sequences were distinguished with respect to their duration and complexity. We initially had three categories, "brief", "elongated", and "complex". "Brief" and "elongated" referred to sequences containing only one marker, and they were distinguished on the basis of their duration (with a minimum duration threshold of 600 ms for the brief sequence category), and complex sequences referred to combined fluencemes, i.e. immediately adjacent fluencemes (following Shriberg, 1994). However, the main issue with the "brief" versus "elongated" categories is that they are distinguished using a duration threshold, which can be highly subjective (i.e. how long is a pause supposed to last to be considered "brief"?). In fact, the most relevant distinction to make in this case is whether fluencemes appeared isolated versus combined, so their duration should not conflate with their complexity[84]. The categories were thus later changed to "simple" versus "complex" (cf section 2.2.2.), without making an a priori distinction on their duration.
- *Fluenceme duration*: In a similar vein, in the first version of our coding scheme, the duration of the sequences was annotated (in ms), by taking into account all the markers which co-occurred within the same sequence. However, this implied

[83] Not to mention terminological changes, with for instance the term *(dis)fluency marker* (instead of *fluenceme*), which has been used in most of our papers. At the very early stages of our work, we also preferred the term hesitation (cf Kosmala & Morgenstern, 2017) over (dis)fluency.
[84] This issue was raised by Ralph Rose during the DiSS 2017 workshop.

looking at both vocal markers (e.g. pauses, lengthening etc.) and morpho-syntactic markers (e.g. repetitions, repairs, interruptions, etc.); but the latter cannot be analyzed with respect to duration, because they involve morpho-syntactic changes (e.g. a truncated word, an interrupted utterance)[85]. Therefore, our initial analysis of the whole sequence's duration (which can contain both vocal and morpho-syntactic markers) is not entirely valid, and can lead to faulty results. It was thus decided to only measure the duration of individual vocal markers, and describe the length of the sequence separately (e.g. the number of fluencemes found within a sequence)[86].

- *Minimum duration threshold for silent pauses*: In line with Derwing et al., (2009) and Tavakoli (2011),we had initially set a 400ms minimum duration threshold for the identification of silent pauses in our data. However, this choice was later criticized[87], and it soon became clear that this threshold was too high for our type of study, given recent work and development on pausing research in SLA (e.g. Kahng, 2014). We thus later opted for a lower threshold (250 ms), in line with De Jong & Bosker (2013), among others, and re-annotated all silences in the data. This point is further developed in section 2.2.1.
- *Presence of inbreaths and clicks within sequences*: In line with most studies on (dis)fluency (e.g. Lickley, 2015; Maclay & Osgood, 1959; Merlo & Mansur, 2004; Shriberg, 1994, etc.) we initially only targeted "typical" fluencemes, which have been widely recognized in the disfluency research literature (cf Chapter 1, section I.1.3). However, it later came to our attention that other vocalizations, such as tongue clicks (e.g. Ogden, 2018; Wright, 2011) and audible inbreaths (e.g. Hoey, 2014) very often co-occurred with fluencemes, and should not be ignored[88], despite not being traditionally categorized as fluencemes[89]. The latter were thus added to our category of fluencemes, and the data was annotated a second time to include them in our analyses.

[85] This further prompted us to distinguish between vocal versus morpho-syntactic markers (cf section 2.2.1.)
[86] A similar comment was also made by Ludivine Crible, who advised not to conflate sequence configuration with length, see section 2.2.2.
[87] This issue was raised by Jürgen Trouvain at the SEFOS Conference (Kosmala, 2019).
[88] This was pointed out by Marina Cantarutti after our talk at the GESPIN conference (Kosmala et al. 2019) who noticed an audible inbreath in the excerpt we were showing, which had been completely left out from the analysis.
[89] In fact, Clark & Fox Tree (2002) did include tongue clicks in their category of collateral signals (cf Chap. 1, section 2.2.) This is further developed in section II. 2.2.1 of this chapter.

- *Annotation of discourse markers*: In line with Crible (2018), who included discourse markers in her category of fluencemes, we initially decided to annotate discourse markers in the early versions of our protocol. However, Crible's analysis of discourse markers is the result of in-depth collaborative work on a fine-grained quantitative annotation model, which has been thoroughly developed by a group of annotators over the years (e.g. Crible & Degand, 2019; Crible et al., 2019). The annotation of discourse markers within a corpus-based (dis)fluency framework thus requires a comprehensive knowledge of their model which took several years to be implemented. We initially tested their model on a small sample of SITAF as part of a pilot test, in order to evaluate the feasibility of the annotation. The data was later coded by a second annotator, but we failed to achieve intercoder reliability (only 30% of agreement). After hours of discussion, we eventually reached intercoder agreement[90], but it was eventually decided not to annotate discourse markers in the rest of the data under study, as it would have been too time-costly to perfectly acquire the annotation techniques of their model to carry out all the analyses. Therefore, we came to the conclusion that it would be more relevant to focus on our own annotation scheme and consolidate it, based on our own theoretical and methodological orientations.
- *Annotation of functions*: In line with the disfluency literature, we initially assigned different functions to fluencemes and included them in our annotation grid, such as "planning" (e.g. Tottie, 2014), "reformulating" (Hieke, 1981) "displaying uncertainty" (Smith & Clark, 1993). However, these categories were soon criticized for being created on an ad hoc basis[91]. After several attempts, we managed to create a new labeling system with novel categories and a set of criteria (cf Appendix 2), and checked for intracoder reliability. Approximately 15% of the data was coded twice by the same annotator within a year interval, and received a Cohen's Kappa coefficient of 0.68 and 78% of agreement. While the overall percentage was fairly high for some of the categories (e.g. 85% for *speech management, discursive, and uncertainty*) it was quite low for one of them (51% for *interactive/communicative*). A few months later, another 15% of the data was coded a second time by a second annotator to check intercoder reliability, but Cohen's Kappa was still relatively low

[90] The annotated data was later used for a small study on filled pauses and discourse markers (Kosmala & Crible, 2021)
[91] This was pointed out by Petra Wagner after my talk in their lab in 2018, who claimed that it was very difficult to tell whether a speaker was intentionally using a fluenceme for "planning" purposes.

(**κ**=0.68), with only 60% of agreement for the *interactive/communicative* category. Given the non-lexical nature of fluencemes, and the fact that their use is essentially shaped by context, it can be highly difficult to annotate their functions at a quantitative level, as they cannot easily be categorized with a limited set of categories. This requires an in-depth analysis of the context, followed by a discussion during a *data session* (cf section III.3.1). Therefore, it was later decided to use the categories introduced by Allwood et al., (1990), *own communication management* (OCM) and *interactive communication management* (ICM) (cf Chapter 1. Section II. 2.2.1), which are more easily distinguishable. Indeed, 15% of the same data was annotated a third time by the same two annotators, using these new categories, and resulted in a Kappa score of 0.78. This score would be characterized as "substantial" agreement (cf section II. 2.3.1.).

- *Gesture annotation*: The first version of our annotation protocol was rather incomplete in terms of gesture analysis, as we initially solely focused on the annotation of gesture phases during fluencemes. While this kind of annotation is relevant to study the relationship between speech suspension and gesture suspension, it only gives a partial picture of the phenomena. Therefore, it was later decided to analyze not only gesture phrases, but also gesture types[92] (cf section 2.2.3) to get an idea of the types of gestures that most frequently co-occurred with fluencemes. In addition, gestures were not only annotated during "disfluent" stretches of speech (i.e. during fluencemes), but during fluent stretches[93] as well, in line with Graziano & Gullberg (2018).

To conclude, multiple changes have been adopted over the past four years to obtain the final version of our annotation scheme. These changes were motivated by our wish to create a consistent and robust quantitative model which could later be replicated on a larger dataset. We also sought to create a model which was suitable for our own research purposes, which further justifies the deletion of old categories, or addition of new ones, in order to better adjust them to our model. Many of these changes are also presented as a result of several discussions with members of the scientific community,

[92] This was first suggested by the reviewers of our GESPIN paper (Kosmala et al., 2019) who asked us to give information about gesture types for our final version.
[93] This was suggested multiple times by Aliyah Morgenstern in the course of my PhD; this was also pointed out by Susanne Fuchs at the GESPIN 2019 conference, who asked what types of gestures the participants produced other than during fluencemes.

through conference talks, paper reviews, and inter-coder reliability tests. We shall now turn to the description of our finalized annotation scheme.

2.2. (Dis)fluency annotation

In this section, we describe the final version of our annotation scheme designed for quantitative treatments. The annotation of (dis)fluency required different levels of analysis, mainly (1) the individual fluenceme level, (2) the sequence level, and (3) the visuo-gestural level, as illustrated in the figure below. The fluenceme abbreviations (e.g. UP, IR etc.) are provided in the next section. It should be noted that Fig. 26 only illustrates the utterance level of (dis)fluency (including verbal, vocal, and visual-gestural dimensions), but it does not take into account the interactional level, which will be further accounted for in the qualitative analyses (cf section III). Once again, this further motivates our choice to carry out qualitative analyses in tandem with quantitative treatments.

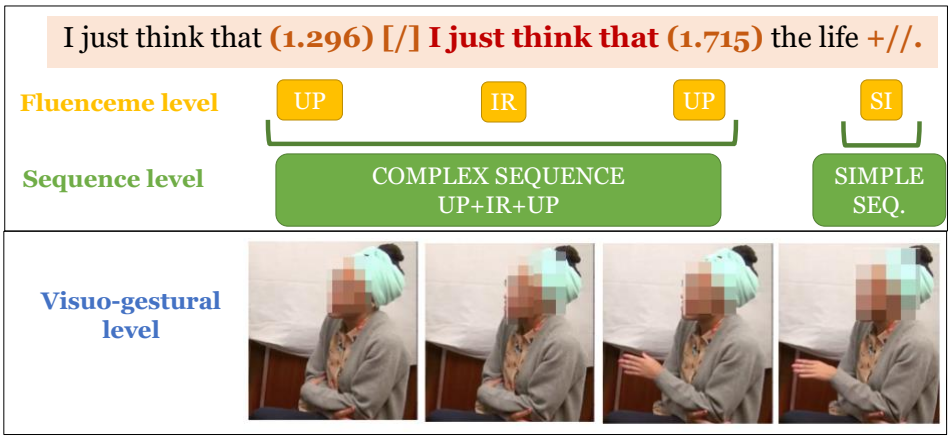


***Figure 26**. Multi-level illustration of (dis)fluency (utterance level)*

2.2.1. Fluenceme level

As explained in the previous chapter (Chap 1, section I. 1.3; and section IV.4.3), the present empirical and corpus-based study of inter-(dis)fluency is based on earlier work in disfluency research, psycholinguistics, SLA, and corpus-based linguistics. Our fluenceme classification is adapted from the ones used by number of experts in the field, such as Shriberg (1994), Crible et al. (2019), Lickley (2015), Candea (2000), Ginzburg et al. (2014), Eklund (2004), Pallaud et al. (2019), Götz (2003), Meteer et al. (1995) among others. However, unlike most previous work on disfluency phenomena, we did not make any *a priori* judgement on the potential "disfluent" functions of fluencemes to annotate them (i.e. distinguish between "significant" and "non-

significant" repetitions, cf Maclay & Osgood, 1959), and we mostly used acoustic (i.e. duration) and morpho-syntactic criteria for their identification. Therefore, their annotation was initially made on formal grounds, i.e., by relying exclusively on their form, from a production perspective. This is the first step of our analysis (fluenceme level). In line with Guaïtella (1993) who spoke of "vocal hesitations", and Pallaud et al. (2019) who studied "morpho-syntactic markers", we distinguished between *vocal fluencemes*, which have perceptible vocal and acoustic features (e.g. decreasing of pitch, diminution of glottal pressure, lengthening of a vowel), and *morpho-syntactic fluencemes,* which mark a break or an interruption in the syntagmatic channel. This distinction was preferred over "covert" and "overt" (Levelt, 1989), or "forward-looking" versus "backward-looking" (e.g. Ginzburg et al., 2014) since repetitions are typically included in the *covert/forward-looking* category (along with pauses, lengthening, and the like), while we believe that a repetition and a pause are conceptually and formally different, and should not belong to the same category. In addition, we also annotated non-lexical vocalizations, also known as *non-lexical sounds*, or *liminal signs*, (e.g. tongue clicks, inbreath, laughter, creaky voice, sigh) in line with Ginzburg et al., (2004), Wright (2011), Ogden (2018), Ward (2006), and Dingemanse (2020)[94]. These markers, which have not typically been categorized as "disfluency" markers in the disfluency literature, were annotated whenever they appeared within a fluenceme sequence[95], in the exception of inbreaths and tongue clicks, which were also annotated outside sequences. For this reason, non-lexical sounds were included in the *peripheral marker* category (in line with Crible, 2018), which refers to markers that very often occur within the vicinity of fluencemes. In this category are also included *explicit editing terms* (Shriberg, 1994). A total of 9 different fluencemes were identified, which are all listed below. They are marked in bold in the examples provided within the list, and their transcription follow the CHAT convention format described in section I.1.5.

[94] Filled pauses are also typically included in the category of non-lexical sounds, but since they are also widely known as vocal hesitations, we preferred to include them in the vocal fluenceme category. Non-lexical sounds are, to a larger extent, vocal markers too; but they are not generally recognized as typical "disfluency" markers, and we wanted to remain consistent with previous classifications in disfluency research.

[95] We chose to annotate them only within fluenceme sequences as they are very frequent in speech and their analysis would require a different type of investigation. We made an exception with clicks and inbreaths because they co-occur very frequently together, so we conducted a study that targeted these markers specifically (Kosmala, 2020b).

VOCAL MARKERS (VOC) (Guaïtella, 1993)

The duration of each vocal marker was annotated in milliseconds with the software ELAN (cf section 2.3). In addition, filled pauses were distinguished on the basis of their form, whether they were realized with a central vowel ("uh"/"euh") or with a nasal consonant ("um"/"eum")[96].

1. ***Filled pause (FP)*** (Clark & Fox Tree, 2002; Candea et al., 2005; Rose, 1998; Vasilescu & Adda-Decker, 2007) also known as "autonomous fillers", they are defined as "the insertion at any moment within spontaneous speech of a long and stable vocalic segment, defined as a type of filler" (Candea et al., 2005, p. 10). They have also been called filled pauses in contrast to silent pauses (Goldman-Eisler, 1968; Maclay & Osgood, 1959). They usually consist of a centralized schwa vowel [ə] and a nasal variant [əm] in English (Clark & Fox Tree, 2002). In French, they mainly consist of a central vowel [ø] (Duez, 2001a), but the nasal variant can also be realized [øm]. Back-channeling, such as "uh-uh" or "mhm" is not included in this category. The term *filled pause* is adopted in this thesis, to avoid ambiguity with the term *filler* that can also refer to filler syllables in language acquisition (cf Peters, 2001).

   like when I was a teenager my life was very **um** (1.429) +//.
   (SITAF Corpus, speaker A18)

   les réseaux sociaux c'est **euh** tout ce qui est **euh** facebook.
   (SITAF Corpus, speaker F07)

   je le regarde toujours **eum.**
   (DisReg Corpus, speaker E1)

2. ***Prolongations (PR):*** also known as lengthening, or drawl (Betz & Wagner, 2016; Clark, 2006; Eklund, 2001; Lickley, 2001; Merlo & Barbosa, 2010; Rohr, 2016; Fox Tree & Clark, 1997). Instances of syllable or word prolongations, resulting in above-average syllable and word duration (Betz & Wagner, 2016, p. 1). *Disfluent lengthening* is often distinguished from *phrase-final lengthening* which is used as a cue for perceiving phrase boundaries (Betz & Wagner, 2016), while

[96] It should be noted that no phonological distinction was made between English-sounding *uh/um* and French-sounding *euh/eum*, even though some speakers transferred their own pronunciation of *(e)uh/m* in their second language. The goal here was simply to distinguish between the two variants, but not to elaborate on pronunciational features. The distinction between *uh* and *euh* is only used for transcription clarity and coherence (i.e. one is more specific to English and the other one to French).

disfluent lengthening is said to be used with no prosodic intention (Merlo & Mansur, 2004). These types of lengthening can also modify the pronunciation of certain items, such as "the" pronounced with a non-reduced vowel [i] (Fox Tree & Clark, 1997). In line with Rohr (2016), and Eklund (2001) we adopt the term *prolongation* in the present thesis. No minimum duration threshold was adopted for our identification of prolongations, and they were identified entirely based on our own auditory judgment, without making an a priori distinction between "disfluent" and "phrase-final"[97].

et donc t'as tout ç**a:a** t'**a:as** après [/] t'as des intrigues plus larges.

(DisReg Corpus, speaker F1)

becau**s:s**e if they (a)re going towards the teache**r:r**s which I am assuming is what they mean.

(SITAF Corpus, speaker A02)

o:on [/] on dirait neuf peut-être.

(SITAF Corpus, speaker A07)

3. ***Unfilled pauses (UP),*** also known as silences or silent pauses (Campione & Véronis, 2002; Cenoz, 1998; De Jong, 2016b; Duez, 1982; Eklund, 2004; Maclay & Osgood, 1959; Nicholson, 2007)[98], they are defined as "silent periods between vocalizations" (Cenoz, 1998). "Hesitation" pauses are often distinguished from "articulatory" pauses (Goldman-Eisler, 1968) e.g. physiological micro-pauses occurring before plosives (De Jong & Bosker, 2013). Researchers have opted for different minimum duration thresholds to exclude articulatory pauses from their analyses, ranging from 180 ms (Duez, 1982), 200 ms (Candea, 2000; Kormos & Denès, Candea) to 400 ms (Derwing et al., 2009; Tavakoli, 2011). Others, on the

[97] We are aware that our method for identifying prolongations (i.e. purely based on perception) is limited to a certain extent, as it can be very difficult to tell exactly when a segment is prolonged (see Eklund, 2001). However, this subjective identification can still be effective as Rohr (2016) argued: the human ear is in fact said to be quite efficient at making predictions based on a speech performance (read Rohr, 2016, p. 72 for more details).

[98] This category is particularly difficult to define, because there are many issues concerning the detection and classification of silences, see Betz (2020, p. 15) for review. Pauses can also serve many different functions, such as marking a syntactic boundary, or gaining time during planning etc. (cf Cenoz, 1998; Nicholson, 2007). It is thus very difficult to tell whether pauses reflect "disfluency" or not (Eklund, 2004). For this reason, we preferred to label all of them "unfilled/silent pauses" without making any a priori judgment on the basis of their duration or function. We still wanted to use a minimum duration threshold, in line with L2 fluency research.

other hand, such as Campione & Véronis (2002) chose not to use a threshold at all. In this thesis, we adopt Kaghn's (2014) and De Jong & Bosker's (2013) minimum threshold of 250 ms, which is a popular choice in L2 fluency studies, and is considered an optimal cut-off point for L2 research (see De Jong & Bosker, 2013). This threshold avoids missing too many pauses and losing potentially key information.

I thought I just had that at the **(0.750)** [/] the French speaker.
(SITAF Corpus, speaker F13)

mais c'est juste euh **(0.950)** tu as dit tu as tort.
(SITAF Corpus, speaker A02)

(1.356) ouais moi aussi j'adore ce genre de jeu.
(DisReg Corpus, speaker A1)

MORPHO-SYNTACTIC MARKERS (MS) (Pallaud et al., 2019)

4. ***Self-repairs, (SR),*** also known as self-corrections, or substitutions (Eklund, 2004; Fox et al., 2013; W. J. Levelt, 1983, 1989) self-repairs refer to corrections (when a string of words is replaced by another) made by the speaker and not the interlocutor. They consist of three parts, (1) the original utterance which contains the item to be repaired; (2) the moment of interruption, and (3) the repair, defined as "the correct version of what was wrong before" (Levelt, 1989, p. 44). Self-repairs are further distinguished between several subcategories, based on Levelt (1983): (a) syntactic repairs (change in the syntax) (b) morphological repairs (change of morpheme), and (c) lexical repairs (change of lexical term). Repaired elements further include (d) additions (added lexical elements or phrases i.e,. Shriberg, 1994; Crible et al., 2019; Eklund, 2004).

[!] so **to me [//] for me** it's free. (a)
(SITAF Corpus, speaker F13)

You're not completely (0.689) **unha [//] unhappy**. (b)
(SITAF Corpus, speaker F07)

(i)l y avait des **choses euh [//] des trucs** vibrants. (c)
(DisReg Corpus, speaker F2)

(en)fin **il eum [//] je trouve qu'il** apprécie plus le pays. (d)
(SITAF Corpus, speaker F03)

5. ***Self-interruptions (SI)*** also known as false-starts, self-breaks, or deletions (Maclay & Osgood, 1959; Pallaud et al., 2013, 2019; Shriberg, 1996), they refer to syntactically or semantically incomplete utterances. This applies to cases when a speaker interrupts their own utterance mid-way, and formulates a new one that shares no syntactic nor semantic link with the previous one. In this thesis, we adopt the term *self-interruption*.

   A13: and good professors are not necessarily:y **+//.**
   A13: if you pay more you don't necessarily get better professors.
   (SITAF Corpus)

   F07: c'est [//] ça parait euh [//] c'est pas euh **+//.**
   F07 : on se sent éloigné en fait.
   (SITAF Corpus)

   D1 : et eum (0.437) (i)l y a un moment (en)fin en fait c'est **+//.**
   D1 : i i [//] ils suivent une famille.
   (DisReg Corpus)

6. ***Truncated words, (TR)*** (Eklund, 2004; Crible, 2017; Shriberg, 1994) Truncation or cut-off of a word before completing its articulation, defined as "linguistic items that are not fully executed/finished, whether or not they are finished later" (Eklund, 2004, p. 164). These items very often co-occur with morphological repairs.

   ou (en)fi:in oui mais j'ai pas regardé.
   (DisReg Corpus, speaker A2)

   you [/] you (a)re growing **hu** [//] you'(a)re growing up.
   (SITAF Corpus, speaker F07)

   oui c'est [/] c'est pas **im** [//] important.
   (SITAF Corpus, speaker A17)

7. ***Identical Repetitions (IR)*** (Candea, 2000; Clark & Wasow, 1998; Crible et al., 2019; Foster et al., 2000; Lickley, 2015; Maclay & Osgood, 1959; Shriberg, 1995).

Identical repetitions of a word, a phrase or a clause that was previously uttered in the speech channel. "Disfluent" repetitions (Shriberg, 1995) are often considered "non-significant", as they have no emphatic stress. In fact, Dumont (2018, p. 47) noted the distinction between language-based repetitions and speech-related repetitions, following Candea (2000). Language-based repetitions, such as oratory repetitions, are often excluded from disfluency analysis because they are commonly used for rhetorical purposes (e.g. marking emphasis) and are thus said to be produced voluntarily by speakers. Just like unfilled pauses, filled pauses, and prolongations, we did not make an a priori judgement on their function, and thus annotated all instances of identical repetitions in the data. Such repetitions can be made more than two times, but they were still annotated as a single repetition of the same item.

les [/] les choses **un peu:u (0.508) [/] un peu** tabou.
(SITAF Corpus, speaker F13)

I [/] I can't say ok let's see tomorrow.
(SITAF Corpus, speaker F09)

bah [/] bah [/] bah voilà y'avait ça.
(DisReg Corpus, speaker A2)

PERIPHERAL MARKERS (Crible, 2018)

8. ***Explicit Editing Terms (EDT)*** (Crible, 2017; Eklund, 2004; Meteer et al., 1995; Shriberg, 1994); words and expressions used by speakers to "compose, and edit and prompt themselves for their verbal behavior" (Shames & Sherrick, 1963, p. 8); this includes words such as "oops", "sorry", "wrong", but also any lexical expression "by which the speaker signals some production trouble" (Crible, 2017, p. 108) such as "I don't know", or "what's the word".

à un moment elle hhh. (en)fin mm **comment dire** euh [/] un moment elle lui demande de prendre son traitement.
(DisReg Corpus, speaker D2)

but I think um (1.460) **how do I say it** yes it's the same person.
(SITAF Corpus, speaker A03)

euh (1.658) moi j(e) dirai:i **je sais pas** un ou:u +/.
(SITAF Corpus, speaker A09)

9. ***Non-lexical sounds*** **(NL)** also known as vocalizations, sound objects, or liminal signs (Dingemanse, 2020; Hoey, 2020; Keevallik & Ogden, 2020; Ogden, 2018; Ward, 2006; Wright, 2011). Sounds that have typically been consigned to "the margins of language" (Dingemanse, 2020, p. 191), but which can nonetheless display interactional work. They are described as non-lexical vocalizations[99], but some of their meaning is conveyed by prosody (e.g. a sigh is typically associated with relief or tiredness). Their category includes respiratory conduct, such as inbreaths, sighs, and laughter, as well as vocalizations (*mm*, and tongue clicks).

[!] so does more money mean a better education?
(SITAF Corpus, speaker A13)

um (4.240) **((sighs))** maybe?
(SITAF Corpus, speaker A02)

hhh. et il se rend compte que:e Louison et Angélique euh lui cachent des choses.
(DisReg Corpus, speaker C1)

2.2.2. Sequence level

We shall now turn to the sequence level of (dis)fluency analysis. As the examples have shown above, fluencemes very often appear in combination, so they are better understood in terms of sequences, or units, constituting simple or complex constructions (Crible et al., 2019; Shriberg, 1994). For instance, truncated words tend to gravitate around morphological repairs, and filled pauses very often cluster with unfilled pauses (Benus et al., 2006; Betz & Kosmala, 2019; Candea, 2000; De Leeuw, 2007; Grosjean & Deschamps, 1972). Once the first level of analysis is completed (i.e. identification of individual fluencemes), our goal is to study their pattern of co-occurrence more closely. This second level of analysis can shed light on their potentially fluent or disfluent contribution at the utterance level, in line with studies in SLA, in psycholinguistics, computational linguistics, and corpus-based linguistics. The description of our annotation scheme is provided in Table 9[100] below.

[99] However, the prefix "non" has been criticized by Dingemanse (2020) who suggested the term "liminal sign" instead.
[100] "Entry" refers to the different values added in the controlled vocabulary in ELAN and Excel (cf section 2.3.2.). Examples of the annotated items are marked in bold.

Table 9. *Annotation scheme (sequence level)*

Tag	Tier definition/description	Entry	Example
Type	Type of sequence, whether it is made of a single fluenceme (simple) or a cluster (complex). Complex sequences only apply to immediately adjacent fluencemes.	"simple"	**euh** la légitimité de leur technique et de leur style (A1, DisReg)
		"complex"	donne je te le lis et puis après si tu veux **le:e (0.580) [/] si tu veux le** lire. (F13, SITAF)
Sequence list	List of all the fluencemes found within the sequence.	"FP" "UP+NL" "FP+MR+UP+TR" etc.	and you're always you know **um (0.689) um (1.290)** being careful (**"FP+UP+FP+UP")** (F07, SITAF)
			things like that **it is [//] I still** have time **("SR")** (F07, SITAF)
Sequence lenght	Number of markers combined within a complex sequence.	"2" "3" "4" "5" "6" etc.	**I was um (1.201) [!] [/] I was** walking with a friend **("4")** (A18, SITAF)
			on est pas **&ob [//] obligé de:e (1.099) [!] [//] pas vraiment de:e (1.138) [/] de** défendre ("10") (A13, SITAF)
Sequence configuration	Pattern of co-occurrence between the different kinds of fluencemes (in no specific order). Includes the combination of two, or three different kinds, or of the same one.	"VOC+VOC" (two or more vocal markers)	**(0.659) um (0.758)** well I've only not been a teenager for one year (A18, SITAF)
		"MS+MS" (two or more morpho-syntactic markers)	et ouais le **&pr le [/] le [//] celui** qui a gagné (A2, DisReg)
		"NL+NL" (two or more non-lexical sounds)	donc **[!] hhh**. en tant que suivante (C2, DisReg)
		"VOC+MS" (one or more vocal marker and a morpho-syntactic marker)	that's **like (2.082) [/] like** seven thousands dollars right there (A13, SITAF)
		"MS+NL"(one or more morpho-syntactic markers and non-lexical sounds)	**hhh. &no &n [//] not** large but significant (F10, SITAF)
		VOC+NL (one or more vocal markers with morpho-syntactic markers)	en effet **eum [!]** le premier mot du poème était une aphostrophe (A1, DisReg)
		VOC+MS+NL (one or more of the three kinds)	au début je trouvais ça **un peu:u ((sighs)) [/] un peu** gros (F2, DisReg)
		MIX (any other configuration; contains an EDT)	**(0.742) [!] euh qu'est ce que je voulais dire hhh.** oui donc l'anecdote (D1, DisReg)
Sequence position	Position of the fluenceme sequence within the utterance.	"initial" (near the beginning of the utterance)	**euh** mais tu les vois pas (A7, SITAF)
		"medial" (in the middle of the utterance)	and one day **(0.420)** maybe if they're not **(0.850)** killed there (F03, SITAF)
		"final" (near the end of the utterance)	mais c'est un p**eu:u euh +//.** (A09, SITAF)
		"standalone" (constitutes an utterance on its own)	**o:or +//.** (A02, SITAF)
		"interrupted" by other participant, so their position is unclear)	en faisant le mort tu sais **euh** +/. (C1, DisReg)
Communication management (Allwood et al. 1990)	Level of communicativeness of the fluenceme sequences, whether they are used to manage the production of a speaker's own production (Own communication management) or to manage the interaction (ICM)	"OCM" (own communication management)	**(0.573) [/] o:on** on savoure toute la sonorité du nom Ronsard (A1, DisReg)
		"ICM" (interactive communication management)	C1: tu trouves? C2: **ba:ah** elle est très maternelle avec lui. (C2, DisReg)

This table contains the name tag of the categories, the different entries, their description, and an example from the data. The present model is largely adapted from Crible et al., (2019), but also borrows from Allwood et al. (1990)'s categories. The table contains 6 different categories.

The first one (*Type*) gives information about the type of sequence (isolated versus combined), and the second (*Seq. List*) gives information about which individual fluenceme is found within a sequence; it provides a list (in the order in which they appeared) to get a precise idea of their combination pattern. As Crible (2018)[101] pointed out, this level of analysis can be further abstracted into smaller schematic entities, which enables us to set a finite number of configurations. This notion of schema echoes assumptions of Cognitive Grammar (Langacker, 1987) whereby a specific unit can be instantiated into a more abstract schema. In our case, a unit composed of different fluencemes (e.g. FP+UP+TR+MR) can be abstracted into a recurring structure (e.g. VOC+MS). Because a considerable number of different combinations are possible (depending on their order, the fluenceme used, the number of markers within the same sequence etc.), it was necessary to first distinguish between the length of a sequence and its pattern, but also to narrow the number of sequence configurations to only 8 possible configurations. The latter were determined by grouping formally and conceptually similar types of markers (vocal markers, morpho-syntactic markers, and non-lexical markers[102]), in order to get a clearer idea of the most frequent and recurrent patterns[103].

In addition, the position of the sequences was annotated (similar to Shriberg, 1994) based on their location in the utterance (but without using syntactic criteria). For the annotation of the *final* position (which can easily be confused with *initial* position, depending on the choice of utterance delimitation), two criteria were used. The first one relied on the identification of self-interruptions. As we have seen earlier (cf section 2.2.1) self-interruptions mark the end of an utterance, so they systematically occur in final position. Therefore, any fluenceme sequence that

[101] Read Crible (2017, pp 34-35) for more information.
[102] Explicit editing phrases are not included in this list because they are very rare, which is why they belong to the MIX entry. We also did not want to mix EDTs with NLs, despite being in the same "peripheral marker" category in the previous section. Non-lexical sounds were distinguished from vocal markers (even though they both rely on vocal features), for reasons already described above, mainly that they behave quite differently in discourse.
[103] This type of analysis was inspired by the work of Crible (2017, p. 119) who identified a limited number of possible configurations by focusing on discourse markers and their neighboring fluencemes, following a hierarchical system.

contained a self-interruption was considered *final*. In addition, fluenceme sequences that occurred at a *transition relevant place* were coded as *final*, and this second criterion was used in order to disambiguate between *final* versus *initial* position when no self-interruption was present. This is exemplified in the example below.

```
1 *D1: (0.452) quand j'étais allée à sa soirée du quatorze juill e:et euh.
                                                  ((smiles, looks at D2))
2 *D2 : ah mais oui c'était toi mais oui:i.
           *******
      ((points towards D1))
      ((smiles, looks at D1))
3*D1 : ouais ouais.
```

In this example, taken from the DisReg Corpus, D1 is talking about a famous party that she went to, and that D2 had already heard about in the past. In line 1, she is referring to that particular party using the noun phrase "sa soirée du quatorze juillet", and lengthens the final syllable of the final word, followed by a filled pause. The position in which the fluenceme sequence occurs can be considered a *transition relevant place*, in the sense that speaker change is made relevant in this specific sequential context, as D1 is expecting her partner to take the floor and assess her previous utterance (which she does in line 2). In this case, the fluenceme sequence was thus considered to occur in final position, and projected the completion of the ongoing utterance.

Another related issue concerns the annotation of isolated unfilled pauses. In CA, periods of silence are usually split into different categories, mainly *pauses*, *gaps*, and *lapses* (Sacks et al., 1974). Pauses refer to intra-speaker silences, which occur inside a speaker's turn; gaps on the other hand, refer to silences that occur between turns and after a possible completion point; lastly, lapses designate very long periods of silence during which participants stop talking. The distinction between pauses and gaps is made on the basis of speaker selection (Bowers et al., 1996): when a speaker has selected a next-to-speak, the following silence (hence "pause") is attributable to the following speaker, as it marks the delay of their subsequent action (which can potentially project a dispreferred answer). However, in cases when the previous speaker has not designated a next-to-speak, the silence is not attributable to anyone, and is hence considered a "gap", as it is not uniquely "owned". These assumptions pose a number of issues for the present study of (dis)fluency. First, in order to study the

overall rate of fluencemes across speakers from a production perspective, one needs to know which pause is "owned" by which speaker, in order to observe potential differences in intra-speaker variability. Second, if we decide to leave out all these non-attributable silences from our quantitative analyses, we would potentially lose a lot of information. For this reason, we did not distinguish between these different types of silences (cf section 2.2.1), except for lapses (i.e. long periods of silence when speakers were both looking at the piece of paper after finishing a speaking task,) which were excluded from the quantitative analysis. In line with Edlund et al., (2009), we annotated "mutual silences" (silences between turns) as follows: the "instigator" of the silent pause is the speaker who last spoke before the silence occurred, or who last spoke alone in case of a simultaneous end of speech; and therefore the "owner" of the silent pause is the one who breaks the silence. The silent pause owned by a next speaker (after the previous speaker ended their turn) thus occurs in *initial* position. On the other hand, the silence which occurs at a transition relevant place (and thus initiated and owned by a prior speaker to designate a next speaker) occurs in *final* position. This is illustrated in the examples below:

Example (A) Pair D, DisReg Corpus

```
1     *D1 : elle avait une jolie robe à fleur.
2     *D2 : (0.573) à sa soirée ?
3     *D1 : (0.409) à sa soirée oui.
```

Example (B) Pair 11, SITAF Corpus

```
1     *A11 : quand tu as des responsabilité:és faire des choses.
2     *F11 : oui.
3     *A11 : quand tu veux c'est (0.930) +...
                                    ((gazes at F11))
4     *F11 : c'est bien dans le sens où t'as pas de responsabilités.
5     *A11 : oui.
```

In example (A), the silent pause in line 2 would typically be labeled a "gap" in CA, as it does not occur at a transition relevant place (unlike in example B in line 3 when the speaker is trailing off, which leads to speaker change). However, if we follow Edlund et al's (2009) method, we consider that this pause was instigated by D1, who spoke first, and then "owned" by D2, who broke the silence and took her turn. This pause

thus occurred in initial position (within D2's utterance). In Example B, on the other hand, the pause occurred in final position (within A11's utterance), at a transition relevant place. These criteria were used systematically to disambiguate the position of silences, thus relying on both conversation-analytic and speech production-based methods.

Lastly, the final category (*communication management*) lies at the frontier between quantitative production-based methods and qualitative conversation-analytic analyses. We first looked at speech processes related to verbal planning, lexical search, repairs etc., to identify cases in which speakers worked on their own production (OCM) while producing fluencemes. In addition, we relied on conversation-analytic methods (cf section III) to find whether fluencemes were more other-oriented and pertained to the development of the interactional exchange (ICM), by looking at cases of adjacency pairs, intersubjectivity, stance taking, progressivity, preference structure, and the like. This category certainly reflects our mixed-method approach, as well as our integrated theoretical framework, and illustrates the functionally and interactionally ambivalence of fluencemes. More examples are provided in section III.

2.2.3. Visuo-gestural level

In the previous section, we focused on the speech level of (dis)fluency, which only provided a partial picture of these phenomena. We shall now move to their visuo-gestural dimension, which further gives an account of their multimodal deployment in discourse. As specified throughout this thesis, the general aim of the present research is to explore the functional and interactional ambivalence of fluencemes, and this can be further achieved by looking at their gestural behavior and the presence of visible bodily conduct in discourse. While strictly *verbal* or *vocal* fluencemes essentially lack semantic and propositional content (unlike other parts of speech such as discourse markers), their accompanying bodily behavior can call attention to their potential meaning and function. For instance, we have seen that a silent pause, which, in essence, is "unfilled", and thus devoid of any lexical meaning *a priori*, can still convey a great deal of information (such as the projection of a piece of information in discourse) if we pay close attention to its co-occurring gestural behavior (cf Fig. 21, section I,1.5.1). Inter-(dis)fluency is thus better understood in terms of *multimodal*

constructions, which resort to a multiplicity of vocal, verbal, and visual-gestural actions in order to (co)-construct meaning in discourse. For this reason, we decided to target a functional classification of gestures for our quantitative annotation, as to uncover their potential functions. In addition, we also sought to compare gesture productions in "disfluent"[104] versus "fluent" stretches of speech, to examine the relationship between gesture production and speech production, in line with Graziano & Gullberg (2018). Even though functional gesture classifications can be relatively subjective, and miss out on other central aspects of gesture production such as shape, configuration, direction, and movement, they rely on a limited number of well-defined categories, which can in turn be used efficiently for quantitative treatments. Additionally, it was essential to choose a consistent gesture classification that did not conflate formal and functional categories (cf Chapter 1, section 3.3.3.) as to provide a reliable overview of gesture distribution across the data. In sum, the goal of this 3rd level of analysis (above the identification of fluencemes and their sequence) is twofold: (1) to get a general idea of the types of gestures that typically co-occur or do *not* co-occur with fluencemes through quantitative analyses, and (2) to examine more closely the multimodal deployment of embodied fluencemes in discourse, by looking more specifically at the formal features of co-occuring gestures and their ecology within the environment, in qualitative analyses (see Chapter 5). Once more, the present study stresses the importance of combining qualitative and quantitative methods which continuously complement one another, in order to shed light on different but complementary aspects of our investigation.

Other relevant points to consider when studying the overall visuo-gestural manifestation of fluencemes is the analysis of *gesture phases* (Kendon, 2004, Seyfeddinipur, 2006; cf Chapter 1 section 3.3.4), as well as the interplay of (dis)fluency and gaze direction (e.g., Jehoul, 2019), so these aspects were also included in our annotations. In summary, our quantitative annotation model of (dis)fluency targeted three specific aspects at the visuo-gestural level, mainly (1) gesture phrase, (2) gesture functional type, and (3) gaze direction. Other instances of bodily behavior, such as facial expressions, body movement, body orientation, and the like, were examined in

[104] We are not exactly pleased with the terms "disfluent" versus "fluent stretch of speech" offered by Graziano & Gullberg (2018), but it is one way to refer to the comparison of gestures that typically co-occur with fluencemes with the gestures that do not.

our qualitative analyses, through *collection studies* (cf section III.3.1.). It should be noted that only (2) and (3) were annotated in all the data (i.e. during fluencemes and outside fluencemes), while information contained in (1) was only annotated during fluenceme sequences. This is further explained in Table 10.

1. ***Gesture phase***:

This annotation is based on the analysis of phases of gestural movement within gesture units (cf Kendon, 2004), and adopts similar categories used by Seyfeddinipur (2006) and Graziano & Gullberg (2018), which are listed below[105].

Rest position	Preparation phase	Stroke	Hold	Retraction

Figure 27 further illustrates an example of gesture retraction during a fluenceme (FP), taken from Kosmala et al., (2019).

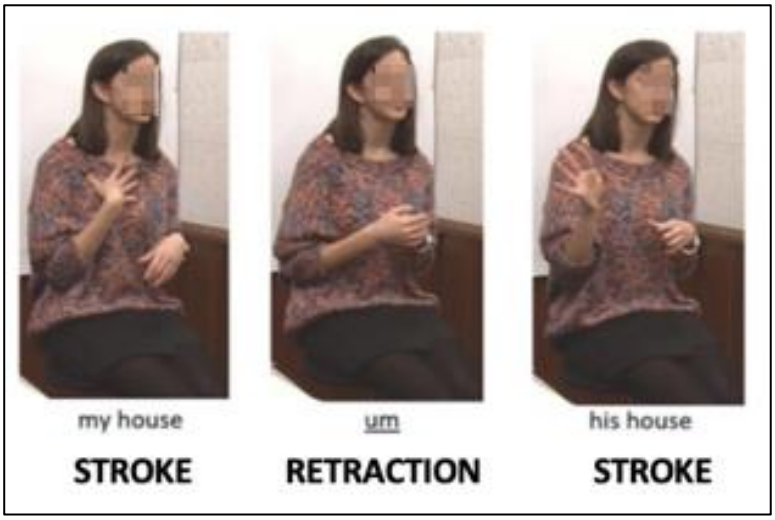


Figure 27*. Example of gesture retraction (Kosmala et al., 2019)*

15% of the data was annotated by a second annotator to measure inter-coder reliability on the recognition of gesture phrases, and received a Kappa score of 0.84.

2. ***Gesture type:***

This tier was initially added to the *fluenceme sequence* category, to annotate the types of gestures that accompanied fluencemes, but it was later added outside this category to annotate all gestures in the video recordings. We first distinguished between two

[105] Read Chapter 1, section 3.3.4. for more details.

general classes of gestures, mainly (a) referential gestures and (b) pragmatic gestures (in line with Kendon, 2004). Referential gestures comprise two subtypes: representational, and deictic-anaphoric gestures. Pragmatic gestures fall into 3 subcategories: (1) parsing/discursive, (2) interactive/performative, and (3) thinking gestures. Our classification system is largely adapted from Kendon (2004), Cienki (2004), Müller (1998), Streeck (2008), and Bavelas et al., (1992), and is summarized in the figure below.

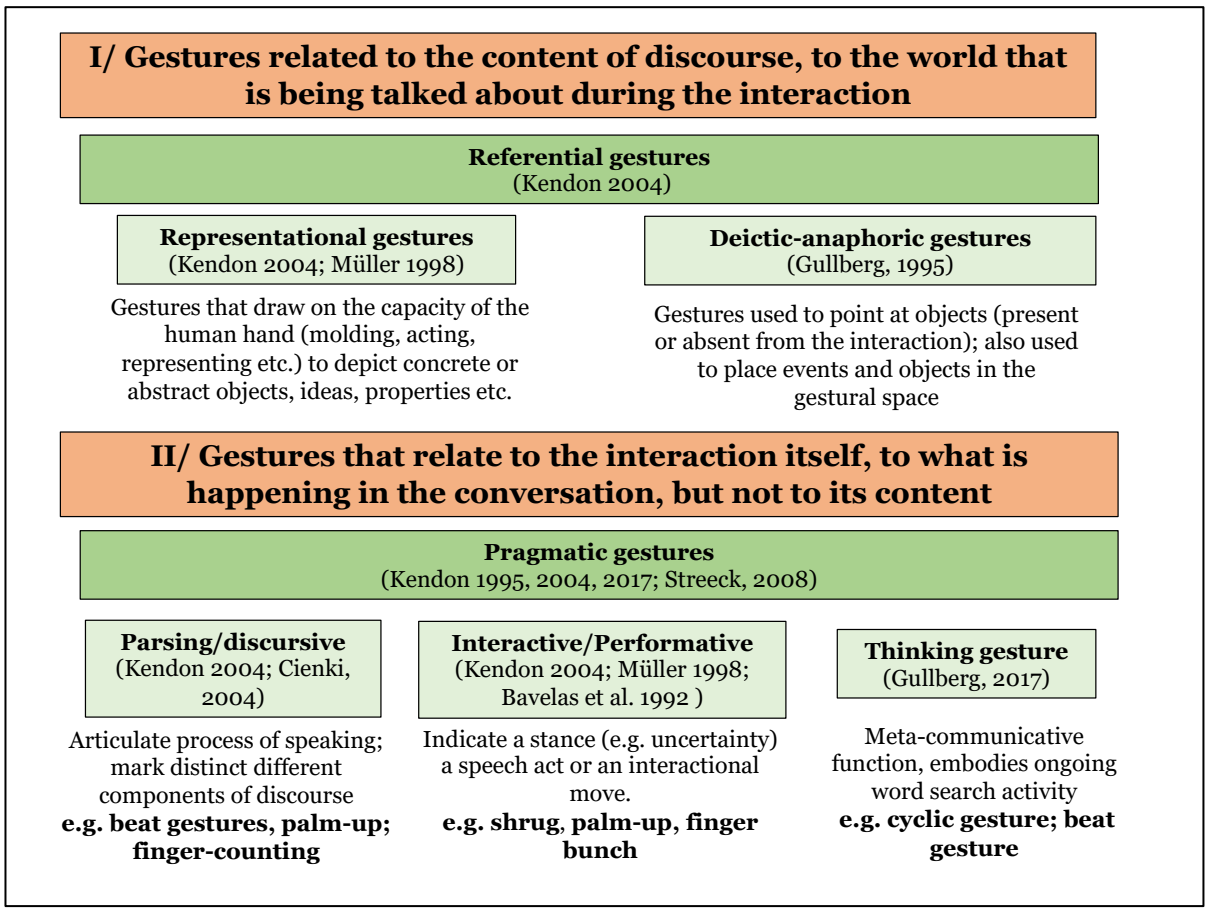
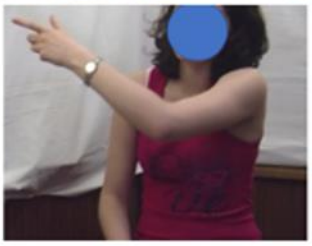
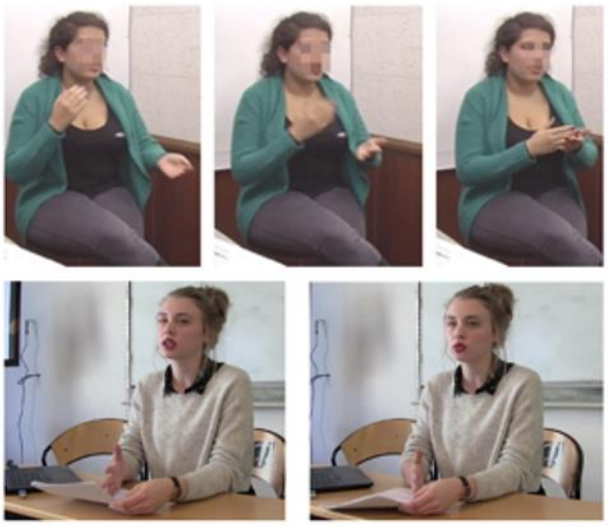


Figure 28. *Gesture classification*



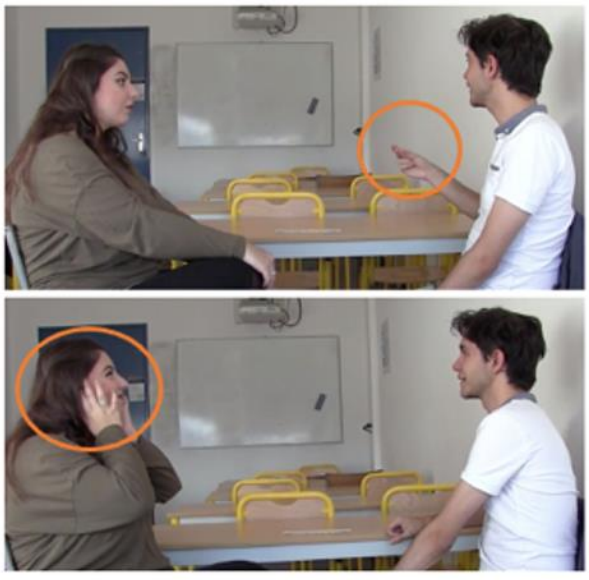
Referential gestures refer to gestures that convey meaning related to the referential content of discourse. Representational and deictic-anaphoric gestures are similar conceptually, in the sense that they both relate to the content of discourse, but they behave differently, and perform slightly different functions, as **representational gestures** rely on depictive methods that draw on practical abilities of the human hand (cf Müller, 1998, Streeck, 2008) to depict an object or event, real or imagined, abstract, or concrete; while **deictic-anaphoric gestures** make use of space to point (with finger, hand, palm, head, or foot etc.) to a location, person, or event, to draw potential relationships between referents, or to place events and objects of imaginary discourse in the gestural space.

Pragmatic gestures, on the other hand, do not pertain to the propositional content of discourse, but rather enact pragmatic actions. **Interactive/performative** gestures have expressive features (e.g. *quotation marks*, *finger bunch*, *grapollo*, etc., Kendon, 2004), and enact interactional moves, or actions (dismissing, turning down, requesting etc.). This subcategory can be distinguished from other gestures with regard to the orientation and direction of the hand – if it is oriented towards the interlocutor, then it is considered interactive (as opposed to when it is oriented to the left or right) as it is directly addressed to the interlocutor (see Bavelas et al., 1992). These kinds of gestures convey information about "the process of conversing with another person" (Bavelas et al., 1992: 473). Gestures associated with these functions are included in Kendon's (2017) categories of *interactional regulators*, *performative*, and *operational* (cf Chapter 1, section. 3.3.3).



Parsing/discursive gestures are used to mark discourse segments, present an idea or an argument, or emphasize parts of discourse; they thus serve typical discursive functions in discourse such as emphasis, structuring, linking, and the like (Cienki, 2004). They are not directly oriented towards the interlocutor, but they can still be used to facilitate comprehension or capture the interlocutor's attention (cf *comprehension-oriented* gestures, Stam & Tellier, 2017).

Lastly, **thinking gestures**, also known as *word searching* gestures (Stam & Tellier, 2017) or "*Butterworth*[106]*"* (Tellier & Stam, 2012), enact a planning, thinking, or word searching activity. This category is not traditionally labelled under the category of pragmatic gestures, although it has been mentioned by a number of authors, such as Ladewig (2014) in her paper on cyclic gestures, or Goodwin & Goodwin (1986) in their paper on co-participation within a word searching activity. Gullberg (2011, p. 143) labels them *thinking gestures*, and they are described as metapragmatic gestures produced during communicative breakdowns as a comment on the silence. These types of gestures are very often accompanied with a salient thinking face, and overt verbal manifestations of lexical search ("oh what's the word*"*). They can also be manifested through finger snaps (cf first picture above), which will be further analyzed in Chapter 5.



While these gestures share recurrent structural, formational, and conventional features (cf Müller et al. 2013), some of which can help with the disambiguation of

[106] This gesture was named after the author Brian Butterworth.

different categories (e.g. a pointing gesture directed towards the interlocutor versus towards the left or right; shrugs, finger bunch, and palm-up open hands gestures are typically recognized as conventionalized pragmatic gestures), specific attention was especially paid to the **ecology** of the gestures, and the **context of use** in which they appeared (in line with Kendon, 2004, 2017). This type of identification is very context-specific and thus relies on subjective judgements, so 15% of the whole data was subjected to a second annotation by a different annotator, to measure inter-coder reliability. The Kappa score obtained was **κ** = 0.85 for the gesture types, and **κ** = 0.78 for gesture subtypes. Our annotation protocol is further described below.

There were three dependent tiers: (1) gesture phrase, (2) gesture type, and (3) gesture subtype. The slots were filled with different entries from the *controlled vocabularies* in ELAN (cf section 2.3.2.), but they could be left empty (N/A); if for example a gesture was not fully completed during a fluenceme, the 2nd and 3rd tiers did not need to be filled. An example from our coding is provided in the table below.

***Table 10**. Gesture coding (during fluencemes and outside fluencemes)*

Gesture phrase	**Gesture type**	**Gest. sub-type**
If gesture stroke is not fully completed (only during fluencemes)		
"rest position" "preparation phase" "hold" "retraction"	N/A	N/A
If gesture stroke is completed (during fluencemes and outside fluencemes)		
"completed gesture" (default)	"referential gesture"	"representational" "deictic-anaphoric"
	"pragmatic gesture"	"parsing/discursive" "interactive/performative" "thinking gesture"

3. Gaze direction

Each change of gaze direction was annotated as either "towards interlocutor", "away" (from the interlocutor), "towards paper" (i.e. the students' notes, books, laptop; the piece of paper with the instructions, etc.), "in different directions", and a few times "towards camera" (only in DisReg). This tier was initially only dependent upon the

fluenceme seq. and *gesture* tier, but it was later annotated independently on the whole data.

2.3. Tools

2.3.1. Statistical tests

Our annotation scheme includes both numerical variables (i.e. any value with a finite or infinite interval, e.g. length, duration) and nominal variables (i.e. categorical variables that can take only a limited number of values, e.g. "simple" and "complex") so different types of statistical tests were used to analyze our data. The tests, which are summarized in the list below, were conducted with various online calculators[107].

- We ran **log-likelihood tests** to measure frequency differences across corpora (e.g. rate of fluencemes per hundred words in L1 versus L2, or in class versus conversation. The higher the log-likelihood value is, the higher the difference is between two frequency scores. For instance, a value of 3.84 or higher is significant at the level of $p < 0.05$, and a value of 6.6 or higher is significant at $p < 0.01$.

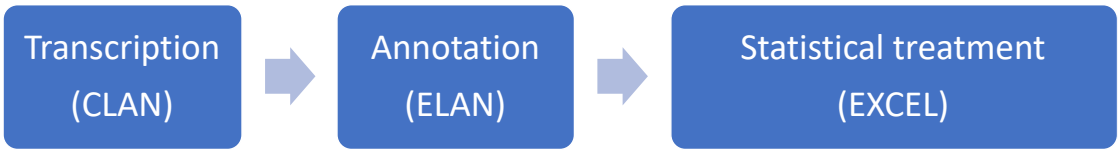
[107] All the tests were performed using the following links: (last retrieved on August 26th 2021)
https://biostatgv.sentiweb.fr/?module=tests (t-test, Pearson, and Wilcoxon)
http://vassarstats.net/propdiff_ind.html (z statistic)
http://ucrel.lancs.ac.uk/llwizard.html (log-likelihood)
https://www.graphpad.com/quickcalcs/kappa1/ (kappa)
https://www.socscistatistics.com/tests/ (chi square)
https://www.statskingdom.com/320ShapiroWilk.html (Shapiro-Wilk)

- In order to check whether the numerical variables were normally distributed, a **Shapiro-Wilk Test** was first performed, obtaining a W value. A small W value (i.e. below the accepted range 0.99-1.00) indicates that the sample is not normally distributed. If the data is normally distributed, **T-tests** can be used to compare means of numerical variables (e.g. duration, number of markers within a sequence) to examine whether there are significant differences between the two populations. The lower the *p*-value is ($p < 0.01$), the more likely the differences found between the two groups are representative of the whole population and not due to individual differences. If the data is not normally distributed, then a non-parametric test is used (**the Wilcoxon Signed-Ranks test**), to evaluate the differences between the two treatments.
- We also computed **Z-scores** to assess the significance of differences between proportions (e.g. proportion of fluencemes in complex versus simple sequences, proportion of gestures in "fluent" versus "disfluent" stretches of speech, etc.). Z score values give information about the number of standard deviations above the mean. A positive Z score indicates that the raw score is higher than the mean average, and a negative score reveals that the raw score is below the mean average. Z tests can also yield significance levels with *p*-values, and the smaller the *p*-value is, the more unlikely the differences between proportions are the result of random processes, so the latter can be considered significant. High or low negative scores associated with small *p*-values thus typically indicate that the differences found between proportions are unlikely due to chance.
- We performed a **chi-square test** of independence to measure the association between two categorical variables, whether the values of one variable relates in some way to the values of others, e.g. whether gesture types relate to speech type, or whether communication management is associated with sequence type. A low value for chi-square means that there is a high correlation between the two sets of data.
- We ran **Pearson's correlation coefficient** to measure the statistical relationship between two numerical variables, mainly between language proficiency score and (dis)fluency rate in SITAF, and mean length of utterance and (dis)fluency rate in DisReg (Cf Chapters 3 and 4). The value $r = 1$ means perfect positive correlation, and the value $r = -1$ means a perfect negative correlation.

- Lastly, we measured inter-rater and intra-rater reliability for the annotation of some of our qualitative categories, i.e., *gesture phrase*, *gesture type*, and *gesture subtype* with **Cohen's Kappa**. Kappa scores are interpreted on a scale from 0 to 1.00, in which values between 0 and 0.20 indicate non to slight agreement, 0.21 to 0.40 indicate fair, 0.41 to 0.60, moderate, 0.60, 0.61 to 0.80 indicate substantial, and values between 0.81 and 1.00 show almost perfect agreement (Landis & Koch, 1977).

2.3.2. CLAN, ELAN, and Excel

As specified earlier (cf section 1.5.3.) our video recordings were transcribed with the software **CLAN**, then transferred and converted to **ELAN** (EUDICO Linguistic Annotator) (Sloetjes & Wittenburg, 2008). ELAN is a multipurpose and multilayer annotation tool which was developed at the Max Planck Institute for Psycholinguistics to provide a technological basis for the annotation of multi-media recordings. ELAN allowed us to work on annotations in multiple time-aligned tiers, which identified specific segments, such as individual fluencemes. The order of the operations is summarized in the pipeline below.



All the fluencemes were manually annotated directly on ELAN, and their value was added on one tier ("marker"), using controlled vocabularies (cf Fig. 29 below). A controlled vocabulary is defined as: "a user definable list of values that are likely to be related in some way and that the user plans to apply to annotations on one or more tiers" (Sloetjes & Wittenburg, 2008, p. 2). Controlled vocabularies are particularly useful, because they avoid typing errors, and improve annotation consistency. A second tier ("marker form") was added to annotate the form of filled pauses (either *(e)uh* or *(e)um*, see section 2.2.1), and of non-lexical sounds (*tongue clicks, inbreath, laughter, sigh, mm*, and *creaks*). In addition, a different tier ("duration") was aligned with the vocal *marker* tier (prolongations, filled pauses, and unfilled pauses) to annotate their duration in milliseconds. Then, 8 other tiers were added (*type, sequence, position, gaze-direction, gesture phrase, gesture type* and *gesture subtype*)

and corresponded to the exact time alignment of the individual fluencemes grouped together, in order to analyze the *fluenceme sequence* level. Once again, a controlled vocabulary was created for each tier (except for the *sequence* tier which generates a considerable number of different values):

- *Type*: “simple” “complex”
- *Position*: “initial” “medial” “final” “interrupted” “standalone”
- *Gaze direction*: “away” “towards interlocutor” “towards paper”
- *Gesture phrase*: “rest position” “completed gesture” “hold” “retraction” “preparation phase”
- *Gesture type*: “referential” “pragmatic”
- *Gesture subtype*: “representational” “deictic-anaphoric” “discursive” “interactional” “thinking gesture”

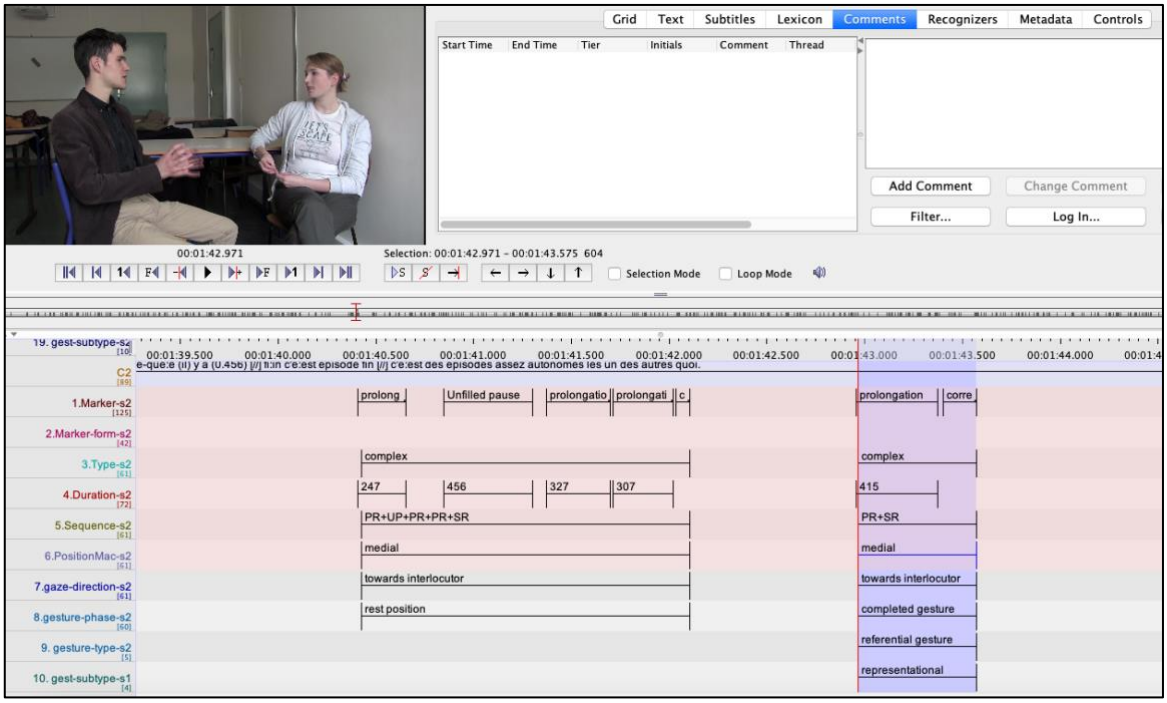


Figure 29. *Annotation grid, ELAN window*

The annotations were then exported to **Excel**, which resulted in an excel worksheet where each row corresponded to a tier from ELAN. In addition, we added two columns for metadata: “condition” (either “spontaneous” or “prepared” for DisReg, or “Tandem EN” or “Tandem FR” for SITAF) and “language” (L1 or L2) for the SITAF Corpus. We created two separate sheets, one with tiers exclusively from the *fluenceme* level of analysis which includes “marker”, “marker form”, and “marker duration” (Fig. 30). In

this sheet, an additional category, "marker type" was added directly on Excel, to distinguish between the different types of markers (cf section 2.2.1). This category includes three values, mainly "VOC", "MS", and "NL".

The second sheet includes all tiers from the *sequence* level of analysis (*type, sequence, position, gaze direction, gesture type, gesture subtype*), and we added 3 more categories which were directly annotated on Excel (cf Fig. 31): *sequence configuration*, *number of markers combined*, and *communication management*. These categories were added later, following the multiple changes that had been adopted on our annotation grid (cf section 2.1.). In sum, these two different Excel sheets, which were obtained through our annotations on ELAN, enabled us to treat *fluenceme* rate and *fluenceme sequence* rate separately. It was necessary to do so, since the two categories cannot be merged together; for example, a complex sequence is counted as a single token, but it can be made of 4 different individual fluencemes, which will count for 4 different tokens. These frequencies could thus not be mingled within the same sheet as it would have led to faulty findings.

	A	B	C	D	E	F	G
1	Participant	Condition	Utterance	Marker	Marker type	Marker form	Marker Duration
2	A2	spontaneous	si on connait rien &euh à l'Écosse ou quoi c'est pas grave ?	Filled pause	VOC	uh	120
3	A2	spontaneous	+< parce-que sinon &euh c'est un peu plat quoi fin moi les sÈries historiques &euh bof bof.	Filled pause	VOC	um	105
4	A2	spontaneous	+< parce-que sinon &euh c'est un peu plat quoi fin moi les sÈries historiques &euh bof bof.	Filled pause	VOC	uh	172
5	A2	spontaneous	&ou fi:in j'ai déjà entendu mais j'ai pas regardé.	Truncated word	MS		206
6	A2	spontaneous	&ou fi:in j'ai déjà entendu mais j'ai pas regardé.	prolongation	VOC		206
7	A2	spontaneous	et est-ce que:e +//.	prolongation	VOC		105
8	A2	spontaneous	et est-ce que:e +//.	Self-interruption	MS		
9	A2	spontaneous	(0.520) parce-que si:i +/.	prolongation	VOC		392
10	A2	spontaneous	(0.520) parce-que si:i +/.	Unfilled pause	VOC		520
11	A2	spontaneous	est-ce que c'est comme &euh Games of Thrones ou y'a genre &euh (1.256) douze à rattraper? &=laughs.	Filled pause	VOC	uh	305
12	A2	spontaneous	est-ce que c'est comme &euh Games of Thrones ou y'a genre &euh (1.256) douze à rattraper? &=laughs.	Filled pause	VOC	uh	156
13	A2	spontaneous	est-ce que c'est comme &euh Games of Thrones ou y'a genre &euh (1.256) douze à rattraper? &=laughs.	Unfilled pause	VOC		1256

***Figure 30**. Excel sheet for the fluenceme level of analysis (from DisReg)*

	A	B	C	D	E	F	G	H	I	L	M	P	Q	R
1	Participant	Condition	Language	Utterance	Type	Sequence pattern	Nb of markers combined	Sequence configuration	Position	Communication management	Gaze-direction	Gesture-phase	Gesture type	Gest subtype
32	F13	Tandem EN	L2	I thought I just had that at the (0.750) [/] the French speaker.	complex	<UP-IR>	2	VOC+MS	medial	OCM	towards interlocutor	return-rest-position		
33	F13	Tandem EN	L2	but &um (...) to me:e (0.415) it's not really a question of money .	complex	FP+UP	2	VOC+VOC	intial	OCM	away	rest position		
34	F13	Tandem EN	L2		complex	PR+UP	2	VOC+VOC	medial	OCM	away	rest position		
35	F13	Tandem EN	L2	a:and not being in (0.541) a great or (0.475) &mm better university or highschool or (0.476) whatever but [c] just maybe .	simple	PR	1		intial	OCM	away	rest position		
36	F13	Tandem EN	L2		simple	UP	1		medial	OCM	away	rest position		
37	F13	Tandem EN	L2		complex	UP+NL	2	VOC+NL	medial	OCM	towards interlocutor	preparation phase		
38	F13	Tandem EN	L2		complex	UP+UP+NL	4	VOC+NL	medial	ICM	in different directions	completed gesture	pragmatic gesture	interactive
39	F13	Tandem EN	L2	maybe &um (1.033) if you pay like I don't know (0.607) &um [c] on:ne thousand dollars a year or I don't know why .	complex	FP+UP	2	VOC+VOC	intial	OCM	away	completed gesture	pragmatic gesture	discursive
40	F13	Tandem EN	L2		complex	FP+PR	2	VOC+VOC	medial	OCM	away	completed gesture	pragmatic gesture	thinking gesture
41	F13	Tandem EN	L2	but &um if you pay that price .	simple	FP	1		intial	OCM	away	preparation phase		
42	F13	Tandem EN	L2	maybe &uh the university can (0.935) bring you all the (1.115) great teachers in all the world or &uh (0.508) I don't know in all the country .	simple	FP	1		intial	OCM	away	completed gesture	pragmatic gesture	discursive
43	F13	Tandem EN	L2		simple	UP	1		medial	OCM	away	completed gesture	pragmatic gesture	discursive
44	F13	Tandem EN	L2		simple	UP	1		medial	OCM	away	completed gesture	pragmatic gesture	thinking gesture
45	F13	Tandem EN	L2		complex	FP+UP	2	VOC+VOC	medial	ICM	towards interlocutor	completed gesture	pragmatic gesture	interactive

Figure 31. *Excel sheet for the sequence level of analysis (from SITAF)*

Lastly, the values contained in the excel sheets were then used to create Pivot Tables to summarize the data obtained in a more extensive table, and which included statistics such as sums, averages, and percentages. We also used **CLAN** to run two specific commands directly on the software, mainly *MLU* (mean length of utterance), and *freq* (number of words per participant); we used this information to calculate the rate of fluencemes per hundred words, and to measure the relationship between MLU and frequency rates (cf Chapters 3 and 4).

III. Methods for qualitative analyses

We shall now finish this chapter with the description of our methods for the qualitative analyses. The latter were carried out on a selection of excerpts from all the data, which were chosen to highlight several instances of the same phenomena, or similar phenomena in different contexts, to provide an interactive multimodal frame of analysis, in line with interactional and conversation-analytic methods cf Chapter 1, section III.3.2).

3.1. Conversation-analytic methods

In line with CA, we paid specific attention to the following practices in relation to inter-(dis)fluency behavior, which are typical of social interactions (cf Chapter 1, section III.3.2. and 3.3.3):

- *Turn-taking mechanisms*; how fluencemes are organized within turns, if they occur at a TRP, if they project upcoming talk, allocate or yield a turn, etc.
- *Episodes of repair*; when they are co-achieved, and not only self-initiated.
- *Preference structure;* if the fluenceme sequence prefaces or delays talk to embody a disaffiliative or dispreferred action.
- *Adjacency pairs*; whether fluencemes occur during adjacent actions, ordered as a first part and a second part (i.e. greeting sequence, question/answer, etc.), or within insertion sequences (i.e. a sequence of turns that intervenes between the first and second parts of an adjacency pair).
- *Alignment*; how fluencemes may be used to anticipate misunderstanding or disagreement, or on the other hand to restore alignment and mutual understanding.
- *Participation Framework*; the status of participants in the interaction (e.g. speaker versus recipient), and the different interactional roles they may play, whether they engage or disengage from the interaction.

An application of some of these methods is briefly illustrated in our short analysis of the following "data fragment" (Ten Have, 2007, p. 126). Data fragments can be used to exemplify a local phenomenon from a specific context, taken from an excerpt of the recording. In this excerpt, taken from Pair D in DisReg, the two friends are talking about a novel (*Moby Dick*) that one of them, Tina,[108] had presumably read for class. While there are many fluencemes in this excerpt (11 in total), we will focus specifically on the ones that appeared in lines 2 to 4.

Example Pair E (DisReg) Moby Dick

1 LEA : hhh. et euh ah oui par rapport à l'anglais tu [/] tu devais pas lire u:un livre euh Moby Dick?

→ 2 **TINA : (0.400) ouais [/] ouais bah Moby Dick euh +/.**
((raised eyebrows))

3 **LEA : e:et du coup tu:u [/] tu l'as fini ?**

4 **TINA : (2.087) <ça:a> ((laughs)).**
((smiles, looks away and tilts her head in different directions))

[108] Thank you, Elinor Ochs, for kindly suggesting to give actual names to the pseudonymized participants, instead of using labels. This makes the rendering of the exchange more natural.

5 LEA : +< ouais.
6 LEA : nan c'est trop dur ((smiles)).
7 TINA : si [/] si nan mais je l'ai fini mais je l'ai lu en français parce-que:e il était trop gros.
8 TINA : et déjà en français je comprenais pas un mot sur deux.

In this specific sequence, Lea (E2) is projecting an action in the form of a question which typically expects a yes/no answer. As the transcript shows in line 2, Tina's (E1) answer to Lea's question is delayed by a silent pause clustered with a repetition of the agreement marker "ouais". Her answer is then interrupted by Lea, who asks a second question, following the initial adjacency pair initiated in line 1. This time, Tina delays her answer for a significant amount of time, with an unfilled pause lasting up to 2 seconds, during which she tilts her head in different directions, and smiles. Even though Tina does not provide a full answer to the question, her conversational partner seems to immediately pick up on it, and in fact answers the question herself, with downgraded assessment in line 6. Tina's answer is eventually provided in lines 7 and 8, following Lea's intervention during the *insertion sequence*. In sum, this example illustrates different parts of a question-answer pair, embedded within the sequential development of the exchange, where the interactants rely on sequential, vocal, and bodily resources to co-construct responsive actions. The turn-initial fluenceme sequences found in lines 2 and 4 are shown to play a crucial role in the development of the turn-at-talk, as they signal Lea's delayed response to the pending question, which is later taken up by Tina, who offers to participate in the question-answer sequence she had first initiated. These types of fluenceme sequences thus typically reflect processes of *interactive communication management*, as opposed to the fluenceme found in line 7 in the form of a prolongation (parce-qu*e:e*), which rather deals with *own communication management*, whereby the speaker may be elaborating on her reason for reading the novel in French, but without straightforwardly doing interactional work. This type of qualitative analysis, drawing on CA methods and the *Participation Framework*, thus provide relevant tools for the interpretation of fluencemes in context.

However, this type of analysis is also based on subjective judgements of the situation, often limited to the interpretation of a single observer. Therefore, researchers may gather at *data sessions* to discuss and share their ideas on the same

video recordings. A data session can be defined as the following (Ten Have, 2007, p. 140):

> an informal get-together of researchers in order to discuss some "data" – recordings and transcripts. The group may consist of a more or less permanent team of people working together on a project or in related projects, or a had hoc meeting of independent researchers.

A data session is often organized as follows: one of the group members brings raw data, i.e. an excerpt from a video recording often accompanied by a transcript, and gives background information. Then, he or she plays the whole recording, or smaller excerpts, and the participants of the session are later invited to spontaneously react, and share their ideas and comments on specific parts of the recording. As Ten Have (2007) further noted, collective explorations of the data in data session practices rely not only on the researcher's individual interpretation, but on *shared* understandings of it. This can be highly beneficial for the observer, who can be challenged, or even "forced" (Ten Have, 2007, p. 141) to go beyond his or her own individualistic and impressionistic judgments. Data sessions are also often comprised of researchers who are experts in a particular field, so the observer can gain insight from their expertise. Ten Have (2007, p. 141) spoke very highly of data sessions, based on his own experience:

> Data sessions are an excellent setting for learning the craft of CA, as when novices, after having mastered some of the basic methodological and theoretical ideas, can participate in data sessions with more experienced CA researchers. I would probably never have become a CA practitioner if I had not had the opportunity to participate in data sessions with Manny Schegloff and Gail Jefferson. And, on the other hand, having later to explicate my impressions and ideas to colleagues with different backgrounds, both novices and experts, helped me to be more clear about those ideas, methodologically, theoretically, and substantively.

While Ten Have talked about data sessions specifically within a CA framework, the latter can also be grounded in a more mixed theoretical and methodological structure. For instance, the *Gesture and Multimodality Group (GMG)* held at UC Berkeley in the linguistics department is comprised of various members in gesture studies, cognitive linguistics, and educational technology. Similarly, the *Co-Operative Action Lab (Co-*

Op) from the department of anthropology at UCLA, gathers various participants from ethnomethodology, conversation analysis, linguistic anthropology, sociology, cognitive science, and more. These kinds of gatherings thus promote collaborative work and seek diversity, and this has played a fundamental role in the construction of our integrated framework, as we have constantly been confronted (at *GMG*, *Co-Op*, or *SeSyLiA*) with different perspectives from diverse fields, which in turn deepened our understanding of inter-(dis)fluency. While data sessions are quite informal and can be freely organized by just three or four people outside a research group, they nonetheless offer a sense of validity in qualitative research.

To conclude, our qualitative methodology, greatly inspired by methods in CA, (among other fields such as interactional linguistics to a larger extent) is strongly data-driven. In addition to our quantitative findings, our methodology relies on two types of qualitative studies, mainly *single case analysis and collection study* (Mazeland, 2006; cf Chapter 1, section IV.3). In a single case analysis, we focus on a single episode with respect to a relevant aspect of inter-(dis)fluency phenomena (cf. excerpt above), and in a collection study, we generalize collections of a particular instance, often drawing from single case studies, but also from the findings generated in our quantitative treatments. This enables us to shed light on specific practices that are not otherwise observable in statistical analyses.

3.2. Multimodal analysis: use of PRAAT for the vocal dimension

As noted several times before, our analysis of inter-(dis)fluency is multi-level, and thus relies on the analysis of several dimensions, mainly *speech*, *visual-gestural*, and *interactional* (cf Chapter 1, section IV.3). These dimensions are explicitly explored in our qualitative analyses, by uncovering the progressive mobilization of multiple semiotic resources in the course of the talk (Mondada, 2016, Morgenstern, 2014, Ogden, 2020). In this section, we explain the way we described the *vocal* level of inter-(dis)fluency, with the help of the software PRAAT.

The PRAAT acoustic software (Boersma, 2001) was designed by Paul Boersma and David Weenink from the University of Amsterdam. It comprises a wide range of features, such as spectral analysis, pitch analysis, formant analysis, intensity analysis, etc., and can be used for labelling, segmentation, filtering, and speech synthesis. Just like ELAN, it allows for multi-level analyses, by aligning audio data with textual or

phonetical tiers. While the present thesis does not linger on phonetic and acoustic aspects of (dis)fluency, the vocal dimension of these phenomena can be further illustrated in our qualitative analyses with regard to pitch and duration. These two features are part of a larger set of measures that can be carried out thanks to PRAAT, and are found on a spectrogram (cf Fig. 32 below). A spectrogram is a visual way of representing three main dimensions of speech and prosody, mainly, *fundamental frequency* (i.e., f0, or pitch curve on the vertical axis), *duration* (i.e., period of time, shown on the horizontal axis), and *energy* (using dark bands to stand for formants on the vertical slice). These specific features can contribute to the present analysis of inter-(dis)fluency as a *multimodal flow*. For instance, they can illustrate how the acoustic signal is momentarily *suspended* or *interrupted,* and how this suspension may synchronize with sudden shifts in eye gaze or head movement. This is illustrated in the example below.

In this excerpt, taken from pair A, Jessica is talking about a board game she had played with her brother and her boyfriend. As the transcript shows in line 2, Jessica produced a complex fluenceme sequence comprised of a prolongation and a self-interruption, which marked the interruption of her current utterance and the initiation of a new one (l.3) in which she changed the course of her narrative. As the picture within the transcript further illustrates, the speaker also moved her head towards her interlocutor and produced a kind of interactional phatic gesture (in the form of a pointing gesture directed upwards) to highlight a piece of information and present it to her partner (i.e. that they had won the game at the same time). What is interesting to note here is the multi-level and multi-modal interruption of the current activity, embodied within: (1) the interruption of the speech flow (interrupted utterance), (2) an interruption in the narrative flow (sudden change in the narrative), and (3) a sudden shift in visible bodily behavior.

Example Pair A (DisReg), Board Games

1 JESS: du coup je pensais gagner et à la fin (il) y a mon frère e:et mon copain qui se sont alliés contre moi.

→ 2 JESS: et du coup bah j'**ai:i +//.**

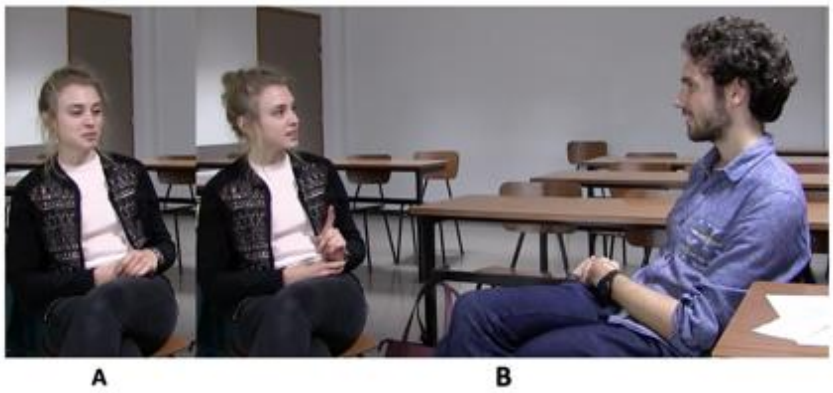
((gazes away)) A

3 JESS: nan on a gagné en même temps!

****

((turns her head and gazes towards interlocutor)) B

((index finger pointed upwards))



4 JESS: on a gagné exactement en même temps.

In addition to the written transcript and pictorial illustration, this multimodal analysis can be further enriched with a PRAAT window, provided below. The window shows a spectrogram and a waveform, as well as a TextGrid, taken from the original transcript. Here we are focusing specifically on the deployment of the fluenceme sequence (prolongation and self-interruption) and its subsequent utterance (*nan on a gagné en même temps*). The line represents the pitch curve, and as we can see, Jessica suddenly changed the pitch of her voice right after producing the fluenceme sequence, and at the start of her new utterance, when she changed the course of her narrative (marked by the negation marker "nan" in initial position). This is clearly visible in the pitch window, where the line is shown moving upwards, which enables us to visualize voice frequency measures. When the speaker produced the prolongation, her voice frequency was approximately 197 hz, while it increased up to 340 hz at the beginning of her following utterance[109].

[109] It should be noted that this analysis is only made possible with a formant ceiling, which refers to the maximum and minimum frequency of the formant search range. An average adult female speaker has a vocal tract length that requires an average ceiling of 5500 Hz (Praat's standard value), and an average adult male speaker has an average ceiling of 5000 Hz. There is, however, large variation between speakers, and if we choose too high a ceiling, we may end up with too few formants in the low frequency region. Cf Praat tutorial for more information https://www.fon.hum.uva.nl/praat/manual/Intro.html (last retrieved on August 26th 2021)

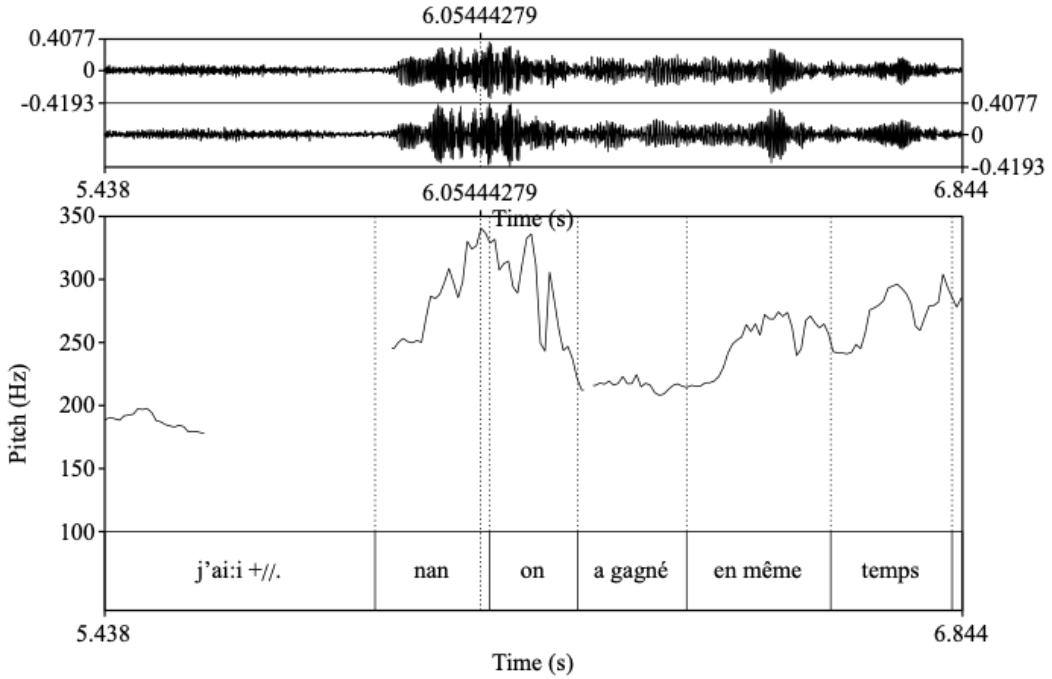


Figure 32. *Pitch analysis. Praat Window*

In sum, this brief acoustic analysis of inter-(dis)fluency phenomena further gives support to their multimodality and multidimensionality; while other acoustic features such as timbre (cf Vasilescu & Adda-Decker, 2007) have been thoroughly analyzed by phoneticians in previous disfluency research, the present study rather focuses on local durational and intonational patterns and the way they may synchronize with other types of bodily behavior. Here, the local change in pitch pattern synchronized with a change in the bodily activity, as well as a change in the current speaking activity.

Conclusion to the chapter

This chapter presented various aspects of our methodology, which includes practices around the act of transcription, our choice of data, the development of our annotation grid, and the application of interactionist and conversation-analytic methods for our multimodal qualitative analyses. Our methodology essentially reflects our integrated theoretical framework, blending different levels of analysis (vocal, verbal, visual-gestural, interactional), and methods (quantitative and qualitative). While we selected a relatively "small" video corpus to conduct our analyses (26,000 words, 2h30), we made sure that the choice of our sample was relevant for the implementation of our annotation scheme (i.e. similar corpus design and size between DisReg and SITAF). We further argued in favor of "small" corpora, which, we believe, can legitimately be

thoroughly exploited for both quantitative and qualitative analyses in the field of multimodal and pragmatic research, in line with Vaughan & Clancy (2013).

For our transcription method, we selected a multimodal unit of segmentation, thus relying on a wide range of criteria (i.e. acoustic, syntactic, semantic, interactional, gestural etc.) in line with Kendon (2004) and Debras (2013). We further developed a set of transcription conventions which included a mixture of different major transcription systems (i.e. CA, CHAT, and multimodal systems) in order to better represent the multimodal deployment of embodied fluencemes in situated discourse.

In line with previous work on disfluency research (e.g. Shriberg, 1994; Pallaud et al., 2019; Ginzburg et al., 2014; Crible et al., 2019, Candea, 2000) we implemented a multi-level annotation scheme of inter-(dis)fluency, which went through various technical and methodological changes, to better address our research questions and hypotheses. This model reflects different levels of analysis (i.e. individual marker, fluenceme sequence, and visual-gestural level), and the quantitative and statistical analyses were carried out with different tools (e.g. ELAN, CLAN, Excel, and statistical tools). They were further completed with qualitative analyses of the data, which rely on a careful observation of the timely social and institutional practices embedded within talk-in-interaction, in line with conversation-analytic methods. Our multimodal analyses were also enriched with the help of the Praat software.

Highlights of Chapter 2:

- We relied on semi-naturalistic data to conduct our corpus-based study on inter-(dis)fluency, using a video-taped dataset comprised of two comparable data samples.
- We argued in favor of "small" specialized corpora, which can still be used efficiently for quantitative and qualitative analyses.
- We used a multimodal transcription system to transcribe the data, relying on different criteria (intonational, syntactic, semantic, interactional, and visual-gestural).
- The present study applies a mixed-method methodology, drawing on quantitative and qualitative analyses.
- Our quantitative annotation scheme, which is adapted from previous work in (dis)fluency research, was further developed to include a multi-level analysis, including the level of individual fluencemes, fluenceme sequence, and visual-gestural level.
- All our analyses were conducted using several annotation and statistical tools, mainly CLAN, ELAN, Excel, and statistical tests.
- We used a reliable gesture functional classification system in order to provide a consistent overview of the overall gestural distribution in the dataset.
- For our qualitative analyses, we relied on conversation-analytic methods, drawing from analytic tools used in Conversation Analysis (single-case analyses, collection studies, data sessions). The software PRAAT was also used as an addition to our multimodal analyses.

Chapter 3. Inter-(dis)fluency in native and non-native discourse

Introduction to the chapter

The present chapter deals with the analysis of inter-(dis)fluency[110] and the distribution of ambivalent fluencemes in native and non-native discourse, based on the SITAF Corpus (cf Chap. 2, section 1.2), thus targeting aspects of L1 versus L2 uses in French and English. While a lot of research in L2 fluency has focused on the relationship between fluency and proficiency by examining the frequency of temporal variables in L2 versus L1 speech, the present study does not linger on proficiency measures specifically, but rather pays attention to the interplay of the different prominent features surrounding the construct of fluency, mainly gesture, gaze, and interactional dynamics. These features, which are essential aspects of our analysis (cf Chapter 2, section II and III) will be examined with respect to L1 and L2[111] fluency in our quantitative and qualitative analyses. Specific attention will also be paid to the situatedness of L2 discourse as grounded within tandem interactions, by taking into account the interactants' methods of participation in the course of their language practices. The general aim of this chapter is to introduce new methods for evaluating the degree of inter-(dis)fluency based on a multi-level scale (reported in the *General Conclusion*), with respect to fluenceme rate, visual-gestural behavior, and interactional dynamics. We further defend our view of inter-(dis)fluency as the result of multimodal and multidimensional processes (cf Chapter I, section IV), not restricted to temporal variables or speech errors, but as an interplay of vocal, visual-gestural, and interactional strategies.

This chapter is structured as follows: we begin with a review of the L2 Fluency research literature, and present our research questions and hypotheses (section I); then we report on our corpus-based findings regarding the distribution of fluencemes in native and non-native discourse, by integrating different levels of analysis

[110] While the term *inter-(dis)fluency* covers all the main aspects of this study (with the notion of interactional, gestural and functional ambivalence) the core term *fluency* will often be used in this chapter with respect to the field of L2 fluency research in SLA.
[111] *L1/L2* and *native/non-native* are interchangeable terms, as both of them are used very frequently in the literature.

(fluenceme, sequence, and gesture/gaze), extracted from our annotations (section II.2.1). These findings are then further exploited with fine-grained qualitative analyses of the data (section II.2.2.), drawing potential relations between the notions of *fluency*, *language proficiency, pedagogical intention*, *speakers' multimodal communication strategies*, and *interactional competence*. Lastly, we end this section with a discussion of our findings, addressing our research questions (section III). This chapter is also largely based on Kosmala (2019, 2020a, 2021) and Kosmala et al. (2019).

I. Literature Review

In Chapter I, we discussed the different definitions underlying the terms *fluency* and *disfluency*, and emphasized the fact that terminological differences ultimately reflected different theoretical orientations. As we have seen, the notion of fluency in Second Language Acquisition has often been associated with the "smoothness" of speech, as well as the ability to talk "fluently" in a second language (Chap. I, section II. 2.1 and 2.2), but these definitions also depend on whether the researcher adopts a "broad" or "narrow" view of fluency (Lennon, 1990), the former focusing on general aspects of oral proficiency, and the latter on one component of oral proficiency.

The present section will review a collection of studies in SLA and L2 fluency research relevant to our study of inter-(dis)fluency. Most researchers in SLA have been interested in finding objective measures of a learner's speech fluency, based on temporal variables, in order to evaluate their level of language proficiency. But more recently, others have also re-examined the notion of fluency by drawing on multimodal and interactional aspects of L2 use, thus viewing fluency as a collaborative problem-solving activity linked to communication strategies. Especially in a learning environment, L1-L2 discourse can, to a larger extent, be viewed as *pedagogical discourse*, in the sense that it relies on multiple pedagogical strategies. This is further elaborated in section 1.1. In section 1.2., we identify the different contributions from the L2 fluency research literature in order to present our research questions and hypotheses (1.3.). As we shall see, several of our hypotheses are also based on assumptions following Cognitive and Usage-Based Grammar introduced earlier (see Chap. 1, section 3.1.) as well as on previous research in (dis)fluency and gesture (see Chap. 1, section 3.3.4).

1.1. Tandem interactions and the notion of pedagogical discourse

In Chapter 2, we pointed out that tandem settings provided a relevant environment for language learning through practices of cooperation and socialization (cf Chap. 2, section I.1.2). During native and non-native face-to-face spoken interactions, meaning often has to be elaborated, adjusted, and co-constructed between the interactants. These interactions can be considered *asymmetrical* (Kurhila, 2001) as they involve one expert (a native speaker) and one novice of the language (a non-native speaker), which may result in interactional difficulties, and a "perceived imbalance" between the two interlocutors (Kurhila, 2001, p. 1088). In *tandem* interactions, where both speakers alternate between their native and non-native status, the language expertise of the participants is essentially contextual and temporary, and not institutionally defined as in teacher-class settings (Debras et al., 2020). These interactions are thus based on mutual solidarity, where both participants genuinely wish to learn their partner's native language (Horgues & Scheuer, 2015). One of the primary goals of the speakers is to achieve mutual understanding, and in order to do so, they may rely on several strategies, such as *foreigner talk* (Ferguson, 1975): when native speakers adapt their speech to the non-native speakers to make it easier to understand (e.g. slower speech rate, simpler vocabulary, louder speech, etc.). They can also use different types of gestures that are more adapted to second language learners (Adams, 1998). For instance, Adams (1998) found that native speakers produced more deictic and iconic gestures when addressing non-native speakers, as a way to promote inter-comprehension. Non-native speakers, on the other hand, may rely on *communication strategies* (Tarone, 1980), which are defined as possible solutions to lexical, grammatical and interaction-related problems. Such strategies include paraphrasing, substitution, appeal for assistance, etc., and speakers can also rely on additional multimodal resources to resolve these difficulties (Gullberg, 2011). This is further developed in section 1.3.

In sum, as native speakers constantly adapt and adjust their body and talk to facilitate production and/or comprehension, and non-native speakers rely on several strategies to deal with their own production, tandem interactions can be considered *pedagogical* to a certain extent. This relates to the notion of *secondary didacticity* (*didacticité seconde*), explored by Moirand (1993), which refers to discourse that is

not pedagogical by nature but motivated by a pedagogical intention. In this sense, the speakers' pedagogical intentions (i.e. to facilitate comprehension, or to deal with problem-solving and lexical related difficulties) may as well emerge in the context of tandem interactions. This stresses the idea that *pedagogical gestures* (Tellier, 2006, 2008a), which are commonly used by instructors within a classroom or training environment during (in)comprehension sequences (e.g. Holt et al., 2015), may also be mobilized specifically during tandem interactions. These kinds of gestures may also show different degrees of didacticity (Azaoui, 2015), depending on the pedagogical intention of the speaker, the context, the type of gesture, and the direction of gaze. Smotrova & Lantolf (2013, p. 398) further talked about the "pedagogical relevance of gesture both as an interactional tool between teachers and students and as a self-directed tool for thinking", and this notion can also be applied to native speakers' gestures. Once more, this idea of a continuum, from "fluent", communicative, pedagogical actions, to more "disfluent", self-oriented ones, is relevant to the present study of inter-(dis)fluency, which will also examine the different degrees of pedagogical and communicative intentions found during the deployment of ambivalent fluencemes. Tandem settings thus provide a relevant interactive framework for the analysis of inter-(dis)fluency, and this will be illustrated in our qualitative analyses. We shall now turn to the review of a number of studies in L2 fluency research in the following subsection.

1.2. Research on L2 fluency

1.2.1. L2 fluency, accuracy, and proficiency

As Eguchi (2016) pointed out, research on L2 fluency can be classified into two types: (1) the task-based language teaching framework, which examines fluency as dependent variables with independent effects of task manipulations (cf Skehan, 2003) and (2) the investigation of L2 fluency, and how the temporal differences between L1 and L2 speech may correlate with L2 proficiency (e.g. de Jong, 2016a) which could be a possible cause of disfluency rate. Previous research has investigated how L1 and L2 fluency may differ, and which temporal aspects of speech may determine overall perceptions of fluency. For instance, pausing phenomena in non-native speech is more likely to be related to vocabulary than other linguistic problems, suggesting that lexical retrieval is one of the biggest "obstacles" of L2 speech fluency (De Jong, 2016a; Witton-Davies, 2014). In line with previous research (e.g. Hilton, 2008; Witton-Davies, 2014),

Eguchi (2016) conducted a longitudinal study on three EFL learners in a Japanese university, and investigated the relationship between "breakdown fluency" (i.e. rate of filled and unfilled pauses, see Skehan, 2003[112]) and vocabulary, by paying specific attention to "lexical pauses" (see Cenoz, 1998 below), i.e. pauses related to lexical retrieval, which were found to be the most frequent, as opposed to other types of breakdown associated with syntactic or morphological errors.

However, while some researchers have noted language-specific features of (dis)fluency (e.g. Clark & Fox Tree, 2002; Maclay & Osgood, 1959), there are, to our knowledge, few empirical crosslinguistic studies on fluencemes which compared fluency rates in native and non-native productions *across languages*, except for Grosjean & Deschamps (1975), De Leeuw, (2007), Candea et al. (2005), Hai, (2017), and Peltonen (2020). For instance, De Leeuw (2007) compared the duration and frequency of vocalic hesitation markers in English, German, and Dutch, and found differences across language groups, with Dutch speakers who produced 10.1 hesitations per minute, versus Germans who produced an average of 6.3 per minute. German speakers also produced more nasal hesitations than English and Dutch speakers. In another study, Candea et al. (2005) compared the production of vocalic fillers in eight languages (Arabic, Mandarin Chinese, French, German, Italian, European Portuguese, American English and Latin American Spanish), and looked at three acoustic parameters (duration, pitch, and timbre). Their results showed timbre differences across languages, with for example Spanish which used a mid-closed vowel, while English used low-central vowels. Similarly, Hai (2017) found differences between Russian native speakers and Chinese non-native speakers with respect to vowel quality. In addition, Grosjean & Deschamps (1975) observed differences between French and English with regard to speech rate (i.e. French speakers spoke faster than English speakers) but also fluenceme distribution (e.g. more lexical repairs produced by English speakers than French speakers). However, none of these studies (except for Hai, 2017) have compared the rate of fluencemes produced by the same native and non-native speakers in different languages. The present study thus aims to bridge this gap by comparing native and non-native productions of fluencemes both in French and English.

[112] Skehan (2003) and Tavakoli & Skehan (2005) distinguished between three aspects of fluency, "speed fluency" (rate of speech), "breakdown fluency", and "repair fluency" (repairs, repetitions, reformulations, substitutions etc.)

A larger number of studies in SLA have reported differences in fluency rates in native versus non-native speech. For instance, Tavakoli (2011) conducted a study on oral narrative tasks performed by English native speakers and L2 speakers of English, and found differences in the distribution of pauses, as L2 learners tended to produce more pauses in the middle of clauses and fewer end-clause pauses than the native speakers. This showed that native speakers tended to pause more often at discourse boundaries, while non-native speakers paused more frequently within clauses. Tavakoli also distinguished between three types of pauses: (1) Replacement pauses—pauses followed by repetitions and replacements; (2) Reformulating pauses—pauses followed by a restart; (3) Online planning pauses—pauses used for planning. Her results indicated that L2 learners often paused before repeating a lexical item (1), when they abandoned a constituent and replaced it with another (2), and when they were formulating and planning their utterances (3). In sum, the author suggested that it was not the frequency of pauses that distinguished L1 from L2 speech, but rather their position in the utterance. In a similar vein, Rasier & Hiligsmann (2007) emphasized the "erroneous" use of pauses in L2 speech, as L2 speakers were more likely to produce pauses between words in the utterance, i.e. between the adjective and the noun, than native speakers. Moreover, Cenoz (1998) found that L2 speakers produced non-juncture pauses very frequently, which suggested planning problems. She explained that L2 speakers were more likely to use non-juncture pauses because they had to look for words "in a language in which they present limited proficiency" (1998, p. 03). She categorized three types of pauses: (1) lexical pauses— indicating problems in lexical retrieval; (2) morphological pauses—pauses followed by repetitions and self-corrections indicating problems at the morphological level; (3) planning pauses. Their results indicated that a majority of pauses produced by the L2 speakers served planning functions.

Another body of research has provided evidence of a higher rate of pausing, "hesitation", or "error" phenomena in L2 than L1 productions (see Brand & Götz, 2013; Deschamps, 1980; Fehringer & Fry, 2007; Matzinger et al., 2020; Riggenbach, 1991) Fehringer & Fry (2007) have found significant differences in the number of hesitation markers produced by bilingual speakers of German and English, with higher rates in their second language. De Jong (2016) further showed that high-proficiency Dutch learners produced fewer pauses than low-proficiency ones. Similarly, Riazantseva (2001) found that Russian learners paused more frequently in their L2 than in their

L1, and their pauses were also found to be significantly longer in their second language. This was also the case in Kahng's (2014) study of Korean learners, who produced pauses which were almost twice as long as the ones produced by the English native speakers. These studies have shown a strong relation between fluenceme rate/duration and proficiency (cf Riazantseva, 2001). In sum, non-native speakers are said to produce more pauses of longer duration and in mid-clause position before low-frequency words, and this can be explained by their limited proficiency of the language (Cenoz, 1998).

In addition, specific emphasis is laid on temporal aspects of spoken fluency, in other words, how fluency fits into models of spoken production, which includes, in parts, a semantic system and a phonological system (Levelt, 1999, cf Chap. 1, section I). As Hilton (2009) argued, many speech processes revolving around lexical retrieval, morphosyntactic encoding or phonological planning are carried out in L1 "without the need of attentional effort in the executive component of working memory" (Hilton, 2009, p. 645). Therefore, instances of retracing, repetitions, reformulations, pauses, and the like are often interpreted as a sign of encoding difficulties in the speech production process. In L2 production, the "network of automatically available lexical and morphophonological representations" is said to be "limited", as Hilton (2009, p. 646) further argued:

> We may follow similar procedures to structure concepts and discourse as in our L1, but encoding difficulties can provoke disfluency at every step of the formulation process: a concept may not activate the appropriate L2 lemma; the lemma may not activate appropriate syntactic, morphological, or phonological routines; and/mental combinations. The close examination of hesitation structures in L2 speech therefore constitutes a useful tool for identifying which processing components prove most problematic for learners.

Similarly, Dörnyei & Kormos (1998) claimed that many of the problem-solving processes emerging in L2 productions were the result of a *resource deficit*. This refers to the inability to retrieve the right lemma during a lexical search, difficulties in phonological encoding, or incomplete knowledge of the L2. Learners may thus resort to a series of "stalling mechanisms" e.g. "uh"/"um", lengthening, discourse markers (Dornyei & Kormos, 1998) to buy more time in speech, in order to deal with processing time pressure. Once more, the constructs of fluency and disfluency in SLA are deeply

grounded in theories of language production, by relating to "breakdowns", "processing time pressure" or "encoding difficulties" in the L2 speech production apparatus. However, it has been emphasized multiple times throughout this thesis that this *cognitive-burden* view of (dis)fluency (cf Chap. 1 section II. 2.2.1) only gives a partial picture of the phenomena under study. While previous research in SLA has given evidence that (dis)fluency rates could relate to perceived fluency and language proficiency (except for Brand & Götz, 2013 who did not detect any clear correlation), this kind of analysis should not be restricted to error and accuracy measures, but should include other crucial components of L2 performance phenomena, such as the interactive nature of face-to-face interaction (cf section 1.2.2. and 1.2.3.). In addition, the correlation between (dis)fluency rates and proficiency has also been criticized (e.g. Simpson et al., 2013) and it has been hypothesized that fluencemes in the L2 may mirror those produced in the L1, as they could be the result of similar cognitive processes (Zuniga & Simard, 2019). In this view, L1 and L2 fluency are said to be closely related (Derwing et al., 2009). In conclusion, the relationship between fluency, accuracy, and proficiency are not straightforward, and may also be related to other phenomena outside general cognitive processes. While these last aspects (proficiency and accuracy) are not central to the present research, our analysis of fluencemes and their distribution in L1 and L2 discourse may still contribute to the existing field of research in L2 fluency, by examining whether L1 and L2 (dis)fluency patterns strongly differ, and how they may do so, or whether they are closely related, not only on cognitive grounds, but on interactional ones as well; this leads us to the following section.

1.2.2. L2 fluency, interactional competence, and "CA-for-SLA"

Additional research in SLA has examined fluency in second language learning, but without focusing on processing difficulties, in line with psycholinguistic-cognitive approaches, but further grounded in an interactional framework, as to identify the different strategies used by learners to deal with problems in interaction (Gullberg, 2011; Tarone, 1980). As Gullberg (2011) pointed out, fluencemes may also relate to interaction-related difficulties, with for example the potential loss of face and floor, which puts learners at an interactional risk. This may prompt learners to engage in multimodal word searching practices, which involve the display of a thinking face (Goodwin & Goodwin, 1986, see Chap 1, section III.3.3.1) accompanied by a thinking

gesture (see Chap. 2, section II. 2.2.3). Specific attention is also paid to individual differences, which show how learners turn to a number of strategies that are very speaker-specific, reflecting their own communicative style, and which appears to determine L2 fluency behavior. Gullberg (2011) gave the example of a learner who had an issue with the term "prescription" and who produced a high rate of fluencemes, but without exploiting her multimodal resources. Her behavior was in fact found to be quite identical in her native speech. This may indicate that her L2 performance was not a necessarily a sign of limited L2 proficiency skills, but rather a reflection of her own individual preferences. This further justifies the need to conduct qualitative analyses in complement with quantitative observations of the data (see Peltonen, 2020). In addition, the analysis of L2 fluency is not only restricted to temporal variables, but includes other phenomena such as discourse markers, back-channeling and turn-taking. As Gürbüz (2017) argued, while the overuse of such phenomena (even in a native language) may be perceived as "disfluent" or inarticulate, no occurrence of them at all may be perceived as unnatural. Similarly, Gilquin (2008) conducted a corpus study on hesitation markers and "smallwords" (e.g. kind of, well, I mean) produced by French learners of English and native English speakers in interviews. Her study showed that pauses were very frequent among both native speakers and learners, but that the latter produced pauses more frequently overall. One interesting finding is that, while French-speaking learners overused pauses (both filled and unfilled), they did not make use of the full range of smallwords. In fact, they were extremely underused. She gives the example of "like", which was very common in native English speech, but almost absent in French learner speech. She added that filled pauses were crucial to non-native speakers as a conversational strategy, as they could be used to signal production difficulties to their conversational partner, but also to keep the floor or to be more polite, functions that also exist in native use. This functional approach to discourse markers, and more specifically (dis)fluency, which lies at the core of most corpus-based studies, aims to support the *ambivalent* view of fluencemes (cf Chap. 1, section II. 2.2.3.), and regards them as conversational tools. This ambivalence is also reflected in the work conducted on word searches in L2. On the one hand, word searches and their solutions are associated with communication strategies used to solve interactional difficulties (Kasper & Færch, 1983; Rydell, 2019); on the other hand, they are associated with "disfluency", which are treated as a deficit in the L2 (Dörnyei & Kormos, 1998).

A number of studies have also examined the way L2 learners may use fluencemes and *gestures* to deal with interactional related difficulties, with for instance the use of "uh" to start the conversation, or answer a question with the correct words (Azi, 2018), or the use of gestures as a communicative resource during repair practices in ESL conversational tutoring (see Seo & Koshik, 2010). In addition, micro analyses of interactional practices in L2 learning situations further shed light on the intersubjective role of gestures; they may be used to exhibit the learner's active co-participation in the language activity, or display alignment and achieve intersubjectivity through gesture replication or gesture co-production (see Belhiah, 2013). In sum, more and more studies in ESL and SLA have (re)considered the concept of L2 fluency, and now view it as a multimodal resource, or a strategy, rather than a cognitive deficit, with a strong focus on interactional data and collaborative aspects of fluency, or "confluence" (cf McCarthy, 2009; Chap. 1, section II. 2.2.) In line with this concept of confluence, Peltonen (2017, 2019, 2020) offered a *Fluency Resources Framework*, which links L2 fluency analysis to a broader perspective of communication, rather than solely from the perspective of temporal speech fluency. She also incorporated the analysis of gestures to study the way speakers make use of them to maintain fluency in interaction. More and more studies in L2 language testing have foregrounded the concept of *interactional competence* (e.g. Galaczi, 2014; Galaczi & Taylor, 2018) in line with CA and interaction research. The construct of interactional competence is defined as the following (Galaczi & Taylor, 2018, p. 226):

> The ability to co-construct interaction in a purposeful and meaningful way, taking into account sociocultural and pragmatic dimensions of the speech situation and event. This ability is supported by the linguistic and other resources that speakers and listeners leverage at a microlevel of the interaction, namely, aspects of topic management, turn management, interactive listening, break down repair and non-verbal or visual behaviours.

As Pekarek Doehler (2018, 2006) further argued, the notion of L2 competence needs to acknowledge the dynamic and adaptive nature of the linguistic system. In other words, linguistic knowledge is not only stored in a mental inventory, it is the result of a system of adaptive resources which are altogether determined by the local contingencies of the interaction, in line with the framework of *interactional linguistics* (see Chap. 1, section III. 3.2.). Pekarek Doehler further criticized the notion of

competence as individualistic, mentalist, or monologic, and decontextualized from practical actions and concrete situations. The notion of competence is much too often based on the perception of the ideal native speaker from the point of view of his or her production, but without taking into account the co-participant of the interaction. By adopting a conversation-analytic and interactional approach to L2 learning, hence *CA-SLA* (Pekarek Doehler & Pochon-Berger, 2011; Pekarek Doehler, 2006), the notion of L2 competence and fluency are thus "anchored in language use, that is, embedded in the moment unfolding of talk-in-interaction" (Pekarek Doehler & Pochon-Berger, 2011, p. 1). *CA-SLA* thus offers a "socially situated" view of learning (Mondada & Pekarek Doehler, 2004), which implies that the processes of L2 learning are only fully understood when "abstracted from their natural ecology, that is, the practices the learner engages in" (Pekarek Doehler & Pochon-Berger, 2011, p. 3). In sum, the analysis of L2 fluency does not only revolve around internal language processes or linguistic structures alone, but their intricate relation to the organization of actions during language practices. Pekarek Doehler & Pochon-Berger (2011) applied this method to their micro analyses of (dis)agreement sequences in French foreign language classroom interactions in German-speaking Switzerland. Their study identified different features of disagreement with regard to preference structure, turn allocation, but also linguistic properties, and they presented quantitative findings which provided a general picture of the techniques used for doing disagreement, as well as qualitative analyses which illustrated the specific turn construction methods adopted by learners to display disagreement. Our integrated theoretical model of inter-(dis)fluency (cf Chap. 1, section IV) is very much in line with this body of research which offers a valuable contribution to the field of SLA.

To conclude, the present study follows the approaches adopted by researchers in different disciplines from various research fields, from applied linguistics to conversation analysis, which bring together valuable insights on L2 fluency, proficiency, second language testing, and interaction. This combination of studies invites us to (re)consider the construct of L2 fluency without solely focusing on temporal variables or general accuracy rates, in line with our integrated theoretical model. As De Jong (2018, p. 14) further argued, fluency behavior is "in part dependent on personal speaking style". In addition, fluencemes are not only signals of trouble in processing and formulating, but can be "part of communicatively effective speech" (De Jong, 2018, p. 14). Once more, this notion of ambivalence is one of the most central

aspects of this thesis, and will be explored in both our quantitative and qualitative analyses.

1.3. Gesture production in Second Language Acquisition

In Chapter 1, we reviewed a number of studies which looked at the relationship between (dis)fluency and gesture (cf Chap. 1, section III. 3.3.4), some of which are concisely summarized in this section as a reminder. In brief, a number of studies have shown a temporal relationship between speech suspension and gesture suspension, (e.g. Esposito & Marinaro, 2007; Graziano & Gullberg, 2018; Seyfeddinipur, 2006; Yasinnik et al., 2005) with the observation that very few gestures accompany "fluent" speech, and a great deal of them tend to be suspended during "disfluent" stretches of speech. In this section, we will review a number of papers that focus more specifically on the role of gestures in L2 discourse, by drawing on a number of studies in psycholinguistics, second language testing, and interactional linguistics.

As we have seen earlier (cf Chap. 1, section 3.3.2), it has been proposed that iconic and deictic gestures are very often produced when speakers experience lexical problems, and that these gestures may help facilitate word finding (Beattie & Butterworth, 1979; Krauss & Hadar, 1999). Further in line with models of speech production (e.g. Levelt, 1989, 1999) a number of authors, such as Krauss & Hadar (1999) and Krauss et al., (1995) proposed that gestures are triggered by the activation of the spatial representation in the *conceptualizer* (one of the mental procedures involved in the planning of messages, see Levelt & Schriefers, 1987). This body of research is in line with the *cognitive-psychological* approach to gesture presented in Chapter 1 (see Chap. 1, section 3.3.2.), and follows the assumption that gestures serve a *compensatory* role in speech production. Krauss et al., (2000) offered the *Lexical Retrieval Hypothesis* (henceforth LHR), further in support of this model. According to this hypothesis, word findings are said to be more successful when accompanied by iconic gestures, as they facilitate access to lexical memory. In the field of SLA, one related question regarding gesture use is whether it can help learners resolve speech difficulties. Studies have reported a tendency for L2 learners to produce more gestures in their L2 than in their L1 to overcome language difficulties in their target language (e.g. Gullberg, 1998; Kita, 1993; Stam, 2006). Stam (2006, 2008, 2018) further demonstrated that the gestures produced by L2 learners provided an "enhanced window onto their mind through which we can view their thinking and mental

representations" (Stam, 2018, p. 165). In this view, speech and gesture are viewed as a single-integrated system (cf Chap. 1, section 3.3.2.) reflecting processes of L2 speaking, thinking, and learning. In line with Slobin's (1987) theory of *thinking for speaking* (cf Chapter 1., section 3.3.2), McCafferty (1998) investigated the gesture production of Japanese learners of English during a picture narration task. His findings showed that the L2 learners predominantly co-produced gestures with speech when engaged in problem solving activities. He also looked at cases of pausing and repairs (labelled "self-regulation"), and observed an absence of gestures during these instances. He concluded that speakers seemed to "look within themselves" (McCafferty, 1998, p.88) and thus did not "externalize" their thinking processes, as opposed to when they were gesturing. In another paper, following Kita's (2000)'s *Information Packaging Hypothesis* which claims that referential and deictic gestures constitute a spatio-motoric mode of thinking, McCafferty (2004) explored the way L2 learners used gestures to solve intrapersonal problems. He conducted another study based on interactions between Taiwanese learners of English in the United States, and argued that speakers used referential gestures (i.e. deictics and representational gestures) to provide "a greater degree of spatial exactness" (McCafferty, 2004, p. 163): because the learners experienced difficulties when speaking their L2, they resorted to the *spatio-motoric channel for thinking* (Kita, 2000), i.e. representational gestures which activate spatio-dynamic information (Krauss et al., 2000). This helped them to "orchestrate speech production in the L2 and to actionally structure the discourse" (McCafferty, 2004, p. 161). In sum, a large number of studies in SLA research have focused on the *intrapersonal* functions of gestures, as well as their cognitive aspects (read Gullberg & McCafferty 2008, for extensive review), and the role gestures play in L2 developmental processes, leading to a number of linguistic difficulties[113].

But as Lopez-Ozieblo (2019) pointed out, gestures used by L2 speakers do not only relate to lexical related problems, but can be used for turn-taking, repair, or discourse management, among other intersubjective actions (also see Chap. 1, section 3.3.4). In fact, a number of studies on gestures in SLA reject the assumption that L2 learners' gestures are used to overcome lexical shortcomings. For instance, Gullberg (1998, 2011) studied interactions of Swedish and Dutch learners of French and French

[113] A large body of research has also examined the facilitative role of gestures in L2 learning, with for instance the acquistion of L2 vocabulary for children and adults (e.g. Kelly et al., 2009; Tellier, 2008b) or L2 phoneme acquisition (e.g. (Hoetjes & Van Maastricht, 2020; Zhang et al., 2020). Read Hoetjes & Van Maastricht (2020) for an extensive review.

learners of Swedish, and found that while learners did produce gestures to resolve lexical difficulties in their L2, they were also used to elicit help from the interlocutor, and thus relied on multimodal communication strategies to manage the interaction. Solutions to such difficulties were characterized by "active co-constructions in which both participants jointly deploy speech, gaze, and representational gestures in a highly structured fashion" (Gullberg, 2011, p. 141). More recently, Graziano & Gullberg (2013, 2018) supported evidence against the LHR by examining the types of gestures that frequently occurred during fluencemes (cf Chap. 1, section 3.3.4). As we have seen, their study reported a high number of pragmatic gestures during fluencemes, which suggests that gestures produced in L2 were not necessarily related to lexical difficulties but rather to difficult aspects of interaction[114]. This does not support the LHR which expects referential gestures to predominantly occur during fluencemes, as to activate lexical items in the conceptualizer. In addition, for gestures to be truly compensatory, it would mean that they would have to occur *during* speech difficulties (i.e. during fluencemes), but as Graziano & Gullberg (2018) argued, the observation that gestures are more likely to occur with "fluent" rather than "disfluent" stretches of speech makes it difficult to assess theories such as the LHR. Once again, this "internalized" and production-based view of gestures only gives a partial picture of the phenomena. As emphasized throughout this thesis, the present work adopts an *integrated* approach to gesture and inter-(dis)fluency, which focuses on situated language use and interactional dynamics, (see Gullberg & McCafferty, 2008) and views gesture as integral to human communication, both with regard to *intrapersonal* and *interpersonal* processes.

To conclude, we can find a large body of research on gesture and SLA across disciplines and theoretical frameworks, and the studies presented here largely reflect the two dominant views presented earlier, mainly the *psychological-cognitive* and the *functional-communicative* view (cf Chap. 1, section 3.3.2). The present study on fluencemes and gestures in L2 discourse largely follows Graziano & Gullberg's work (2013, 2018) which rejects the compensatory role of gestures in L2 discourse. Further in line with the *functional-communicative* approach, we maintain that gestures perform a wide range of functions that are ultimately shaped by their context of use in the interaction, as well as their degree of expressiveness (Kendon, 2004), and their

[114] These findings were also found in Akhavan et al. (2016), among others (cf Chap 1. Section 3.3.4).

mode of representation (Müller et al., 2013) and that such manual actions directly contribute to gesturers' multimodal utterances (Kendon, 2004, 2014). Therefore, the idea that gestures "compensate" for speech is, in our view, inadequate, as we believe that it is more a matter of *selecting* among the *relevant scope of behaviors* (Cienki, 2015, cf Chap. 1, section III. 3.2.1), whether a gesture is deemed more relevant to enact a specific action in a specific context (e.g. a word searching sequence) rather than speech alone, or the other way around. This selection is essentially shaped by the coordination of the participants' actions and their positionings in an interactional sequence.

1.4. Research questions and working hypotheses for the present study

To sum up, a large body of research in SLA has focused on temporal variables and proficiency measures when studying L2 fluency (e.g. Cucchiarini et al., 2000; Götz, 2013; Hilton, 2009; Lennon, 1990, among others) and have shown differences in fluencemes' distribution between L1 and L2 (e.g. Fehringer & Fry, 2007; Riggenbach, 1991; Tavakoli, 2011). Therefore, a number of researchers have mainly associated L2 fluency with lexical and encoding difficulties (Dornyei & Kormos, 1998, Krauss et al., 2000). However, another body of research has focused on interactional and communicative aspects of L2 fluency as to identify the different multimodal communication strategies learners may mobilize when dealing with interactional difficulties (Gullberg, 2011, Tarone, 1980). In this view, L2 fluency is largely associated with the notion of *interactional competence* (Galaczi, 2014) which focuses on the ability to co-construct interaction rather than to deal with difficulties alone; this idea also foregrounds the notion of *CA-SLA* (Pekarek Doehler & Pochon-Berger, 2011) which associates L2 competence with language use, and elaborates on the notion of interactional competence by embracing conversation-analytic methods (Pekarek Doehler, 2018). Therefore, the notions of L2 fluency, proficiency, or competence are not context-independent, but context-bound, based on observable practices that illustrate speakers' abilities to co-construct meaning during situated tandem interactions, and to a larger extent in *pedagogical* settings (cf section 1.1.) In addition, we reviewed a number of studies on gestures in SLA, and briefly presented the *Lexical Retrieval Hypothesis* (Krauss et al., 2000) which claims that (referential) gesture production facilitates word finding, as well as *the Information Packaging Hypothesis*

(Kita, 2000) which claims that (referential) gestures help speakers organize spatio-motoric information. However, we also presented a different body of research which does not agree with these views, and pays close attention to pragmatic gestures and the way they synchronize with (dis)fluent cycles of speech (Graziano & Gullberg, 2013, 2018).

The selection of studies presented above reflect different views and approaches to L2 fluency and gesture, from psycholinguistic perspectives to conversation-analytic ones, which have led us to the construction of our integrated framework of inter-(dis)fluency (cf Chap. 1, section IV), moving now to our research questions and hypotheses. Our research questions stem from this large body of research, and aim to address the specificities of non-native versus native fluency. Our hypotheses are also based on assumptions following cognitive grammar and usage-based linguistics introduced earlier (cf Chap. 1, section III. 3.1). The general aim of the present research is to measure the overall distribution and general pattern of behavior of fluencemes in native and non-native discourse, and English and French, as to discuss potential differences between L1 and L2 fluency (quantitative analyses), but also to highlight their multimodal deployment in micro-analyses of the data, captured in situated tandem interactions (qualitative analyses). Four central research questions thus emerge, which comprise a series of subquestions, listed below.

> **RQ1:** Are L1 and L2 fluency closely related? If not, how does L2 fluency differ from L1 fluency, and how would these differences be characterized?
> **RQ2**: Is L2 fluency necessarily associated to the learners' proficiency levels?
> **RQ3**: How do L2 learners make use of fluencemes to overcome language difficulties? Do they use gestures to "compensate" for lexical shortcomings?
> **RQ4**: To what extent does visible bodily behavior play a role in the functional and interactional ambivalence of inter-(dis)fluency?

To answer these questions, we present the following assumptions and hypotheses:

RQ1:

- We expect several differences between L1 and L2 fluency behavior, mainly differences in frequency, duration, position (in line with De Jong, 2016b; Ferhinger & Fry, 2007, Gilquin, 2008, Tavakoli, 2010, among others), but also sequence complexity (i.e. sequence configuration and number of markers combined within a sequence).

- Crosslinguistic differences between French and English are also expected, following the assumption that (dis)fluency is language-specific (Grosjean & Deschamps, 1975; De Leeuw, 2007, Candea et al., 2005).
- We also hypothesize that L2 learners will produce fewer recurrent patterns of combination, and more complex fluenceme sequences. In line with Crible (2018), rare combinations are less automatic, less entrenched in speakers' memories, and may therefore be one characteristic of L2 fluency.
- We also expect differences in gestural behavior, with a higher rate of gestures produced in L2 than in L1 both during fluencemes and outside fluencemes. However, contrary to the LHR, we do not expect L2 learners to produce more referential gestures than L1 speakers.

RQ2:

- In line with Brand & Götz (2013), we do not expect a clear correlation between fluency and proficiency, but rather expect a high number of individual differences (De Jong, 2018, Gullberg, 2011).
- In addition, while L1 and L2 fluency may show overall differences (in line with RQ1) it does not mean that a higher rate of fluencemes in L2 is systematically associated with limited proficiency. These differences may also reflect individual strategies (Gullberg, 2011), language specificities (Grosjean & Deschamps, 1975), or different degrees of pedagogical intention (Azaoui, 2015; Moirand, 1993). These differences thus need to take into account not only temporal variables and frequency measures, but interactional factors as well, in line with the notion of *interactional competence* (Galaczi & Taylor, 2018) and *CA-SLA* (Pekarek-Doehler, 2006).

RQ3:

- In line with Cienki's (2015) dynamic *scope of relevant behavior* theory, we expect speakers to use the spoken modality as a default mode for dealing with language difficulties, but there can also be specific instances during which they make use of visible bodily behavior to act on their multimodal communication strategies (Gullberg, 2011, 2014). In sum, we expect L2 learners to mobilize all available relevant resources (i.e. speech, manual gesture, gaze, body movement) along with fluencemes to deal with language difficulties.
- Contrary to the LHR, we also expect L2 speakers to use more pragmatic gestures than referential gestures during fluencemes, either to seek help from

their interlocutor, to enact a speech act, to embody a word search activity, or to mark distinct aspects of discourse, depending on the context.

RQ4:

- The ambivalence of fluencemes, that is to say, the degree of communicativeness, conventionalization, and fluency, will be illustrated in speakers' individual strategies during micro-analyses of interactional sequences. While some speakers will only focus on the spoken and vocal modality of discourse and produce a high number of fluencemes to disengage from the interaction and deal with linguistic problems alone (DISfluency), others will make use of gaze and gestures to display their engagement to the interlocutor (Goodwin & Goodwin, 1986) and invite them to construct meaning in tandem (FLUENCY).

II. Results

2.1. Quantitative findings

This section presents the quantitative findings extracted from our annotations, which were obtained through several statistical treatments. Section 2.1.1 focuses on the fluenceme level of analysis (fluenceme rate, type distribution, and duration of vocal markers), and section 2.1.2. looks into the combination level (simple/complex sequences, number of markers combined within a sequence, sequence configuration, and sequence position). Finally, we finish this section with the visual-gestural level of analysis (in section 2.1.3.), which includes gesture analysis and gaze behavior during fluenceme sequences ("fluent" stretches of speech) and outside fluenceme sequences ("disfluent" stretches of speech, Graziano & Gullberg, 2018). Our analyses compare native versus non-native productions, broken down by language (English versus French) to measure crosslinguistic differences, as well as proficiency levels. Results are provided in raw values[115] and relative/normalized frequency (i.e. a frequency relative to some other value as a proportion of the whole, such as the number of words in the corpus, or a total number of tokens). Our basis of normalization for the rate of

[115] All the statistical analyses were run on raw numbers, but most of the tables and figures provide relative values, for ease of exposition. Raw values can be found in Appendix 3.

fluencemes and gestures is *per hundred words* (henceforth phw) due to the small size of the data (i.e. 141,121 words, see Chap. 2, section I.1.4).

2.1.1. Marker level: rate, form, and duration of individual fluencemes

A total of 3172 fluencemes were annotated and extracted from the SITAF Corpus, with 1173 tokens found in L1 (English and French) and 1999 in L2 (English and French). Figure 33[116] reports the distribution of fluencemes in relative frequency per participant (American and French) and per language proficiency (L1 versus L2). A test of log-likelihood (henceforth LL, see Chap. 2, section II.2.3.1.) was conducted to measure these differences statistically. Results indicate that American speakers produced significantly more fluencemes in their L2 (49.2 phw) than in their L1 (16.1 phw) (LL = 495.45 , p < 0.0001), and this was also the case of French speakers who produced 23 fluencemes phw in their L1 as opposed to 33.1 phw in their L2 (LL= 48.69 ; p < 0.0001). We can also observe differences between American and French speakers, with the American speakers who produced significantly more fluencemes in their L2 (49.2 phw) than French speakers (33.1 phw) (LL = 77.24 ; p > 0.001), while they produced fewer ones in their L1 (16.1 phw) compared to French speakers (23 phw) (LL = 104. 88 ; p > 0.001).

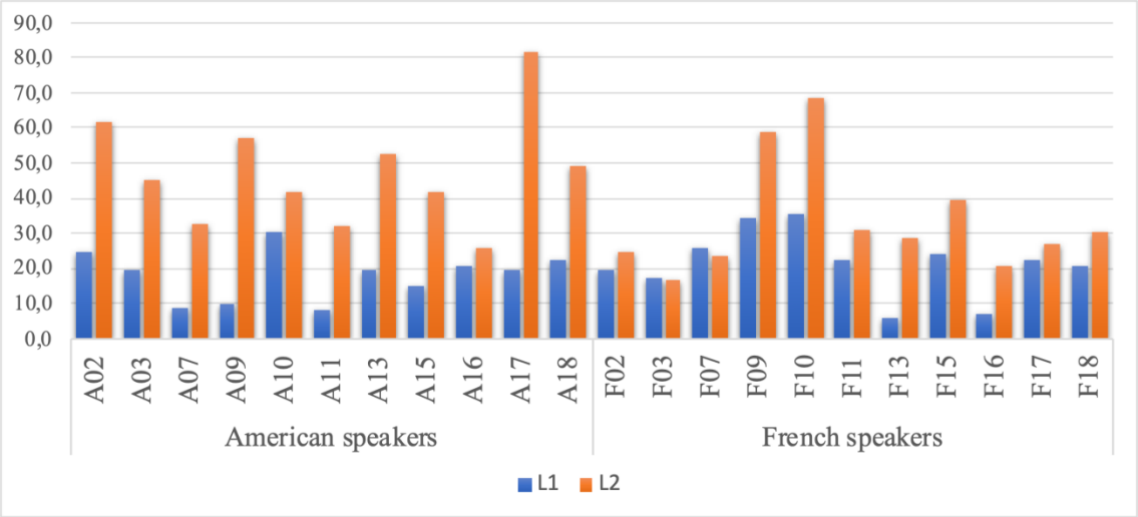


Figure 33. *Rate of individual fluencemes per hundred words*

Overall, results show high differences in frequency between native and non-native productions, as well as between speaker groups. We can note, however, a number of individual differences and instances of speaker variability within the two groups. For

[116] The exact values (relative and raw) are found in Table 51 in Appendix 3.

instance, A07 produced 32.6 fluencemes phw in her L2, as opposed to A17, who produced significantly more tokens (81.6 phw). Similarly, F13 only produced 5.9 fluencemes phw in her L1, while F10 produced 35.2.

Table 11. *Proportion of marker types in L1 and L2 (American speakers)*

Marker type	L1 % (raw)		L2 % (raw)		z score	p value
EDT	0.3%	(2)	1%	(6)	N/A[117]	
MS	29.4%	(152)	28%	(330)	0.68	0.4
NL	9.3%	(48)	10%	(114)	-0.2	0.8
VOC	60.8%	(314)	62%	(736)	-0.47	0.6

Table 12. *Proportion of marker types in L1 and L2 (French speakers)*

Marker type	L1 % (raw)		L2 % (raw)		Z score	p value
EDT	0.3%	(2)	1%	(8)	N/A	
MS	29.6%	(195)	24%	(196)	2.24	0.01*[118]
NL	7.3%	(48)	12%	(94)	-2.74	0.006*
VOC	62.7%	(412)	63%	(515)	-0.25	0.8

Tables 11 and 12 compare the distribution of fluencemes by marker type: vocal markers (VOC), morphosyntactic markers (MS), and peripheral markers, which include explicit editing terms (EDT) and non-lexical sounds (NL) (see Chapter 2. section 2.2.1)

It is interesting to note that, despite significant differences in frequency between L1 and L2, no differences were found in the proportion of marker types for the Americans, as the *z* tests (see Chap. 2, section II.2.3.1.) yielded no significance between native and non-native productions (see Table 11). For the French speakers however, two differences were found: they produced slightly more morpho-syntactic markers in their L1 (29.6%) than in their L2 (24%), and they produced almost twice as many non-lexical sounds in their L2 (12%) than in their L1 (7.3%).

A more detailed account of the proportion of marker types is provided in Table 13 and 14 which identify all the fluencemes that were annotated in the data (excluding EDTs and additions which were extremely rare overall).

[117] Z tests cannot be conducted on values no greater than 5.
[118] An asterisk is added whenever the *p* value is below 0.04 and therefore shows significance.

Table 13. *Proportion of fluencemes in L1 and L2 (American speakers)*[119]

	L1 % (raw)		L2 % (raw)		Z score	p value
morpho-syntactic markers						
lexical repair	1.7%	(9)	0.9%	(9)	1.43	0.1
morphological repair	2.5%	(13)	2,5%	(25)	0.02	0.9
syntactic repair	2.7%	(14)	4,8%	(48)	-1.94	0.05
identical repetition	11.7%	(60)	18,2%	(181)	-3.27	0.001*
self-interruption	4.7%	(24)	1,7%	(17)	3.35	0.0008*
truncated word	5.7%	(29)	4,9%	(49)	0.59	0.5
vocal markers						
filled pause	9.6%	(49)	19,4%	(193)	-4.94	< 0.002*
prolongation	14 %	(72)	28,4%	(283)	-6.25	< 0.002*
unfilled pause	37.9%	(193)	26,1%	(260)	-4.58	< 0.002*
peripheral markers						
NL sound	9.4%	(48)	11,4%	(114)	-1.26	0.2

Results show that, for the American speakers, differences were found in the proportion of certain fluencemes in their L1 and L2, especially morpho-syntactic and vocal markers. They produced more identical repetitions in their L2 (18.2%) than in their L1 (11.7%), but more self-interruptions in their L1 (4.7%) than in their L2 (1.7%). In addition, they produced more filled pauses in their L2 (19.4%) than in their L1 (9.6%) as well as more prolongations (28.4% versus 14%). However, they produced more unfilled pauses in their L1 (37.9%) than in their L2 (26.1%). No significant differences were found for the rest of the markers.

For the French speakers (Table 14), on the other hand, fewer significant differences were found. As Table 14 reports, French speakers produced slightly more self-interruptions in their L1 (3.6%) than in their L2 (1.8%), but contrary to the American speakers, they produced more prolongations in their L1 (18.6%) than in their L2 (12.1%), and fewer unfilled pauses in their L1 (23%) than in their L2 (29.1%). They also produced more NL sounds in their L2 (11.6%) than in their L1 (7.3%). We can thus find a number of crosslinguistic differences between the two groups: while American

[119] Rates per hundred words can be found in Appendix 3, Table 53 for the French speakers, and Table 52 for the American speakers. Tables 13 and 14 of this section look at the proportion of fluencemes, to get an idea of their frequency relative to one another in a specific language.

speakers produced a relatively high proportion of unfilled pauses in their L1 (37.9%) it was not the case for French speakers (21.6%) (z = 5.48 ; p < 0.002) ; on the other hand, French speakers produced significantly more filled pauses in their L1 (21.6%) than American speakers (9.6%) (z = -5.50 ; p < 0.002).

Table 14. *Proportion of fluencemes in L1 and L2 (French speakers)*

	L1 % (raw)		L2 % (raw)		Z score	p value
Morpho-syntactic markers						
lexical repair	1.6%	(11)	0.6%	(5)	1.94	0.05
morphological repair	1.6%	(11)	3.1%	(25)	-1.72	0.08
syntactic repair	4.1%	(27)	2.4%	(20)	-1.79	0.07
identical repetition	13.5%	(88)	11.8%	(95)	0.99	0.3
self-interruption	3.6%	(24)	1.8%	(15)	2.14	0.03*
truncated word	4.4%	(29)	4.4%	(36)	-0.01	0.9
Vocal markers						
filled pause	21.6%	(141)	22.6%	(182)	-0.41	0.6
prolongation	18.6%	(121)	12.1%	(98)	3.41	0.0006*
unfilled pause	23%	(150)	29.1%	(235)	-2.62	0.008*
Peripheral markers						
NL sound	7.3%	(48)	11.6%	(94)	-2.74	0.006*

The duration of the vocal markers was also analyzed for all speakers of both groups. Results show that, despite differences between the two groups on average, a great number of individual differences prevented the former from being statistically significant (cf Table 54, 55, and 56 in Appendix 3 for more details). For the American speakers, filled pauses were on average longer in their L1 (M = 658ms ; SD = 238ms) than in their L2 (M = 514ms ; SD = 192,3ms ; p = 0.1), as well as unfilled pauses which had a longer duration on average in their L1 (M = 754ms ; SD = 444ms) than in their L2 (M = 683ms ; SD = 300ms ; p = 0.2). For the French speakers, it was quite the opposite: their filled pauses were longer in L2 (M = 465ms ; SD = 190ms) than in L1 (M = 371ms ; SD = 214ms ; p = 0.1), and this was also the case for their prolongations (M = 383ms ; SD = 122ms in L1; M = 459ms ; SD = 187ms in L2 ; p = 0.01) and unfilled pauses (M = 629ms ; SD = 292ms in L1; M = 717ms ; SD = 443ms in L2 ; p = 0.4).

As the Shapiro-Wilk test revealed, neither filled pause, prolongation, nor unfilled pause duration were normally distributed (W= 0.97 ; W = 0.80 ; W = 0.86),

so the duration values (aggregated per speaker) were submitted to a Wilcoxon test for comparison of means (cf Chap. 2, section II.2.3.1.). Given the lack of statistical evidence (except for prolongations in French), hardly any conclusion regarding duration can be reached at this point. This may suggest that duration is not necessarily a reliable indicator of (dis)fluency, contrary to what previous studies have shown (e.g. Kahng, 2014; Riazantseva, 2001). This is more in line with Cucchiarini et al. (2000) who found that the difference between native and non-native speakers was more determined by a greater number of pauses than a longer duration. These findings also further suggest high individual variation, as shown in the boxplots in Figures 34 and 35 which give information about the variability and dispersion of the data. The lower part of the box displays the first quartile, the higher part shows the third quartile, and the line dividing them is the median (the middle value of the dataset). The upper and lower whiskers represent scores outside the middle 50%, and the data points that are located outside the whiskers show observations that are distant from the rest of the data. In this case, the data shows a skewed distribution (where the median cuts the box into two unequal pieces, e.g. prolongations in the French group), with some data points that are located further outside the upper quartiles (e.g. for the unfilled pauses for the French and American group).

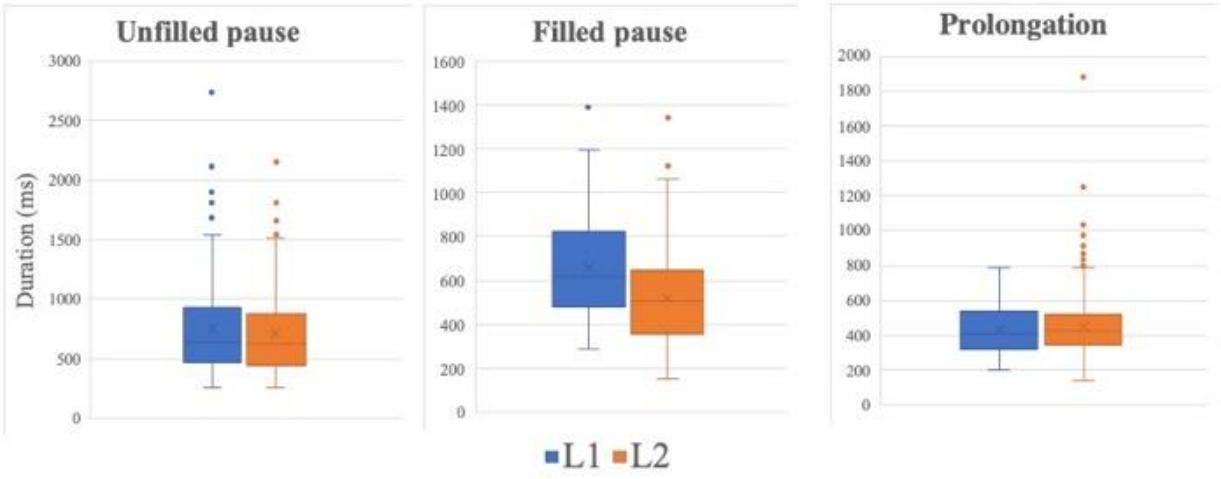


Figure 34. *Duration of vocal markers in L1 and L2 (American group)*

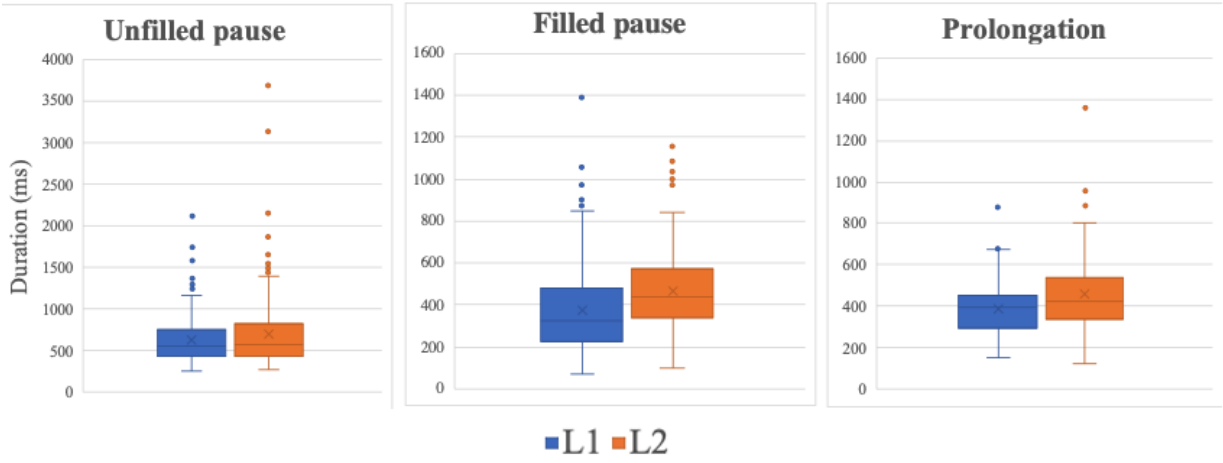


Figure 35. *Duration of vocal markers in L1 and L2 (French group)*

Turning now to the form of the filled pauses (see Fig. 36), as either produced with a central vowel *(e)uh)* or a nasalized one *(e)um*[120]), results show that American speakers produced more instances of "ums" in their L1 (N= 43/49) than in their L2 (N=62/193) (z = 7.017 ; p = <. 0002), but more "uhs" in their L2 (131/193) than in their L1 (6/49) (z = -7.017 ; p = <. 0002). The French speakers, however, showed an opposite trend: they produced more "euhs" in their L1 (N= 125/141) than in their L2 (N=115/182) (z = 5.195 ; p <. 0002), but more "eums" in their L2 (N = 67/182) than in their L1 (N = 16/141) (z = -5.195 ; p <. 0002). We may wonder whether this result may be linked to language-specific features; while American speakers used "um" a lot in their first language (88%), but French people rarely did (11%), the latter may have produced more "ums" in their L2 to adapt their speech to "sound" more like the Americans, and the other way around (i.e. Americans may have produced more "euhs" in their L2 to adapt it to the French). This kind of phenomenon is known as *phonetic adaptation* (Hwang et al., 2015), and refers to the production of L2-like sounds that are missing in the L1 in order to adapt it to the target language. While this kind of hypothesis needs to be further investigated by running a thorough acoustic and phonetic analysis of filled pauses (which calls for a different type of investigation), it is still interesting to note significant differences in form distribution between the two groups and between L1 and L2.

[120] As explained in Chapter 2, no acoustic differences were made between English-sounding *uh/ums* and French-sounding *euh/eums*, so a bracket is used to avoid making a distinction in writing.

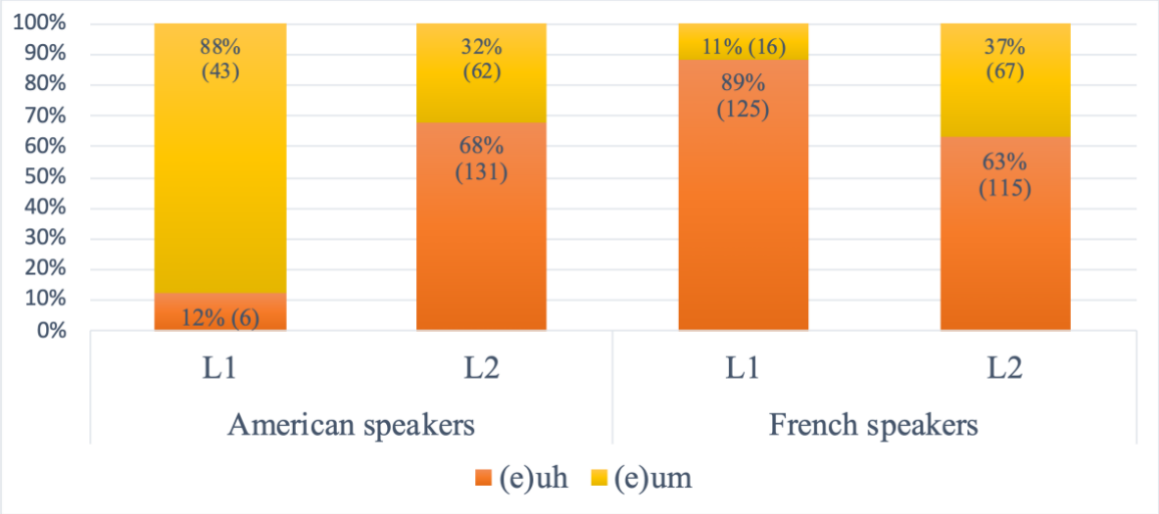


Figure 36. *Proportion of filled pause types in L1 and L2*

As to the types of non-lexical sounds, three main categories emerged: (1) clicks, (2) inbreaths, and (3) other. The latter includes various sounds and sound objects, i.e. sigh, nasal vocalizations, laughter, coughs, creaks and the like. They were grouped together because they did not often co-occur with fluencemes, as opposed to clicks and inbreaths, but their exact distribution can be found in Tables 57 and 58 in Appendix 3. Figure 37 shows the proportion of non-lexical sounds in L1 and in L2. While numerical values seem to suggest some differences (e.g. 27% of clicks in L1 versus 39% in L2 for the Americans), none of them actually reached statistical significance, as reported in Table 59 in Appendix 3 which shows the scores obtained with the z tests.

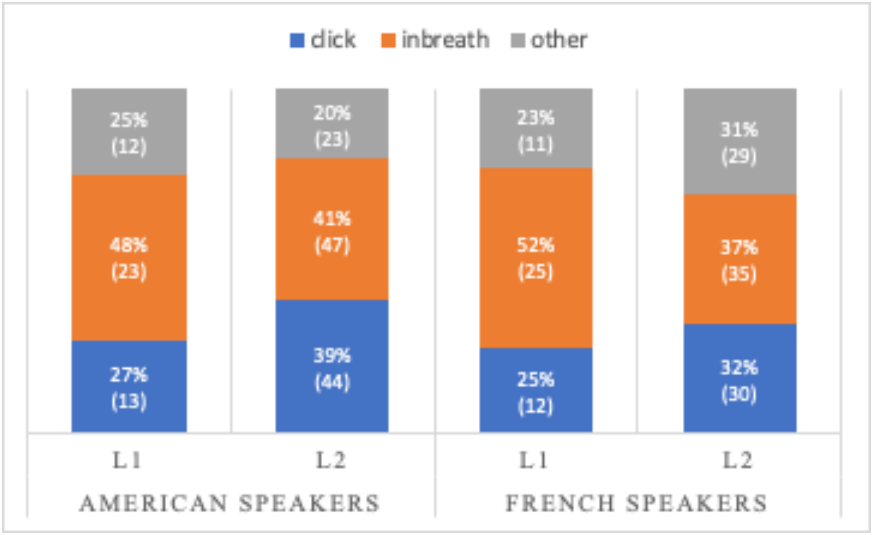


Figure 37. *Percentage distribution of NL sounds in L1 and L2*

Unlike filled pauses, the differences between the two means in L1 and L2 were not significant, and this may be interpreted as a sign that non-lexical sounds are not

necessarily language-specific, as opposed to filled pauses. However, we should be cautious when making this assumption, given the lack of statistical evidence. Once more, these kinds of results call for further investigation in phonetic research, which goes beyond the scope of the present study.

Before moving on to the *sequence* level of analysis in the following section, let us briefly investigate the possible relationship between fluenceme rates and proficiency levels, in line with previous research in SLA on fluency (see section 1.2).

Table 15. *Self-evaluation scores and L2 fluenceme rate*

	Self-evaluation score (oral production)	**Self-evaluation score (listening comprehension)**	**L2 rate (phw)**
A02	6/10	7/10	61.7
A03	6/10	7/10	45
A07	6/10	8/10	32.6
A09	7/10	8/10	56.8
A10	6/10	6/10	41.7
A11	6/10	8/10	32.2
A13	7/10	7/10	52.6
A15	7/10	8/10	41.7
A16	4/10	8/10	25.5
A17	6/10	7/10	81.6
A18	5/10	7/10	48.9
F02	7/10	8/10	24.8
F03	7/10	7/10	16.5
F07	7/10	7/10	23.3
F09	7/10	7/10	58.7
F10	5/10	5/10	68.2
F11	6/10	6/10	31
F13	8/10	7/10	28.8
F15	7/10	7/10	39.4
F16	8/10	7/10	20.9
F17	8/10	7/10	26.9
F18	6/10	7/10	30.4

As explained in Chapter 2 (cf Chap. 2, section I.1.2.3.) the students who took part in the study assessed their own level of proficiency, by rating it on a scale from 0 to 10. Even though these scores are very subjective and by no means provide a reliable

measure of L2 proficiency, it is still interesting to enquire into a potential relationship between perceived proficiency and fluency.

Table 16. Pearson R scores and p values for the correlation tests

	American speakers	**French speakers**
Oral prod. score and L2 rate	r = 0.1315 ; p = 0.2	r = 0.5779 ; p = 0.04
Listening comp. score and L2 rate	r = -0.3956 ; p = 0.2	r = -0626 ; p = 0.05

To that aim, a Pearson's correlation (cf Chap. 2, section II. 2.3.2.) was conducted to measure a possible correlation between fluenceme rates and self-assessed evaluation scores. These values are given in Table 15, as a reminder. It is interesting to note, at first glance, that the American speaker (A16) who was attributed the lowest self-evaluation score for oral production (4/10) actually produced the lowest rate of fluencemes in her group (25.5 phw). In the French group, on the other hand, the French speaker (F10) who was attributed the lowest self-evaluation score for oral production (5/10) produced the highest rate of fluencemes in her group (68.2 phw). This brief look at the data already suggests individual and/or language-specific differences. Indeed, none of the scores found in Table 16 yielded a substantial significant correlation between the two means, according to the Pearson's correlation test. This is further discussed in Section III.3.1.2.

So far, the findings reviewed above suggest that the most striking difference regarding fluenceme distribution in native versus non-native productions is *frequency*, as all the speakers from the two groups showed a tendency to produce considerably more fluencemes in their L2 than in their L1, which was found to be statistically significant. When it comes to the other features (i.e. duration, form, proficiency), the differences between L1 and L2 patterns of behavior are not so straightforward, as they largely depended on other variables, such as language (English versus French) and individual variation. These differences will be further discussed in section III.

2.1.2. Sequence level: type, length, position, and patterns of co-occurrence

1567 sequences (i.e. isolated or combined markers, see Chap. 2, section II.2.2.2.) were identified in total, 821 for the American speakers, and 746 for the French speakers.

Figure 38 reports the proportion of *simple* (isolated tokens) and *complex* (combined markers) sequences in L1 and L2 for the two groups. Numbers show that the American speakers produced more complex sequences in their L2 (N= 302/512) than in their L1 (N=133/309), which was found to be statistically significant (z = -4.435 ; p <. 0.0002).

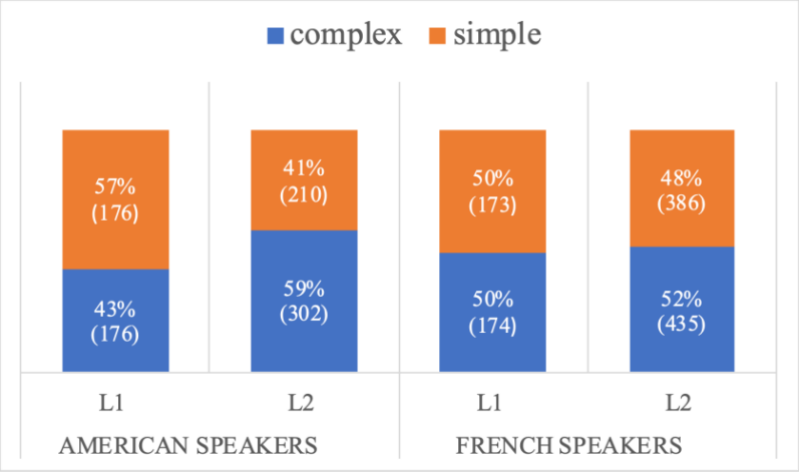


Figure 38. *Proportion of complex and simple sequences in L1 and L2*

For the French speakers, however, no significant difference was found between the proportion of complex sequences in their L1 (N=173/347) and in L2 (N=206/399) (z = -0.0177 ; p =0.6). Once again, these results may indicate that (dis)fluency behavior is not necessarily determined by differences in proficiency, but may also be influenced by language differences and/or individual preferences.

Turning now to the length of sequences, i.e. the number of markers found within a complex sequence, Table 60 (in Appendix 3) gives information about sequence length, i.e. average number of markers combined within a sequence. As the Shapiro-Wilk test revealed, the values were not normally distributed, so a Wilcoxon test was performed to measure differences between L1 and L2. As results showed, American speakers combined 1.7 markers on average in their L1, versus 2.4 in their L2, which was statistically significant (p = 0.005), but French speakers produced 1.9 markers per sequence in their L1 versus 2 in their L2 on average, which does not show significance (p = 0.5). However, the standard deviation values are rather high in L2 for both the American and French speakers, which suggests a large amount of variation within the two groups. This is further illustrated in the boxplots below.

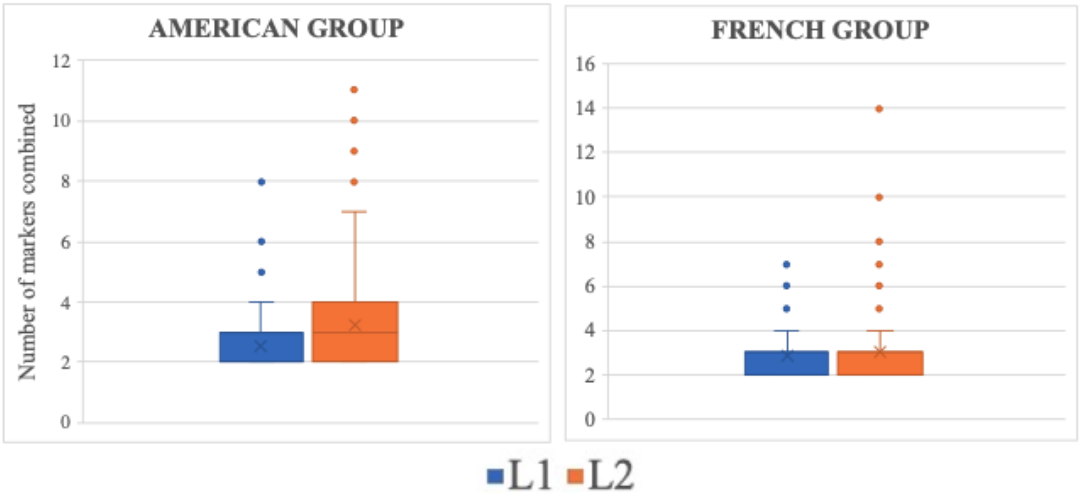


Figure 39. *Range of markers combined in L1 and L2*

The number of markers combined ranged from 2 to 8 in L1 versus 2 to 11 for the Americans, and from 2 to 7 in L1 versus 2 to 14 for the French. Some speakers, such as F13, combined up to 3 markers in her L1, versus 14 in her L2, and on average combined a higher number in her L2 as well (1.6 in L1 versus 2.2 in L2), others, such as F11, combined up to 5 markers in her L1, as opposed to 10 in her L2, but combined roughly the same amount on average (2.3 in L1 vs 2.4 in L2). Similarly, A17 combined up to 10 different markers in her L2, as opposed to 3 in her L1, while A03 combined up to 8 markers in her L1, and 10 in her L2. These different patterns of behavior largely reflect how spread out the data is, but it may also suggest individual preferences, which are not directly observable in overall measures of frequency or tendency. This further justifies the need to illustrate specific instances of the data through qualitative analyses (see section 2.2.).

Tables 17 and 18 show the different sequence configurations (cf Chap. 2, section 2.2.2.) and their distribution in L1 and L2 for the American and French speakers. Once more, the differences are more significant within the American group, as nearly half of their sequences (48%) are made of vocal markers and morphosyntactic markers (VOC+MS) in their L2 as opposed to 29% in their L1. By contrast, 35% of their sequences consisted in combinations of vocal markers (VOC+VOC) in their L1, as opposed to 20% in their L2. In short, these findings suggest that the American speakers made use of mainly two different patterns of co-occurrence in their L1 and their L2, with a slight preference for stalling strategies in their L1 (VOC+VOC) as opposed to a mixture of stalling and repair mechanisms (VOC+MS) in their L2.

Table 17. *Sequence configurations (American group)*

Seq. Conf.	L1 % (raw)		L2 % (raw)		Z score	p value
MIX	2%	(2)	2%	(6)	N/A	
MS+MS	14%	(18)	4%	(11)	3.821	< 0.0002*
MS+NL	1%	(1)	1%	(3)	N/A	
VOC+MS	29%	(38)	48%	(145)	-3.757	< 0.002*
VOC+MS+NL	7%	(9)	9%	(27)	-0.749	0.4
VOC+NL	14%	(19)	17%	(50)	-0.584	0.5
VOC+VOC	35%	(46)	20%	(61)	3.229	0.001*

For the French speakers, however, the differences are not so clear-cut, given the lack of statistical significance between the two proportions. Only one pattern (VOC+MS) reached a significant statistical score (p = 0.01), and shows differences between L1 and L2, with a higher rate in L1 (49%) than in L2 (36%).

Table 18. *Sequence configurations (French group)*

Seq. Conf.	L1 % (raw)		L2 % (raw)		Z score	p value
MIX	1%	(2)	3%	(7)	N/A	
MS+MS	10%	(18)	9%	(19)	0.432	0.6
MS+NL	1%	(1)	2%	(5)	N/A	
NL+NL	1%	(1)	1%	(3)	N/A	
VOC+MS	49%	(85)	36%	(76)	2.516	0.01*
VOC+MS+NL	8%	(13)	6%	(12)	0.697	0.4
VOC+NL	10%	(18)	17%	(36)	-0.06	0.05
VOC+VOC	20%	(35)	24%	(51)	-0.9	0.3

One major difference is also found between the French and American groups. While the French group used the VOC+MS pattern 49% of the time in their L1, the American group only produced it 29% (t = -3.637 ; p < 0.0002), and the two groups showed an opposite tendency in their L2 (fewer VOC+MS combinations in the L2 for the French, as opposed to a higher proportion in the L2 for the Americans). Lastly, the American speakers used the VOC+VOC pattern in their L1 more frequently (35%) than the French speakers (20%), although this difference did not reach much significance (t = -2.822 ; p = 0.04).

The last variable analyzed at the sequence level is utterance position, and this time the two groups showed similar patterns of behavior. Results are reported in Tables 19 and 20 below. As the statistical scores reveal, not many differences were

found between L1 and L2 in both groups, except for medial and final positions: American speakers produced more fluenceme sequences in medial position in their L2 (57%) than in their L1 (48%), and French speakers produced slighly more fluencemes in final position in their L1 (12%) than in their L2 (9%).

Table 19. *Sequence position (American group)*

Seq. Position	L1 % (raw)		L2% (raw)		Z score	p value
final	12%	(38)	10%	(50)	1.136	0.2
interrupted	2%	(5)	3%	(13)	-0.873	0.3
initial	37%	(115)	30%	(155)	2.052	0.04*
medial	48%	(147)	57%	(290)	-2.253	0.01*
standalone	1%	(4)	1%	(4)	N/A	

Table 20. *Sequence position (French group)*

Seq. Position	L1 % (raw)		L2 % (raw)		Z score	p value
final	14%	(48)	9%	(34)	2.313	0.02*
interrupted	1%	(5)	2%	(7)	-0.339	0.7
initial	30%	(105)	33%	(131)	-0.025	0.4
medial	53%	(185)	54%	(216)	-0.0082	0.8
standalone	1%	(4)	3%	(11)	N/A	

Overall, the distributions appear to be largely similar in the two speaker groups and in the two languages, which may suggest that language proficiency has little effect on the position of fluencemes in our data.

So far, our findings have exclusively focused on the verbal and vocal level of native and non-native fluency, analyzing different temporal variables, such as rate, distribution, sequence configuration, position, among others. We shall now move to our third level of analysis (above fluenceme and sequence) involving visible bodily behavior.

2.1.3. Visuo-gestural level: gesture production and gaze behavior

In this section, we will review the distribution of gestures during fluencemes ("fluent" stretches of speech) and outside fluencemes ("disfluent" stretches) in L1 and in L2, in English and in French, in order to investigate the temporal relationship between (dis)fluency and gesture production, in line with Graziano & Gullberg (2018). We will first examine the co-occurrence of fluencemes and phases of gestural action (cf Chap.

1, section III. 3.3.4), measure the rate of gestures that occurred during fluent and disfluent stretches of speech, and finish with the analysis of gaze behavior.

Tables 21 and 22 report the proportion of gesture phases during fluenceme sequences in L1 and in L2, for the American and French groups. As the Tables show, two main differences can be observed between L1 and L2 in the two groups.

Table 21. *Proportion of gesture phases during fluenceme sequences (American group)*

	L1 % (raw)		L2 % (raw)		Z score	p value
stroke	18%	(55)	21%	(107)	-1.081	0.2
hold	9%	(28)	18%	(92)	-3.5	0.0005*
preparation	5%	(15)	8%	(40)	-1.643	0.1
rest position	64%	(197)	48%	(244)	4.482	< 0.0002*
retraction	5%	(14)	6%	(29)	-0.706	0.4

Table 22. *Proportion of gesture phases during fluenceme sequences (French group)*

	L1 % (raw)		L2 % (raw)		Z score	p value
stroke	16%	(57)	24%	(94)	-2.418	0.01*
hold	9%	(32)	15%	(61)	-2.502	0.01*
preparation	6%	(20)	8%	(31)	-1.1083	0.2
rest position	62%	(216)	46%	(185)	4.34	< 0.0002*
retraction	6%	(22)	7%	(28)	-0.369	0.7

First, both American and French speakers kept their hands in rest position more frequently in their L1 (64% for the Americans, and 62% for the French) than in their L2 (48% for the Americans and 46% for the French). This suggests a higher gestural activity in L2 than in L1, but these findings will be further confirmed when we compare gestural rates during fluent and disfluent stretches of speech. Interestingly, the two groups also showed a tendency to hold their hands in a static position more frequently during fluencemes in L2 (18% for the Americans, 15% for the French) than in L1 (9% for the Americans and the French). This finding is further elaborated in Chapter 5. Overall, these results show that gestures rarely co-occur with fluencemes, in line with previous work (cf Chap. 1, section III. 3.3.4, and section 1.3. of this chapter), but a larger proportion of them are produced in non-native productions. This leads us to the analysis of gesture production during fluent and disfluent cycles of speech.

Figure 40 reports the rate of gesture strokes per hundred words in native and non-native productions for the American and French speakers[121]. Numbers indicate that American speakers produced 16 gestures phw in their L2 (N = 376), versus 11 in their L1 (N = 336), which was found to be statistically significant (LL = 27.91 ; p < 0.0001). The same applies to the French speakers, who produced 10 gestures phw in their L1 (N = 375), as opposed to 15 in their L2 (N = 275; LL = 34.34 ; p < 0.0001). These findings confirm our earlier prediction that speakers produce more gestures in their L2 than in their L1 overall. Despite significant differences between L1 and L2 overall, a few exceptions can be noted. For instance, A03 produced more gestures in her L1 (21) than in her L2 (15), as well as F11 (24 in her L1 and 18 in her L2). Some speakers also showed a tendency to gesture a lot in their L2 (e.g. A10 and F11) while others gestured very little in comparison (e.g. A15 and F18).

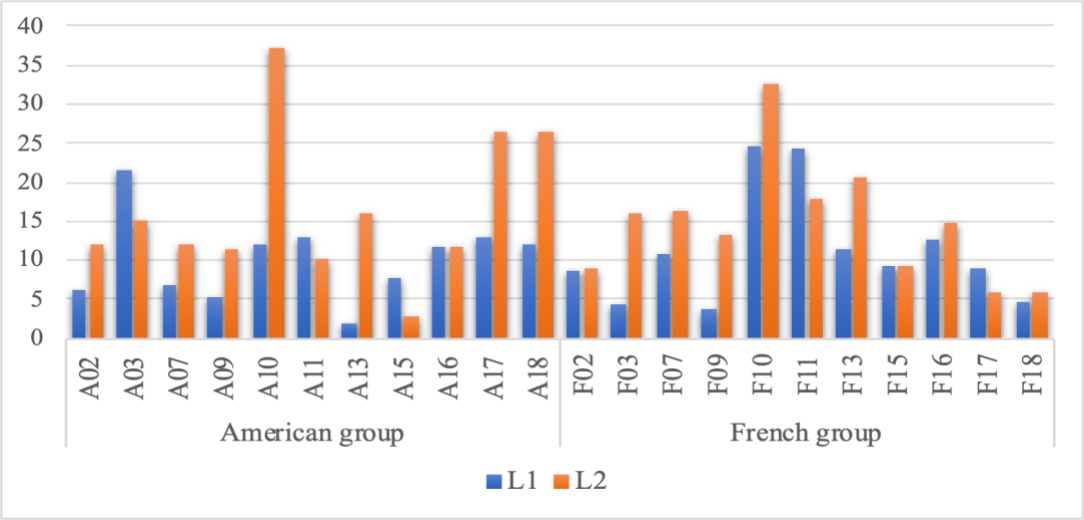


Figure 40. *Rate of gestures (phw) in L1 and L2*

Figure 41 shows the proportion of gestures in fluent versus disfluent stretches of speech (during fluencemes and outside fluencemes). Results show that gestures predominantly occurred without fluencemes, which supports previous work (cf Chap. 1, section III. 3.3.4). This was found to be true for the two groups and in L1 and L2, as the American speakers produced 84% of their gestures (N= 282/336) during fluent stretches of speech, as opposed to 16% (N= 54/336) during disfluent ones (z = -14.59 ; p < 0.0002) in their L1. This was also the case in their L2, with a lower rate of gestures during fluencemes (N=105/376) than outside fluencemes (N=271/376) (z = -12.10 ; p < 0.0002). Similarly, the French speakers produced significantly more gestures in

[121] Raw values can be found in Appendix 3, Table 63.

their L1 during fluent (*N* = 218/275) than disfluent stretches of speech (*N* = 57/275) (z = -13.73 ; p < 0.0002), as well as in their L2, with a lower rate of gestures during fluencemes (N = 94/375) than without them (N= 281/375) (z = -13.65 ; p < 0.0002).

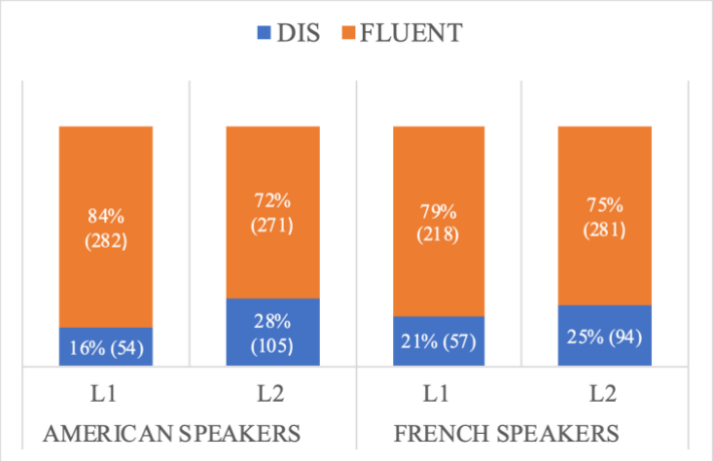


Figure 41. *Proportion of gestures during fluent and disfluent cycles of speech in L1 and L2*

Moving now to the distribution of gesture types and subtypes in L1 and L2. As described in Chapter 2. (cf Chap. 2, section II.2.2.3), we first distinguished between two main classes of gestures, following Kendon (2004), mainly *referential* and *pragmatic* (gesture types), and then added several subcategories (gesture subtypes). As results show (Tables 23 and 24), speakers from both groups produced a higher proportion of referential gestures in their L1 (28% for the Americans and for the French) than in their L2 (21% for the Americans, 20% for the French), and American speakers produced more pragmatic gestures in their L2 (80%) than in their L1 (72%).

Table 23. *Proportion of gesture types and subtypes un L1 and L2 (American group)*

	L1		**L2**		**Z score**	**p value**
referential gestures % (raw)						
	28%	**(95)**	**21%**	**(78)**	**2.33**	**0.01***
representational	10%	(33)	7%	(26)	1.40	0.1
deictic-anaphoric	18%	(62)	14%	(52)	1.67	0.09
pragmatic gestures % (raw)						
	72%	**(241)**	**80%**	**(301)**	**-2.60**	**0.009***
discursive	39%	(131)	44%	(167)	-1.46	0.1
interactive	31%	(104)	31%	(116)	0.02	0.9
thinking	2%	(6)	5%	(18)	-2.25	0.02*

Table 24. *Proportion of gesture types and subtypes un L1 and L2 (French group)*

	L1		**L2**		**Z score**	**p value**
referential gestures % (raw)						
	28%	**(76)**	**20%**	**(76)**	**2.19**	**0.02***
representational	15%	(40)	9%	(35)	2.05	0.03*
deictic-anaphoric	13%	(36)	11%	(41)	0.84	0.8
pragmatic gestures % (raw)						
	71%	**(194)**	**74%**	**(278)**	**-1.01**	**0.3**
discursive	39%	(108)	41%	(152)	-0.32	0.7
interactive	31%	(86)	34%	(126)	-0.62	0.5
thinking	2%	(5)	6%	(21)	-2.43	0.01*

As to the subtypes, the differences were not statistically significant in L1 and L2 in the two groups, except for thinking gestures, which were used slightly more frequently in L2 (5% for the Americans, 6% for the French) than in L1 (2% for the Americans and the French). This is an interesting point to consider, and it could indicate that speakers may need to "flag the fact of an ongoing word search" (Gullberg, 2001, p. 143) more frequently in their L2 than in their L1 as a result of interactional difficulties. A more detailed typology of thinking gestures is provided in Chapter 5.

Figure 42 shows the proportion of pragmatic and referential gestures in L1 and L2, more specifically in fluent and disfluent cycles of speech for the American and French groups. A chi-square test of independence (cf Chap. 2, section II, 2.3.1.) was performed to examine the relation between gesture type (pragmatic or referential) and speech type (fluent or disfluent). The relation between these variables was not significant neither for the American group in their L1 (χ^2 (1, N= 336) = 0.5 , p = 0.4) and their L2 (χ^2 (1, N = 205) = 0.5 , p = 0.4) nor the French group in their L1 (χ^2 (1, N = 275) = 0.3 , p = 0.5) and their L2 (χ^2 (1, N = 376) = 0.7 , p = 0.4). This may suggest that fluencemes have little impact on the speakers' gestural behavior in our data, but this finding may also be due to the limited number of gestures that actually co-occur with them (cf Tables 21 and 22).

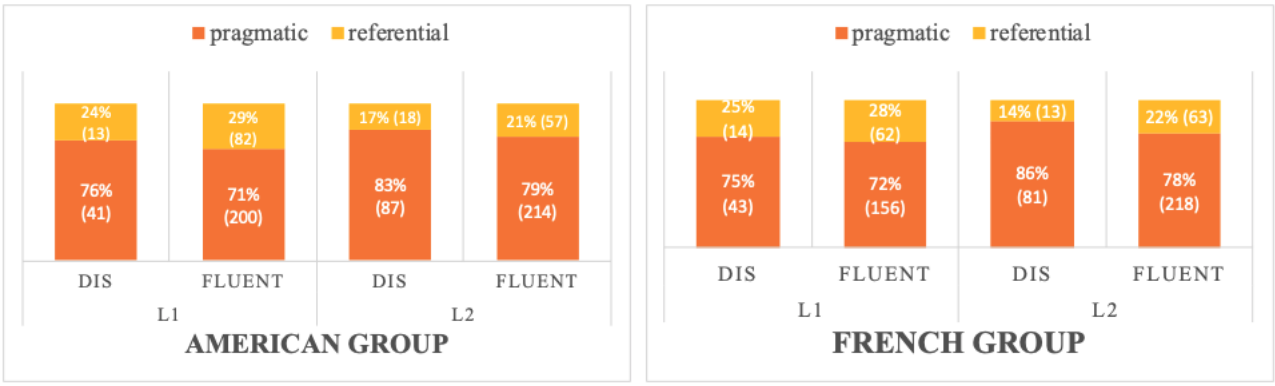


Figure 42. *Proportion of pragmatic and referential gestures in fluent and disfluent cycles of speech*

Tables 25 and 26 further show the distribution of all gesture subtypes in L1 and L2 and in fluent versus disfluent speech, and just like the distribution of gesture types, none of the values found in the two sample proportions reached statistical significance, except for French speakers who produced slightly more interactive gestures outside fluencemes in their L2 (37% during outside fluencemes as opposed to 23% during fluencemes).

Table 25. *Proportion of gesture subtypes in fluent and disfluent speech (American group)*

	L1					L2				
	DIS		**FLUENT**		***Z (p)***	**DIS**		**FLUENT**		***Z (p)***
deictic-anaphoric	15%	8	19%	54	-0.75 (0.4)	10%	11	15%	41	-1.17 (0.2)
representational	9%	5	10%	28	-0.15 (0.8)	7%	7	6%	16	0.2 (0.7)
discursive	35%	19	40%	112	-0.62 (0.5)	43%	45	45%	122	-0.37 (0.7)
interactive	30%	16	31%	88	-0.23 (0.8)	25%	26	33%	90	-1.59 (0.1)
thinking	11%	6	0%	0	N/A	15%	16	1%	2	N/A

Table 26. *Proportion of gesture subtypes in fluent and disfluent speech (French group)*

	L1					L2				
	DIS		**FLUENT**		***Z (p)***	**DIS**		**FLUENT**		***Z (p)***
deictic-anaphoric	12%	7	13%	29	-0.20 (0.8)	9%	8	12%	33	-0.87 (0.3)
representational	12%	7	15%	33	-0.54 (0.5)	5%	5	11%	30	-1.54 (0.1)
discursive	42%	24	39%	84	0.49 (0.6)	43%	40	40%	112	0.46 (0.6)
interactive	28%	16	32%	70	-0.58 (0.5)	23%	22	37%	104	-2.41 (0.01*)
thinking	5%	3	1%	2	N/A	20%	19	1%	2	N/A

Despite the lack of statistical evidence overall, it is still interesting to note that thinking gestures almost never occurred during fluent speech both in L1 and L2 and in the two

groups. Therefore, this finding can be interpreted as a sign that thinking gestures are closely associated to (dis)fluency phenomena, and may thus be an *embodiment* of inter-(dis)fluency behavior. This hypothesis is further discussed in Chapter 5.

We shall now conclude this section with the analysis of gaze direction. As reported in Chapter 2 (cf Chap. 2, section II.2.2.3.) four categories of gaze were used: "towards interlocutor", "away", "towards paper", and "in different directions". Figure 43 reports the proportion of shifts in gaze direction in L1 and in L2 for the two groups[122], but it should be noted that the findings yielded no statistical significance overall (cf Table 61 in Appendix 3), except for the American group who gazed towards the piece of paper more frequently in their L2 (12%) than in their L1 (6%) ($z = 4.36$; $p < 0.002$).

However, when we compare gaze behavior in fluent versus disfluent stretches of speech, a number of significant differences can be found both in L1 and L2 and in the two groups. As Figure 44 shows, there is a higher proportion of gaze withdrawal (*gaze away)* in disfluent than in fluent stretches of speech in L1 (59% vs 34% ; $z = 7.098$; $p > 0.002$), as well as in L2 (50% vs 31% ; $z = 6.364$; $p > 0.0002$). Conversely, there is a higher proportion of mutual gaze (*gaze towards interlocutor*) in fluent than disfluent speech in L1 (54% vs 30% ; $z = -6.56$; $p < 0.0002$) and in L2 (55% vs 28% ; $z = -8.83$; $p < 0.0002$).

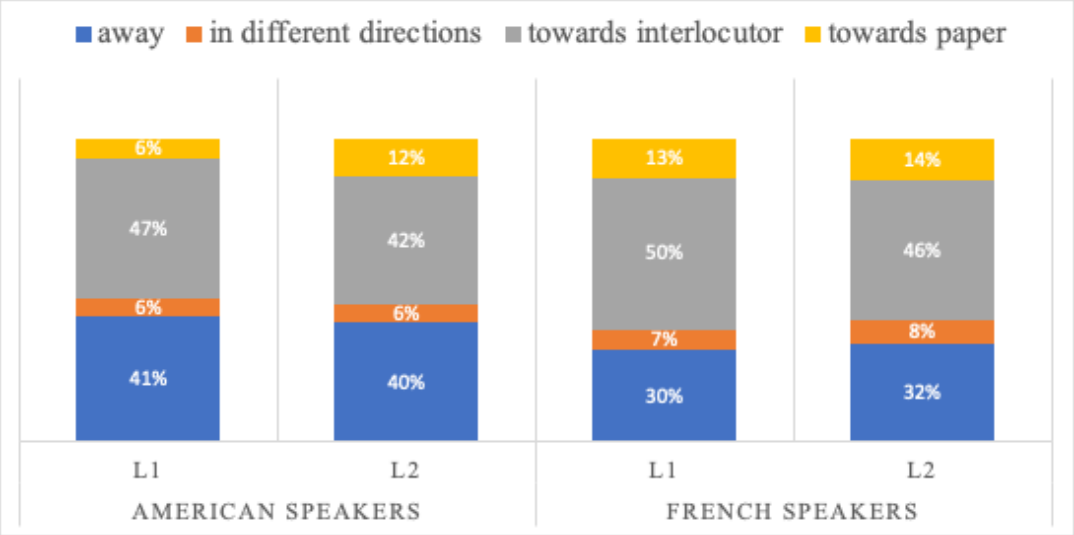


***Figure 43**. Gaze direction in L1 and L2 (American and French group)*

[122] Raw values can be found in Appendix 3, Table 62.

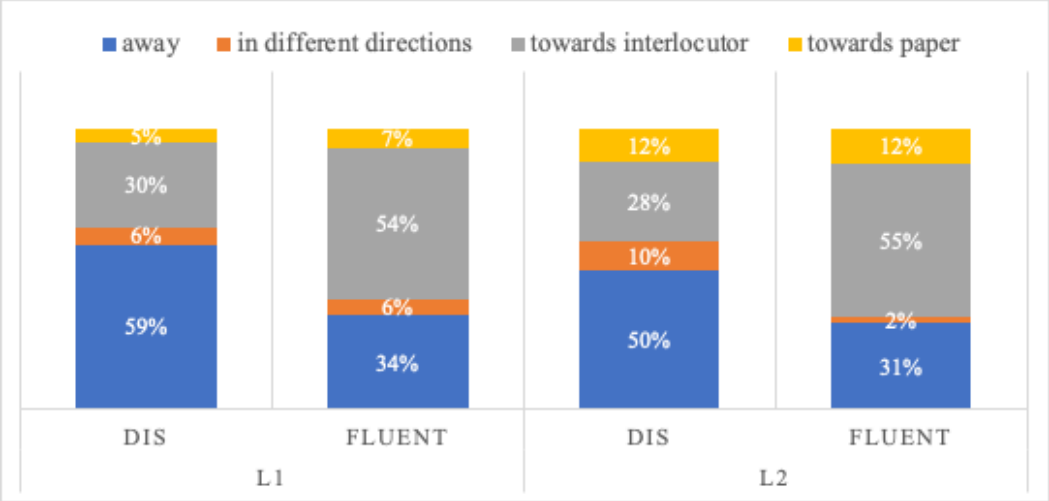


Figure 44. *Gaze direction in fluent and disfluent stretches of speech (American group)*[123]

Similar results are reported in Fig. 45 for the French group, with a higher proportion of gaze withdrawal in disfluent speech both in L1 (39% vs 25% ; z = 4.59 ; p < 0.0002) and in L2 (47% vs 23% ; z = 7.95 ; p < 0.0002), and a higher proportion of mutual gaze in fluent speech both in L1 (58% vs 36% ; z = -6.17 ; p < 0. 0002) and L2 (55% versus 28 % ; z = -9.24 ;p < 0. 0002).

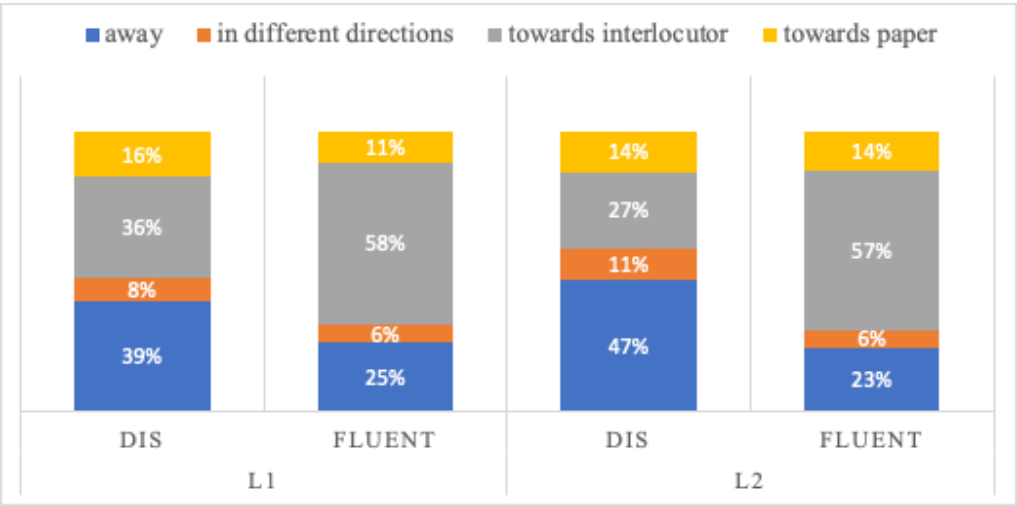


Figure 45. *Gaze direction in fluent and disfluent stretches of speech (French group)*

It is also interesting to note differences between the two groups: French speakers gazed towards the piece of paper 16% of the time in their L1 during fluencemes, as opposed to the Americans who did it only 5% (z = 4.73 ; p < 0.002), while Americans spent considerably more time gazing away during fluencemes in their L1 (59%) than the French (39%) (z = 4.732 ; p < 0.0002).

[123] Raw values for this figure and Figure 43 are found in Appendix 3, Table, 62.

To conclude, the aim of this section was to analyze overall patterns of gestural and gaze behavior that are typical of inter-(dis)fluency phenomena in native and non-native productions. The binary opposition between "fluent" and "disfluent" stretches of speech was made explicit here in order to compare the types of gestures that typically co-occur with fluencemes with the ones that do not. However, it has been pointed out several times throughout this thesis that inter-(dis)fluency phenomena should not be restricted to "disfluent" stretches of speech, but it is interesting to note a number of tendencies in the data, mainly the absence of gestures and mutual gaze during fluencemes, which may indicate a form of disengagement in the gestural and interactional activity. This is one side of DISfluency, characterized by a high rate of verbal fluencemes, complex sequences, and limited gestural activity. The quantitative results observed in this section can be summarized as the following:

- **Higher rates of fluencemes** in L2 than in L1 both within the French and American groups.
- **Differences in fluenceme distribution between L1 and L2:** more identical repetitions and filled pauses in L2 but more self-interruptions and unfilled pauses in L1 for the American group. For the French group, more self-interruptions and prolongations were found in L1, but more non-lexical sounds in L2.
- **Differences in the form of filled pauses**: more ums in L1 than L2 for the Americans, but more uhs in L1 than in L2 for the French.
- **No significant differences in duration** between L1 and L2 and in the two groups.
- **More complex sequences in L2** than in L1 for the American group. But no significant differences for the French.
- **Higher number of markers combined in L2 for the American group**. But no significant differences for the French.
- **Slight differences in utterance position**: more instances of medial position in L2 for the Americans, and more instances of final position in L2 for the French.
- **More instances of held gestures during fluencemes in L2** than in L1 for the two groups.
- **Higher rate of gestures in L2** (both during and outside fluencemes) for the two groups.
- **More pragmatic gestures and thinking gestures during fluencemes** than outside fluencemes both in L1 and in L2 for the two groups.
- **More instances of gaze withdrawal during fluencemes** than outside fluencemes both in L1 and in L2 for the two groups.

These findings, which are further discussed in Section III, give an overall idea of the form, distribution, and co-occurring visible bodily behavior of fluencemes; they do not, however, paint a full picture of the present phenomena, as they do not portray the

multimodal deployment of fluencemes in situated tandem activities. This leads us to the following section, which presents our qualitative analyses.

2.2. Qualitative analyses

In this section, we further explore the interactional ambivalence of fluencemes by illustrating their multimodal quality in situated sequences. We begin with an overview of their distribution with regard to *communication management* in L1 and in L2 (section 2.2.1) by crossing different variables (mainly sequence type and gesture), before presenting 5 micro-analyses of several excerpts taken from the data, in sections 2.2.2. and 2.2.3. As explained in Chapter 2 (cf Chap. 2, section I.1.2.), participants were given pseudonyms specifically in the qualitative analyses to render the exchanges more natural.

2.2.1. Communication management: overview of the data

As mentioned earlier (cf Chap. 1, section II. 2.2.1.), Allwood et al., (2015) distinguished between two types of fluencemes (*communication management* in their terms) based on their function in communication. They presented a model with two main systems, one that is concerned with the speaker's management of his or her "linguistic contributions to communicative interaction" (Allwood et al., 2005, p. 2), in other words, speakers' own planning processes (*own communication management,* henceforth *OCM*), and another that pertains to the interactional exchange and the interactants' turn-taking mechanisms and multimodal feedback (*interactive communication management,* henceforth *ICM).* In this thesis, we adopted a similar labeling system (cf Chap. 2, section II. 2.2.2.) and annotated fluencemes based on whether they were more intrapersonal, in relation to internal production processes (OCM), or whether they were more interpersonal, and contributed to the sequential development of the interaction through the enactment of speech acts and the co-achievement of intersubjectivity (ICM). Our methodology is slightly different from Allwood et al., (2005), as we did not annotate backchanneling devices (such as *yes*, *no mm mm*, head shakes, and the like) but rather chose to focus exclusively on typical "disfluency" markers (cf Chap. 2, section II.2.2.1.) in order to explore how the same a priori "disfluent" forms (i.e., forms marking "disfluency") could perform different functions.

Figure 46 shows the proportion of fluencemes performing ICM and OCM functions in L1 and in L2 and in the two speaker groups. As the numbers show, the proportion of OCM is overwhelmingly greater both in L1 and in L2, and in the two groups. A chi-square test further showed that there was no significant association between language proficiency and functions of fluencemes both for the American group, χ^2 (1, N = 821) = 0.1, p = 0.7) and the French one, χ^2 (1, N = 746) = 0.2 , p = 0.6. This is a very striking result, which, to some extent, gives credit to previous psycholinguistic work on (dis)fluency that focused on their role in the speech production system. It is interesting to note that neither speaker group nor language proficiency seem to have an effect on the distribution of these functions, as all fluencemes predominantly performed the OCM function. However, despite the high proportion of fluencemes associated with intrapersonal processes (about 80-82%), the remaining proportion (18%) is nonetheless of great value for the present work, as it illuminates the interactional nature of fluencemes, which have often been too restricted to the overwhelming 80%. This is further explored in our qualitative analyses.

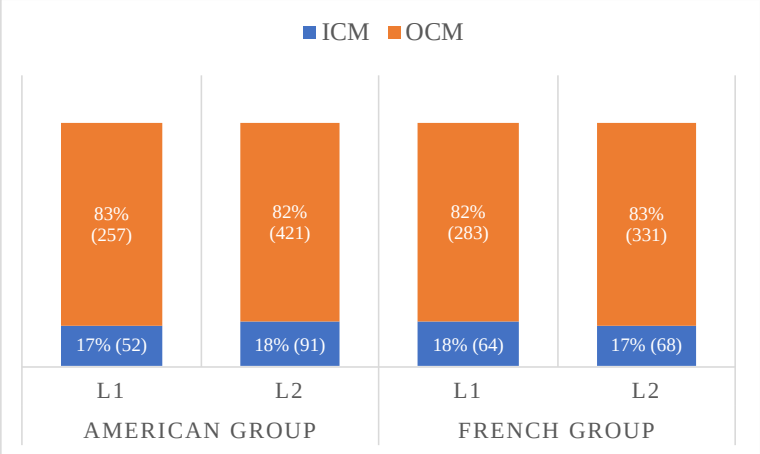


Figure 46. *Proportion of OCM and ICM functions in L1 and L2*

Table 27 reports the proportion of fluencemes performing ICM and OCM functions in simple and complex sequences, thus analyzing the association between fluenceme type and function for the American and French group. Overall, results seem to show a relationship between the two variables, as American speakers were more likely to produce simple sequences when performing ICM in their L2, while they produced more complex ones when they performed OCM functions.

Table 27. *Proportion of simple and complex sequences during ICM and OCM*

AMERICAN GROUP				
	L1 % (raw)		**L2% (raw)**	
	complex	**simple**	**complex**	**simple**
ICM	38% (21)	63% (35)	45% (45)	55% (54)
OCM	45% (133)	55% (163)	60% (271)	40% (180)
χ^2 (p)	$\chi^2 = 1.05$, $p = 0.3$		$\chi^2 = 7.1$, $p = 0.007$*	
FRENCH GROUP				
	L1 % (raw)		**L2% (raw)**	
	complex	**simple**	**complex**	**simple**
ICM	38% (24)	63% (40)	35% (24)	65% (44)
OCM	52% (149)	47% (134)	55% (182)	45% (149)
χ^2 (p)	$\chi^2 = 4.7$, $p = 0.02$*		$\chi^2 = 8.7$, p =0.003*	

No significant differences were found in their L1. Similarly, the French group produced a higher proportion of simple sequences when performing ICM than OCM, both in their L1 and L2, while they showed a tendency to produce more complex sequences when they performed OCM functions. These findings suggest that **the process of working on one's production may require more complex sequences than taking part in an interactional practice.** This is further illustrated at the end of this section.

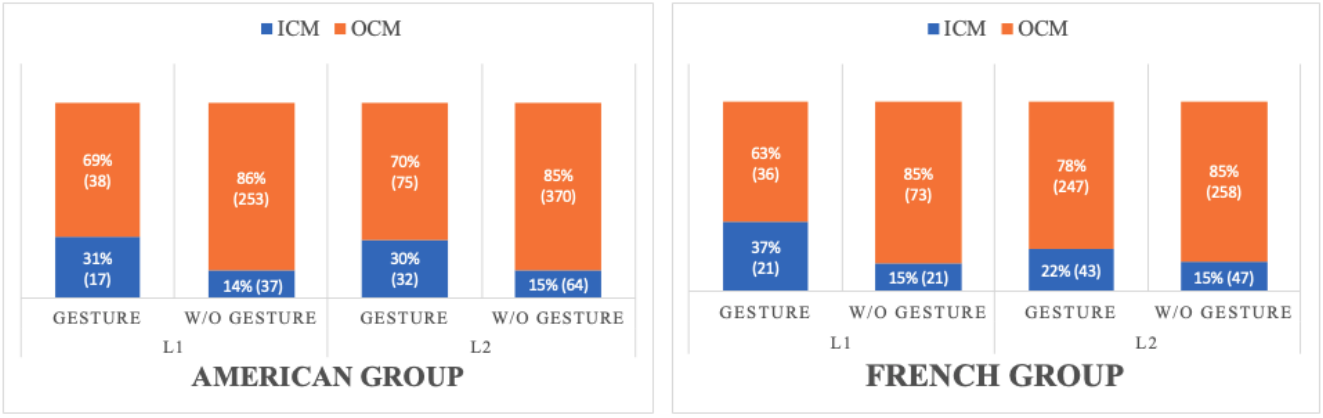


Figure 47. *Proportion of fluenceme sequences that did or did not co-occur with gestures during ICM and OCM*

Figure 47 further reports the proportion of fluencemes that performed the OCM or ICM function whenever they co-occurred or did not co-occur with gesture strokes, to test the relationship between gesture production and fluenceme function. As results

indicate, the American group showed a tendency to produce more gestures when they performed the ICM function than the OCM, and this was statistically significant both in L1, χ^2 (1, N = 309) = 9.4, p = 0.002, and in L2, χ^2 (1, N = 512) = 13.6 , p < 0.002. This was not significant for the French group however, as the relationship between gesture and communication management did not reach significance neither in L1, χ^2 (1, N = 347) = 3.7 , p = 0.05, nor in L2 χ^2 (1, N = 399) = 0.03 , p = 0.8. Once again, the two speaker groups show different tendencies, perhaps reflecting individual communication strategies. This is further illustrated in section 2.2.2. and in section III. OCM and ICM are another way to illustrate the functional and interactional ambivalence of fluencemes; instead of exclusively focusing on their planning processes (OCM) or zooming in on their interactional dimension (ICM), the present study aims to illustrate how the same forms can display different patterns of behavior, depending on their co-occurrence within the micro and macro context, and their accompanying visual-gestural behavior. This is illustrated in the following example, taken from Pair 7 interacting in English.

Excerpt Pair 07 – Teenage years

1 *JUL: and it's a lot [/] a lot about reputation and things like that.

2 *JUL: but on <the other hand> +/.

3 *AM: mm mm.

4 *AM: +< you're not really sure who you are yet.

5 *JUL: +< yeah exactly and you [/] you're growing hu [//] you're growing up.

6 *JUL: you:u're like uh you're not a chi:ild but you're not an adu:ult.

→ **7 *JUL: and you're like (0.326) what am I doing here? ((laughs))**

****

((head shake + gazes away)) ((gazes towards NS))

8 *AM: ((laughs)) yeah.

→ 9 *JUL: on the other hand I th [//] I still think that **(1.011) [!]**
**********- .-.-.-.-.- .-.-.-**********-.-.-.-.-.-
((gazes away))
((frowns))
a teenager you know *yy* you (a)re still uh h um [//] don't have that much [//] that many problems.
10 *AM: yeah.

During this exchange, the two tandem partners were asked to talk about the following topic: "the best years of your life are teenage years". The excerpt is taken from the beginning of the exchange, after the American speaker Amber (A07/Native-Speaker) gave her opinion on the topic, mainly that she thought that the early twenties were the best year of her life and not her teenage years. Julie (F07-Non-Native Speaker) agreed with that, and shared her view.

Here, Julie, the non-native speaker (JUL) is presenting different arguments that are not in favor of the given topic, and as the transcription shows, a number of fluencemes are found in her speech flow, from simple sequences (l. 1 with a repetition, l. 6 with two prolongations) to complex ones (l.5 with a MS+MS pattern which combines a truncated word and a morphological repair, or l.9 with a VOC+MS pattern which combines 4 different markers, FP+TR+FP+SR). A majority of the fluenceme sequences found in this excerpt perform OCM functions, as the speaker is mainly dealing with morphological, lexical, and syntactic difficulties to help her manage her own production, or plan parts of her speech. There is, however, one fluenceme which "stands out", in line 7. As the multimodal transcript (cf Chap. 2, section I. 1.5.3.) further shows, Julie produces an unfilled pause of 326 milliseconds, which is accompanied by a shoulder shrug, a headshake, and a palm lateral gesture. As we have seen earlier (cf Chap. 1, section III. 3.3.1) shrugs can be viewed as *compound enactments* (Streeck, 2009b, p. 189), comprised of a variety of bodily behaviors (i.e. pout, hand activity, shoulder tilt, etc.) which typically enact a stance such as indifference, incapacity, submissiveness, common ground, etc. (see Debras, 2017). Here the shrug was initiated during the vocal fluenceme, and it reached its completion at the end of Julie's utterance ("what am I doing here?"). It is interesting to note that the pause is barely perceptible in the vocal channel; it is of rather short duration (326 ms) compared to the speaker's average duration of pauses in her L2 (about 754 ms, see Table 56 in Appendix 3), but also compared to the other pause produced in line 9 which

is significantly longer (1.011 ms). Most importantly, the pause appears to "give room" to another modality which becomes more relevant in that context, as it enables the speaker to express her attitude towards her utterance, and thus convey a communicative intention. This is what Goodwin calls *contextual configuration*[124], defined as the following (Goodwin, 2000, p. 1490):

> As action unfolds, new semiotic fields can be added, while others are treated as no longer relevant, with the effect that the contextual configurations which frame, make visible, and constitute the actions of the moment undergo a continuous process of change.

In this case, we can hardly speak of "disfluency", but rather of inter-fluency, where vocal and bodily behaviors are coordinated and interact with one another, building a communicative and *fluent* multimodal flow. This further highlights the need to view fluencemes as embodied within their multimodal environment (cf Chap. 1, section III). This multimodal inter-fluency process thus illustrates an instance of ICM, characterized by its pragmatic and communicative dimension. This pause thus presents very different characteristics from the one co-produced in line 9 with a tongue click during which the speaker seems to be searching for the next word or phrase. This difference is also displayed in accompanying visible behavior (i.e. she is frowning and is keeping her hands in rest position). In this case, she rather seems to be "looking within herself" (McCafferty, 1998, p.88) and not enacting a speech act (as she previously did in line 7). Pauses have in fact shown to be highly multifunctional, and their perception is to a great extent determined by the kind of approach taken (Dodane & Hirsch, 2018). From a strictly formal point of view, pauses may be defined as an interruption in the acoustic channel, and from a speech production perspective, they can be viewed as physiological processes marking the boundary of a breath group, or a prosodic unit. From an interactionist perspective, they may be regarded as significant delays, marking a dispreferred answer in the next turn-at-talk. Lastly, from a visual-gestural perspective, they may be used to give room for gestures which function as a window onto cognitive and interactional processes. Once again, the construct of (dis)fluency needs to take into account different dimensions (speech, interaction and gesture) in order to be fully grasped. The different degrees of inter-(dis)fluency can thus be measured on the basis of (1) duration— shorter vocal

[124] This is further explored in Chapter 5.

fluencemes associated with fluency vs longer ones associated with DISfluency, (2) sequence combination – simple fluencemes associated with fluency and complex ones to DISfluency, and (3) gestural behavior – interactive gestures displaying fluency vs no gestural activity displaying DISfluency, and (4) communication management – OCM reflecting DISfluency and ICM reflecting fluency. This multi-level scale is further developed in the *General Conclusion* (cf Fig. 71).

2.2.2. Non-native speakers' multimodal communication strategies

Gullberg (2011) distinguished between three major types of difficulties experienced by L2 learners during their non-native productions, mainly lexical, grammatical and interactional related difficulties. In her paper on multimodal communication strategies (cf section 1.2.2. and 1.3.), she aimed to investigate whether the different types of communicative difficulties would yield different types of multimodal behavior, and explored the role of individual communicative style, by presenting several micro-analyses from her data. For instance, when dealing with grammatical difficulties, e.g. tense marking, she showed that learners tried to resolve these problems by using temporal adverbials ("yesterday", "tomorrow") or by making use of their surrounding gesture space. Regarding interactional related difficulties, Gullberg showed that L2 learners relied on several multimodal resources to manage problems that resulted from their "non fluent hesitant productions" (Gullberg, 2011, p. 139). She further noted that "every disfluency is a potential locus for loss of face and of floor" (p. 143). However, we do not exactly agree with this view, as we have pointed out before that fluencemes are not necessarily the *result* of interactional difficulties, but that some of them, when coordinated with visible bodily behavior, may in fact be used to *resolve* such difficulties. We thus claim that certain fluencemes can also *act* as communication strategies to a certain extent, and this is illustrated in the following qualitative analyses, each one reflecting a speaker's strategy. In line with Gullberg (2011) a lot of emphasis is laid on individual communicative styles, to examine whether speakers engage or DISengage from their interlocutor in the course of the interaction, and whether this is reflected in their visual-gestural activity. In this section, we identify 3 different multimodal strategies mobilized by non-native speakers as an attempt to solve lexical, grammatical, and/or interactional difficulties. Most of the analyses are taken from Kosmala (2019, 2020, 2021).

Strategy 1. Use of an explicit editing phrase and a combination of visual-gestural activities: Projecting the progressivity of the word search

During embodied word searching activities, non-native speakers are often prompted to mobilize a number of vocal and visual-gestural resources to deal with their communication difficulties while trying to keep the conversational floor. They may simultaneously combine different verbal, facial, and gestural practices, with for instance the display of a thinking face, the production of a thinking gesture to offer a metapragmatic comment on their utterance (see section I. 1.3), or a self-addressed question signaling production trouble (Hayashi, 2003). The two following examples illustrate instances of these coordinated resources mobilized by the non-native speakers in the course of their search.

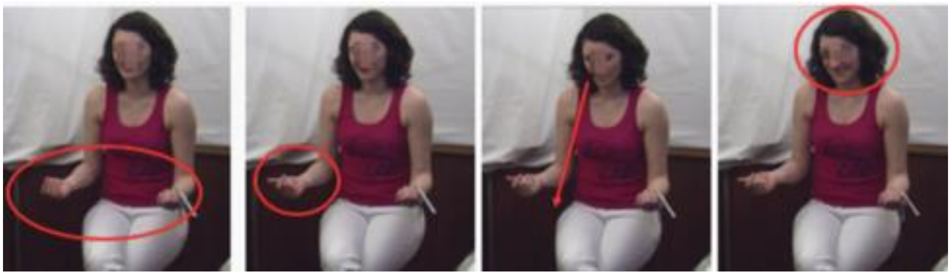
The first excerpt is taken from the interaction in English between Sally (F11, the French speaker) and Harry (A11, the American speaker). In this particular sequence, Sally is talking about the people who were once part of a reality TV show and then had to go on with their lives after the show ended. As the transcript shows, the non-native speaker is experiencing a number of lexical and grammatical difficulties, and seems unable to retrieve the right words or expression to convey what she means, as reflected by the number of fluencemes produced in lines 3 and 4. This prompts her tandem partner to assist her by completing her utterance with the right lexical item (l.6), which is typical of tandem interactions where native speakers provide the L2 learners with corrective feedback to help them with their production (see Debras et al., 2015, 2020; Horgues & Scheuer, 2020).

Excerpt 1A – Pair 11.

```
1     *SAL: and (1.030) the problem is when you [/] you stop the TV
show
            *********
      ((hands held together + gazes away))
      and want to go back to work to [/] to live with your family.
                           ***********
                  ((pointing gesture to the left, gaze on HAR))
2   *HAR:   mm mm.
            ((gaze towards SAL))
3   *SAL: (0.900) everybody saw um [/] saw you in this tv show so they
   can't have                          *********
((looks up, palm open hand gesture; raised eyebrows))
```

→ ***SAL: um (1.250) [!] um (2.160) I don't know how the word uh (1.490)**

****



((looks away, hands held together, finger snap gesture, smile))

they can't trust on you.

4 *HAR: ok.

((head nod, gaze towards SAL))

5 *SAL: on you:ur um (0.510) [/] on <you> +/.

****

((both hands coming together, palms facing upwards; looks away; pouts))

6 *HAR: +< uh reliance.

((gazes at SAL))

In line 3, Sally (French NNS) produces a fairly long fluenceme sequence comprised of 9 different markers (a filled pause, an unfilled pause, a tongue click, a second unfilled pause, an explicit editing phrase, a second filled pause, and a third unfilled pause, following the *MIX* combination). The length[125] of this sequence (9 markers) is considerably greater than the average of her group (1.9 markers), as well as her own (2.1), which underlines the degree of variation found in fluenceme use (as previously illustrated in the boxplot, Figure 39, section 2.1.2). In this particular case, Sally is experiencing lexical difficulties, as she is looking for a specific lexical item and does not have the word for it, and this is overtly expressed in her explicit editing phrase ("I don't know how the word"). Explicit editing phrases (EDT), which can be classified as peripheral markers of fluencemes (cf Chap. 2, section II. 2.2.1), refer to lexical expressions by which the speaker signals some production trouble (Crible 2017, p. 108). EDTs are rather relatively fixed chunks that are stored in speakers' memories (e.g. formulaic units, see Gürbüz, 2017, section 1.2.1.) and can easily be retrieved without any processing time (Wood, 2001). Here it is interesting to note that, even though the speaker failed to produce the correct multi-word expression in her target

[125] Note that the term "length" is borrowed from Crible (2017) to refer to the number of markers combined in a sequence, not actual duration (see Chap. 2, section II).

language, she still used it competently for pragmatic purposes. Indeed, she relied on this EDT to signal production difficulties to her partner, but also perhaps to keep the floor as not to be interrupted (Maclay & Osgood, 1959), and display the progressivity of her word search (Goodwin & Goodwin, 1986; Hayashi, 2003). These cues are in fact understood by her partner, who does not interrupt her right away, but rather provides verbal ("ok") and visual ("head nod") backchannel (l.4) before offering assistance (l.6). Therefore, the production of this ungrammatical verbal expression, does by no means impede the overall interactional flow, as Sally manages to keep her turn while dealing with her lexical difficulties until Harry offers his help. This further identifies the notions of *grammatical* versus *interactional* competence (see section 1.2.2.) displaying different dimensions of fluency (see Chap 1., section II.2.1.).

In addition, as illustrated in the pictorial illustration within the transcript, the L2 speaker does not only rely on verbal and vocal fluencemes to work on her production, but on several multimodal resources as well. The complexity of her sequence, composed of different clustered fluencemes, is also **visible in her gestural activity**, as she first holds her arms and hands in the same position during the production of the filled pause, then produces a finger snap gesture with her right hand during her unfilled pause, and then holds both her arms and hands again while looking down, and finally she starts smiling when she produces the EDT. This example shows a synchrony between the complexity of the fluenceme sequence and the complexity of the gestural activity, which further underlines the multimodal dimension of inter-(dis)fluency. This point is further developed in Chapter 5. Her multimodal strategy thus initially consisted in suspending speech for a very long time, probably because she needed to buy time while looking for a specific word (as indicated in the finger snap gesture, see Poggi 2001)[126], but since she failed to find it, she then explicitly signaled it (l.3), perhaps in order to save face (Smith & Clark 1993), and to hold the floor, as her partner did not interrupt her, and kept gazing at her. This shows that her fluenceme sequence is not only related to internal speech processes (*own communication management*) but that it may also have some interactional dimension as well. By producing this very complex sequence and relying on several visual-gestural resources, the non-native speaker simultaneously manages to (1) buy time while planning what to say next, (2) offer metacognitive information to her partner to

[126] Finger snap gestures are further analyzed in Chapter 5.

signal that she is looking for a specific word through the finger snap gesture (*thinking gesture*) (3) hold the floor until she acknowledges that she has not found the target word, and finally (4) return to a more "fluent" delivery in the verbal flow ("they can't trust on you"). This underlines the multidimensional aspect of fluencemes and their functional ambivalence: while her utterance is very "disfluent" from a speech production perspective because of its length and complexity (*speech fluency*), the fluenceme sequence was also used as a tool to hold the floor (*interactional fluency*), while displaying the current state of her search through visible bodily behavior (*visual-gestural fluency*).

The second excerpt is taken from Pair 13 in English between Elena (F13- NNS) and Francis (A13- NS) during which the two speakers are talking about University tuiton fees. Just like Sally, Elena is experiencing a number of lexical and grammatical difficulties, as indicated by the series of fluencemes (in her first turn, in bold) but also by her facial expressions, gaze behavior, gestural and head movement.

Excerpt 1B – Pair 13.

→ 1 ***ELEN: but I [/] I'm not sure (be)cause here um (0.768) [!] [/] here (0.889) if you:u uh I ain't got the w word here**
((thinking face a.)) ((looks up; smiles b.))

a. here (.) if you:u uh b.I ain't got the w word here

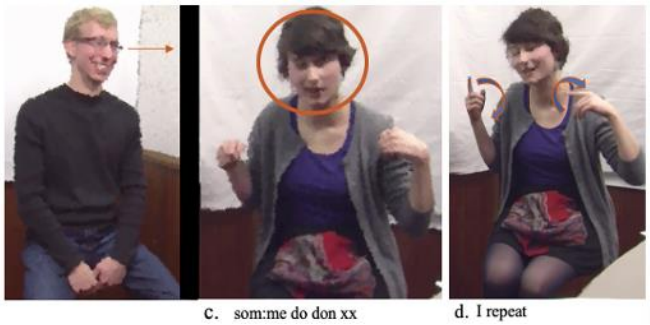
eh hhh. um if the <state> didn't
((looks towards Francis))

2 *FRAN: +< mm mm.
((head nod))

→ 3 ***ELEN: give you som:me do(llars) don do xxx +//.**
((thinking face c.))
***ELEN: ((smiles))**

4 ***FRAN: I repeat.**
****
((cyclic gesture+ eyes closed d.))

5 *ELEN: if the state doesn't give you money.
((looks towards Francis))



c. som:me do don xx d. I repeat

6 *FRAN: mm mm.
7 *ELEN: you have to pay uh four hundred (0.569) euros for a year. but I don't.
8 *FRAN: mm mm.
9 *ELEN: so to me [//] for me it's free.
10 *FRAN: yeah.
11 *ELEN: a:and my teachers are (0.632) really great so (0.735) +…
12 *ELEN: I don't think that you have to pay to have a great education.
13 *FRAN: four hundred euros a year man.

Here Elena is enacting a lexical search activity by coordinating vocal fluencemes and bodily actions which allow her to project the current progressivity of her search. As she is trying to make a point (that students do not have to pay a lot of tuition fees to get good education) she first displays a state of uncertainty with a thinking face (picture a.), while suspending the course of her utterance (with an "um", a pause, a tongue click, and a lengthening, turn 1) followed by another EDT "I ain't got the word" (also ungrammatical, just like Sally in the previous example) which makes her word search explicit. She makes her current activity even more visible and almost theatrical by raising her head, looking up, and smiling (picture b.), as if the words were going to fall from the sky. She then initiates a new segment "if the state" (turn 1) and gazes towards her partner to display her tentative lexical retrieval success, but then produces a series of truncated words (turn 2) accompanied by a second thinking face which makes her abandon her current utterance and start a new one ("I repeat") which states her current re-adjustment towards the completion of the segment ("if the state doesn't give you money"). This re-adjustment is also embodied in a cyclic gesture[127] in which both hands are rotating as to convey the process of starting over (picture d.). Once again,

[127] This gesture (along with other cyclic gestures) are more thoroughly analyzed in Chapter 5.

she makes this process visible to her partner, and the notions of suspending, interrupting and restarting, which are inherent to fluencemes, are embodied in several visible activities. These embodied fluencemes, coordinated with different head, gestural, and bodily movements constitute actions that are relevant to her lexical search activity, but they also allow her to keep the floor. Her tandem partner seems to attend to her actions attentively, as he coordinates his behavior with her by punctuating the interaction with several backchanneling devices and tokens of agreement ("yeah" "mm" and head nods) without interrupting her. It is only after the completion of Elena's lexical search activity that he shifts his participation status of "hearer" from "speaker" (Goodwin, 1980), and makes an assessment ("four hundred euros a year man", turn 13). Just like Sally, Elena's utterances are highly "disfluent" from a strictly verbal perspective, but it doesn't stop her from pursuing her word search activity without her partner's assistance (as opposed to Sally). She also actively provided information about the progress of her search, from verbally expressing her uncertainty (turn 1) to re-shaping the outcome of the search with a self-interruption following her production difficulties (turn 3). The process of starting over was further projected and made readily available to her partner with a cyclic gesture, indicating that her search was still in progress.

To conclude, these two examples have shown that the production of vocal and verbal fluencemes does not only signal that a speaker is currently experiencing trouble, but it also provides solutions as to how to resolve lexical problems, with the help of co-occurring visual gestural resources. Following Goodwin & Goodwin (1986), Rydell (2019) and Hayashi (2003) we further argue that fluencemes embodying word searching activities are not only manifestations of internal cognitive processes, or "symptoms" of a L2 resource deficit, but relevant displays of an ongoing search, marking a shift in the current speaking activity, which, as Hayashi (2003, p. 114) put it:

> Specifically, these publicly observable displays of trouble in producing a next item due mark a shift in the activity that participants engage in at the moment, from whatever has been going on (e.g., storytelling) to one in which a solution to word-finding trouble is pursued. It invokes a different participation framework in which collaborative participation by recipients in the solution of the speaker's word-finding trouble might become relevant.

The different postures adopted by the two speakers during their word search also further reflected embodied displays of "doing thinking" (Heller, 2021); this practice is further developed in Chapter 5, along with the analysis of thinking faces and thinking gestures. We shall now move on to the next strategy whereby non-native speakers invite their partner to take part in the collaborative word search.

Strategy 2. Mutual gaze and concurrent gesture: visibly requesting help from the interlocutor

Many authors have emphasized the crucial role of gaze in interaction (e.g. Goodwin, 1981; Goodwin & Goodwin, 1986; Gullberg, 2011; Sweetser & Stec, 2016, among others). Gaze enables speakers to embody multiple viewpoints when engaged in storytelling activities, and perform a series of discourse and cognitive functions such as visually "checking" for the interlocutor's approval, or finding access to memory space (Sweetser & Stec, 2016). During word searching sequences, gaze shifts can further signal whether a search is self-directed or other directed (Goodwin & Goodwin, 1986; Hayashi, 2003, Rydell, 2010): speakers may wish to look straight at their partner and perform gestures that are relevant for the solutions of problems (other directed), or they may also withdraw their gaze and display a thinking gesture (self-directed, as illustrated in excerpts 1a and 1b). In this section, we explore the role of mutual gaze during multimodal communication strategies, and illustrate how speakers visibly request help from their partner in ways that are relevant to the current activities they are engaged in.

The first example is taken from Pair 3 in French, with the American speaker Julia (A03-NNS) and the French speaker Marina (F03-NS). In this excerpt, the pair was asked to talk about the differences between being a traveler and a tourist.

Excerpt 2A – Pair 3

```
      1     *JUL: e:et peut-être les voyageurs par contre euh (0.790) [!]
            a:a [/] a l'opportunité de:e rester
                                ((gazes away))

→           un plus lointain peu (0.580) +. . .
      ((hands held in the same position and slightly move down))
                  ((moves her head and gazes towards Marina))
                        ************************
```


2 *MAR: plus longtemps.
((gazes at Julia))
3 *JUL: plus longtemps.
*MAR: ((nods, smiles))

In this brief example, Julia is experiencing difficulties with the pronunciation of the adverb "longtemps", as she mispronounces it (the spelled word "lointain" reflects the initial mispronunciation, but it does not necessarily represent her initial intention). Her mispronunciation is related to difficulties in phonological encoding: she produces a sequence of four words (understood as "un peu plus longtemps") which may be challenging for a non-native speaker because of the three rounded vowels in French (*un* /œ̃/ *peu* /ø/ *plus* /y/.)[128]. The comparative in French can also be quite challenging for the American speaker as it requires the use of an additional word, as opposed to the "er" suffix in English. Julia quickly realized that she mispronounced the word, and as the transcript shows, she paused for 580 ms (which is a little below her average of 629 ms in her L2), held her arms and hands in the same position, and with her gaze fixed on her interlocutor, she slightly moved her head in the direction of her partner, which could be interpreted as requesting for help. The pause thus marks a transition relevant place, where speaker change is made relevant: Marina understands her partner's request, and takes the floor to provide a phonological repair, but she also skips the quantifier ("un peu"), perhaps to facilitate the phonetic realization of the vowels and focus on the target adjective ("longtemps"). In the subsequent turn, Julia repeats the target word, this time with the right pronunciation, which projects the end of her current utterance. This type of activity is known as *doing pronunciation* (Brouwer, 2004, p. 93), which is a specific type of repair sequence in which a L1 speaker typically corrects a L2 speaker at the phonetic level. As Brouwer further noted, during such repair episodes, the conversation is momentarily "put on hold" (Brouwer,

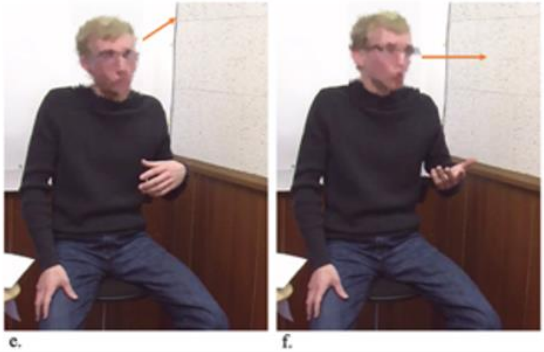
[128] This excerpt was presented at the LSPPC6 Conference in June 2021. Thank you Céline Horgues, Sylwia Scheuer, and Christelle Exare for your help on the phonetic analysis.

2009; p. 93) and the participants are then oriented to matters related to language competence. The fluenceme here thus functions as an *initiation technique* (Schegloff et al., 1977, p. 369), signaling trouble with regard to the delivery of the turn, during which the non-native speaker implicitly solicits her partner's correction. Following this brief repair sequence, Julia then goes on saying that travelers, unlike tourists, tend to stay in a foreign country for a longer amount of time (omitted from the transcription). Contrary to what the previous examples have shown, Julia did not verbally convey her production problems (with an EDT) but solely relied on sequential, vocal and visual-gestural strategies to request help. Unlike Sally and Elena, she looked straight at her interlocutor, and it functioned as an indication that her partner's participation was now considered relevant for the current activity. This further demonstrates that multimodal strategies can also be mobilized by learners when they experience pronunciation-related difficulties, along with lexical ones.

The second example is taken from Pair 13 in French, where this time Francis (A03) is the non-native speaker and Elena (F03) the native speaker. Here Francis is talking about the kinds of sensitive topics that friends can have during a conversation, and he does not exactly find the words for it.

Excerpt 2B – Pair 13 (FR)

```
      1  *FRAN:    um mais (1.278) en même temps on peu:ut vraiment
                   si [/] si:i (1.033) dans une groupe euh
                  **********************************
                  ((left hand held; looks up e.))
                  qui:i [/] qui discutons de:es des choses <politiques>.
                  ************************************************
                  ((left hand rotating))
         *ELEN:                                                +< mm mm
                                                                 ((head nod))
→     2  *FRAN:   ou des choses euh (1.655) <quoi tu [/] tu>.
                  ****************************************
            ((left hand held e.))   ((left open palm extended f.))
```



e. f.

3 *ELEN: <religieuses politiques> les [/] les choses
****
((left hand rotating + looks at FRAN g.))
un peu:u [/] un peu tabou.

4 *FRAN: +< oui tout ça.
((shoulder shrug and palm up open hands g.))

Contrary to Elena in the previous exchange (Excerpt 1b), and just like Julia, Francis does not explicitly signal to his tandem partner that he is looking for a word, but he still displays that his talk is currently being suspended, with the held gesture (picture e.), and the combination of different vocal fluencemes (prolongations "ca:an"," i:if", unfilled and filled pause). It first appears that the L2 learner is mainly concerned with buying time to work on his production, given the high duration of his pauses (over one second, l.1 and l.2, which is above his average of 834ms in his L2). After retrieving one noun phrase ("choses politiques" l. 1) he initiates another one ("des choses" l.2) and eventually shifts from his solitary word search project to a joint one by inviting Elena to take part in it. He does so by gazing, and extending his left open palm (which was previously held) towards her (example of a *Palm Up Open Hand Gesture*, cf Müller, 2017; picture f.). This "offering" gesture (Streeck, 2009) appears to metaphorically hand over Francis' current search to his partner, who joins in and offers a new lexical item, "(choses) religieuses politiques", followed by an elaboration further enhancing Francis' initial idea: "les choses un peu [/] un peu:u tabou" (l.3). She also produces a cyclic gesture at the same time (picture g.). As seen earlier (Chap. 1, section III. 3.3.2.), these gestures can be used to express duration, continuity and process (Ladewig, 2014) and it appears here that she is producing it in order to ensure continuity between Francis' previous utterance and her own. This is further developed in Chapter 5 where we present a typology of cyclic gestures and their relationship to inter-(dis)fluency. A state of mutual understanding is then accomplished when Francis, almost

immediately after Elena's prior turn, offers a positive assessment "oui tout ça" (l. 4), accompanied by a shoulder shrug, which further displays his affiliation. In this case, the fluencemes emerged in a context of co-construction, which further supports the idea that word searches are not only internal activities associated with speech difficulties, but also collaborative ones that can be co-achieved (Rydell, 2019).

These two examples have demonstrated the role of mutual gaze during word searching or repair sequences, which, along with other vocal and gestural cues, signal that a state of co-participation to the joint project may become relevant. This co-participation reflects the interactional dimension of fluencemes, which do not only display internal production processes (self-oriented, OCM) but function as relevant interactional resources (other-oriented, ICM).

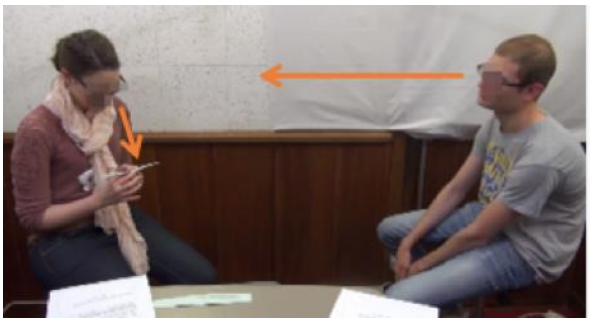
Strategy 3. Gaze towards the piece of paper: disengaging from the current activity

The last communication strategy selected for this section illustrates very different modes of behavior, further reflecting the interactional ambivalence of fluencemes. The first example is taken from Pair 09 in which the French participant Emilie (F09-NNS) is interacting in English with her American Partner Arthur (A09-NS). Just like all the other argumentative tasks, the participants were asked to discuss a topic that was written on a piece of paper, and then decide on their level of agreement at the end of the discussion. Here the topic was: "the tourist sees what he sees and the tourist sees what he has come to see." Right from the beginning of the conversation, Emilie had a difficult time grasping the meaning of the topic and expressing herself, which led to interactional-related difficulties. While this context is a little similar to the previous examples, as Emilie is experiencing production difficulties due to her lack of vocabulary in her second language, her behavior is radically different from the previous speakers. The interaction between Emilie and Arthur is organized in terms of a two-pair part exchange involving two adjacency pairs (question-answer) where Arthur, who is actively participating in the interaction, invites his partner to speak by asking her questions (l. 1 and 3), completes her utterances (l.5), and puts an end to the exchange (l.8).

Excerpt 3A – Pair 9.

1 *ART: (0.410) but um (0.440) what do you think about that distinction in general?
((gazes at EMI))

2 *EMI: um (1.870) um +/.
((gaze fixed on the paper))



3 *ART: do you think it makes sense?

4 *EMI: (0.590) &y [//] yes of course because uh the traveler um (0.965) e [//] he doesn't &uh (0.520) decide and uh he:e [/] he see what uh (0.930) [/] what uh (0.580) +...
((gazes at the paper; slight head movement then gazes at ART))

5 *ART: whatever is there.

6 *EMI: yeah.

7 *EMI: (0.490) a:and uh whereas the [/] the [/] the tourist um (0.570) [!] see what's he wants to see so.
((gazes at the paper))

8 *ART: ok yeah agreed I'd say.

As it is shown in the transcription, Arthur first invites his partner to elaborate on the topic by asking her a question that is quite straightforward (l.1). But Emilie fails to produce a fluent verbal delivery, as the latter is filled with a complex fluenceme sequence, comprised of two filled pauses and one unfilled pause of nearly two seconds (l.2), which is a significant delay (the average duration of her pauses is 450 ms in her L2). As opposed to Sally and Elena in the previous examples who managed to hold the floor while looking for a specific lexical item, Emilie fails to do so, as Arthur interrupts her to rephrase his initial question (l.3). This loss of floor can be explained by the fact that Emilie is totally disengaged from the interaction, as her eyes are fixed on the piece of paper she is holding in her hand. As the quantitative results indicated, French speakers spent half of their time gazing away while producing fluencemes (57%) but

also sometimes towards the piece of paper (14%.) Emilie (F09) actually spent 38% of her time gazing towards the piece of paper while producing fluencemes in her L2, which is significantly higher than the average of her group. But it should be noted that she also did it quite frequently in her L1 (32%). This shows that in either language Emilie constantly needed to rely on the piece of paper in order to interact with her partner, which illustrates an idiosyncratic feature. It seems that her strategy in this case is to suspend speech for a fairly long time in order to look at the piece of paper and deal with her lexical difficulties. This may help her to better perform in the interactional task (i.e. answer questions, give her opinion) while thinking about what to say next, and how to say it, with the help of the piece of paper. She actually repeats the phrase that was initially written on the piece of paper ("see what he wants to see") to finish her argument (l.7). The fluencemes found in this context are thus very different from the ones analyzed previously, as they merely contribute to the flow of the interaction, since Emilie failed to keep the floor and provide a satisfactory answer, which prompted Arthur to end the sequence ("ok agreed I'd say", l. 8).

A similar instance of gazing is found in Pair 18 in French between the American speaker Rosie (A18-NNS) and the French speaker Sophie (F18-NS). In this excerpt, the two speakers are talking about social media (same topic as in Excerpt 1A, "le paradoxe des réseaux sociaux c'est que cela rend les gens plus seuls"). After reading the piece of paper aloud to her partner, Sophie first initiated the exchange, and gave her opinion on the matter. The selected excerpt starts from here.

Excerpt 3B – Pair 18

1 *SOP: (0.500) j(e) pense que:e (0.513) bah ça peut permettre de rester en contact avec les gens.

2 *ROS: mm mm.

3 *SOP: parce-que par exemple sur facebook on peut retrouver des personnes qu'on avai:it pas vu depuis longtemps.

4 *SOP: (1.275) et voilà.

5 *ROS: mm mm.

→ **6 *ROS: (0.550) j'ai besoin du vocabulaire.**
((reaches toward the piece of paper))

```
7    *SOP:  (laughs).
8    *ROS:  (0.587) ok.
       ((gazes towards paper))
9   *ROS:   (0.925) j je ne suis pas complètement (0.725) d'accord
parce-que je pense que les gens [/] euh les gens choisissent d'être
euh plus seuls.
10  *ROS:   eum je pense que:e le:es réseaux sociaux um (1.200) hhh.
limitent l'interaction entre des gens.
11   *SOP:  mm.
```

In line 4, Sophie projects the end of her turn with a sequence-final "voilà", which gives the floor to Rosie. It is now Rosie's turn to give her opinion, but as shown in the transcription, her turn is delayed by two vocal fluencemes: one during which she leans forward to reach the piece of paper (an unfilled pause of 550 milliseconds), and a second one (an unfilled pause of 587 ms), during which she inspects the piece of paper. These actions thus momentarily disrupt the progressivity of the exchange, as the L2 speaker is not oriented to the current argumentative task anymore, but to her own expressive difficulties. This is overtly expressed in line 6: "j'ai besoin du vocabulaire", which further justifies her need to perform these actions, which are now deemed relevant in order to pursue the exchange. The piece of paper thus becomes a relevant *pedagogical* tool, which provides the right expressions and vocabulary in the target language. As depicted in Fig. 48, Rosie constantly shifts her gaze, alternating between the piece of paper and her interlocutor, and swinging back and forth between her lack of vocabulary and the interactional task at hand. In this case, the L2 speaker's strategy is to delay parts of her delivery and rely on the piece of paper, in order to provide parts of her answer that are written on the piece of paper (with lexical chunks such as "réseaux sociaux" or "être plus seul"). This strategy was in fact common among both native and non-native speakers, who often looked back at the piece of paper to get a precise idea of the topic, not necessarily because of lexical difficulties (see Excerpt 4a in 2.2.3.). These examples show the importance of material objects in multimodal communication (Boutet, 2018, Goodwin, 2003, Streeck et al., 2011, see Chap. 1, section III. 3.3.1.), and how they may influence speakers' actions. This is further discussed in Chapter 4 when we report the results from the DisReg Corpus.

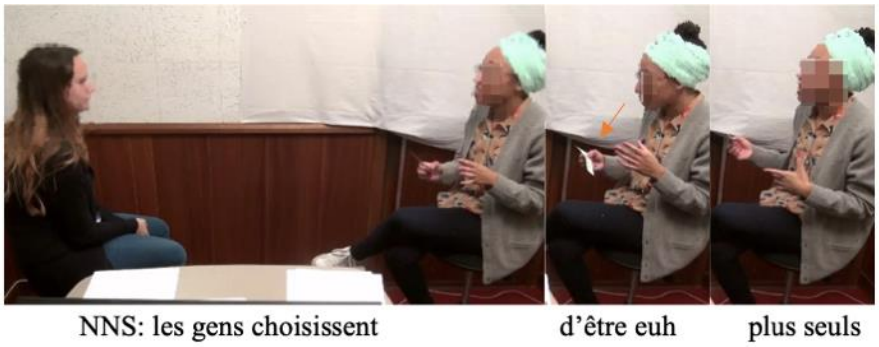


Figure 48. *Gaze shifts by the non-native speaker (NNS)*

In brief, it is primordial to consider the social, multimodal, and material structure of the environment to fully grasp the deployment of fluencemes during communication strategies. In the course of face-to-face interactions, and further embodied within a *Participation framework* (Goffman, 1981; Goodwin, 1980; Goodwin & Goodwin, 2004, see Chap. 1, section III. 3.2.1.), interactants are constantly invited to display forms of engagement or disengagement in their emerging talk. Word searching or repair episodes present a number of possibilities for the learners to either put the interaction on hold to display embodied thinking and the progressivity of the word search (Strategy 1), request help from their interlocutor (Strategy 2), or retreat into a more solitary activity with the help of an external object (Strategy 3). This diversity of behaviors reflects individual preferences, in line with Gullberg (2011), but it also further demonstrates the multifunctionality of fluencemes, which can either create "fluency" or "disfluency", depending on the point of view taken. While they may disrupt the flow of speech by inserting significant delays in the acoustic channel, these delays can embody an interactional process, thus ensuring continuity between the interactants' co-actions (or on the contrary, they may also momentarily disrupt the continuity of the exchange, cf excerpts 3a and 3b). In the following subsection, we further emphasize the role of fluencemes during the negotiation and co-construction of meaning in situated practices.

2.2.3. Inter-(dis)fluency and the co-construction of meaning in situated pedagogical practices

While the previous section focused on the speakers' individual strategies, the present one takes into account the joint productions of the interactants, and the way their actions may be situated within a larger pedagogical setting. Further in line with the notion of *secondary didacticity* (see section I.1.1.) the present section explores the different speakers' pedagogical intentions, and focuses on three selected examples

from the same pair (Pair 11) interacting in English and in French. This pair was selected because of the quality of their exchanges: the two participants took their native speaker role very seriously, and tried to give each other as much corrective feedback as possible[129]. Most of the following analyses are taken from Kosmala (2020a) and explore the way interactants co-construct and negotiate meaning in tandem.

Readjusting meaning through talk and gesture

The first sequence (4a) is taken from the same interaction in English as excerpt 1A (section 2.2.2.), in which Sally (F11), the French speaker is the non-native speaker and Harry (A11), the American speaker, is the native speaker. During this argumentative task, the speakers were asked to discuss the following topic: "Paradoxically, social media make people more lonely". This sequence takes place shortly after sequence 1A, during which Sally is talking about reality television shows.

Excerpt 4a – Pair 11 French

1 *SAL: because you [/] you [/] you go to TV show and you say oh what a [//] all I want is to [/] to have my <swimming pool> etc.
2 *HAR: +< yeah.
→ **3 *SAL: o(ne) +//.**
****
((points towards the piece of paper))
4 *HAR: I agree.
****
((points towards Sally))
5 *HAR: one thing though is I think they're asking about like facebook.
((takes the piece of paper from Sally))
6 *SAL: +< yeah lonely <so lonely so> +/.
7 *HAR: +< they're asking about like face [/] they're asking about facebook and twitter and stuff.
8 *SAL: oh yeah.
→ **9 *HAR: so like social media is like (1.150) [/] is like the internet.**
****

[129] Sylwia Scheuer, who collected the SITAF corpus with Céline Horgues, actually pointed out during a SeSyLiA seminar that A11 (Harry) and F11 (Sally) were a really engaging pair and that they were perfect for the analysis of corrective feedback (Debras et al., 2015, 2020).

((right hand rotating; gazes at Sally))

It turns out that Sally has misunderstood the topic, as she began to talk about reality television shows (which portray individuals in real-life situations on TV), and the people that appeared on them. She started by criticizing them for being very dramatic and very proud on TV, and then pointed out the problems they encounter when the show ends and they have to go on with their lives (cf Excerpt 1a, section 2.2.2). Sally seems very eager to discuss this topic, but she has not yet realized that it was not the one indicated on the piece of paper. In line 3, Harry initiates a turn at a transition relevant place, following Sally's prior utterance, and attempts to shift the topic of conversation to lead her in the right direction. He does so by first pointing towards the piece of paper, as shown in the following illustration:

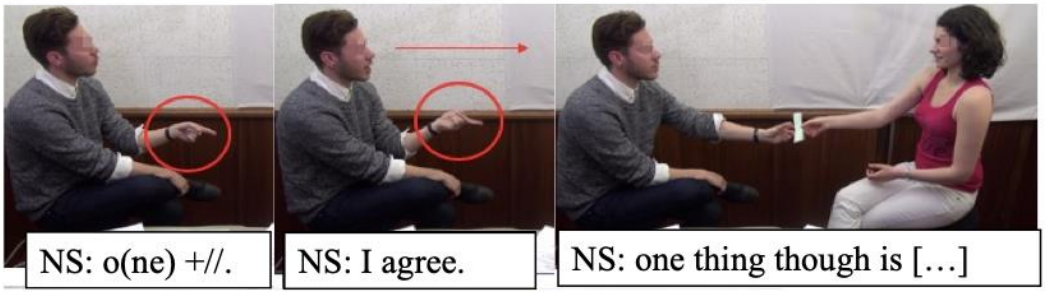


Figure 49. *Deictic gesture during a fluenceme sequence performed by the native speaker (NS)*

When Harry produces the deictic gesture directed towards the piece of paper, he also produces a truncated word and a self-interruption at the same time (line 3). It is interesting to note that he immediately reshaped the course of his talk to express his agreement ("I agree"). He probably initially meant to tell her upfront that it was not the right topic (which he does in his subsequent utterance "one thing though is I think they're talking about like facebook"), but instead he decided to indicate his agreement. He may have carefully chosen not to interrupt her abruptly, as to display his orientation towards her stance. Once more, the fluenceme sequence produced here is thus by no means a sign of DISfluency per se, as Harry is not experiencing any production difficulty (as opposed to Sally who produces a series of very complex

fluenceme sequences); it is actually used pragmatically to align with his interlocutor, and to help her save her face, as well as his own (Goffman, 1955). Interrupting her in the midst of her talk to tell her that she was wrong would have threatened his face as an understanding tandem partner. A close examination of eye gaze and gesture is also revealing of the current interactional practice: when the native speaker first produces the fluenceme sequence and the deictic gesture, his gaze is fixed on the piece of paper, but after interrupting himself and indicating his agreement (second picture) he quickly gazes towards his interlocutor, moves his palm and finger upwards, and slightly orients it towards her. This change of orientation offers an additional interactive dimension to the gesture (Bavelas et al. 1992). Once more, the piece of paper is used as a medium to negotiate meaning and modify the course of the interaction, in ways that are relevant to Harry's pedagogical actions.

The native speaker thus plays two roles here: he first fulfills the role of the co-participant in a conversation in which he is oriented towards his partner, pays attention to what she is saying, makes sure not to interrupt, and displays his stance; but he also plays the role of the native speaker, who has to adapt his speech to the non-native speaker, and make sure that his interlocutor understands the topic. This illustrates the fact that linguistic abilities are tightly linked to the different social identities of the speakers, and that the "expert" or "novice" status of the co-participants are constructed locally within the course of interaction (Pekarek Doehler, 2006). In this case, it was both relevant for Harry to play the role of the "expert" native speaker in order to clarify the misunderstanding of the exchange (and he had the authority to do so, as a native speaker) but also to play the role of a cooperative hearer. He manages to play the two roles at the same time by relying on several semiotic resources. Instead of verbally asking his interlocutor to give him the piece of paper (which would completely interrupt the course of the conversation) he relies on a deictic gesture, which requires no overt verbalization as the gesture is already semantically transparent (McNeil 1985). He then quickly checks whether the topic was really about social media and gives the paper back to her so that she can read it again. But Sally has still not grasped the meaning of it, as she seems convinced that he gave her back the paper to mention the part where it says that it makes people lonely, so she starts mentioning it (line 6). Harry then interrupts her, repeats himself, and adds another piece of information related to social media: "they're asking about twitter and facebook and stuff" (line 7). Sally finally understands his point, as indicated by her oh-prefaced

declarative "oh yeah" (line 8), which displays a change of state (Atkinson & Heritage, 1984). Then, in line 9, Harry produces another fluenceme sequence that is made of a fairly long unfilled pause of 1150 milliseconds and a repetition of the discourse marker "like". As the pictorial illustration within the transcript shows, the speaker also produces another cyclic gesture synchronized with the fluenceme sequence (further analyzed in Chapter 5), and his gaze is fixed on his interlocutor. In this case, this gesture may have been used to encourage his interlocutor to speak (Ladewig 2014). Since his gaze is fixed on his partner, and Harry produces a very long unfilled pause, it may suggest that he is inviting his interlocutor to continue speaking, but this time on the right topic. The speaker thus attempted to construct the meaning around the noun phrase "social media" with his interlocutor, and encouraged her to take part in it, or at least to capture her attention (similar to teachers' gestures; see Tellier 2008) which highlights another potential pedagogical intention.

This sequence has shown two cases in which meaning had to be readjusted and elaborated by the native speaker in order to ensure continuity in discourse and between the two participants. This process of readjustment was first initiated by the deictic gesture, and then by a cyclic gesture which co-occurred with a fluenceme sequence, stressing the interactional dimension of fluencemes[130]. Moreover, these gestures also expressed the pedagogical intention of the native speaker, whose purpose was to overcome the misunderstanding in the conversation. In fact, some studies have suggested that in native/non-native interactions, native speakers use numerous deictic gestures to facilitate comprehension, promote communication and overcome inadequacies in the conversation (see Adams 1998 for a review). As we have seen (cf section I. 1.3.) a number of studies in SLA have also pointed out the use of representational gestures produced by native speakers and non-native speakers, (Gullberg 2014; Adams, 1998; Stam 2001) which will be discussed in the following analysis.

[130] These types of gestures, i.e. deictic, cyclic, thinking gestures etc., will be further analyzed in Chapter 5 where we document their uses across corpora (in both SITAF and DisReg) by taking into account their specific formational and spatial features. For the sake of clarity, the present section focused more specifically on the effect of tandem settings on fluencemes and gestures in order to address our research questions and hypotheses (cf section 1.4.).

Co-construction of meaning through parallel gesturing

In this second example, taken from the same recording, the same speaker, Sally (F11) is talking with her partner Harry (A11) about the time spent on social media which does not reflect what people actually do in real life. The two speakers very much agree with one another on the topic, as indicated in the tokens of agreement ("yeah", lines 3 and 6, "I agree" line 4; head nods, l.5). At some point in the interaction, Sally produces a simple and fairly short vocal fluenceme, a prolongation (380 ms) of the determiner "the" (line 7 in bold) before retrieving the noun "twitter".

Excerpt 4b– Pair 11 (EN)

1 *HAR: people just post what's interesting
((right palm open hand gesture))
not actually the fact that they spend like eight hours at home alone.
((right hand retracted; left palm open hand gesture))
2 *HAR: they just post the one picture
((left open hand extended outward to the left))
that's just like this is really c [//] this is the one cool thing that happened to me.
((head tilt; left open hand does a circling movement; palm is then held out upwards to the left))
3 *SAL: yeah (nods).
4 *HAR: but doesn't r [//] actually reflect what is happening to you which I agree [/] I agree with.
((both open hands held out to the left))
5 *SAL: mm mm
((head nod))
6 *SAL: yeah.
→ 7 ***SAL: they're more on the:e twitter and <facebook>**
****
((right hand gesture mimicking the action of typing of one's phone))
8 *HAR: +< they're mm yeah.
****
((similar gesture of typing on one's phone; shoulder shrug))

9 *SAL: than with all the real people that just [/] just great news I'm writing on twitter ok (0.440) it's <xx>.
10 *HAR: +< yeah.

As the illustration shows, Sally produces a representational gesture at the same time as her vocal fluenceme, and gazes towards her interlocutor. Her gesture is depicting the action of typing on one's phone, in order to convey meaning related to twitter and social media. As we have seen (section I.1.3) representational gestures in L2 discourse are typically used to elicit lexical help from the interlocutor, or to compensate for speakers' lack of vocabulary (Gullberg, 2014). However, it seems unlikely in this case that she was requesting help from her partner, since she produced the target word ("twitter") quite quickly, following the short prolongation (as opposed to Excerpt 1A where she produced a series of rather long vocal fluencemes). In addition, the relationship between the target word and the gesture is redundant (e.g. Alibali et al., 2000) in the sense that the gesture does not carry much additional semantic information about the noun "twitter", which supports our initial remark. Given the iconic property of the gesture, she still may have produced it to highlight the word "twitter" and the action of typing on one's phone (with the rhythmic movement of the gesture). This context is thus very different from Excerpt 1A, as Sally does not seem to be experiencing any production difficulty, nor is she trying to look for a specific word. In this case, she is rather displaying active participation in the interaction and seems to be seeking confirmation and agreement from her interlocutor. In fact, her partner produces a similar referential gesture while expressing his agreement (l. 18). A case of parallel gesturing (Graziano et al., 2011) is found here, where the next speaker of the conversation repeats a gesture made by the preceding speaker, which shows mutual understanding. This example further illustrates that fluencemes are not always associated with production problems, and even though they are more frequent in L2 discourse, it does not mean that L2 speakers are necessarily more "disfluent" in the interactional sense.

The final excerpt from Pair 11 (in French) shows a similar instance of parallel gesturing. This time, the speakers were asked to discuss the following topic: "l'adolescence la période la plus heureuse de ta vie?" Similarly to their exchange in French, the non-native speaker (Harry, A11) opened the conversation by giving his opinion. He first pointed out that adolescence was definitely not the happiest time of his life for many reasons, and finished by overtly indicating his stance, line 1, showing that what he has just said must be considered as personal opinion, not fact. This is where the selected sequence begins.

Excerpt 4C – Pair 11 (FR)

1 *HAR: ça c'est m(on) mo:on av [//] opinion.
2 *SAL: bah moi j(e) suis d'accord ((laughs)).
3 *HAR: oui ? ((laughs))
4 *SAL: si [/] si.
5 *SAL: nan [/] nan l'adolescence euh (1.320) y'a [//] t'as plein de boutons déjà ((laughs)).
6 *HAR: &ah oui (laughs)
7 *SAL: nan +//.
8 *SAL: oui mais t'a:as [//] tu [/] tu découvres plein de choses euh le:es [/] <les gens sont> +/.
9 *HAR: +< oui.
10 *HAR: les hormones.
11 *HAR: les hormones ((laughs)) tout à fait.
12 *SAL: c'est [/] c'est très problématique.
→ **13 *HAR: toujours raison de:e ((laughs)) +/.**

****

((left hand moving up and down like a wave, gazes at his hand))

14 *SAL: oui t'as des [/] des hauts et des bas euh

****

((produces a similar gesture; looks at Harry))

les gens deviennent gentils deviennent méchants.

15 *HAR: ah oui.

16 *SAL: ils te:e +/.

17 *HAR: oui c'est [/] c'est horrible.

18 *SAL: s c'est [/] c'e:est un passage difficile entre le monde des bisounours le monde des enfants et le monde des adultes.

19 *HAR: ah oui.

The two speakers are in perfect agreement and finish each other's sentences, which illustrates cases of joint sentence production (Sacks 1992) and *dialogic syntax* (Du Bois 2007), which is again typical of face-to-face spoken interactions. At some point in the interaction (line 13), Harry produces a complex fluenceme made of a prolongation ("de:e") and laughter[131], and he also produces a wave-like gesture at the same time. Once again, the non-native speaker relies on several semiotic resources to build the meaning of his multimodal utterance. Instead of verbally expressing the notion of ups and downs, which his partner does right after (line 14), he produces a representational[132] gesture that metaphorically conveys the meaning of ups and downs by enacting sea wave movements. He may have done it for several reasons: (1) he is experiencing lexical difficulties and needs to rely on a referential gesture to compensate for this lexical deficit, which could be an example of a communication strategy (Gullberg, 2011); (2) he is eliciting lexical help from his interlocutor (Gullberg 2014), and thus inviting her to take part in the joint word search (Goodwin & Goodwin 1986). These two explanations could apply, but what happens right after could provide a good indication of his initial intention. As shown in line 14, Sally, the native speaker, completes her partner's utterance and verbally expresses the notion of ups and downs,

[131] Note that laughter was included in our category of non-lexical sounds whenever they immediately co-occurred with fluencemes (cf Chap. 2, section II.2.2.1.)

[132] Note that this type of gesture would typically be labeled as "metaphoric" in McNeill's (1985) typology (see Chap. 1, section III. 3.3.2. and 3.3.3).

but she also repeats his wave-like gesture before he finished producing it, and the two speakers both gaze at each other during this moment.

All of these elements thus illustrate another case of co-construction, but this time with no readjustment (cf Sequence 4B). Both speakers jointly deployed speech, gaze and referential gestures in tandem to negotiate meaning. Gaze plays a key role in this example. When Harry initially produced his referential gesture (line 13) he did not look back at his interlocutor to signal that he was in trouble, but he kept looking at his gesture instead. This is what Streeck (2008) calls "depicting by gesture", when the gesturer "attends to the gestures, glances at the hands every so often during a depiction episode, and so does the recipient" (Streeck, 2008, p. 289). Then, Sally produced a similar wave-like gesture, but not to assist him or help him, but rather to demonstrate a visible form of engagement by taking part in the co-construction. And when they produced the same gesture in tandem, they also gazed at each other. This is closely related to the notion of interactional synchrony (Wallbott, 1995) where we find social congruence and a sense of co-operation, and it also shows another instance of parallel gesturing (cf Sequence 4b) where speakers repeat each other's gestures. A similar instance was also found in Debras et al., (2020) in which they showed the way a native speaker (also from the SITAF Corpus, taken from Task 1, cf Chap. 2, section I.1.2.) displayed her understanding of the non-native speaker's prior turn by repeating the same lexical item and gesture initially produced by her partner. The authors pointed out the prominent role of representational gestures in maintaining mutual understanding.

This sense of mutual cooperation is also shown in the different tokens of agreement found throughout the interaction ("ah oui", "si", "oui"). As opposed to Excerpt 4a, where the interaction was potentially in danger because of the misunderstanding, here the co-participants are perfectly aligned with each other. However, it is still possible that the non-native speaker did not know the word for "hauts et bas", which could be the reason why he produced the referential gesture, but it was not deemed relevant in this context, because he did not overtly seek help from his interlocutor. He may have intentionally chosen to use a gesture which has more expressive features and visual properties to convey the meaning of ups and downs; in any case this was found to be successful as his interlocutor repeated the same gesture and later elaborated on the meaning ("un passage difficile entre le monde des bisounours le monde des enfants et le monde des adultes", line 18). Therefore, the

meaning around ups and downs was co-constructed in tandem, and this activity reinforced the mutual understanding of the two speakers. This further argues against the view of representational gestures as being simply *compensatory* (in line with the LHR, section 1.3), as they can also be used to display forms of engagement.

The native and non-native roles were less strictly defined in this case (as opposed to Excerpt 4a), but the speakers both shared a similar pedagogical intention and they used referential gestures in order to be understood. This stresses the idea that pedagogical gestures (Tellier, 2008) can be used outside the class environment to serve pedagogical purposes (*secondary didacticity*), and that they can show different degrees of didacticity (Azaoui 2015) depending on the pedagogical intention, the context, the type of gesture, and the direction of gaze. The three examples examined above illustrate this point. In Sequence 4a, the deictic and the cyclic gesture used by the native speaker conveyed a stronger pedagogical intention in the sense that he wanted to overcome his partner's misunderstanding of the topic by directing her attention to the piece of paper, and then by inviting her to elaborate on the meaning of social media. In Excerpts 4b and 4c, the referential gestures used by the two speakers reinforced their mutual understanding, but Sally, the native speaker, still elaborated on the meaning of the referential gesture by verbalizing it (in sequence 4b), which may show that she still intended to take part in the co-construction of the expression. This type of intention can still be considered pedagogical, but to a lesser extent than Harry's in Sequence 4a.

To conclude, our detailed qualitative analyses of the data have exemplified how L2 (dis)fluency is not necessarily associated with a "lack of skill" or a "resource deficit", but is largely embedded within a set of interactional practices involving the gesturers' active co-participation in the talk. The degree of (dis)fluency found in those markers is neither fixed nor systematic and is highly determined by their context of use, as well as their accompanying visual-gestural features. Since fluencemes carry little semantic or pragmatic information, their accompanying visible behavior can further determine whether they display instances of *interactive communication management* (ICM) or *own communication management* (OCM), and their status is constantly being re-shaped in the course of the multimodal talk, depending on the gesturers' available resources. The aim of this section was to bridge the gap between quantitative and qualitative findings in order to better illustrate the ambivalent nature of fluencemes

which can both be the result of internal speech processes and social interactional practices.

III. Discussion

The present section addresses our research questions formulated earlier (cf section I.1.4) by drawing on a selection of findings obtained from the statistical treatments (section II. 2.1.) as well as our qualitative analyses (section II. 2.2.) This section is structured as follows: we first report on the potential differences between L1 and L2 fluency in the American and French groups (3.1.), discuss the role of fluency in situated tandem discourse and the multimodal strategies yielded by non-native speakers to deal with language difficulties in their L2 (3.2) and conclude on the crucial role of gaze and gesture in the multi-level ambivalence of inter-(dis)fluency.

3.1. Specificities of L1 and L2 Fluency

One of the main questions this chapter sought to answer was whether L2 fluency differed significantly from L1 fluency, or whether the two were correlated (RQ1). As we have seen, these differences can be measured by looking at temporal variables and fluency rates (e.g. Tavakoli, 2011, Riggenbach, 1991). In addition to temporal variables, we also looked more specifically at sequential (i.e. patterns of co-occurrence), positional (utterance position), and visual-gestural features.

3.1.1. Fluenceme rate, distribution, and patterns of co-occurrence in the American and French groups

In line with previous studies in SLA (e.g. De Jong, 2016, Fehringer & Fry, 2007, Gilquin, 2008) our results confirm earlier predictions that non-native speakers produce significantly more fluencemes in their L2 than in their L1, and this was true for both the American and French groups. However, fluency is not only assessed by frequency measures, but durational features as well; unlike previous studies (e.g. Cenoz, 1998 ; Rasier & Hiligsmann, 2007), no significant differences were found in the average duration of the vocal markers (i.e. filled and unfilled pauses and prolongations) in L1 and L2 for the French and American groups. This can be explained by the high degree of variability and dispersion found in the data, reflecting individual differences.

At the sequence level, a number of differences were found. American speakers were more likely to produce complex sequences in their L2 than in their L1, and their sequences contained a higher number of markers in their L2. This finding may indicate one salient feature of L2 fluency, characterized by a higher rate of long and complex sequences. However, this was not true for the French group where no significant differences were found, so this difference may also be language-specific. Differences were also found in the sequence configurations: American speakers showed a tendency to produce sequences which mainly consisted in the *VOC+MS* (vocal marker + morpho-syntactic marker) configuration in their L2, and the *VOC+VOC* configuration in their L1, suggesting preferences for stalling strategies in the L1, as opposed to a mixture of stalling and repair mechanisms in the L2. For the French group, the *VOC+MS* pattern was used more frequently in their L1 than in their L2, showing the opposite tendency. Overall, these results seem to suggest that some specific patterns are more prominent than others, especially *VOC+MS* and *VOC+VOC*[133], but their use is not systematically determined by levels of proficiency. This does not support our initial hypothesis which claimed that non-native speakers would produce fewer recurrent patterns of combination (cf RQ1), since the *VOC+MS* pattern was used 50% of the time by American speakers in their L2. It is still interesting to note, however, that the speakers used different types of patterns in their L1 and L2. This further emphasizes the idea that fluencemes are made of dynamic and flexible structures with different possible configurations, which are more or less fixed depending on language and other contextual features, following assumptions from Cognitive Grammar and usage-based linguistics (cf Chap. 1, section III. 3.1.2.). For instance, our results showed a correlation between sequence complexity and communication management (section II.2.2.1.), with simple sequences associated with *Interactive Communication management* (ICM) and complex ones associated with *Own Communication Management* (OCM). This was also illustrated in our qualitative analyses, where we showed how highly complex sequences occurred during specific instances of word searching trouble (e.g. excerpts 1a and 1b).

As to the utterance position of the fluenceme sequences, not many significant differences were found between L1 and L2 in the two groups, except for the medial and

[133] These combinations could include a more fined-grained level of abstraction by identifying the specific types of vocal markers that typically occur with specific types of morpho-syntactic markers (e.g. Crible, 2018), but this was not the primary goal of this chapter.

final position. The American group was found to produce more sequences in medial position in their L2 than in their L1, and the French group produced slightly more sequences in final position in their L1 and their L2. However, contrary to previous studies (e.g. Cenoz, 1998, Rasier & Hiligsmann, 2007) the position of the sequences was not annotated with regard to syntactic structure (e.g. intra-clausal versus inter-clausal, or inter-propositional versus intra-propositional) or at the level of the word or morpheme, but more in terms of the overall stream of speech, in line with our notion of multidimensional flow (cf Chap. 1, section IV.4.2.).

In addition, several crosslinguistic differences were found between the two groups, perhaps reflecting language-specific or cultural-specific preferences, to name but a few: while the American speakers produced a significant number of unfilled pauses compared to French speakers in their L1, French speakers produced more filled pauses than American speakers in their L2. Another interesting finding was the realization of the filled pause ("(e)uh" versus "(e)um"), which showed opposite tendencies in the two groups. While American speakers used predominantly the um-type filled pause in their L1 as opposed to the uh-type in their L2, French speakers did exactly the opposite, with a strong preference for uh-type filled pauses in their L1 and um-type filled pauses in their L2. This finding validates cross-linguistic preferences, further in line with Clark & Fox Tree's (2002) and Candea et al.'s (2005) argument that filled pauses are language-specific. It is thus important to note that L2 fluency is not only determined by overall differences in language proficiency, as it can also be speaker-, and to a larger extent, language-specific.

3.1.2. Fluency and language proficiency

A second closely related question concerns the correlation between fluency and language proficiency (RQ2). While the present study does not aim to target these aspects specifically, it was still deemed relevant to investigate a potential relation between fluency and proficiency, in line with previous work (e.g. De Jong, 2016, Riggenbach, 1991). Even though the proficiency levels of the participants were not officially assessed but self-evaluated by the students themselves, it is still interesting to study the relationship between *perceived proficiency* and fluency rates.

As findings indicated, no significant correlation was found between the two variables in the two groups. This is consistent with Brand & Götz (2013) who found no trend for a correlation of accuracy (i.e. levels of oral competence in terms of

grammatical, lexical, and phonological proficiency) and fluency, but this was perhaps due to the limited size of their sample (only 5 speakers). Similarly, the size of our selected sample is quite small, which makes it difficult to draw general conclusions. However, this finding does show to a certain extent that L2 fluency is not systematically associated with (perceived) proficiency. This further questions the extent to which fluency rates are a valid indicator of L2 proficiency, as previously challenged by a number of authors (see De Jong et al., 2015, Derwing et al., 2010; Zuniga & Simard, 2019).

3.1.3. Gestural and gaze behavior

One of the major contributions of the present study is to analyze inter-(dis)fluency in terms of gaze and gestural behavior, in order to go beyond the traditional view of L2 fluency in SLA which has too often been restricted to temporal variables. Following previous work (e.g. Gullberg, 1998, Kita, 1993, Stam 2006), we expected a higher rate of gestures in L2 than in L1, and our findings confirmed this prediction, as the two speaker groups produced significantly more gestures in their L2 than in their L1, both in fluent and disfluent cycles of speech (even though gestures did not frequently co-occur during fluencemes), which demonstrates a higher gestural activity in the L2. In addition, speakers showed a tendency to hold their gestures during the production of fluencemes, which further confirms the relationship between speech suspension and gesture suspension (e.g. Esposito & Marinaro, 2007, Graziano & Gullberg, 2018). This is further developed in Chapter 5. In addition, both the American and French speakers were found to hold their hands during fluencemes more frequently in their L2 than in their L1, which may reflect a need in the L2 to hold different modalities (vocal and gestural) at the same time in order to put the interaction on hold and buy more time. This observation is also consistent with the prominence of the *VOC+VOC* pattern in the Americans' L1, which further reflects a different type of time-buying strategy at the vocal level.

Additionally, results showed that the two speaker groups produced a higher proportion of referential gestures in their L1 than in their L2 overall, which challenges the idea that speakers produce more referential gestures in their L2 to deal with lexical difficulties (e.g. Stam, 2001). This is further discussed in section 3.2. Moreover, a large majority of the gestures found during fluencemes were pragmatic, and not referential, contrary to what the Lexical Retrieval Hypothesis (Krauss et al., 2000) suggests

(discussed in section III.3.2.). In addition, it is worth noting that speakers produced a higher proportion of thinking gestures in their L2 than in their L1 (and almost exclusively during fluencemes), which may reflect one prominent feature of L2 (dis)fluency as a display of *doing thinking* (Heller, 2021, cf Chap. 5). Such gestures, along with thinking faces, were also examined in our qualitative analyses, and showed how they were used as an interactional practice to display the progressivity of a word search (excerpts 1b and 2B). When it comes to gazing behavior, the two groups showed a tendency to withdraw their gaze during fluencemes (i.e. gaze away, or towards the piece of paper), and this was the case both in L1 and in L2, which demonstrates a notable feature of (dis)fluency in general, regardless of language proficiency.

To conclude, many inter-related features need to be taken into account when examining the relation between L1 and L2 fluency, from temporal variables and sequential features to visual-gestural behavior during fluencemes and outside fluencemes. Table 28 summarizes the different variables used to analyze the distribution of fluencemes and fluenceme sequences, as well as gaze and gestural behavior, across the two speaker groups. In sum, many differences can be found between the two speaker groups, as well as instances of individual variability which prevented some variables (i.e. duration, position, type, and length) to reach significance. What these results suggest is that specific aspects of L1 and L2 fluency do differ on some levels (i.e. frequency, form, type, length, configuration, gesture phrase and gesture rate), but not systematically across the two speaker groups, suggesting speaker-and language-specific patterns of behavior, in line with Gullberg (2011) and De Jong (2018). The most salient feature remains frequency, as nearly all speakers from both groups produced significantly more fluencemes in their L2 than their L1, as well as more gestures. This difference in frequency does not mean, however, that L2 fluency is necessarily associated with a lack of proficiency in the L2, but it can be further interpreted as a sign that non-native speakers require more stalling and repair strategies in their L2 to perform a variety of actions such as word search, planning, turn-taking or time-buying.

Table 28. *Summary of (dis)fluency variables for the American and French group*

		American group	French group	Inter-group differences
Fluenceme level	**Rate**	Higher rate in L2 than in L1	Higher rate in L2 than in L1	AM speakers produced a higher rate in their L2 than the FR
	Marker Type	More IR and FP in L2, more SI and UP in L1	More SI and PR in their L1, and more NL sounds in their L2	AM speakers produced more UP than FR speakers in L1, but FR speakers produced more FP in their L1 than AM speakers
	Form	More ums in L1 than L2	More uhs in L1 than L2	AM used um more often than the FR in their L1
	Duration	no significant differences	no significant differences	no significant differences
Sequence level	**Type**	more complex sequences in L2	no significant differences	no significant differences
	Lenght	higher number of markers combined in L2	no significant differences	no significant differences
	Configuration	more stalling strategies in L2 (VOC+VOC) and a combination of stalling and repair in L1 (VOC+MS)	higher proportion of VOC+MS pattern in L1	no significant differences
	Position	more instances of medial position in L2	more instances of final position in L1	no significant differences
Visual-gestural level	**Gesture phrase**	more holds in L2 than L1	more holds in L2 than L1	no significant differences
	Gesture production	more gestures in L2	more gestures in L2	no significant differences
	Gesture types and subtypes	more pragmatic gestures during fluencemes both L1 and L2, and more thinking gestures in L2	more pragmatic gestures during fluencemes both L1 and L2, and more thinking gestures in L2	no significant differences
	Gaze direction	more instances of gaze withdrawal during fluencemes (both L1 and L2)	more instances of gaze withdrawal during fluencemes (both L1 and L2)	FR speakers gazed towards paper more often than AM in L1, while AM speakers gazed away more frequently than the FR

Some of these communication strategies, as we have seen, may lead to multimodal gestalts of "doing thinking" and to a larger extent, *doing fluency*, i.e. managing the fluency of the speakers' multimodal flow through instances of gestural and vocal suspensions (i.e. gesture holds and pausing strategies), or embodied displays of thinking (gaze withdrawal, thinking face, and thinking gesture), along with other relevant interactional actions. This point is further elaborated in Chapter 5.

3.2. How L2 learners deal with language difficulties: beyond lexical retrieval

3.2.1. L2 fluency anchored in language use

It has been stressed out multiple times in this thesis that the concept of L2 fluency needs to be situated within actual language use, by taking into account the multimodal, social, and material environment in which fluencemes emerge. Further in line with the framework of *CA-SLA* (Pekarek-Doehler, 2006) and the notion of interactional competence (Galaczi & Taylor, 2018) we claim that L2 competence, and to a larger extent L2 inter-(dis)fluency, is not only the result of individual and internal cognitive processes related to encoding difficulties (e.g. Hilton, 2009) but may also constitute a relevant interactional tool for maintaining the fluency of the exchange (Peltonen, 2020). In this view, L2 speakers make use of a variety of features to deal with language difficulties in the course of the interaction. Even though vocal and verbal fluencemes remain the default mode (according to the scope of relevant behaviors (SRB) theory, cf Cienki, 2015) for dealing with such difficulties (since they occur predominantly without gestures), we have shown specific instances (cf section II. 2.2.) during which speakers mobilized a combination of vocal and visual-gestural features to perform a series of actions, such as requesting help, putting the current speaking activity on hold, displaying the progressivity of the talk, and the like.

In addition, we paid specific attention to the *pedagogical* dimension of tandem interactions (cf sections I. 1.1. and II.2.2.3.) which present a number of specificities. First, the relationship between the participants is asymmetrical (Kurhila, 2001) as they involve one expert (a native speaker) and one novice of the language (a non-native speaker). Consequently, this may have prompted native speakers to take on different discourse roles, i.e. the role of co-participant and the role of expert, and we have seen that they very often alternated between these two (cf section II.2.2.3) in order to give assistance to their non-native partner. In addition, the fact that the two partners alternated between their native and non-native status also further reinforced mutual solidarity and a sense of cooperation, which offers an additional interactive frame of analysis.

3.2.2. The interplay of vocal, verbal, and visual-gestural resources

As argued above, the combination of resources mobilized by speakers in the course of the exchange enabled them to maintain *inter-fluency*, which emphasizes the communicative and interactional dimension of fluencemes, which have too often been restricted to episodes of trouble or encoding difficulties. In line with Gullberg (2008) and McCafferty (2002), we believe that gestures further provide an interactional effect on fluencemes. Following Graziano & Gullberg (2013, 2018) our results showed that speakers produced predominantly more pragmatic gestures during fluencemes than referential ones both in L1 and L2, which makes it difficult to assess theories like the Lexical Retrieval Hypothesis (LRH) which claims that referential gestures facilitate access to lexical memory and compensate for speech failure. In addition, our qualitative analyses have shown that representational gestures were not necessarily used to overcome language difficulties or to deal with intrapersonal problems, but rather to display forms of engagement in the interactional task at hand, leading to multimodal joint productions (excerpts 4b and 4c). In this view, we do not believe that learners make use of referential gestures to compensate for a lack of skill or to activate a spatio-motoric mode of thinking (e.g. Kita, 2000), which would imply that such gestures accompanying fluencemes are exclusively based on internal processes reflecting models of speech production, thus completely disregarding interactional dynamics. Therefore, we remain doubtful about these claims.

While a lot of emphasis has been laid on the role of gestures as a window onto thought and cognition, facilitating access to mental spaces, working memory, or leading towards a different mode of thinking (e.g. Kita, 2000; McNeill, 1992; Stam, 2018; Sweetser, 2007) the present study focused more specifically on the communicative aspects of gestures, and their emergence in situated interaction (i.e. tandem settings), which can further our understanding of fluencemes and their interactional ambivalence.

3.3. The importance of visible bodily behavior in the multi-level ambivalence of fluencemes

As our quantitative findings further showed, a majority of the fluenceme sequences performed Own Communication Management (henceforth OCM) both in L1 and L2, with only a small proportion of Interactive Communication Management (henceforth ICM) overall. This result may explain why a great number of studies on (dis)fluency

have not explored their interactional quality in detail and rather focused on internal and cognitive processes, as the latter uncover a wider range of fluencemes. However, the fact that the interactional dimension of fluency has often been overlooked in past studies is also largely explained by the differences found in theoretical and methodological frameworks (cf Chap. 1). In line with McCarthy (2009), Peltonen (2020), and Pekarek-Doehler (2006), and further grounded in our multidisciplinary framework (cf Chap. 1, section IV) the present study aims to bridge the gap between quantitative production-based (mostly psycholinguistic or phonetic) studies conducted on disfluency and usage-based, interactional, multimodal approaches to social interaction, to integrate different levels of analysis. In particular, we have seen that visual-gestural features played a certain role in the type of communication management, with a higher proportion of gestures found during ICM than OCM in L2 for the American group (however, no differences were found for the French group). This further questions the extent to which gaze and gesture may reflect the multi-layered ambivalence of fluencemes (RQ4): while certain highly complex fluencemes mark a significant suspension or interruption in the speech flow because of their length and complexity (*speech DISfluency*), they may also display relevant interactional work signaling that a search is currently being performed, and that it may require the partner's co-participation (*interactional fluency*). Such interactional displays are made visible with the help of mutual gaze and co-occuring gestures performing interactive functions (i.e. enact a stance, give the floor, perform a speech act, see excerpts 2a and 2b). On the other hand, short vocal fluencemes that do not severely interrupt the flow of speech may in fact DISrupt the progressivity of the exchange (e.g. excerpts 3a and 3b) if speakers are not oriented towards it. Once again, our analysis of inter-(dis)fluency vouches for a multi-layered, multi-dimensional and multi-level approach, thus going beyond temporal features of L1 or L2 fluency. This is further discussed in the *General Conclusion* (cf Fig. 71, section III). In addition, this ambivalence was reflected in the different learners' strategies for dealing with language difficulties: while some speakers momentarily suspended the course of their multimodal production to signal production difficulties and display the progressivity of their search (excerpts 1a and 2b), some relied on mutual gaze and other gestural activities to invite their interlocutor to participate in the current search (excerpts 2a and 2b), others momentarily retreated from the current activity by displaying forms of disengagement and orientating towards the piece of paper (excerpts 3a and 3b).

Conclusion to the chapter

To conclude, while the general aim of this chapter was to address differences between L1 and L2 fluency overall, in line with the SLA literature, the present study has shown several aspects of inter-(dis)fluency that go beyond the notions of L2 competence, proficiency, and accuracy, by integrating different variables (form, position, sequence complexity, visual-gestural behavior) and levels of analysis (speech, interaction, and gesture). Our study corroborates previous findings in SLA, mainly that fluencemes are significantly more frequent in L2 than in L1, but we have also seen that they tend to form more complex sequences (for the American group), and that specific fluencemes or patterns are preferred in the L2, depending on the speaker group. Overall, our findings have shown individual and language-specific differences, which calls for further investigation in other fields of research, such as phonetics, but which also stresses the impact of individual speaking style, in line with Gullberg (2011) and De Jong (2018). Given the size of the data, however, no general conclusions can be drawn, and the number of tendencies which were reported within fluency behavior may only be specific to our type of data (i.e. semi-guided interactions in tandem settings at a French university).

In addition, unlike previous work on gesture in SLA, the goal of this chapter was not to study the role of gestures in language learning specifically (i.e. Adams, 1998; Gullberg, 1998; Tellier, 2008) or their role in cognition (i.e. Alibali et al., 2000; Kita, 2003, Stam, 2001), but rather to present ambivalent fluencemes as potential interactional components of L1 and L2 multimodal fluency, by examining their co-occurrence with visual and gestural behavior. In line with Graziano & Gullberg (2013, 2018) we have shown that even though fluencemes tend not to co-occur with gestures, a majority of gestures accompanying fluencemes perform pragmatic functions, which can be used by speakers to regulate the interaction and provide metacommunicative comments on their performance. Such actions were illustrated in our qualitative analyses by taking into account their context of use in pedagogical settings, thus stressing the need to combine quantitative and qualitative methods in order to shed light on different but inter-related aspects of our investigation. While only a small percentage of fluencemes occurred in interactional contexts (18% approx.) it is still essential to provide a closer examination of such fluencemes in situated discourse in order to understand how they may contribute to the interactive flow. Qualitative

examples have shown that they can be used as individual communication strategies, or in joint productions to negotiate meaning, and they are deployed differently by speakers, who make use of vocal fluencemes as well as gestures to resolve language difficulties and/or to maintain a speaker-hearer relationship, which is in line with Peltonen (2019) and McCarthy (2009). Once again, this underlines the interactional ambivalence of fluencemes and the fact that their usage is highly contextual and depends on a number of situational features.

However, some limitations in our study should be noted. As stated earlier, (e.g. Chap. 2, section I.1.4.) the size of our sample is relatively small compared to most corpus-based studies on L2 fluency, which makes it difficult to draw general conclusions, so our findings must be interpreted with caution. In addition, our statistical methods were rather "basic" as opposed to previous authors who used mixed-linear regression models (e.g. Graziano & Gullberg, 2018) or multiple correspondence analysis (e.g. Grosman et al., 2018) which make use of fixed and random effects (in the case of the linear mixed model) and further display the relationship between different variables (in the case of the multiple correspondence analysis). For future work, a more thorough statistical methodology should be applied to a larger sample of the data to perform a more robust quantitative analysis. However, it should be noted that the aim of the present study was to combine quantitative and qualitative analyses (cf Stivers, 2015), which has not been systematically done before in (dis)fluency research. This further motivated our choice to work on a small sample (cf Chap. 2, section I.1.4.). In addition, the present study focused more specifically on fluenceme sequences to explore their patterns of co-occurrence and accompanying visual-gestural behavior, but it did not analyze individual fluencemes in detail (except for the duration and form of certain markers). This drawback was already pointed out by Lopez-Ozieblo (2019) in a recent paper which examined different types of fluencemes (i.e. cuts-offs) and their co-occuring gestures. It will be helpful for future work to compare the production of the different fluencemes and their accompanying gestures in order to better capture the differences in L1 and L2 behavior.

Highlights of Chapter 3:

- L2 fluency is inherently a multimodal process anchored in language use, which should not be decontextualized from practical actions.
- L2 fluency does not necessarily reflect cognitive processes and language difficulties, but may also embody the notion of interactional competence, and thus work as a resource to maintain fluency in discourse.
- In the SITAF Corpus, where speakers interact in tandem settings, we have shown the way speakers mobilized a number of relevant resources in the course of the talk to deploy multimodal communication strategies and display their (dis)engagement to the ongoing talk.
- L1 and L2 fluency differ on a number of levels, mainly at the level of speech production, visual-gestural behavior, and interactional dynamics, but these differences have shown nuanced findings, given the range of variation found within speakers and the two speaker groups. This shows that L2 fluency is not only affected by differences in proficiency but also differences in language and individual speaker style.
- Fluencemes are dynamic and fluid markers of (dis)fluency, both pertaining to internal cognitive processes associated with planning and repair, and to interactional dynamics and interpersonal relations.
- Quantitative and qualitative analyses work in tandem to measure the degree of (dis)fluency found in native and non-native productions and show different but complementary aspects of our investigation.

Chapter 4. Inter-(dis)fluency across communication settings

Introduction to the chapter

The present chapter presents the findings obtained from the DisReg Corpus, covering "ordinary" versus "institutional" aspects of multimodal talk, comparing productions of French students in two different language styles and communication settings (i.e. class presentations versus face-to-face interactions). As we shall see, the notions of "setting" and "style" encompass a wide array of dimensions, ranging from the type of delivery (i.e. monologue versus dialogue), the degree of preparation (prepared versus spontaneous) to other sociolinguistic factors (i.e. register, the speaker's discourse identity, or the type of addressee). The main research question addressed in this chapter is whether all these inter-related factors do have an impact on the distribution of fluencemes and gestures, and, if it is the case, how it is manifested in both the vocal/verbal and visual-gestural channel.

This chapter is structured as follows: In section I, we explore these dimensions by reviewing a number of studies in which the different effects affecting (dis)fluency rates in speech have been investigated. In our own model, we ground these effects in a larger interactional framework. We also address a gap in the literature regarding gesture analysis with respect to language style, which has, to our knowledge, been underexplored. We conclude with our research questions and hypotheses, some of which stem from the ones formulated in Chapters 1 (section IV. 4.3) and 3 (section I.1.4) following our integrated approach to inter-(dis)fluency. In Sections II and III we present the quantitative and qualitative findings extracted from our annotations of the data, which, similarly to our study of SITAF, integrates different levels of analysis (speech, visuo-gestural, and interactional) and mixes statistical and conversation-analytical methods. These findings are then discussed in Section III with regard to participation frameworks, face-work, audience design, and common ground.

I. Literature Review

In the previous chapter, we presented a large existing body of research in Second Language Acquisition, gesture studies, and corpus linguistics, to better understand

differences between L1 and L2 fluency. Language proficiency (i.e. native versus non-native speech) was thus the primary binary variable that was used systematically in our analyses to compare other sets of variables (i.e. position, sequence complexity, gesture production, etc.) within our two groups of speakers. Similarly, in the DisReg Corpus, we seek to compare (dis)fluency and gesture rates in two different conditions, styles, or settings (i.e. students giving a formal class presentation versus exchanging during a casual conversation, see Chap.2, section I.1.3). The identification of these conditions is not straightforward, as it involves multiple factors, which have been recognized in radically different ways in the literature, accounting, once more, for different approaches. We briefly summarize and discuss these approaches in the first section (1.1) before reviewing a number of studies on (dis)fluency and gesture research (sections 1.2. and 1.3.). This review will serve as a basis for our research questions and hypotheses addressed in section 1.4.

1.1. Identification of the relevant factors

In this section, we present a number of factors that distinguish formal prepared class presentations from face-to-face spontaneous conversations, based on a short review of the literature in different disciplines (i.e. psycholinguistics, second language acquisition, gesture studies, and conversation analysis). These factors include type of delivery (1.1.1.), mode of speech (1.1.2), language style (1.1.3.) and social practice (1.1.4).

1.1.1. Type of delivery: spontaneous versus read speech

In early psycholinguistic work on disfluency, (cf Chap. 1, section I.) great emphasis was laid on the role of disfluencies in *spontaneous* speech. Note, for instance, the name of the recurrent DiSS workshop (cf *Introduction*) *Disfluency in Spontaneous Speech,* which focuses on the recognition of these markers in spontaneous speech. A clear distinction was thus often made between "spontaneous" and "read" "laboratory", or "planned" speech, with the former containing significant rates of disfluencies. Such evidence was supported by a number of researchers in the field (e.g. Hirschberg, 2000; O'Shaughnessy, 1992; Shriberg, 2001, 1999). Unlike spontaneous speech, which is characterized by multiple non-linear processes which constantly compel speakers to work on their online production (see Chap. 1, section I.), read speech is said to be

highly "constrained" by its existing written content, and thus requires minimal complex processing.

Witton-Davies (2014) elaborated on the degree of preparation time required by different speaking tasks, by identifying three types: (1) scripted— read aloud or memorized speech; (2) prepared – speech prepared beforehand, and (3) spontaneous – no preparation at all. As a lot of researchers have shown (e.g. Bailey & Feirrera, Lickley, 2015; Shriberg, 1994, among others, see section Chap. 1, section I. 1.2.) unplanned, unprepared, and spontaneous speech deliveries necessarily give rise to many disfluencies. In fact, Rodríguez et al., (2001, p.1) provided a definition of disfluency that is almost synonymous with spontaneous speech:

> We apply a wide definition of disfluency as any acoustic, lexical or syntactic feature that distinguishes spontaneous from read speech. In fact, we should better refer to them as spontaneous speech events.

This proximate relation between disfluency and spontaneous speech probably dates back to the earliest studies conducted on disfluency research by Goldman-Eisler (1958, 1968, 1972) in which she showed the importance of hesitation phenomena[134] in the encoding and decoding of messages in spontaneous speech. She found that disfluency rates varied according to the degree of linguistic processing required by a speech task. For instance, she compared the distribution of pauses in samples of spontaneous speech (radio talk and discussions between young academics) and in samples of reading (read by subjects who were not the initial speakers of these texts), and found that long pauses within clauses that were present in spontaneous speech were virtually absent from read speech.

This distinction can also be found in papers on speech recognition systems, where prosodic patterns may facilitate the identification of relevant aspects of the speech signal. For instance, Silverman et al., (1992) examined prosodic characteristics of over a hundred utterances, and compared spontaneous and read versions of it. Half of the utterances were taken from a corpus of spontaneous answers to request the name of a city, and the others were the same word strings read by subjects who tried to model authentic dialogue. In their analysis of silent pauses, they found differences between spontaneous and read segments, with 45% of pauses that were

[134] Note again the use of the term "hesitation", which was also used by Maclay & Osgood (1959), cf Chap. 1, section II.2.2.2.)

"ungrammatical" in spontaneous speech (i.e. not located at grammatical boundaries) whereas only 11% were ungrammatical in read speech. In a similar vein, Deese (1984, in O'Shaughnessy, 1992) found that planned speech contained fewer restarts (3.8 phw) than unplanned speech (5 phw). In a more recent study conducted on different speech samples (e.g. reading, conference, news broadcast etc.) for the purposes of developing prosodic detection tools, Goldman et al., (2010) also found that spontaneous speech samples contained a higher amount of hesitations than in read or prepared speech. These findings are consistent with the earlier work conducted by Goldman-Eisler.

In sum, this body of research is largely grounded in a psycholinguistic framework (cf Chap. 1, section I.), where specific attention is paid to the role of disfluencies in speech processing and the marking of syntactic structures. However, the distinction between "spontaneous" and "read" speech remains restricted to the domain of speech analysis, and is thus not suited to the present study. In addition, it is not only the type of delivery that distinguishes these two conditions, but to a larger extent the overall setting: in one case speakers are usually interacting face-to-face, while in another they are most often reading from a text in front of an experimenter (or alone). In the case of DisReg, this distinction hardly applies since, even though the students were indeed *reading* their notes, they were also giving an oral performance to their classmates and teacher, which as we shall see (section 1.1.3 and 1.1.4), has very different implications. This leads us to the following section, which introduces a closely related variable, the speech mode.

1.1.2. Mode of speech: dialogue versus monologue

As Witton-Davies (2014) pointed out, speech samples can be categorized in various ways, depending on the means of elicitation (naturalistic or experimental), the mode (monologue or dialogue), the interface (face-to-face, computer, or laboratory), and the type of task and condition. As we have seen in the previous section, the distinction between "read" and "spontaneous" largely depends on these different categorizations, i.e. read speech is often elicited in an experimental setting, while spontaneous speech tends to be produced in more naturalistic settings (also see Chapter 2, section I.1.1.). These two distinctions may thus be closely related, as dialogues tend to be less controlled and more spontaneous, as opposed to monologues which involve a greater deal of speaker's control on their production.

However, previous research on (dis)fluency in monologue and dialogue have led to contradictory results (read Witton-Davies, 2014, pp. 62-66 for an extensive review). One the one hand, a number of papers have shown that fluency correlates with dialogues, with a faster speech rate, shorter utterances, and a fewer number of pauses in dialogic situations (e.g. Kowal et al., 1983; Michel, 2011); while others have found evidence that speakers tend to be more fluent during monologic tasks than dialogic ones (e.g. Skehan, 2001). However, it should be noted that all these studies were conducted on non-native speakers and examined temporal characteristics of L2 fluency specifically (cf Chap. 3, section, I.) which adds another crosslinguistic variable. As Witton-Davies (2014) observed, other effects may have an impact on (dis)fluency rates such as task complexity, or topic, which is further described in section 1.2. In sum, mode of speech alone (i.e. dialogue versus monologue) is not enough to compare differences between speech samples, and calls for further discussion.

Clark (1996, pp. 9-10) defined dialogue on the basis of the following features: co-presence (the two participants are in the same physical environment), visibility (they can see each other), audibility (they can hear each other), instantaneity (they can see each other instantly), evanescence (their actions quickly fade away), recordlessness (their actions leave no record), simultaneity (they produce and receive speech simultaneously), extemporaneity (their actions are carried out spontaneously in real time), self-determination (they determine their own actions, vs scripted), and self-expression (they engage in actions as themselves, vs roles). The counterpart of dialogues, i.e. monologues, have been defined in a variety of ways, depending on the type of experimental study which was used (read Bavelas et al., 2008, for review), or on the importance given to the audience's passivity; for instance, Clark defines it as: "one person speaks with little or no opportunity for interruption or turns by members of the audience" (Clark, 1996; p. 4). Drawing on existing definitions and experimental studies conducted on gesture production in dialogue and monologue situations (see section 1.3) Bavelas et al. (2014) suggested a continuum of possible contexts that could characterize different degrees of interactivity, from extreme monologue to free dialogue, summarized in Figure 50 below. At the very low-end of the continuum, we can find examples of "interior monologues", which are very unlikely to be recorded during experiments, but which are most likely to be considered as "monologues", as the speaker is alone, and talking to themselves. This seems to be one salient characteristic of monologic talk. Near the middle of the continuum, we find situations

where speakers are talking to an audience of listeners, but they do not display a state of participation. This is another essential feature of monologic discourse.

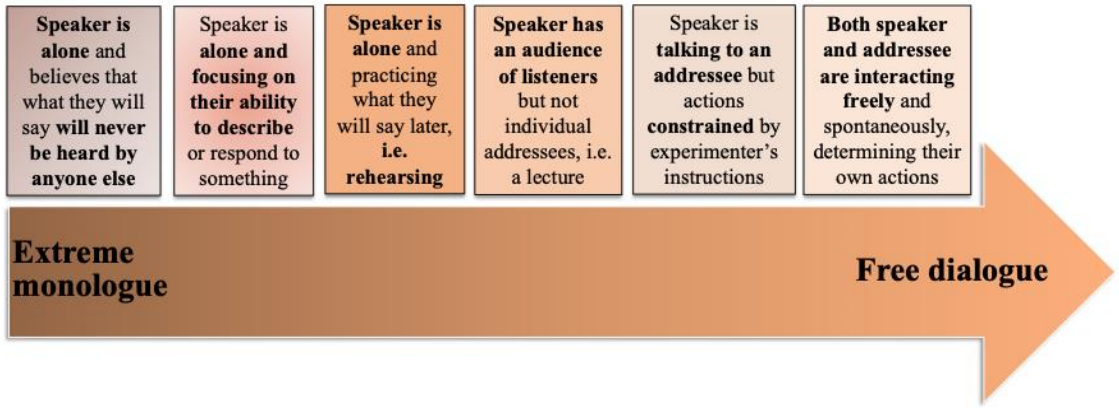


***Figure 50**. Continuum of contexts from monologue to dialogue (adapted from Bavelas et al., 2014, p. 624)*

Free dialogues on the other hand, involve active participation from the two parties, and unlike during experimental procedures where they are usually constrained by the experimenter's instructions, they are free to "act as themselves" and interact in real time in casual, personal, and naturalistic settings (cf Chap. 2, section I.1.1.). In sum, it is essential to consider the overall context and setting to determine the different aspects of monologic and dialogic situations, which will be further discussed in section 1.1.4. We shall now move to another aspect of language variation that takes other sociolinguistic factors into account.

1.1.3. Variation in language style and register

In the field of sociolinguistics, one essential component of language variation research is the study of style, developed by a number of linguists and social psychologists, including the pioneering sociolinguist William Labov, who is considered to be the founder of variationist research. While this section is by no means intended as an exhaustive review of sociolinguistic research, it will draw on a selection of features of language style found in this field to better capture differences between dialogic (spontaneous) and monologic (prepared) speech. There have been a number of different approaches to the study of style since the 1970s (read Bell, 2006 for review), starting with Labov's (1966) traditional *Attention to Speech Model*, followed by others, such as *Speech Accommodation Theory* (Giles & Powesland, 1975), *Audience Design* (Bell, 1984), and *Stylization* (Coupland, 2001), among others. According to Bell (2006), the most common approach to style in sociolinguistics is *audience design*,

which states that speakers' style choices are primarily a response to their audience, and are thus determined by their type of addressee. Bell (2006, p. 993) further defines style[135] as the following:

> Style is what an individual speaker does with a language in relation to other people. Style is essentially interactive and social, marking interpersonal and intergroup relations.

Speakers thus constantly design and accommodate their style to adjust it to their audience; this practice also closely echoes the notion of *recipient design* in Conversation Analysis (Goodwin & Heritage, 1990; Jefferson, 1974), whereby the design of a speaker's specific action is oriented towards the status of their recipient. In this view, speakers and recipients take on a variety of *discourse identities* (Goodwin & Heritage, 1990) which "intersect with a range of social arrangements involving entitlement to knowledge" (Goodwin & Heritage, 1990, p. 293)[136]. In the case of DisReg, these features play an essential role, since it is not solely a matter of mode of communication (i.e. whether the speaker is engaged in a dialogic versus a monologic task) but of audience (i.e. whether the speaker is interacting with a friend or performing in front of a teacher and classmates). We will return to this distinction when we formulate our research questions and hypotheses in section 1.4.

Eskénazi (1993) identified different dimensions, or axes, to language style, which further take into account the overall environment, or setting, surrounding the speaker. He distinguished between three different dimensions, listed below:

- Intelligibility: the degree of clarity that the speaker intends their message to have, the effort to be clear.
- Familiarity: speaker's familiarity with the listener, or audience.
- Social strata: the degree of register which determines the speaker's tone, from colloquial to a more cultivated one (Labov, 1966).

[135] Note that here the notion of *language or speaking style* is very different from the idea of *individual communicative style* (cf Chap. 2) that we explored in SITAF when we described speakers' individual communication strategies, leading to individual variation. In this case, style does not refer to the idiosyncrasy of an individual speaker, but rather a choice of words or register in relation to a specific social situation, i.e. formal style or register. The effect of individual differences will also be explored in this chapter.

[136] This was also explored in the SITAF Corpus where we presented the different native and non-native identities adopted by the participants in the course of the interaction, see Chap. 2, section III.

These axes are further illustrated in the figure below, which includes the position of different task types, speech deliveries, and modes of communication introduced earlier, in order to get a more comprehensive picture of the types of differences we can find across language styles and settings.

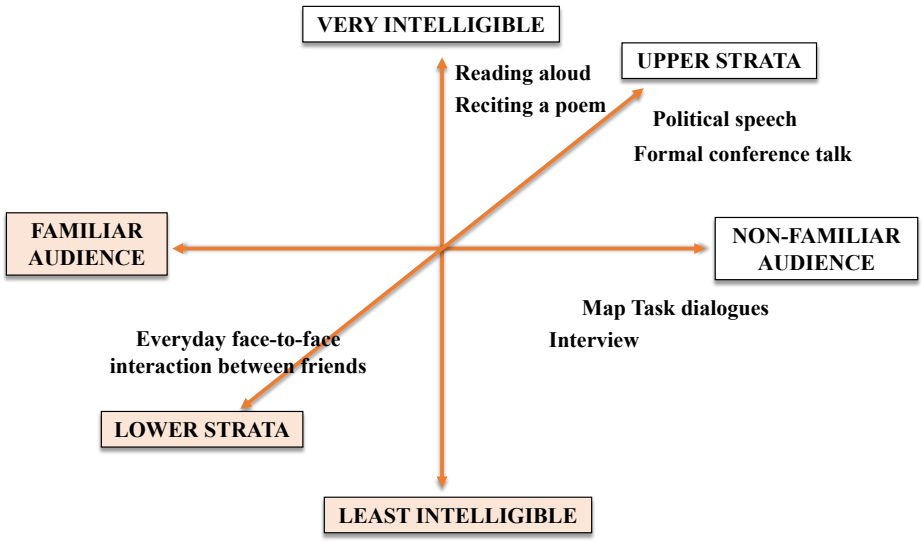


Figure 51. *Dimensions of style, adapted from Eskénazi (1993, p. 503)*

It would thus appear that the notion of "style" is more suited to the present study of the DisReg Corpus, as it is not only restricted to differences in mode (dialogue versus monologue), or delivery (read/prepared versus spontaneous), but includes other dimensions as well. As we have seen earlier (see Chap. 1, section II.2.2.1.) a lot of emphasis has been placed on the role of certain fluencemes, (i.e. filled pauses in American and British English) as *pragmatic* or *sociolinguistic* markers, with the corpus-based evidence that their use is strongly determined by sociolinguistic factors, such as gender, age, setting or register (Bortfeld et al., 2001; Tottie, 2011, 2014). Differences across language styles are thus highly expected, which, once again, vouches for the flexible and dynamic nature of fluencemes, altogether shaped by a number of situational and contextual features (see Chap. 1, section 3.1. and 3.2.). We will return to this point in section 1.2.

1.1.4. Kind of social practice: ordinary and institutional talk

Still in tune with the notions of environment, context, and setting, we shall now turn to the conversation-analytic framework of talk-in-interaction to discuss its implications for the present study of the DisReg Corpus. In CA, a distinction is often drawn between "ordinary" conversation and "institutional" talk. Although this distinction is not clear-cut, and does not apply to all instances of interactional events

found in the two practices (Drew & Heritage, 1992), several researchers have attempted to pin down distinctive features of institutional talk, i.e. news interviews, courtroom proceedings, medical consultations, classroom instructions, which are very different from everyday conversations. These features can be summarized as the following (Heritage, 1997):

- Turn-taking organization: institutional talk is characterized by specific turn-taking procedures which may involve the restriction of one party to speak, in contexts where a large audience is present and it is thus necessary to restrict their initiative to interact (i.e. speech deployed by a defense attorney in a courtroom, or by a lecturer at a conference talk).
- Overall structural organization: some types of institutional talk may be governed by a specific structural organization in which specific actions emerge in a particular order, i.e. a doctor's appointment usually follows a particular sequential organization: the patient presents the reason for the visit, the doctor examines the patient, then evaluates their diagnosis, and offers details of treatment.
- Lexical choice: The choice of specific words or phrases can reflect a speaker's stance in a particular setting, e.g. during a lecture the speaker may use the pronoun "we" instead of "I" to involve the audience (Goffman, 1981).

In addition, the participant's orientation to the talk in institutional settings presents a number of specificities, as it involves particular constraints on the speakers. They are often oriented to a core goal, task, or identity; for instance, during an emergency call, the core objective of the caller is to describe the type of emergency and its location. In the case of DisReg, the main goal of the student is to successfully present his or her assignment to the classroom. This specific orientation has at least two major consequences on the structure of the talk: first, it is not appropriate for the student to discuss anything else that is not related to the presentation (e.g. a personal anecdote, or a recollection, which are typical of conversations); second, the audience's co-participation is not expected during the presentation. This last constraint has a profound impact on the organization of repair (see Chap. 1, section III. 3.2.2.). Students, acting as presenters, when confronted with an episode of trouble, can no longer rely on their partner's participation to a repair sequence, and must therefore find other relevant ways to deal with difficulties alone in the course of their talk, i.e. by looking through their notes for instance (see Chap. 2, section I. 1.3.2.). This may

compel them to mark significant delays in the middle of their production, which are not deemed so appropriate in the context of delivering a formal presentation.

Goffman (1981) dedicated an entire volume to the analysis of different "forms of talk", from everyday ordinary conversations to institutional forms of talk. In one section, he focused on one specific institutional setting, the lecture, which is of interest to our study as it presents a number of similarities with a class presentation. Goffman defines a lecture in the following terms (Goffman, 1981, p. 165):

> A lecture is an institutionalized extended holding of the floor in which one speaker imparts his views on a subject, these thoughts comprising what can be called his "text". The style is typically serious and slightly impersonal (…) a platform arrangement is often involved, underlining the fact that listeners are an "immediate audience". I mean a gathered set of individuals, typically seated, whose numbers can vary greatly without requiring the speaker (typically standing) to change his style, who have the right to hold the whole of the speaker's body in the focus of staring-at-attention.

A notable reference to bodily behavior is noted here, which, so far, has not been explicitly touched upon in this section, (except briefly in section 1.1.2 when we mentioned Bavelas et al's 2008 work). The reason for this is that, to our knowledge, the study of style has much too often been restricted to modes of speaking, i.e. choice of tone, or words, but not of gestures. We further address this gap in section 1.3. Participants' uses of bodies, as emphasized throughout this thesis, are fundamental to the deployment of language activities in the course of multimodal talk, and may exhibit differences across situations.

To conclude, in addition to language style, it is crucial to consider the overall contextual structural organization of the exchange, the participant's orientation to the talk, their gestural and bodily behavior in the surrounding spatial and material environment, as well as the type of social practice the speakers may engage in. Indisputably, it is also essential to consider the type of audience the participant is interacting with in a specific language practice, as it also determines the choice of the speaker's style, or participation status. In this section, we have sought to pinpoint different characteristics of discourse to help us identify the main differences between our two data samples in the DisReg Corpus (class presentations versus conversations), in order to address several assumptions and hypotheses regarding the distribution of

fluencemes and gestures in the two situations (section 1.4). These different factors, mainly the type of delivery, degree of interactivity, mode, audience, style, etc., which emerge from different theoretical frameworks (i.e. psycholinguistics, sociolinguistics and conversation analysis) should not be identified separately (as it has often been the case in the literature), but should interact with one another, by combining aspects of verbal speech processes with multimodal and social resources of talk-in-interaction, following our integrated approach to inter-(dis)fluency. In the next section, we further identify the different effects that may have an influence on the production of fluencemes by reviewing a number of studies in (dis)fluency research.

1.2. Effect of task type, discourse domain, and style on (dis)fluency: evidence from experimental and corpus-based studies

In order to understand the different processes associated with (dis)fluency, and identify the different causes underlying their production, several researchers have conducted experimental and corpus-based studies which manipulated task type, discourse domain, utterance complexity, or topic. In addition to style and setting, other factors, such as discourse domain or familiarity may thus also come into play. For instance, Schatcher et al., (1991) conducted a study on the production of disfluencies in different academic disciplines, based on lectures in the humanities and natural sciences. They argued that undergraduate students were less required to choose among options to structural, "formal" and "factual" lectures like pure sciences, as opposed to humanities lectures, and thus predicted that lecturers in the humanities and social sciences would produce more disfluencies[137] (and more specifically filled pauses) than those in natural sciences. Their results showed indeed a higher number of filled pauses during lectures in humanities (6.46 per minute) than during natural sciences (1.39 per minute), which supported their prediction. They further designed an interview to all the lecturers during which they were asked to talk about graduate training procedures and practices, and also found significant differences in the rate of disfluencies between lectures and interviews, with a higher rate of disfluencies during interviews (5.28) than during lectures (4.85).

[137] Note that the terms "disfluency" and "hesitation" are mostly used in this section to reflect the authors' terminology, even though we do not adopt these terms.

Similarly, Bortfeld et al. (2001) found that speakers produced a higher rate of disfluencies when discussing unfamiliar topics than familiar ones. Based on a large corpus of English conversations, they investigated the different processes associated with the different types of disfluency. They found that disfluency rates increased when speakers were faced with "heavier cognitive demands" (Bortfeld et al. 2001, p. 135), in other words, when the topic was unfamiliar (e.g. discussing tangrams versus discussing children,), or when speakers produced longer turns (in line with Oviatt, 1995, Shriberg 1996, and Beattie 1979, who found that high disfluency rates were associated with longer utterances). This last effect could be explained by the difficulty of planning longer utterances. However, it should be noted that while this was true for most disfluency types, there was quite a different pattern for filled pauses which were not correlated with utterance length, and were found to serve interpersonal functions. This evidence was also supported by Tottie (2011, 2014, 2015, 2016, 2019) who, as we have seen earlier (cf Chap. 1, sections II. 2.2.1. and III. 3.1.2.), argued that filled pauses could be used intentionally for a stylistic purpose, and thus belonged to the same category of "planners" along with discourse markers such as "and" or "but". In a similar vein, Crible (2018) found that filled pauses frequently co-occurred with discourse markers, especially in contexts of low interactivity where speakers produced long stretches of talk, such as lectures or political speeches. Tottie (2016) further showed instances of wide variation within filled pause frequencies, ranging from 1.4 to 22.8 per 1,000 words in the SBC Corpus. Filled pauses were found to be more frequent during task-related interactions taking place in non-private settings than during conversations between family and friends. She hypothesized that long narrative turns or a thoughtful presentation of evidence required more planning than a conversation among relatives and friends, which may indicate that filled pauses were linked to the demands of planning what to say.

More in line with the notion of style and register (see section 1.1.3.), Duez (1982) conducted a study on the distribution of filled and unfilled pauses in the speech of French politicians across three different contexts (political speech, political interviews, and casual interviews) in order to investigate their possible stylistic function. She found that pauses were much more frequent in interviews than in political speeches, but were strikingly longer in the latter. She suggested that the high rate of pauses found in interviews may relate to the fact that the politicians were focusing on planning and production issues when producing spontaneous speech, while the long duration of

pauses in political (prepared) speeches reflected a stylistic function, namely to emphasize what was being said. This is in contradiction with Tottie (2016) who claimed that long narrative turns required more planning than conversations, hence more pauses. It would appear that, in Duez's case, the effect of register and formality (which she called "political power", Duez, 1997) was much stronger than the effect of speech mode and turn-taking since French politicians (unlike common speakers) are trained oral performers who are expected to deliver convincing speeches with very few hesitation marks. Conversely, Moniz (2019) and Moniz et al., (2014) found that lectures contained more disfluencies than conversations. They conducted a corpus-based study based on two speech samples, one which includes recordings of university courses in the presence of the lecturer and the students, and another one which contains map-task dialogues between two participants. While they found a considerable range of speaker variation within both dialogues and academic presentations, their results showed significant differences in the distribution of disfluency types overall: dialogues showed twice as much fragments as lectures but fewer additions, and filled pauses were the most frequent ones in both corpora. They also found more instances of silent pauses in lectures than in dialogues, and more complex sequences made of repetitions and substitutions used for lexical searches. These findings are thus not consistent with Schachter et al., (1991) who found a higher rate of disfluencies during interviews than during lectures. This lack of consistency in the literature can be explained by the different types of speech samples used: while Schatcher et al., (1991) worked on interviews, Tottie's (2016) work was based on casual conversations between friends and family members, and Moniz et al., (2014) conducted their study on map-task dialogues. While these three speaking tasks are all "dialogic" ones (as opposed to lectures and narratives), they exhibit substantial differences in style and setting, which makes it difficult to compare their findings.

Taken together, these corpus studies have shown the impact of degree of preparation and formality on the distribution of fluencemes. Other factors, such as anxiety, may also come into play. For instance, Christenfeld and Creager (1996) investigated the relationship between filled pauses and anxiety in a production experiment with undergraduate students. They found significant differences between the low anxiety and high anxiety conditions, with an average of 7 filled pauses per minute in the latter and 4 in the former. They concluded that the use of filled pauses was *not* necessarily a by- product of anxiety, but a sign that students were more self-

conscious of their speech (cf. Broen & Siegel, 1972). The role of such self-monitoring can also explain Tottie's (2014) corpus findings that showed a higher frequency of filled pauses in task-oriented contexts (deliberation, presentation of evidence), where there can be more instances of professional pressure and/or important outcomes at stake than in casual conversation, where speakers might not be very self-conscious. This variable could thus also apply to the DisReg Corpus, where students probably experienced a certain degree of stress (although this is not directly observable) while delivering their presentations.

In sum, these experimental and corpus-based studies on (dis)fluency have shown that not all fluencemes are "equal", i.e. filled pauses have a more lexicalized status and tend to serve more interpersonal functions than other markers. In addition, different variables, such as topic of conversation, anxiety, degree of formality, audience design, mode of speech, etc. all seem to have an effect on the distribution and frequency of fluencemes. However, the relationship between these variables is not a simple one, as it involves multiple factors and processes (cognitive, interpersonal, social, etc.), which have led to contradictory results in the literature. As Bortfeld et al., (2001) further noted:

> Disfluencies may arise from quite different processes or within quite different situations. As we proposed earlier, perhaps some disfluencies serve an interpersonal coordination function, such as displaying a speaker's intentional or metacognitive state to a partner, while others simply represent casualties of an overworked production system.

In line with Bortfeld et al. (2001), and further situated in our functionally and interactionally ambivalent approach to inter-(dis)fluency (cf Chap. 1, section IV), the present study is motivated by the assumption that cognitive, social, and situational factors may all interact to affect speech and gesture production. As Bortfeld et al., (2001) further criticized, the issue with a majority of the past studies conducted on (dis)fluency is that fluenceme rates have often been compared in different corpora which were collected under very different conditions with different tasks and different samples of speakers. This is not the case of DisReg, which includes samples of the same speakers engaged in two distinct speaking tasks. In addition, we pay specific attention to the overall setting and environment in which the participants perform these tasks, mainly casual versus institutional settings (see section 1.1.4), which, we believe, have

a strong influence on (dis)fluency, but also gesture production. This leads us to the next section.

1.3. Effect of style and setting on gestures: a gap in the literature

While a considerable amount of research has been conducted on the different effects influencing (dis)fluency rates, very few studies have targeted these aspects specifically in gesture research. There have not been, to our knowledge, many studies that investigated the effect of setting or style on the production of gestures, except perhaps for Bavelas et al., (2008, 2014) who studied the independent effects of dialogue on gestures, drawing on a number of papers in experimental research. These papers (read Bavelas et al., 2008, p. 497 for review), described different visibility experiments in which speakers were asked to perform a task in two different conditions, (i) visible, i.e., face-to-face and (ii) not visible, telephone conversation, or listening to a tape-recorder. The aim of these studies was to measure the effect of visibility on gesture rates. However, the task was the same one in both conditions (i.e. giving directions to a location, giving an opinion, retelling a cartoon, etc.), and the only difference was whether the addressee was visible or not. As Bavelas et al., (2008) noted, these experimental studies have often overlooked the possibility that dialogue itself could have an effect on gesture. They further stated (Bavelas et al., 2008, p. 499):

> We propose that visibility is one aspect of the speaker's communicative context and that speakers adapt their communicative choices to the parameters of their particular communicative context. Even holding constant what they are going to convey, there may be different situational resources or constraints that determine how they can do so. Some of these social parameters are social, for example, whether there is an addressee, as noted above, or whether the addressee shares common ground with the speaker.

In order to measure the effect of dialogue on gesture independent from visibility, they designed an experiment which consisted in describing the picture of an 18th century dress, in three different conditions: (1) two participants talking face-to-face (dialogue/visibility), (2) two participants on the phone (dialogue/ no visibility), (3) one participant talking to a tape recorder (monologue/ no visibility). Their findings showed that dialogue had a significant effect on the speakers' rate of gesturing, which

was independent of the effect of visibility, as it was consistently higher in the telephone condition than the tape recorder. A majority of these gestures were referential[138], and they were much more frequent in the dialogue condition overall (13.9 phw in the visible dialogue condition, 10.3 in the dialogue not-visible condition, and 3.77 in the not-visible monologue condition). Interactive gestures, although less frequent than referential ones, were also found to be more frequent in the dialogue condition (0.77 phw in the visible dialogue condition, 0.26 phw in the not-visible dialogue condition, and 0.36 phw in the not-visible monologue condition). This confirmed their prediction that visibility was not the only variable affecting gesture rates, and that gestures were highly sensitive to situational and social factors, regardless of whether they were seen or not. However, it should be noted that the gestures analyzed in their study were elicited in a controlled setting, i.e. as part of an experiment, so it does not truly reflect the situated ecology of gestures and the context in which they occur (Goodwin, 2000; Mondada, 2018; see Chap. 2, section I.1.2.). In line with previous experimental studies, it was again the same task which was performed in the three conditions, so their study only targeted differences in speech mode (i.e. monologue vs dialogue) and visibility, but it did not account for differences in style or setting. Conversely, the present study conducted on the DisReg corpus does not only compare aspects of monologic and dialogic situations, but takes into account a wide array of inter-related factors introduced in this chapter, i.e. language style, register, audience design, social setting etc. We will return to this point in section 1.4.

In addition, we can find several papers in the gesture literature that focus on the analysis of gesturing behavior in specific contexts, such as during lectures (Sweetser & Sizemore, 2008) in political communication (Streeck, 2008b), or during auctions of fine arts (Heath & Luff, 2011), among others; but the gestures analyzed in these institutional settings have not been systematically compared to other conversational ones. For instance, Sweetser & Sizemore (2008) noted that lecturers would most likely use gestures differently from conversational participants as they would often take up a larger personal gesture space to keep the attention of an audience who is sitting further away. However, they also found that some kinds of gestures, such as deictic or interactive ones, could show very similar patterns in the two situations. In their paper on gestural space, they analyzed a few instances of a

[138] "Topic gesture" in their terms, see Chap. 1, section III. 3.3.3.

lecturer's gestures during a lecture at a Colloquium, and described the way he directly addressed the audience by pointing towards them. Even though the lecturer was somewhat engaged in a "monologue", he was in fact "addressing their silent partners in the exchange" (Sweetser & Sizemore, 2008, p 48) which accounts for the dialogic dimension of the lecture. Once more, this example shows that the line between "monologue" and "dialogue" is not so clear, and should thus include other parameters. Sweetser (1998) also documented the use of two specific hand gestures (the B "barrier hand" and the Palm-Up gesture) used by lecturers, but did not offer a systematic comparison of these gestures produced in different communication settings. This limitation was pointed out in Sweetser & Sizemore's paper (2008) in which they strongly suggested to conduct more comparative analyses. The present study on the DisReg Corpus thus aims to bridge this gap by offering a quantitative and qualitative analysis of gestures in two communication settings.

Lastly, we can also find a number of studies specialized in teachers' gestures in classroom environments (to name but a few: Alibali & Nathan, 2007; Azaoui, 2015; Holt et al., 2015; Moro et al., 2020; Tellier, 2008) in which they stressed the importance of gestures to increase learning and understanding. We can also find papers on students' gestures recorded in the classroom when trying to understand algebraic concepts (e.g. Dwijayanti et al., 2019) or negotiating schemas when learning mathematics (e.g. Abrahamson & Bakker, 2016). These papers, however, which are more pedagogical-oriented, do not systematically compare the types of gestures produced by students in the classroom with the ones produced in a conversation. We believe that the field of gesture research could benefit from the present corpus-based study in order to understand more precisely how speakers differ across situations, and how they adapt their gestures to the audience, by taking into account the social environment in which they interact, as well as other parameters such as style and register. These differences can both be measured quantitatively and qualitatively by drawing on statistical evidence, coupled with micro-analyses of the data.

1.4. Research Questions and Hypotheses

In the previous sections, we painted a complex picture of the various features characterizing different speech situations in order to address our research questions and hypotheses regarding the distribution of fluencemes and gestures across our two data samples of the DisReg Corpus. As we have seen, speech situations cannot solely

be identified on the basis of spontaneity or interactivity alone, but must include a range of parameters that all interact with one another, such as style and social setting, drawing on different theoretical frameworks (i.e. psycholinguistics, sociolinguistics, corpus-based linguistics, conversation analysis, and gesture studies). For lack of a better word, we will primarily talk about *language style* and *communication setting* in this chapter to describe the sets of variants that characterize conversations and institutional talk, i.e. degree of register, type of audience (audience design), and setting (spatial and material). In this view, style reflects the impact of the *multimodal* environment or setting on one person or a group of individuals. In line with Eskénazi (1993), we offer a three-dimensional representation of style in the DisReg Corpus, which also includes the role of setting, as well as other variables described in this chapter, illustrated in the figure below[139]:

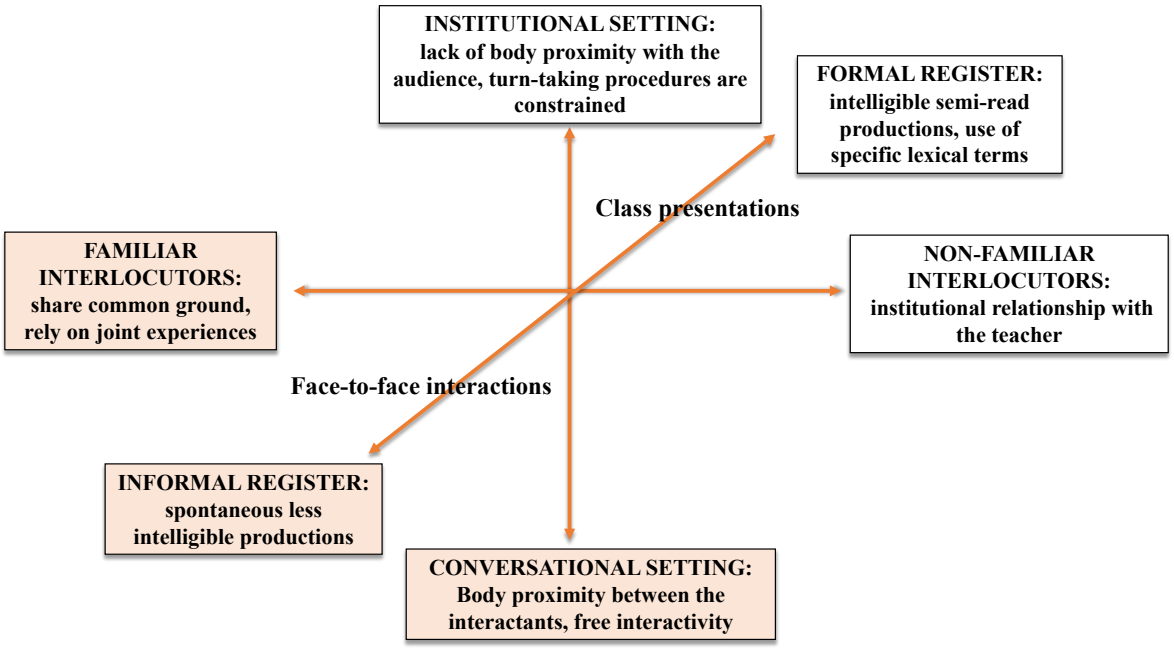


***Figure 52**. Three-dimensional representation of style in the DisReg Corpus*

Class presentations take place in an institutional setting, characterized by a number of institutional constraints (no interaction with the audience, the student needs to present a specific assignment prepared at home) as well as spatial ones (the student is

[139] This also echoes the work of Iriskhanova & Cienki (2018) who offered a semiotic continuum of gestures based on a multiplicity of semiotic parameters (e.g. conventionality, semanticity, social and cultural import etc.,) to characterize the high variability of gestures in context. Similarly, our research takes into account the multiple variables at different levels (register, audience, and setting) affecting fluenceme and gesture use in different situations.

alone[140], sitting or standing in front of a desk, facing a group of students and the teacher). This audience is non-familiar, and their relationship is primarily institutional, as it includes the student's teacher who will later examine their work and give them a grade. In addition, the style of the student must be formal, intelligible, and to do so they may rely extensively on their notes to deliver the presentation. On the contrary, face-to-face interactions are more informal and do not require the participants to talk for a limited amount of time on a restricted topic (even though they were given a piece of paper with a choice of topics to start the conversation, they were free to talk about anything they wanted). In addition, the two participants know each other, and thus share common ground, which facilitates active displays of intersubjectivity.

All of the features identified above thus go beyond differences in mode or delivery, and are not independent but constantly interact with one another. For instance, during a class presentation, the fact that the audience never interacts with the presenter is bound to the type of institutional setting, where they are not expected to interact. Our primary assumption is thus that all of these features have an effect on (dis)fluency and gesture production; even though it is not possible in this case to identify them separately, there may be some factors that have a stronger effect than others, as suggested by the contradictory results found in the literature. As we have seen, some studies have shown higher rates of fluencemes in dialogic situations than monologic ones (i.e. Duez, 1982, Schatcher et al., 1991) while others have shown the opposite tendency (i.e. Moniz, 2014, Skehan, 2001). As suggested earlier, this lack of consistency in the literature may be explained by the different types of procedures and designs used in these studies, which led to different findings. Some of our research questions and hypotheses, however, stem from this body of research, and follows assumptions from Cognitive Grammar and Interactional Linguistics presented in Chapter 1. We do not wish to treat these variables separately, but we consider it essential to properly identify them in order to better grasp the differences in (dis)fluency and gestural behavior across the two communication settings. Our research questions are addressed below, followed by our hypotheses:

[140] It should be noted that in one case (Pair D) the students presented their assignment together, but they took their turn respectively and did not interact with one another.

RQ1: Do style and setting play a role on (dis)fluency and gesture production?

- **H1**: In line with previous work on (dis)fluency (cf section 1.2.) and further grounded in our functionally and interactionally ambivalent approach to inter-(dis)fluency, we claim that fluencemes are dynamic systems which are highly sensitive to situational factors, and are thus inevitably affected by style and setting. While they may also be affected by other variables (i.e. topic difficulty or anxiety), these factors cannot be directly observed in the corpus since we did not manipulate (dis)fluency with task difficulty (as opposed to previous experimental studies, e.g. Hartsuiker & Notebaert, 2010, see Chap. 1, section II. 2.2.1.), so our main focus will be on the effect of style and setting, as well as individual differences across speakers.
- **H2**: Following theories of audience and recipient design, and further grounded in an interactional and multimodal approach to social action, we expect speakers to constantly adjust their body and talk for their audience, and rely on a multiplicity of semiotic resources and diverse media to build meaning (stream of speech, body, material objects, etc.,) which will inarguably have an impact on their visual-gestural behavior.

RQ2: If (dis)fluency and gestures are influenced by these factors, how are these differences characterized?

- **H3**: Given the lack of interactivity and the temporal constraints imposed on oral presentations, more fluencemes are expected in institutional settings than in conversational ones. We do not believe that the distinction between "read" versus "spontaneous" speech will affect (dis)fluencies positively, i.e. read speech associated with fewer fluencemes (e.g. Goldman et al., 2010). Even though speakers were indeed reading their notes and their speech was prepared beforehand, they were not asked to simply "read aloud", but to give an actual performance, which still requires the student to be "spontaneous" at times, and deal with the planning of their discourse (Tottie, 2016).
- **H4**: We further hypothesize that the fluencemes produced during class presentations will be closely associated with planning and repair processes, with more instances of pauses (of longer duration). Since the presenters pay

great attention to their own production, we also expect them to produce more morphological repairs (i.e., if they misread a word, they may wish to correct it).

- **H5**: We also predict that the mean length of utterance (MLU) would be longer in class presentations than in conversations, and we expect it to correlate with higher fluenceme rates (in line with Shriberg, 1994 and Oviatt, 1995).
- **H6**: Since gestures are one salient characteristic of social interaction, we expect fewer gestures in class presentations than conversations overall (following Bavelas et al., 2008), as well as differences in distribution, with more interactive gestures in conversation and more discursive gestures in class.
- **H7**: Fewer instances of mutual gaze are also expected in class presentations. We further predict that the students will predominantly gaze towards their notes written on a piece of paper, or laptop, which will be used as resources to maintain the continuity of their presentation (cf H9).

RQ3: Can we identify specific multimodal social practices that reflect these potential differences in style and setting? If yes, how would it influence the use of fluencemes?

- **H8:** In conversations, interactants rely on joint productions. Fluencemes will therefore occur more frequently in contexts of turn-taking where they embody visible displays of active participation in the talk, and occur in specific interpersonal contexts where speakers rely on shared experience, such as storytelling. Several of these fluencemes, we believe, will reflect instances of Interactive Communication Management (ICM), hence reflecting the more communicative, fluent, side of inter-(dis)fluency.
- **H9:** In class presentations on the other hand, fluencemes will be used as a resource to deal with trouble in the talk, as students must find ways to deal with the temporality of their presentation within their material and spatial environment (their notes, their book, and the audience), as well as how to switch from different modes (reading and talking). Therefore, most of these fluencemes will reflect instances of Own Communication Management (OCM) reflecting the more production-oriented, DISfluent, side of inter-(dis)fluency.

Our hypotheses are further summarized in the table below, describing different features distinguishing class presentations and face-to-face conversations with regard

to (dis)fluency and gesture, grouped together according to their potential underlying factors. In sum, the aim of this study is to consider many interrelated aspects of discourse and their potential effect on (dis)fluency and gesture behavior. In line with previous work on (dis)fluency, we seek to uncover the different effects affecting their use, but by incorporating co-occurring gestural behavior, which has received little attention in the literature. Unlike previous studies which focused on specific factors without necessarily considering others (e.g. comparing "dialogue" versus "monologue" without taking into account audience design, or comparing "prepared" versus "spontaneous" without considering the type of setting, etc.,), this study will reveal the complexity of human communication, which cannot easily be broken down into fixed categories.

Table 29. *Summary of the hypotheses presented in this chapter*

Features characterizing the two situations	Potential underlying factors
More fluencemes in presentations than conversations (H3) Fewer instances of mutual gaze in class presentations, but high instances of gazing towards the notes (H7) In presentations, fluencemes will be used as a resource to deal with trouble in the talk, pertaining to own communication management (H9)	Lack of interactivity and institutional constraints during presentations (temporal and turn-taking constraints) with a non-familiar audience
Fluencemes produced in class presentations associated with planning and repair with instances of longer pauses (H4) Longer MLU in class presentations correlated with higher fluenceme rates (H5)	Formal register, intelligible semi-read productions in class presentations
More interactive gestures in conversations than presentations (H6) In conversations, fluencemes will rather be associated with interactive communication management in interpersonal contexts (H8)	High interactivity in social conversations between familiar interlocutors

II. Results

2.1. Quantitative findings

This section presents the quantitative findings extracted from our annotations of the DisReg Corpus. It is structured the same way as the Results section in Chapter 3, and applies the same statistical methods to the data (cf Chap. 3, section II). Our analyses

compare fluenceme rates, gesture distribution, and gaze behavior across two language styles and communication settings (formal class presentations in front of a classroom versus face-to-face dyadic casual conversations between friends). Just like Chapter 3, values are provided in raw/absolute and relative frequency (e.g. proportion or ratio), and our basis of normalization for the rate of fluencemes and gestures is per hundred words (henceforth phw).

2.1.1. Marker level: rate, form, and duration of individual fluencemes

A total of 2870 fluencemes were annotated in the data, with 1472 tokens during class presentations, versus 1398 during face-to-face conversations. Figure 53[141] reports the rate of fluencemes per hundred words and per speaker in the two settings, and a test of log-likelihood indicated that these differences were significant statistically: the French students produced significantly more fluencemes overall during class presentations (28.4 phw) than during conversations (20.4 phw) (LL = 325.38 , p < 0.0001).

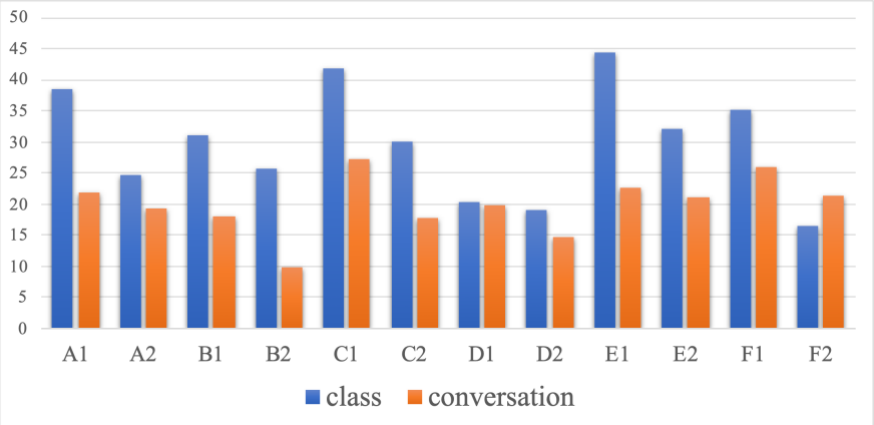


Figure 53. *Rate of individual fluencemes per hundred words*

Some individual differences can be noted. For instance, F2 is the only speaker who actually produced fewer fluencemes during the class presentation (16.5) than the conversation (21.4), and D1 produced about the same amount in the two situations (20.2 and 19.9). Others, like B2, produced considerably more fluencemes during the presentation (25.6) than when she was engaged in the interaction (9.7). In fact, she produced relatively few fluencemes during the conversation in comparison to the average in her group (20.4). These individual differences will be further explored in the qualitative analyses in Section 2.2. Table 30 compares the distribution of

[141] Raw frequencies are reported in Appendix 4, Table 60.

fluencemes by marker type: vocal markers (VOC), morphosyntactic markers (MS), and peripheral markers which include explicit editing terms (EDT) and non-lexical sounds (NL). Results show that in the two situations, vocal markers were the most frequent overall (about 60%) without presenting significant differences. Differences are found, however, in the proportion of morpho-syntactic markers and non-lexical sounds. The latter were more frequent during presentations (20%) than during conversations (10%), while morpho-syntactic markers were more frequent during conversations (27%) than presentations (20%).

Table 30. *Proportion of marker types in class presentations and conversations*

Marker type	class (raw)		conversation (raw)		Z score	p value
EDT	1%	(8)	1%	(11)	-0.804	0.4
MS	20%	(292)	27%	(373)	-4.344	<0.0002*[142]
NL	19%	(278)	10%	(134)	7.103	<0.0002*
VOC	61%	(894)	63%	(880)	-1.22	0.2

A more detailed description of the proportion of marker types is provided in Table 31 which identifies all the fluencemes that were annotated in the data (excluding additions which were extremely rare overall, *N*= 5). Results further show that the proportion of identical repetitions[143], self-interruptions, and prolongations was significantly higher in conversations than class presentations, but the proportion of filled pauses and non-lexical sounds was higher during class presentations. Additionally, morphological repairs were slightly more frequent in class (0.9 per hundred words) than in conversation (0.4 per hundred words, *LL* = 9.39 , *p* = 0.01, cf Table 68 in Appendix 4), as we expected (cf H4) and unfilled pauses were found to have a higher rate per hundred words in class (6.7) than in conversation (4.7; *LL* = 21.15 , *p* = 0. 001) despite no differences in proportion. Overall, these differences in distribution confirm most of our hypotheses, and may reveal characteristics of

[142] Just like Chapter 3, an asterisk is added whenever the *p* value is below 0.05 as to highlight its significance.

[143] Note that the proportion of identical repetitions (based on the total number of fluencemes) is higher in conversations than in presentations, but they showed no differences in overall rate per hundred words (about 2.4 in both situations, cf Table 68 in Appendix 4). This shows that both situations may require about the same amount of fluencemes globally, but more fine-grained differences can be found within the structure of (dis)fluency.

institutional and conversational talk with regard to (dis)fluency; we will return to this point in Section III.

Table 31. *Proportion of fluencemes in class presentations and conversations*[144]

	class (raw)		conversation (raw)		Z score	p value
morpho-syntactic markers						
lexical repair	1.3%	(19)	1.1%	(16)	0.346	0.7
morphological repair	3.1%	(45)	2.1%	(29)	1.645	0.1
syntactic repair	3.4%	(44)	4%	(56)	-1.503	0.1
identical repetition	8.2%	(121)	11.6%	(161)	-2.997	0.002*
self-interruption	0.5%	(7)	3.2%	(45)	-5.521	< 0.0002*
truncated word	3.8%	(56)	4.5%	(63)	-0.963	0.3
vocal markers						
filled pause	26.3%	(387)	20.2%	(281)	3.871	< 0.0002*
prolongation	11%	(162)	20.2%	(280)	-6.736	< 0.0002*
unfilled pause	23.4%	(345)	23%	(319)	0.341	0.7
peripheral markers						
NL sound	18.9%	(278)	9.6%	(134)	7.103	< 0.0002*
explicit editing phrase	0.5%	(8)	0.8%	(11)	-1.22	0.2

Turning now to the duration of the vocal markers, the Shapiro-Wilk test (cf Chap. 2, section II. 2.3.1.) revealed that neither filled pauses, unfilled pauses, nor prolongations had a normal distribution (W= 0.86; W= 0.59; W= 0.76), so the duration values (aggregated per speaker) were submitted to a Wilcoxon test for comparison of means. Results show that, despite numerical evidence, no significant differences were found between the average duration of filled pauses during class presentations (M = 412 ms, SD = 240ms) and conversations (M = 340 ms, SD = 199 ms ; p = 0.06). Similarly, no significant differences were found for prolongations (M = 351 ms, SD = 142 ms in class; M = 350 ms, SD = 155 ms in conversation; p = 0.4). Unfilled pauses, however, were found to be of longer duration in classroom settings (M = 695 ms , SD = 543 ms) than conversational ones (M = 594 ms , SD = 323 ms ; p = 0.01). It is interesting to note that, despite no significant differences in proportion (they represent about 23% of the total fluencemes both in class and in conversations), unfilled pauses show significant

[144] Just like Chapter 3, the rate of markers per hundred words can also be found in Appendix 4, Table 68. This table focuses on the proportion of fluencemes to highlight the different ways they are distributed compared to one another.

differences in duration and frequency. This may reflect one specific temporal feature of class presentations during which students rely on time-buying strategies to deal with their production. We will return to this point when we describe the different combinatory patterns used in the two situations (cf Table 34). Just like the vocal fluencemes in SITAF (cf Chap. 2, section II.2.1.1.), the data shows a wide range of variation in the duration values[145], as illustrated in the boxplots from Figure 54 below.

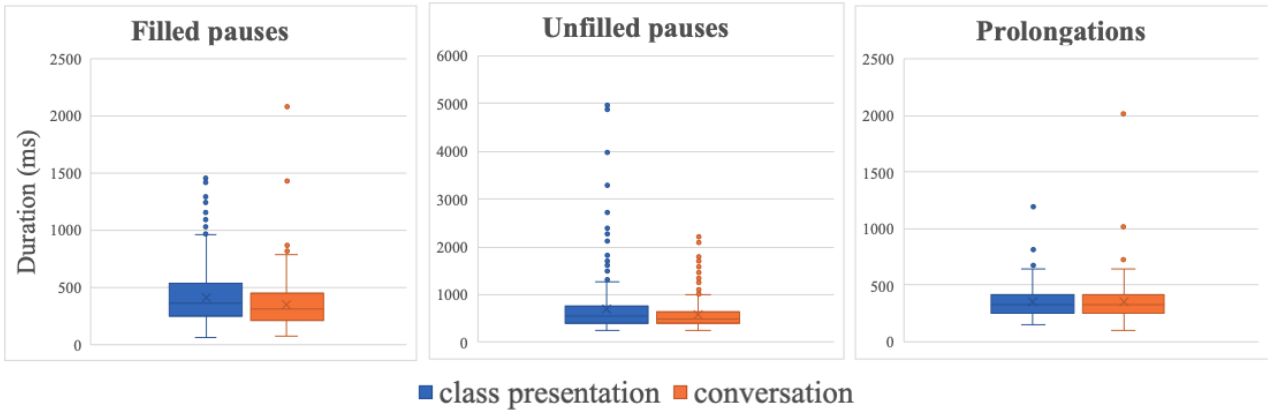


Figure 54. *Duration of vocal markers during class presentations and conversations*

As we have seen earlier (cf Chap. 2, section II. 2.1.1.), boxplots give information about the variability and dispersion of the data. The lower part of the box displays the first quartile, the higher part shows the third quartile, and the line dividing them is the median (the middle value of the dataset). The upper and lower whiskers represent scores outside the middle 50%, and the data points that are located outside the whiskers show observations that are distant from the rest of the data. As the boxplots show, filled pauses' duration values ranged from 60 ms to 1455 ms during the presentations, and 72 ms to 2085 ms during the conversations, with two outliers (i.e. data points) situated at a great distance from the median[146]. For the unfilled pauses, the values ranged from 255 ms (lowest outlier) to 4965 ms (highest outlier) during presentations, whereas during conversations the largest outlier is situated at 2220, which is almost 2000 ms below. Lastly, for the prolongations, a wider distribution is found during conversations, with values ranging from 105 ms to 2016 ms, as opposed to presentations which have a lower range (156 ms – 1196 ms). In sum, filled pauses and prolongations show a much wider distribution during presentations, which may explain the lack of statistical significance between the two means, but unfilled pauses

[145] The exact values are found in Tables 69, 70, and 71 in Appendix 4.
[146] A second Wilcoxon test was conducted after removing these two extreme values, but the differences between the two means still did not reach significance (p = 0.06).

show almost the opposite tendency during presentations, with extremely high maximum values (up to 5 seconds, which is very far from the mean and the median). As we shall see, these very long silences, may reflect a *relevant absence of talk* (Hoey, 2015), during which students may go through their notes in a way that is relevant to pursue the next steps of their presentation. This is further elaborated in section 2.2.

Turning now to the distribution of filled pause types, Figure 55 reports the proportion of "euh" and "eum" across the two situations. Overall, the *euh* variant was the most widely used in both situations, but it was more frequent in conversation (z = -5.144, p < 0.0002), as opposed to "eum" which was over twice as frequent in class (z = 5.144, p < 0.0002).

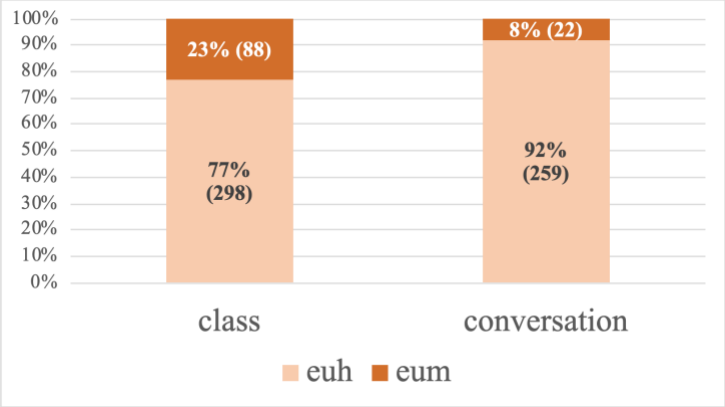


Figure 55. *Proportion of filled pause types ("euh"/"eum") in class and conversation*

In the previous Chapter, we observed crosslinguistic differences in form distribution, which supported the view that filled pauses were essentially language-specific (cf Chap. 3, section II.2.1.1), and this finding provides additional evidence that their form is also sensitive to the type of situation.

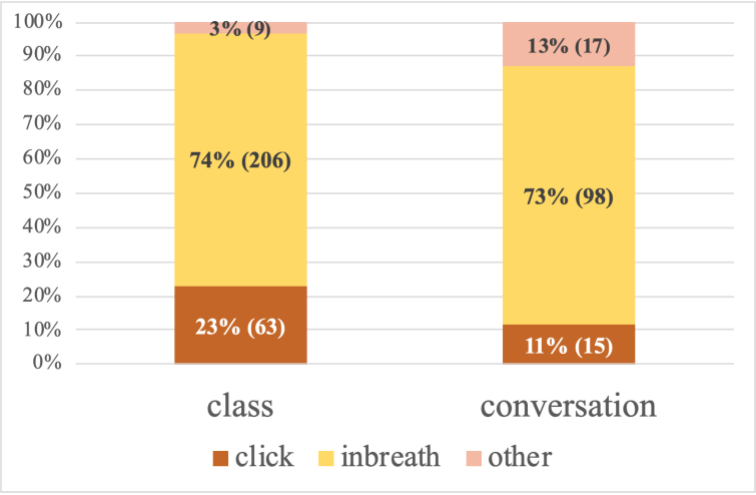


Figure 56. *Proportion of NL sounds in class and conversation*

Figure 56 illustrates the proportion of the different non-lexical sounds per type (audible inbreaths, tongue clicks, and other, see Chap. 3 section II.2.1.1.). As it shows, there is a greater proportion of tongue clicks in class (23%) than in conversation (11%), and this was found to be statistically significant (z = 2.783 ; p = 0.005). In addition, the proportion of non-lexical sounds, i.e. laughter, sigh, creaks (see Table 75 in Appendix 4) was greater in conversation than in class (z = -3.695 ; p = 0.002). The proportion of audible inbreaths, however, did not differ significantly across the two situations (z = 0.048, p = 0.9). These differences in distribution may further reflect specificities of oral presentations versus face-to-face interactions; while inbreaths are common across the two situations, the proportion of clicks was found to be higher in class, while other types of non-lexical sounds, which are perhaps more typical of social interactions, were more frequent in conversation. Tongue clicks and inbreaths, which have been regarded as meaningful interactional resources in the literature (e.g. Hoey, 2014; Ogden, 2013; Torreira et al., 2016; Wright, 2005, 2011) serve a variety of functions, from the display of stance or affect (e.g. disapproval, annoyance) to the handling of sequence management, such as marking a word search or indexing a new sequence of talk. These differences can be measured qualitatively with micro-analyses of the data, by taking into account their phonetic (Wright, 2005) and kinetic (Ogden, 2013) features, which is done in section 2.2.

Before moving on to the combinatory level of fluencemes (cf section 2.1.2.), we shall now conclude this section by analyzing the relationship between utterance length and fluenceme rate. Table 32 reports the mean length of utterance (MLU) per speaker in the two settings. The data showed a normal distribution (W= 0.96), so a two-paired t-test was conducted to compare the differences between the two samples. Results show that speakers produced significantly longer utterances during their presentations than during the conversations ($t(11)$ = 5.308, p = 0.0002), which can be explained by the lack of interactivity in presentations which leads to fewer interruptions (cf Table 32), and may hence result in shorter utterances. This may also explain the high rate of fluencemes found in presentations, which could reflect the effect of utterance complexity on (dis)fluency (in line with McLaughlin & Cullinan, 1989). However, a Pearson Correlation Coefficient (cf Chap. 2, section II. 2.3.1.) was computed to test the potential positive relationship between fluenceme rate and utterance length, but it did not reach significance (r = 0.37 , p = 0.07).

***Table 32**. Mean length of utterance per speaker in class and conversation*

	class presentations		conversations	
Speaker	MLU	SD	MLU	SD
A1	12.6	5	10.2	7.8
A2	15.9	6.8	8.3	5.9
B1	11.9	7.1	10	8.5
B2	13.8	5.8	12.3	7.9
C1	15.5	9.3	8.2	6.8
C2	16.8	9.7	10.6	8.3
D1	19.5	8.3	8.9	11
D2	14.7	7.4	6.5	5.7
E1	13.3	6.2	9.4	9.3
E2	12.3	6.5	13.2	8.1
F1	20	11.1	10.8	8.9
F2	13.1	6.2	7.3	7.9
Total	14.7	7.4	9.6	8

This shows that despite a higher rate of fluencemes and a longer length of utterances in class, these two variables are not necessarily related in the data. Given the significant differences found in the duration of unfilled pauses (cf Fig. 54), an additional test was conducted to measure the relationship between the mean duration of unfilled pauses and the mean length of utterance, and this time, a moderate positive correlation (r = 0.52 , p = 0.0008) was found between utterance length and unfilled pause duration, as illustrated in the scatter plot below.

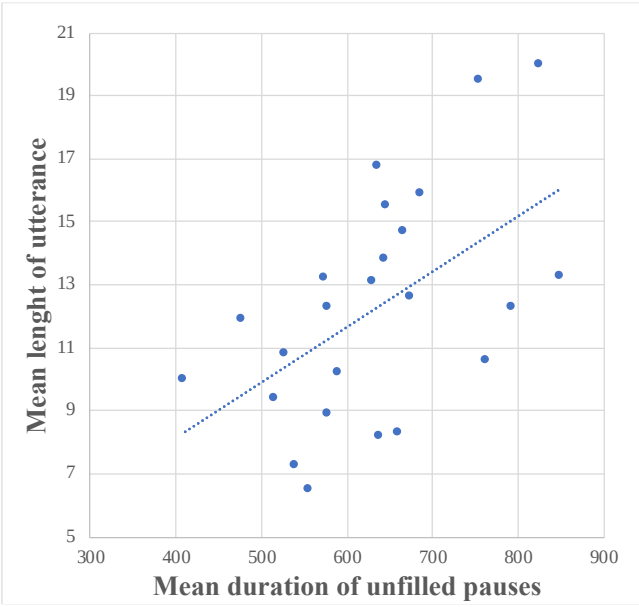


***Figure 57**. Relation between MLU and mean duration of unfilled pauses*

Although it does not show a perfect positive correlation, the data still indicates a tendency for unfilled pauses to last longer when speakers produce longer utterances. This further reflects specific temporal characteristics of class conversations, characterized by lengthy utterances and long pauses, among other processes described in the following sections.

2.1.2. Sequence level: type, length, position, and patterns of co-occurrence

A total of 1576 sequences was identified, with 759 in class presentations, and 817 in conversations. Figure 58 illustrates the proportion of simple versus complex sequences in the two settings.

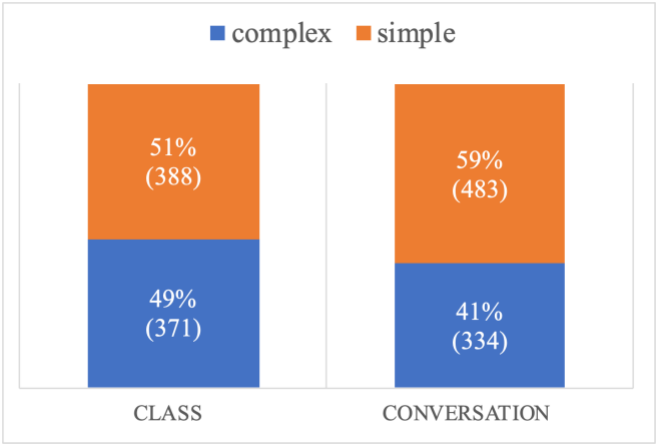


Figure 58. *Proportion of simple and complex sequences in class and conversation*

As results show, there is a slightly higher proportion of complex sequences in class (49%) than in conversation (41%) and as the *Z* test further reveals, these differences are significant ($z = 3.91$, $p = 0.001$). This indicates that, in addition to longer utterances and pauses, class conversations are also characterized by more complex sequences, in line with Moniz (2014), and this variable is shown to have an effect on setting; $\chi^2(1, N = 1576) = 10.1, p = 0.001$. However, no significant differences are found in the average number of markers found within these sequences, as illustrated in the boxplot[147] in Figure 59. As the Shapiro-Wilk test revealed, the values were not normally distributed ($W = 0.67$) so a Wilcoxon test was performed to measure differences between the two samples. On average, speakers combined 2.9 markers ($SD = 1.3$) during their class presentations, ranging from 2 to 11, and during the conversations they combined 2.7 markers on average ($SD = 1.1$), ranging from 2 to 9, which is not a significant difference ($p = 0.6$).

[147] The exact values are found in Table 60, Appendix 4.

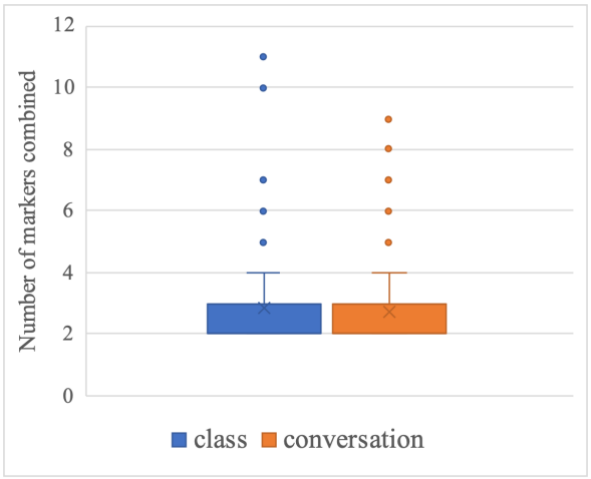


***Figure 59**. Range of markers combined in class and conversation*

However, a few individual differences can be found: for instance, A1 produced fewer clustered markers during his presentation (2.8) than when engaged in the interaction (3.7), while F1 did the opposite, and produced longer sequences during her presentation (3.3) than the conversation (2.7). Once again, students exhibit a number of individual preferences, which calls for in-depth qualitative analyses of the data (see Section 2.2.).

Table 33 reports the proportion of the different sequence configurations across the two settings. Two main patterns emerge, and reveal significant differences in the two situations, mainly *VOC+MS* and *VOC+NL*. Once more, class presentations and face-to-face conversations exhibit differences within the structure of (dis)fluency, with half of the fluenceme sequences composed of vocal and morphosyntactic markers in conversation, as opposed to 30% in class.

***Table 33.** Proportion of sequence configurations in class and conversation*

Seq. Conf.	class (raw)		conversation (raw)		Z score	p value
MIX	2%	7	3%	10	-0.957	0.3
MS+MS	6%	23	7%	22	-0.21	0.8
MS+NL	1%	5	1%	4	N/A	
NL+NL	1%	4	0%	0	N/A	
VOC+MS	30%	113	49%	164	-5.061	<0.0002*
VOC+MS+NL	8%	28	6%	21	0.657	0.5
VOC+NL	32%	118	13%	43	5.979	<0.0002*
VOC+VOC	20%	73	21%	70	-0.423	0.6

This pattern seems to be fairly recurrent, and may reflect a tendency in French conversation to combine stalling and repair strategies in order to (co)-construct

(dis)fluency. In fact, it is interesting to note that this pattern was also used about 50% of the time by the French speakers in their L1 in the SITAF Corpus (See Chap. 3, section II. 2.1.2. Table 18) More discussion regarding (di)similarities in fluency behavior across the two corpora is provided in the *General Conclusion*.

During class presentations, however, speakers equally made use of two patterns, mainly *VOC+MS* and *VOC+NL*, and the latter has a much higher proportion in class (32%) than in conversation (13%). This is a striking result, which further accounts for the prevalence of vocalizations, such as clicks and inbreaths, in class presentations. The latter, along with other vocal markers, such as (un)filled pauses and prolongations, seem to play a major role in the fluency of presentations; they may be used to mark discourse structure, index a new sequence of talk, or project an upcoming delay. This is further explored in our qualitative analyses (section 2.2.).

Lastly, Table 34 shows the different utterance positions of fluenceme sequences across the two situations. Fluencemes occurred slightly more frequently in initial position in class (43%) than in conversation (37%), but they were more frequent in final position in conversation. This further reflects the degree of interactivity and turn-taking mechanisms found in face-to-face interactions, where speakers are more likely to produce utterance-final fluencemes to cede the floor. During presentations, on the other hand, speakers may prefer to produce utterance-initial fluencemes for planning purposes at the macro level.

Table 34. *Utterance position of fluenceme sequences in class and conversation*

Seq. Position	class (raw)		conversation (raw)		Z score	p value
final	3%	22	11%	91	-6.335	< 0.002*
interrupted	0%	0	2%	19	N/A	
initial	43%	326	37%	302	2.246	0.01*
medial	54%	410	48%	396	2.202	0.02*
standalone	0%	1	1%	9	N/A	

In the previous sections, we presented our findings with regard to the distribution of verbal and vocal fluencemes. So far, we have noted a number of significant differences across the two situations, mainly a higher rate of fluencemes, longer pauses, a slightly higher proportion of complex sequences, as well as specific combinatory and positional patterns. We will summarize these findings at the end of this section. We shall now shift to the visuo-gestural level of analysis.

2.1.3. Visuo-gestural level: gesture production and gaze behavior

Table 35 reports the rate of fluencemes during phases of gestural action (Kendon, 2004) in class and conversation. As results show, speakers overwhelmingly kept their hand in rest position when they produced fluencemes in both situations (about 67% of the time), which is consistent with our findings from the SITAF corpus (cf Chap. 3, section II. 2.1.3.). This is further confirmed by the proportion of gestures found during utterances produced with fluencemes and outside fluencemes, which shows that a majority of gestures were produced without fluencemes both in class and in conversation, as illustrated in Figure 60.

Table 35. *Distribution of fluenceme sequences during gesture phases in class and conversation*

	class (raw)		conversation (raw)		Z score	p value
stroke	14%	107	15%	120	-0.545	0.5
hold	12%	88	9%	71	1.913	0.05
preparation	8%	58	6%	46	1.607	0.1
rest position	63%	478	67%	551	-1.86	0.06
retraction	4%	29	3%	27	0.553	0.5

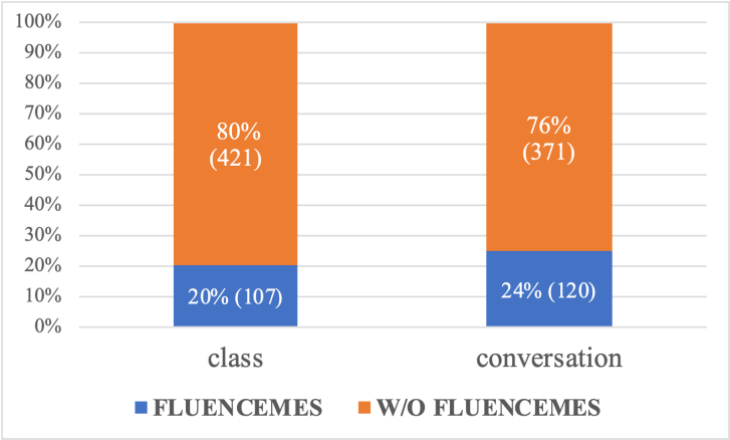


Figure 60. *Proportion of gestures that occurred in utterances with or without fluencemes in class and conversation*

However, unlike our previous study on native and non-native productions, no significant differences are found in the distribution of gestures and gesture phases between the two settings, whether they co-occurred with fluencemes or not. This may reveal that while language proficiency may have an effect on the type of gestural behavior accompanying fluencemes, setting and style seem to have very little. However, this table only gives information about the types of gesture phases that typically co-occur with fluencemes, which only offers a partial view of the gestural

practices characterizing class presentations and face-to-face interactions. This needs to be completed with information regarding the distribution of all gestures, regardless of fluency, which is provided in Figure 61 below.

As results show, students produced 10.2 gestures phw on average during their class presentations (N= 528) and 7.6 phw during the conversations (N= 491), which is a significant difference (LL = 20.46 , p < 0.0001). Contrary to our expectations (H5, cf Section I. 1.4.) students produced significantly more gestures when they performed a monologic task than a dialogic one, which further refutes Bavelas et al., (2008)'s findings that speakers produce more gestures in dialogue than in monologue.

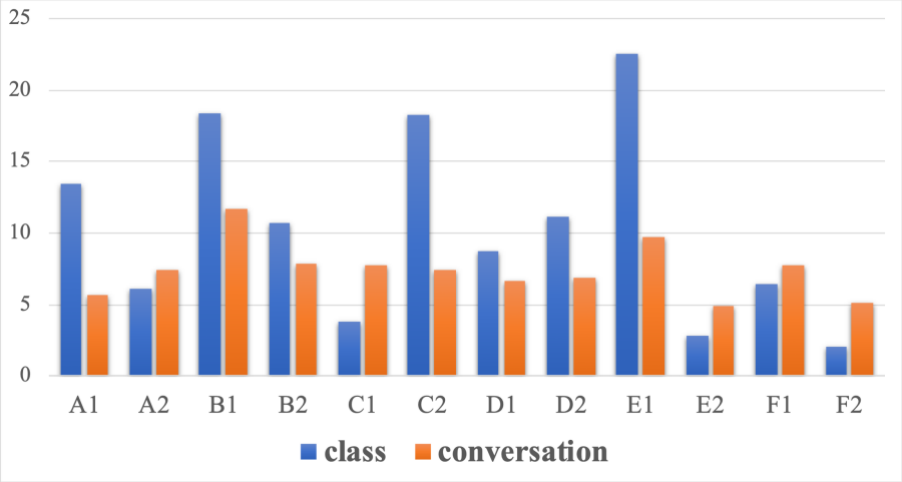


Figure 61. *Rate of gesture strokes (phw) in class and conversation*

Once again, this further demonstrates that mode of speech alone (see Section I. 1.1.2) is not sufficient to characterize differences between formal class presentations and casual interactions. Unlike Bavelas et al.'s (2008) study, the students from the DisReg Corpus did have an audience, and despite the lack of interactivity between the speaker and the group of hearers, it appears that a majority of students still produced gestures like they would do in a conversation, which is more in line with Sweetser & Sizemore (2006). However, this needs to be confirmed by looking at the specific types of gestures that were most frequently deployed in the two situations, which is provided in Table 37 further below. In addition, as Fig. 61 shows, a number of individual differences were found, as students display very different gesturing strategies: while some (i.e. A1, B1, C2, and E1) mobilized a significant amount of gestures to deliver their presentations, others (i.e. F2, E2 and C2), showed the opposite tendency. This further demonstrates that speakers have their own stylistic idiosyncrasies, which reflects their own individual language style when it comes to fluency and gesturing behavior, regardless

of setting or register. Once more, this calls for further investigation in more detailed qualitative analyses of the data (Section 2.2.).

Table 36 provides more information about the distribution of gesture types and subtypes in class and conversation. As results show, a number of significant differences can be found in the two settings. Overall, students produced significantly more pragmatic gestures in class than in conversation, especially discursive ones (i.e., gestures used for presentation and emphasis) which represent 83% of all their gestures. Conversely, they mobilized a higher proportion of referential gestures in conversation than in class, and especially representational ones (17% in conversation versus 4% in class).

Table 36. *Proportion of gesture types and subtypes in class and conversation*

	class		**conversation**		**Z score**	**p value**
referential gestures % (raw)						
	17%	**(89)**	**36%**	**(175)**	**-6.822**	**< 0.002***
representational	4%	(21)	17%	(85)	-0.1333	< 0.002*
deictic-anaphoric	13%	(68)	18%	(90)	-2.389	0.01*
pragmatic gestures % (raw)						
	83%	**(439)**	**64%**	**(316)**	**6.9**	**< 0.002***
discursive	69%	(363)	22%	(108)	14.992	< 0.002*
interactive	11%	(60)	39%	(191)	-10.178	< 0.002*
thinking	3%	(16)	3%	(17)	0.384	0.7

Additionally, students produced considerably more interactive gestures when they were engaged in the conversational task, which is not surprising, given the lack of interactivity in class presentations, which offers little room for communicative gestures. Interestingly, students produced an equal amount of thinking gestures in the two situations, which represent a very small proportion (3%) of all the gestures overall. Similarly, in the SITAF Corpus, thinking gestures amounted to about 3-4% of all the gestures in L1 in the two speaker groups, but significant differences were found in their distribution in L1 and L2. It would appear that in the case of DisReg, setting, unlike language proficiency, did not have a strong effect on these gestures. A more detailed typology of thinking gestures is provided in Chapter 5.

Table 37 gives more information about the distribution of gesture subtypes[148] based on whether they occurred with or without fluencemes, across the two settings. Here it is interesting to note that, despite no significant differences in the distribution of fluencemes across gesture phases (cf Table 35 and Fig. 60), some differences can be noted regarding the proportion of gesture types with respect to fluency.

Table 37. *Proportion of gesture subtypes with or without fluencemes in class and conversation*

	CLASS					CONVERSATION				
	Fluencemes		**W/O**		**Z (*p*)**	**Fluencemes**		**W/O**		**Z (*p*)**
deictic-anaphoric	8%	9	14%	59	-1.51 *(0.1)*	12%	14	20%	76	-2.17 (0.03*)
representational	5%	5	4%	16	0.41 *(0.6)*	28%	33	14%	52	3.39 *(0.0007*)*
discursive	69%	73	69%	299	0.04 *(0.9)*	20%	24	23%	84	-0.60 *(0.5)*
interactive	4%	4	13%	56	N/A	27%	32	43%	159	-3.16 (0.001*)
thinking gesture	14%	15	0%	1	N/A	14%	17	0%	0	N/A

As the Table shows, not many differences are found in the distribution of gestures during fluencemes and outside fluencemes in class presentations, except for thinking gestures that virtually never occurred outside of fluencemes. In conversations, however, we find a number of differences, with a higher proportion of deictic-anaphoric gestures without fluencemes (20%) than with them (12%), but more representational gestures during fluencemes (28%) than without them (14%). Lastly, a higher proportion of interactive gestures are found outside fluencemes. Overall, these findings reveal a significant association between gesture type and setting; χ^2 (1, N = 1019) = 247.1 , p < 0.00001).

In addition, if we further compare the two settings, results show that students produced more interactive gestures during fluencemes in conversation than in class, as well as more representational gestures (z = -4.57 , p < 0.0002), but they produced more discursive gestures during fluencemes in class (z = 7.407 , p < 0.0002). This is the same pattern of distribution described earlier (cf Table 37, when we reviewed the proportion of gestures found in presentations and conversations. This further confirms that, regardless of (dis)fluency, a certain category/type of gestures is

[148] Information regarding the distribution of the two main gesture types (referential and pragmatic) can be found in Figure 76, Appendix 4.

preferred in a specific setting, reflecting a number of situational factors, such as audience design (interactants are more likely to deploy interactive gestures in a conversation because of their familiar relationship), or style (students need to deliver a clear presentation and they may do so with the help of discursive gestures). We will return to this point in Section III.

We shall now conclude this section with the analysis of gaze behavior. Figure 62 reports the proportion of different gaze directions in class and conversation, excluding "in different directions" and "towards camera"[149] which were too rare overall (the exact values are found in Table 75, Appendix 4). As the table shows, the differences between the two situations are highly significant. Students spent nearly 70% of their time gazing towards their notes (or laptop, book etc.,) during the class presentations, and only looked in direction of their audience 27% of the time. Conversely, during the interactions, speakers gazed considerably more frequently towards their interlocutor (58%, $z = 43.798$, $p < 0.0002$). It is also interesting to note that they gazed more frequently away during the interactions (27%) than the presentations (4% , $z = -24.98$, $p < 0.0002$). This is a surprising result, as it is very common to withdraw one's gaze when engaged in an interactional practice, as to display a state of disengagement, a word search, or the end of a sequence, among other things (e.g. Goodwin & Goodwin, 1986; Kendon, 1967; Rossano, 2013; Streeck, 2014).

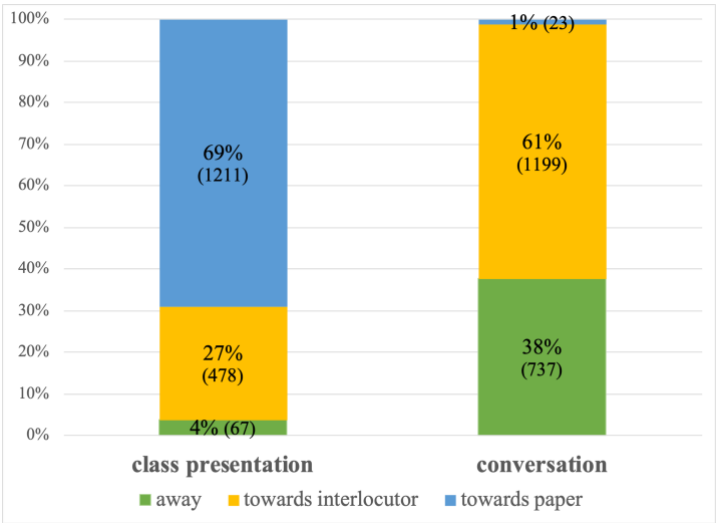


***Figure 62**. Proportion of gaze direction in class and conversation*

[149] It is interesting to note that 25 instances of gazing towards the camera were found during the class presentations, while it happened only once during the conversations (see Table 75 in Appendix 4). The students were asked explicitly not to look in direction of the investigator (who was holding the camera) to avoid being distracted, but it appeared that some students could not help doing so while they were delivering their presentation; perhaps it was a way for them to include the investigator in the audience.

This finding may reveal that looking away is a pattern of gazing more commonly found in face-to-face conversations rather than formal class presentations. Perhaps presenters did not want to be seen gazing away by their audience, which could potentially reflect a loss of face or control over their presentation; or perhaps they did not find it necessary to gaze away, since they were extensively relying on their notes.

Overall, there seems to be a mismatch between the students' gesturing and gazing behavior. While they gestured quite frequently during their presentations, and even more frequently than they did during the conversations, they constantly gazed towards their notes, suggesting that they were barely orienting to their audience and were much more focused on performing the task at hand. The latter seems to have a significant effect on gazing behavior, and this was confirmed by a chi-square test of independence (cf Chap. 2, section II. 2.2.1.) which showed a significant relationship between gaze direction and setting; χ^2 (1, N = 3715) = 206.9 , p < 0.00001.

The next figure illustrates the proportion of gaze direction within and outside fluencemes. In class presentations, we find a higher proportion of gazing towards the piece of paper during fluencemes (z = 5.602 , p < 0.0002), but a higher proportion of gazing towards the interlocutor without them (z = -6.542, p < 0.0002). Overall, these findings account for a strong association between fluency and gaze direction (χ^2 (1, N = 1756 , p < 0.0001). Similarly, in conversations, a slightly higher proportion of gazing towards the interlocutor is found without fluencemes (z = -6.26 ; p < 0.0002) but the proportion of gaze withdrawals is greater during fluencemes (z = 6.407 , p < 0.0002).

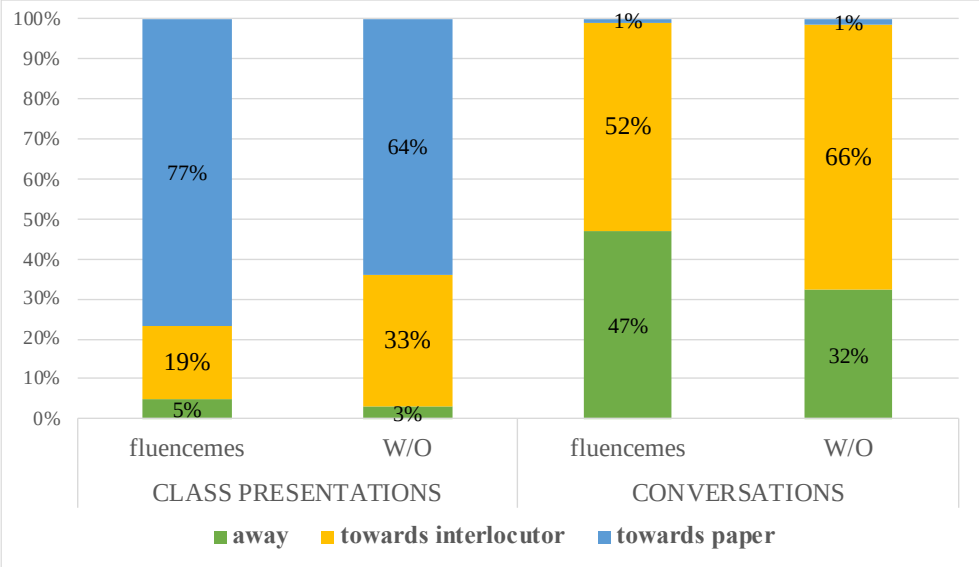


***Figure 63**. Proportion of gaze direction with and without fluencemes in class and conversation*

Overall, these findings show that speakers were more likely *not* to establish eye contact when they produced fluencemes in both situations, which is consistent with our findings from the SITAF Corpus where the two language groups were found to withdraw their gaze more frequently during fluencemes both in their L1 and L2 (see Chap. 3, section II. 2.1.3.). This further emphasizes the fact that gazing away is a very common practice of (dis)fluency, as it enables speakers to momentarily retreat from the current activity to attend to other relevant ones, such as retrieving an item from memory, looking for a specific word, checking for a sentence in a book, etc. Additionally, it is interesting to note differences in gazing behavior during fluencemes across the two situations: Speakers gazed towards their interlocutor significantly more frequently during fluencemes in conversation than in class (z = -12.952, p < 0.0002). This further shows that, even though gazing away is a common activity during fluencemes, it varies significantly depending on the situation. Once more, this accounts for the dynamic and fluid nature of fluencemes that are constantly being (re)shaped by local and global situational features.

To conclude, this section has outlined several differences regarding (dis)fluency and visible bodily behavior in two distinct communication settings and language styles, i.e. formal class presentations and casual interactions. As we have seen, these two situations cannot be solely distinguished on the basis of type of delivery (read versus spontaneous) or mode of speech (dialogue versus monologue) since our findings have shown a greater rate of fluencemes and gestures in class presentations, which reveals that the latter is not necessarily characterized by carefully read speech and a total lack of interactivity, as is often expected of "monologues" (see section I. 1.1.).

In sum, many significant differences were found in the two situations, which are briefly summarized in Table 38. These findings, which are further discussed in Section III, give an overall idea of the rate, form, and distribution of fluencemes, as well as the frequency and distribution of gestures and gazing behavior. In addition, many individual differences were found across the two settings, further reflecting individual speaking styles and idiosyncrasies. These differences are further highlighted in the following section.

Table 38. *Summary of the quantitative findings*

	class presentations	**conversations**
VOCAL-VERBAL FEATURES (FLUENCEMES)		
Fluenceme rate	Higher rate in class presentations than conversations	
Distribution	more NL sounds, filled pauses and longer unfilled pauses (correlated with utterance lenght)	more repetitions, interruptions and prolongations
Filled pause type	more eum-type filled pauses	more euh-type filled pauses
Combination type	more complex sequences	more simple sequences
Utterance position	more instances of utterance-initial fluencemes	more utterance-final fluencemes
VISUAL-GESTURAL FEATURES (WITHIN AND OUTSIDE FLUENCEMES)		
Gesture rate	Higher rate in class presentations than conversations	
Gesture distribution	more pragmatic gestures overall, and more discursive gestures (with and without fluencemes)	more referential gestures overall, and more interactional gestures (with and without fluencemes)
	no significant differences for thinking gestures, but occurred almost exclusively during fluencemes	
Gaze behavior	more instances of gazing towards piece of paper (with and without fluencemes)	more instances of gazing away and towards the interlocutor (with and without fluencemes)

2.2. Qualitative analyses

In this section, we further explore the differences between formal class presentations and casual face-to-face conversations in relation to inter-(dis)fluency by presenting a number of micro-analyses from the data. Just like Chapter 3, we begin with an overview of their distribution with regard to *communication management* in the two situations, and further explore the functional and interactional ambivalence of fluencemes across the two situations (section 2.2.1). We then present several analyses from selected excerpts of the data to further illustrate the multimodality and multifunctionality of inter-(dis)fluency with regard to audience design, common ground and participation framework. Just like Chapter 3, the participants were assigned pseudonyms specifically in this section to render the exchanges more authentic (cf Chap. 2).

2.2.1. Overview of Communication Management in the two situations

As emphasized several times before throughout this thesis, we view inter-(dis)fluency as a dynamic process containing fluid categories, whose degree of interactivity,

lexicality, and to a larger extent, *fluency*, is determined by a number of contextual and situational features. In this sense, the same a priori "disfluent" non-lexical forms can serve production-oriented functions related to planning processes (*own communication management*, OCM), but also more communicative functions related to intersubjectivity and turn management (*interactive communication management*, ICM, see Chap. 1., section II. 2.2.1, and Chap. 3, section II.2.2.1.). Figure 64 reports the proportion of fluencemes with ICM and OCM functions in class and conversation. As the numbers show, only 1% of the fluencemes performed the ICM function during the class presentations, as opposed to 27% in the conversations. This result is very striking, although not surprising, given the interactional constraints imposed on the presentations. Unlike our previous study on the SITAF Corpus which showed no significant differences between L1 and L2, here the data suggests a strong effect of setting on the functions of fluencemes $\chi^2(N(1) = 1756 , p < 00001)$, thus giving more support to their ambivalent nature.

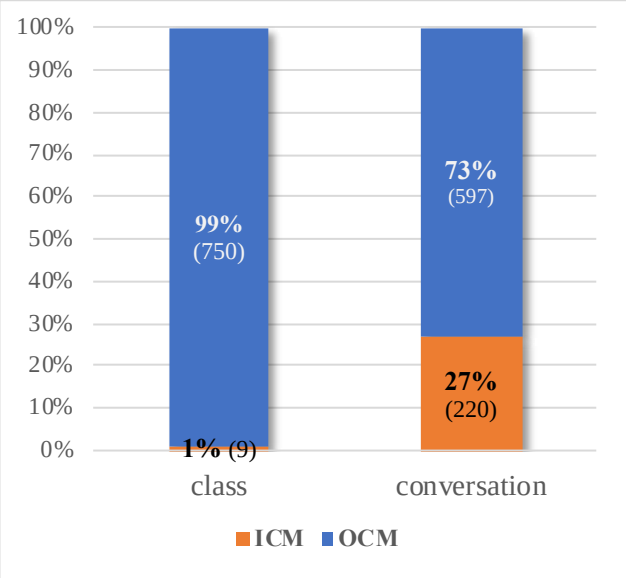


***Figure 64**. Proportion of OCM and ICM functions in class and conversation*

There is, however, one important point to make. Even though the class presentations were a priori "monologic" and did not involve participation from the interlocutors, it should not stop presenters from *addressing* their audience and maintaining intersubjectivity, as we have seen in Sweetser & Sizemore's (2006) paper in which they described a lecturer's pointing gesture oriented towards the audience (cf section I.1.3.). Although the students did address their audience occasionally during their presentations, they almost never used fluencemes to display a stance or maintain mutual understanding, like they would more often do in a conversation (see section

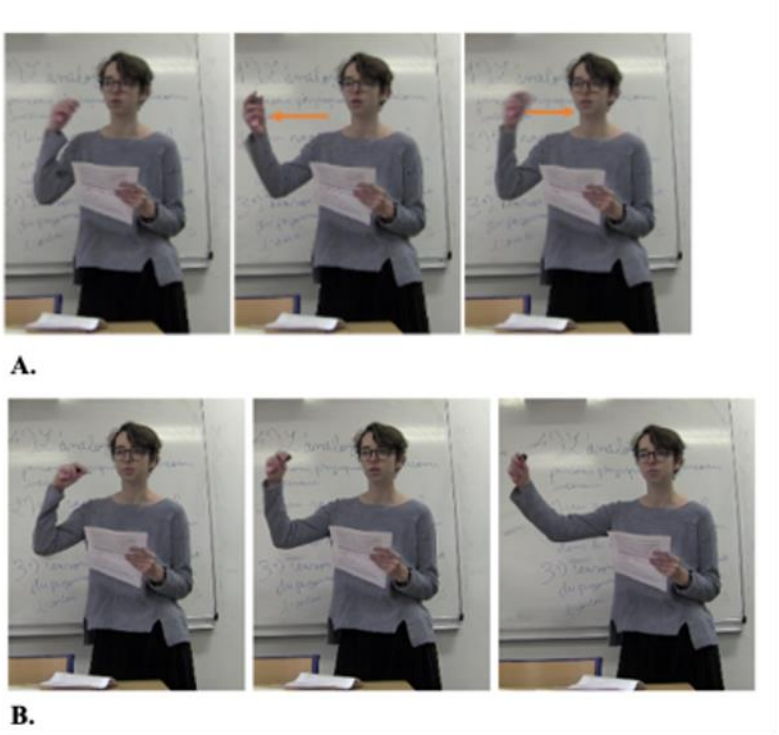
2.2.2.). As Figure 64 suggests, fluencemes were more frequently used by the students as time-buying tools to handle aspects of planning or repair, thus focusing on their own performance. There are, however, some exceptions that deserve to be noted, as is illustrated in the brief excerpt below.

Excerpt CLASS 1.1

This excerpt is taken from one recording in which Alex (Participant D1) is presenting her assignment to the class with her partner Jenny (D2)[150]. In this excerpt, the presenter is talking about a character from a French novel who is waiting for a phone call from Europe in a wooden cabin.

```
1  (0.703) eum on peut lire
2  euh +/ ce bruit très bref (0.420) le mettait lui qui se préparait à la
   conversation"/.
        ((reads a quote from the book))
3  c'est à dire le bruit de (0.468) vous savez de quand ça fait bi:ip bip

   ~~ ~ ~~ ~ ~~ ~ ~~ ~~~ ~ ~~ ******.-.--.-.--.-.--.-.~~~~~~~~~~**********
   [_____________________________________________] [____________________]
                GESTURE UNIT A.                          GESTURE UNIT B.
```



[150] This is the only example from the corpus where two students are presenting together and not alone.

fin voilà quand il attend le:e [/] la réponse de l'Europe.
~~ ~ ~~ ~~ ~ ~~ ~~ ~*********-.-.--.-.--.-.--.-.--.-.--.-.--.-.
[_____]

GESTURE UNIT C.

C.

4 (0.602) [!] +/ dans un état d'excitation sans nom et à la lettre il appellait de ce bout si dès le premier mot
((resumes her reading activity, gazes towards notes))
celui qui se trouvait à l'autre bout "/.
((gazes towards the audience))

At the beginning of the excerpt, she is reading a passage from the book which refers to a specific noise on the telephone ("ce bruit très bref" l.2), while gazing towards her notes, holding the piece of paper with both her hands. In line 3, she interrupts her current reading activity to address the audience specifically ("vous savez") and establish intersubjectivity, which is manifested in her vocal and bodily behavior. First, she initiates a noun phrase "le bruit de", which is followed by an unfilled pause of 468ms. As we have seen earlier, (cf section 2.1.1.) pauses were very common during class presentations, but this one, unlike the pauses produced in her previous utterances in lines 1 and 2, is accompanied by a gestural movement. As the illustration (A) further shows, she rapidly extends her right arm back and forth, and with her fist slightly closed, she moves her thumb and index fingers apart as she initiates the movement. As the multimodal transcription indicates (cf Chap. 2, section I. 1.5.3.) the stroke of this gesture is perfectly synchronized with her vocal fluenceme. This is another example of *contextual configuration* (Goodwin, 2000; cf Chap. 3, section 2.2.1, excerpt Pair 07) where a different semiotic field is deemed more relevant to pursue the current action; here she used her body to convey meaning, and in order to do so, she momentarily delayed the unfolding of her verbal utterance, which was treated as no longer relevant. This is also a reference to the dynamic *scope of relevant behavior* theory (Cienki, 2015) introduced several times in this thesis, whereby

speakers shift the focus of their current expressive behavior, from vocal to visual-gestural resources. Here the two resources overlap with one another, which enables the speaker to delay her current speaking activity while expressing communicative content to her audience. Once again, we can hardly speak of "disfluency" in this case, but rather multimodal *inter-fluency*.

The same gesture is then repeated with the same hand configuration, although slightly modified, as this time she extends her right arm farther away in her gesture space, and her movement is synchronized with a syllable prolongation of the onomatopoeia "bi:ip", and is marked by a slight beat motion, as she repeats the same item. It appears that the first gesture (A), initiated during the pause, foreshadowed or projected the course of her next manual action (Streeck, 2009), which was then further elaborated during the prolongation and repetition of the onomatopoeia. And, as just described, the extended movement and beat of the gesture were timely synchronized with prosodic patterns of the onomatopoeia[151]. These two gestures, which can be considered *representational* in this case, as they convey meaning related to a specific content (the type of noise made on the phone) through a gestural mode of representation (Müller et al., 2013, cf Chap. 5) were extremely rare in class presentations overall (about 4%, see section 2.1.2.), but here it appears that the presenter wanted to make sure that the audience understood what this "bruit très bref" referred to, so she deployed a combination of vocal and gestural resources to construct meaning specifically for her audience. She is not concerned with her own performance anymore, i.e. which phrase to read, what to say next etc., but seems fully oriented towards her audience.

These two fluenceme sequences i.e., the simple vocal fluenceme made of an unfilled pause, and the complex one made of a prolongation and a repetition (following the *VOC+MS* pattern), were both rather frequent in class presentations overall; but here they serve entirely different functions from the rest of fluencemes. Take for instance, the other unfilled pause found in this excerpt in line 4 clustered with a tongue click, following the *VOC+NL* pattern, which was also very common in class presentations (about 32% of all patterns, see section 2.1.2.); this fluenceme sequence, which occurs in utterance-initial position, during which the speaker is once again looking through her notes, does not have any communicative value, and adds no

[151] The temporal relationship between gesture and (dis)fluency is further elaborated in Chapter 5.

propositional content to the speaker's utterance. Similarly, in line 3, almost immediately after gestures A and B, the speaker produces a third one (picture C) during which she extends her right arm to about the same position as the previous one within the gesture space, but this time with a Palm-Up Open Hand. As her arm and hand return to rest position, she also initiates another fluenceme sequence following the *VOC+MS* pattern ("le:e [/] la") during which she first lengthens the masculine determiner "le" and then corrects it to the feminine gender ("la"). Unlike the two previous fluenceme sequences in line 3, this one does not co-occur with a gesture stroke, but is in fact produced at the completion of the speaker's gestural activity. She also gazes back towards her notes at that exact same moment, thus resuming her reading practice.

In sum, this short excerpt has illustrated how the same a priori "disfluent" forms or patterns of co-occurrence can serve radically different functions, depending on their co-occurring visual-gestural behavior. This is consistent with the qualitative analyses of the SITAF Corpus we presented in the previous chapter (cf Chap. 3, section 2.2.), which further supports the dynamic nature of fluencemes across and within languages and settings. Further in line with our ambivalent approach to inter-(dis)fluency, we shall now present two other excerpts from the same pair of speakers (Pair D) which illustrate different uses of *tongue clicks*.

2.2.2. The case of tongue clicks: blending vocal and kinetic behaviors

Tongue clicks (*tsk, ttut*) can be described phonetically as "a click articulated with the tongue tip, with central release which is generally slow and affricated" (Ogden, 2013, p. 302). Clicks, which have often been assigned to the margins of language, belong to a larger class of vocalizations, also known as *sound objects* (Reber & Couper-Kuhlen, 2020), *peripheral linguistic objects* (Ogden, 2018), or *liminal signs* (Dingemanse, 2020). Just like other vocal fluencemes such as filled pauses or unfilled pauses, they essentially lack semantic weight; however, a number of researchers in socio-interactional research (to name but a few: Hoey, 2014; Ogden, 2013, 2018, 2020; Schegloff, 1996; Ward, 2006; Wright, 2005, 2011) have recognized their relevant contribution to the accomplishment of talk-in-interaction, and have documented several recurrent social practices involving their use. They can be further defined as "sounds made in the vocal tract alongside speech, not as part of the lexical content of the language, but *clearly as a resource for making meaning*" (Ogden, 2013, p. 299,

our emphasis). While they are often used to display stance or affect (e.g., disapproval, annoyance, irritation, impatience, sympathy, see Wright, 2011, p. 208), they can also handle aspects of sequence management, such as projecting a new sequence of talk, closing down a current topic (Ogden, 2013, Wright, 2007), or marking a word search (Pinto & Vigil, 2019, Wright, 2005). Ogden (2013) in fact identified three main functions of clicks: (1) *marking incipient speakership* – clicks in turn-initial position used to project speech and mark the transition from listener to incipient speaker; (2) *handling sequence management* – when they occur during word-searches or index new sequences of talk; (3) *displaying a stance* – when they project a stance (annoyance, impatience etc.). Additionally, in face-to-face interactions, clicks are often associated with a number of visible and kinetic behaviors, such as eyebrow flashes, or swallowing (Ogden, 2020, 2018) and manual gestures (Pinto & Vigil, 2019), which further accounts for their multimodality. However, while a lot of work has been done on tongue clicks in Conversation Analysis and phonetics, their analysis is virtually absent from (dis)fluency research, as they have not traditionally been labeled as fluencemes. Yet, our results have shown that tongue clicks, among other vocalizations and breathing phenomena, often cluster with other fluencemes, so they should not be overlooked. In the next examples, we further explore their functional ambivalence, and the different ways they may contribute to the fluency of multimodal discourse.

Excerpt 1.2. CLASS[152]

In the previous excerpt, we already illustrated an instance of utterance-initial click clustered with an unfilled pause, used to project a new sequence of talk, and more specifically to mark the resumption of the reading activity. In the following excerpt, taken from Jenny's (D2) presentation in class, we present a similar practice involving a tongue click, along with other vocal and visible behaviors. Following her partner's intervention (cf Excerpt 1.1.), it is now her turn to present her part and talk about the notions of "point de départ et point d'arrivée" in another French novel which tells the story of a man, named Barnaba, who has become fond of paintings after becoming blind. Here she is analyzing the ways one particular painting (*Le tableau de David*, not in the transcription) becomes a key figure in the life of Barnaba. The transcription includes pictorial illustrations of her visual-gestural behavior, as well as a PRAAT

[152] This excerpt was presented at the *Laughter and Other Non-Verbal Vocalisations Workshop* (Kosmala, 2020b).

picture, showing the waveform sampled from the fluenceme sequence (in bold in the transcription).

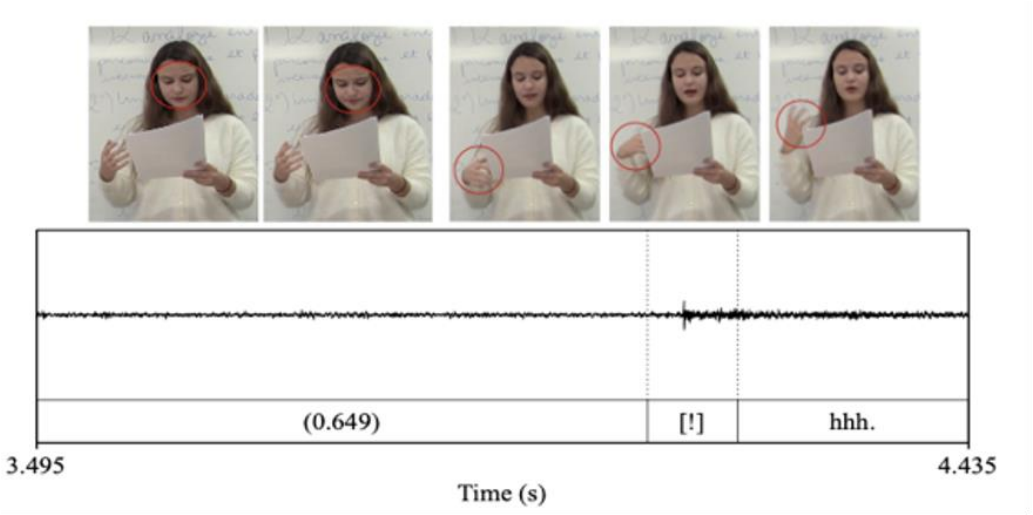
1 hhh. (0.400) et donc euh ce tableau (0.400) constitue donc u:un leitmotiv dans le texte dans le récit hhh. et dans le parcours de Barnaba

2 c:c'est une figure presque obsessionelle à laquelle i [//] il revient toujours.

→ **3 (0.649) [!] hhh.** c'est donc un point de départ dans le musée

~~ ~ ~~~~ ~ ~~~~

((parted lips, eyebrow flash, swallowing activity, open mouth, right hand preparation))



4 car c'est bien la première étape du parcours

The tongue click occurs in line 3, clustered with an unfilled pause of 649ms and an audible inbreath. This complex fluenceme sequence, following the *VOC+NL* pattern, is produced in initial position, and projects a new sequence of talk ("c'est donc un point de départ dans le musée") which marks the conclusion of her current argument (the fact that this painting is a recurrent theme in the character's life). What is interesting to note is that the projection of this new stretch of talk is also made visible in her visual-gestural behavior, as we can see her opening her mouth and moving her right hand in preparation during the audible inbreath, following the tongue click. The latter is clearly visible in the waveform, as clicks tend to be relatively "loud transient sounds" (Ogden, 2013, p. 307)[153]. It is also surrounded by silence and an inhalation, during

[153] As pointed out several times in this thesis (cf Chap. 2, section III.3.), the present study does not aim to extensively explore the phonetic aspects of tongue clicks and other fluencemes, but rather focuses on

which the speaker deploys several physical actions; we can see her swallowing, flashing her eyebrows, and slightly frowning prior to the production of the click. As Ogden (2013) explained, clicks can be regarded as the culmination of a swallow, which, along with breathing, are basic activities that are closely associated to their production in English conversation. A very similar pattern of distribution was also found here, during a French oral presentation, which calls for more crosslinguistic comparisons.

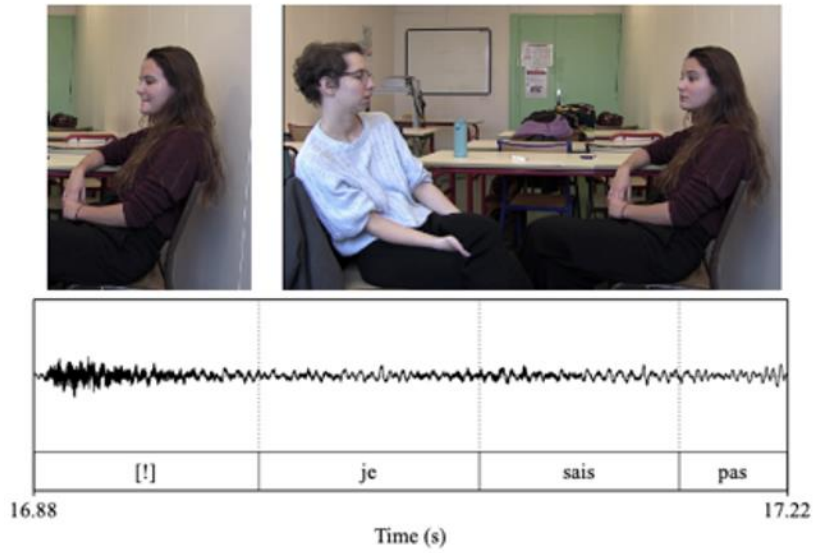
In the following excerpt, with the same speakers (Jenny and Alex, pair D) engaged in the interactional task, we find another instance of a tongue click, but this time initiated at the beginning of a turn, in response to a prior one.

Excerpt 2.1. Conversation

In this excerpt, the two friends are talking about a TV show that Jenny highly recommends to Alex, although she has not watched it yet. They are also talking about two actors, Yvan Attal (not included in the transcription) and Neil Schneider, who both play in the show.

	1	*JEN:	regarde là fi:in ça a l'air +//.
	2	*JEN:	fin j'ai pas trop regardé mais ça a l'air vraiment bien.
	3	*JEN:	(0.621) e:et euh [/] et lui il est vraiment vachement bien dedans.
	4	*ALX:	(0.668) ah ouais.
	5	*JEN:	e:et euh [/] e:et Neil Schneider je l'ai pas encore vu mais euh (0.929) +...
	6	*ALX:	parce-qu'il joue dedans aussi lui?
	7	*JEN:	(0.531) hhh. euh ouais.
	8	*ALX:	(0.631) mais il a un rôle euh (0.924) +...
→	**9**	***JEN:**	**[!] je sais pas je l'ai pas vu encore.**
	((gazes away, smacks her lips))((gazes towards ALX; raised eyebrows))		

local durational, phonetic, or intonational patterns and the way they synchronize with other types of bodily behavior.



10 *JEN: je vais regarder les premiers.
11 *JEN: (0.564) j'ai pas vu.
12 *ALX: mm.

Jenny seems very eager to watch this TV show, and makes a series of positive assessments, as indicated by the number of intensifiers found in her utterances; “ça a l’air *vraiment* bien”, (l. 2); “il est *vraiment vachement* bien” (l.3), among others (not in the transcription). As Pomerantz (1984) explained, assessments can be viewed as products of participation whereby participants claim knowledge of what they are asserting (cf Chap. 1, section III. 3.2.3.). Here there seems to be a mismatch between Jenny’s positive assessment towards the TV show and the actors who play in it, and her actual lack of knowledge (since she has not watched the show yet). In fact, she keeps inserting second downgraded assessments (“j’ai pas trop regardé”, l. 2 or “je l’ai pas encore vu” l. 5) as a response to her subsequent positive ones. This imbalance is further reflected in the sequential development of the exchange, where Alex keeps initiating questions about the show (l.4 and 8), and Jenny answers with turn-initial vocal fluencemes, which function as delaying devices (l. 4 and 7), as a way to display ignorance, thus disturbing the overall progressivity of the exchange. The tongue click occurs in line 9 in turn-initial position, following Alex’s invitation to take the turn and elaborate on the show[154]. Here the click prefaces Jenny’s lack of knowledge in the form of a non-response “je sais pas”, suggesting a dispreferred next action. This is further displayed in her visible behavior, as she is seen raising her eyebrows, and slightly

[154] As explained earlier, (cf Chap. 2, section 2.2.2.) we consider the second unfilled pause in line 8 to be clustered with the filled pause as part of Alex’s turn, where it occurs at a *transition relevant place* in utterance-final position.

pouting, shortly after producing the non-response. In this case, the click displayed epistemic stance, but also indexed the next turn-at-talk, following a transition relevant place in the prior turn. Once again, this further shows that the same forms can follow different patterns of distribution, and occur at different positions within the turns-at-talk, hence displaying radically different functions. In the two previous excerpts from the class presentations, the clicks were mostly used to index a new sequence of talk while the presenters were going through their notes and presenting their next argument; in this excerpt, however, the click was used as a response to a prior turn, hence more oriented towards the ongoing exchange. While the different types of practices underlying the uses of clicks in conversation have already been documented several times before (e.g. Ogden, 2013, 2018; Wright, 2007; 2011), they have almost never been analyzed within the scope of (dis)fluency. Just like other fluencemes, clicks have the potential to serve both production-oriented (OCM) and more interaction-oriented (ICM) functions, depending on the setting and style used. The interactional contribution of fluencemes is further described in the following section.

2.2.3. Embodied displays of intersubjectivity in storytelling: the interactive dimension of fluencemes

As described earlier, (cf Chap.2, section I.1.3.1.) in the conversation-sessions, the students were given a sheet of paper with a list of topics written on it beforehand, to help them start the conversation. One of them said "funny anecdote at university", and while this topic of conversation was not covered by all the students, two pairs (Pairs B and D) found it relevant to bring it up in the course of their interaction. The following excerpts are thus taken from the recordings of these pairs. The two extracts are rather long, so several lines are omitted from the transcription. The first analysis is mostly based on Kosmala (2020), and the second one was presented at a data session at the *Co-Operative Action Lab* in UCLA in February 2020.

Excerpt 2.2. The funny-looking shoes

In this extract, Paul (B1) is retelling a funny encounter he had with a staff member of the university at the beginning of his undergraduate program, who was wearing five FiveFinger shoes ("gants de pieds") at his office. These shoes, which are typically found in outdoor activities, look very unusual and rather ludicrous, as they are designed in a way that shows all individual toes. What is interesting about this excerpt is the way Paul takes on different viewpoints through his gaze behavior and body movement to

make the scene quite dramatic and very humorous. He begins by setting the scene where the event took place by describing where he was sitting and how he came to see these funny looking shoes (l 17-27). He then acts out the entire encounter as if it was a play, by re-enacting, and re-voicing (Goffman, 1981) the dialogue between himself and the "protagonist" of his story (l. 22-27), and personifying the characters' actions and movements (i.e. walking down the hall). This is what Goffman (1981) calls "animation"; when a speaker animates themselves or someone else as a character. Paul's funny story is in fact a success, as demonstrated by Paula's (B2) positive assessment at the end of this excerpt, l. 20 "c'est tellement drôle" which marks the completion of Paul's narrative task. He does not only retell a personal anecdote, but adds dramatic elements, reporting a *problematic* event (Ochs & Capps, 2001), i.e. the focal point of his narration, which presents an out-of-the ordinary circumstance (staff wearing funny-looking shoes). While there are many fluencemes and gestures in this excerpt, we will focus on four particular moments in the interaction (marked by the arrows in the transcription) that are of interest for the present section.

1 *PAUL: hhh. eum le [/] la [/] la personne avec qui j'ai dû eum euh faire des bras de fer pour être admis à Paris 3 en L1.

→ **2 *PAUL: euh c'était un secrétaire alors c'était l'année où euh +//.**
((gazes away))

3 *PAUL: le saviez vous (0.286)?
****

((gazes towards Paula, with index finger from his right hand oriented towards Paula))

4 *PAULA : (laughs)

5 *PAUL: l'année dernière ils ont viré tous les (laughs) [/] tous les secrétaires

6 *PAUL: nous a dit u [//] une prof de linguistique euh au [/] au premier semestre.

((lines omitted from the transcription))

14 *PAUL: donc du coup ce mec qui malheureusement a été viré après **parce-qu'il était génial s'appelait XX.**

→ **15 *PAUL: [!] (0.400) salut.**

****

((looks down, index finger from his right hand pointed upwards))

16 *PAULA: ((laughs))

→ **17 *PAUL: euh il euh [/] il se baladait avec des eum [/] des euh je sais pas si tu vois les [/] le:es [//] des gants de pieds.**

****

((left POH facing down, extended opposite him, gazes towards his hand))

19 *PAUL: +< c'est des chaussures avec **le:es euh voilà**

****

((extends index finger to draw a series of circles in the air while moving to the left; smiles & gazes at Laura))

avec les orteils.

18 *PAULA: +< ah oui je vois très bien c'est celles avec les orteils.

20 *PAUL: et il avait des [//] ces chaussures-là.

21 *PAUL: alors j (0.430) au début je l'avais vu que à son bureau donc il était à son bureau etc donc je voyais son [/] son buste comme ça et on parlait et tout.

22 *PAUL: et un jour il est [//] il m'a dit

23 *PAUL: +/ oh ben accompagnez moi dans le couloir euh machin je vais faire un truc et tout on va parler "/.

→ **24 *PAUL: et il s'est levé il a fait le tour de son bureau**
et j'ai vu qu'il avait
de:es [//] les [/] les ga:ants de pieds là un espèce de plastique xx.
****

25 *PAULA: ((laughs))

25 *PAUL: (laughs) genre ah (laughs) garder ma dignité surtout.

26 *PAUL: (laughs) et du coup j'ai continué à parler

27 *PAUL: +/ ah oui bien sur mm oui très bien etc mmm t'as des gants de pieds c'est trop bizarre (laughs) "/.

28 *PAULA: c'est tellement drôle.

Paul first introduces the protagonist of his story as "la personne avec qui j'ai dû eum euh faire des bras de fer pour être admis à Paris 3" (l.1), but before describing him in detail, he first digresses from his initial story to offer a contextual frame, i.e. the background of the event. We can thus observe an interruption within his discourse (l.2) at different levels, i.e., at the *narrative* level, as he abruptly changes the course of his story; at the *syntactic* level, as he interrupts the delivery of his verbal utterance (l. 3) with a complex fluenceme sequence made of a filled pause and a self-interruption; and at the *interactional* level, as he re-shapes the trajectory of his current action by addressing his interlocutor and an imaginary audience ("le saviez-vous"?, l. 3). This "catchline", *le saviez-vous* (*Did you know*?) further contributes to the humorous and dramatic dimension of his story. It is a fixed expression which is often found in Trivia

articles[155] advertising jingles, or showcases, and introduces well-known world facts, "mind-blowing" or "fun" facts to a large audience. By uttering this very specific expression, Paul takes on an entirely different discourse identity, by playing the role of a television host at a talk show, addressing an imaginary audience (which includes Paula). This interactional strategy is further reflected in his visual-gestural behavior, as he initiates an interactive gesture immediately after interrupting his verbal utterance, with his index finger slightly oriented towards Paula. This gesture, even though it shares close formational characteristics with deictic gestures (because of the pointing) is considered *interactive* in this case, and belongs to the subcategory of "delivery" gestures, according to Bavelas et al, (1995), i.e. gestures used to "hand over information relevant to his or her main point" (Bavelas et al., 1995, p. 397). The stroke and retraction of the gesture synchronizes with a pause of 286 milliseconds produced in utterance-final position. In sum, the interruption (l.3) or the pause (l.4) are, once more, merely a sign of "disfluency" per se, but constitute relevant interactional actions which contribute to the humorous dimension of his narrative.

In line 15, we find another instance of interactional fluency, marked by the pause and the tongue click. After describing how a lot of staff members were fired (according to Paul) that year (omitted from the transcription), Paul re-introduces the protagonist (who was also reportedly laid off from the university[156]) who plays a major part in his narrative. Once again, Paul is not only retelling an anecdote, he is dramatizing it, by addressing the protagonist directly ("salut") with a solemn tone, as if he was paying a tribute. He produces a second pragmatic gesture with the same hand configuration as before (index finger extended), but this time he is looking down, and raising his hand and finger upward. This gesture has already been documented as part of a repertoire of German recurrent gestures (see Bressem & Müller, 2014, p. 1583) and is said to draw attention of other participants to particular important topics. Once again, this gesture[157] is manifested during a fluenceme sequence, which further reflects the multimodal and pragmatic dimension of inter-(dis)fluency. In addition, the fluenceme sequence marks a shift in the current status of the co-participant, who does not appear to be Paula anymore, but the staff member from the story, hence reflecting

[155] Note a similar expression in English, with for example the headline of this article (https://www.rd.com/list/did-you-know-facts-most-people-dont-know/) *Did-you-know Facts That Are Almost Hard to Believe* (last retrieved on August 26th 2021).
[156] It should be noted that this information is purely based on his interpretation of what his teacher allegedly reported, not on actual fact.
[157] A similar gesture is also found in the *Board Games* example from Pair A in Chapter 2, section III.3.2.

a complex participation framework involving multiple participants (Goodwin & Goodwin, 2004).

Now that Paul has rightly introduced the main character of his story, he shifts back to his status of conversational co-participant in order to build meaning around the funny looking shoes, by extending his left palm open hand downwards, and rapidly wiggling his fingers, as to reproduce the wiggling movement of toes (l. 17). This gestural activity is synchronized with another complex fluenceme sequence, made of multiple fluencemes (i.e. a filled pause, an explicit editing phrase, a repetition, a syllable prolongation, and a syntactic repair). Once more, Paul mobilizes a combination of vocal and visual-gestural actions to establish meaning. By doing so, he manages to make visibly available to his partner the iconic and comical aspect of these atypical shoes, and further relies on common ground (Clark, 1996). These shoes are not very common, so Paul needs to make sure that Paula clearly understands what he is referring to; after placing the referent in the gestural space, Paul further elaborates on the iconic properties of the shoes by drawing a series of circles in the air with his index finger (l. 17), while moving to the left. Paula then takes a look at his gesture, and confirms her understanding ("ah oui je vois très bien c'est celles avec les orteils", l. 18). Another instance of *dialogic syntax* is found here (Du Bois, 2007; cf Chap. 2, section II, 2.2.3.) where the two participants build their utterances in concert with each other. A state of mutual understanding is then achieved, which enables Paul to resume his ongoing narrative activity, which he does immediately after in his subsequent turn (from lines 20 to 27). A similar gesture is then repeated in line 24 as he re-introduces the shoes in his multimodal discourse (l.24), but this time the gesture was produced closer to his body in his personal gesture space, using both palm-open hands flipping over upwards and downwards, followed by a wiggling of his fingers. These gestures share a similar semantic core with the one previously deployed in line 17, mainly the hand configuration, (open palm), and the movement (wiggling of fingers). Perhaps the repetition of this gesture can be interpreted as a sort of *running joke*, which further reinforces the comical and amusing dimension of his story, which turns out to be very successfully done.

In sum, this excerpt has demonstrated how Paul positioned himself as an entertaining storyteller, who used a multiplicity of resources other than talk alone, which were central to his ability to construct a humorous story and engage with his conversational partner. Once more, this further reflects the close relationship between

fluency and gesture, as well as the embodied nature of inter-(dis)fluency and the different pragmatic roles fluencemes may play in interaction. We shall now move to our second excerpt, taken from Pair D.

Excerpt 2.3. Louis Garrel

In the following extract, it is Jenny who initiates the retelling of an amusing narrative, but unlike excerpt 2.1., it is not the story itself that will be of interest to us, but the story *initiation* (Lerner, 1992) i.e., the story preface (Sacks, 1972) leading up to the telling of the actual event. In the previous example, Paul provided some context about the university before retelling the funny anecdote (story preface), and overall, the progressivity of his story was maintained throughout the course of the retelling, without being interrupted by Paula, who closely attended to his talk, displaying tokens of understanding and appreciation. In the following example, however, we will demonstrate how the progressivity of the storytelling becomes immediately disrupted following the story initiation, because of problems in shared understandings. Just like Excerpt 2.2, three particular moments in the interaction are identified for our analyses of inter-(dis)fluency.

1 *JEN: hhh. ma:ais euh la dernière fois j'étais au café (il) y'a une pote (laughs) qui me raconte qu'elle avait une pote (laughs) complètement bourrée qui était dans un bar.

2 *JEN: et elle était en fait euh assise en face de louis garrel.

~ ~~ ****************************

((extends her left arm and hand towards ALX))

((ALX gives an astonishing look))

3 *JEN: hhh. (0.425) parce-qu'il était avec euh des potes e:et voilà donc bref xx xx.

4 *ALX: +< comme nous et le type de les choristes.

****

((ALX extends her POH towards JEN; JEN's hands return to rest position))

→ **(2.220)**

((JEN first looks left, then leans her head forward))

*ALX: mais t'étais pas là ?

*JEN: (0.470) on le connait lui ? (laughs)

*ALX: le type de les choristes Jean Baptiste Maunier au XX.

*JEN: oui [/] oui [/] oui.

*JEN: ah il était au XX ?

((mouth open, raised eyebrows, expression of surprise))

*ALX: +< ben pareil.

*ALX: mais oui mais [/] mais t'étais là quand +//.

*ALX: nan t'étais pas là?

*JEN: +< nan ?

*ALX: nan t'étais pas là?

*JEN: nan j'étais pas là.

*ALX: bon ben il était là

*ALX: bon ben vas y.

*JEN: moi j'ai juste vu les jumeaux machin comment ils s'appellent?

*ALX: +< c'est qui les jumeaux?

→ ***JEN: tu sais les jumeaux horribles là plein de chirurgie esthétique.**

****

((winces, brings both her hands to her forehead, and with her index fingers move down to her chin))

21 *ALX: (0.507) les frères Bodganoff?
((frowns during the pause, then raises her eyebrows))

22 *JEN: ouais voilà et eux ils étaient au XX!
****
((points towards JEN with her little finger from her left hand))
23 *ALX: au XX?
24 *JEN: mais c'est nul genre mais ça n'a aucun intérêt genre alors que pour le coup euh (0.567) lui il est vraiment +/.
((lines ommitted from the transcription))

→ **35 *JEN: (0.742) [!] et euh qu'est-ce que je voulais dire**
((gazes away))
oui donc l'anecdote marrante donc euh bref elle était en face de Louis Garrel. ((slighly snaps her finger and points towards ALX))
36 *ALX: +< ouais Louis Garrel
((POH oriented towards JEN))

Jenny first projects the beginning of her storytelling with turn-initial fluencemes (an inbreath and a prolongation, l.1) and further provides some background about the humorous event. It is important to note that the story is in fact not her own personal narrative (unlike Paul), but a friend of a friend's, so she has not experienced it herself. Just like Paul, Jenny first describes the location and setting where the event took place (at a bar, where the girl in the story was sitting opposite Louis Garrel) through verbal and gestural depictions[158]. However, the progressivity of her story initiation is immediately troubled by Alex's facial reaction (l. 2) who gives an astonishing look after hearing Louis Garrel's name. Louis Garrel is a famous French actor, who is not the kind of person one would casually meet at a bar, which explains Alex's reaction.

As the transcription shows, Alex's visible response to the mention of Louis Garrel momentarily disrupts the progressivity of Jenny's narrative, who delays her upcoming turn (l.3) with a fluenceme sequence made of an inbreath and unfilled

[158] Note that the gestural activity found at the beginning of this excerpt (l.2) is further analyzed in Chapter 5.

pause. She then quickly elaborates on the reason why he may be there, but brushes it aside with a sequence-closing expression "bon bref voilà", as this part of the story, it would seem, is not deemed relevant to her story-in-progress. However, the mention of a French celebrity becomes a relevant topic of conversation for Alex, who retells their supposedly shared experience, in line 4 "comme nous et le mec de les choristes" by extending her left palm up open hand towards Jenny. A long silence of two seconds follows, during which Alex first looks to her left, then leans forward, as to display trouble in understanding. A series of question-answer sequences are then co-produced within the exchange (from lines 5 to 17) to help them clarify the misunderstanding: Alex is talking about a place (transcribed as *XX*) that both her and Jenny went to in their past common experience, and she seems convinced that Jenny was there when she saw another famous French actor, Jean Baptiste Maunier; while Alex understands who she is referring to (marked by her repetition of the agreement marker "oui", l. 8), she in fact did not share this experience with Alex. When the source of trouble is finally *repaired* (Schegloff, 1991), Alex invites Jenny to resume her storytelling activity ("bon ben vas y", l.17). However, the latter decides to retell another experience of her own, which took place at the same place, but that she did not share with Alex; but she also experiences trouble of memory, as she cannot remember the names of the people she saw, and refers to them as "les jumeaux" (l.18) This decontextualized simple noun phrase does not provide enough information for Alex to understand whom she is alluding to, so she asks for clarification (l. 19). Jenny then inserts an additional prepositional phrase "plein de chirurgie esthétique", along with a visual-gestural depiction which further describes the atypical properties of these individuals: she winces, brings both her hands to her forehead, and with her index fingers move down to her chin. After a pause, during which Alex is frowning, she finally understands that Jenny is talking about the Bogdanoff Brothers, who are French television presenters known for their excessive plastic surgery. This second misunderstanding is hence repaired, which gives Alex another opportunity to talk about other celebrities she has met at that place (omitted from the transcription). Jenny's narrative is eventually re-launched in line 35, initiated by a fluenceme sequence made of an unfilled pause, a tongue click, a filled pause and an explicit editing phrase ("qu'est ce que je voulais dire") during which she looks away, her hand resting under her chin (cf Fig. below), displaying another instance of a *thinking face* (cf Chap.5). After retrieving the initial topic (the funny encounter with Louis Garrel) from memory, Jenny thus displays a

shift from her solitary searching activity to a joint production with Alex, as they both remember in tandem (illustrated in Fig. 65 below).

Figure 65. *Shift in Participation*

In sum, this excerpt has shown the different ways through which Alex and Jenny have been trying to map their common experience, based on their knowledge of French celebrities and their physical attributes, deeply rooted in film French culture. This provided opportunities for them to address trouble in understanding, request for clarification, and display mutual orientation and intersubjectivity. Both speakers performed these relevant activities by deploying a number of resources among their relevant scope of behavior, mainly, talk, fluencemes, facial expressions, and gestures, which enabled them to co-construct the continuation of the dialogue.

To conclude, these analyses have illustrated the emergence of fluencemes in larger interactional contexts, where they functioned as byproducts of systematic interactional practices, i.e. humorous personal narratives (Excerpt 2.2.) and displays of shared (mis)understandings to co-construct meaning (Excerpt 2.3.). In the first excerpt, Paul skillfully made use of fluencemes and gestures to pursue the delivery of his humorous narrative quite continuously, and establish a referent by combining vocal and gestural strategies. In the second one, however, fluencemes mainly emerged in contexts of trouble in shared understanding, hence disrupting the progressivity, and to a larger extent, *fluency*, of the storytelling sequence. However, several fluencemes were also used to preface and resume Jenny's storytelling activity, which further illustrates their potential to mark the continuity as well as the DIScontinuity of discourse.

2.2.4. The interplay of vocal and material resources in the course of class presentations

We shall now conclude this section with two short examples from the class presentations which reflect the ways the students dealt with the different material resources they had within reach, to pursue the delivery of their oral assignment. The excerpts are taken from Paul (B1) and Linda (F1).

Excerpt 1.3. Paul's presentation

In this excerpt, Paul is analyzing the rhyming and rhythmic patterns of a poem, by describing specific instances of assonance (the resemblance in sounds of words or syllables between their vowels or consonants).

1 (0.510) eum (0.617) et là ce qui est assez intéressant à noter c'est
((gazes towards notes; brings his hands to his cheeks))
les euh [/] les assonances en début de mot
((gazes towards audience))

→ **2 euh qui se reprennent vraiment d'un vers sur l'autre.**
((gazes towards notes and slightly pushes his book aside))

3 (0.445) euh entre enflammer et entreprise
~~~~~~~~~~~~~~~~~******~~~**********
((gazes towards audience))**

4 (0.589) débriser dessein
**** *****

((produces quick beat movements with his pen while rotating his wrist and slightly moving it on the table))
5 (0.468) ensuite ça s'arrête mais une fois qu'on est partis on peut
****
((performing rotating motion with his right hand while holding the pen))
difficilement s'empêcher de rapprocher ruiner et espérance
**** ******
((produces beat movements with a flick of his wrist))
6 et mettre en pièce et finesse
********** *******
→ 7 qui là sont euh [/] euh sont euh liées avec la:a [/] la rime.
*****-.-.-.-.-.-.-.-.-.-.-.-.-.-**
((gazes towards paper, RH returns to rest position + writes something))

As the multimodal transcription shows, Paul re-arranges the different objects placed opposite him, by grabbing his pen with his dominant hand (right hand), and slightly pushing his book to his left side, with his left hand (l.2). Then, he produces a pause of 445 milliseconds clustered with a filled pause ("euh") of 255 ms; and while his left hand is resting on the book, he moves his dominant hand in preparation while holding the pen and raising it in the air. Once more, the preparation of this gesture is synchronized with the fluenceme sequence, illustrating an instance of discourse preparation at the multimodal level (preparing speech and gesture). Paul then deploys a sequence of gestures (from lines 4 to 6), characterized by a series of beat movements and flicks of the wrist, still carried out with his dominant hand holding the pen in the air, as he moves across the table in different positions. Each time, the stroke of the gesture is synchronized with specific key words from the poem ("enflammer" "entreprise", "débriser", "dessein" "mettre en pièce", and "finesse"), as he places them in his discourse space, i.e. the desk, marking different components of his verbal discourse. This series of manual operations on relevant objects i.e., the sheet of paper

and the book, which are part of Paul's embodied environment (the classroom), closely resembles Goodwin's (2007) example of a homework activity between a father and his daughter, where answers to a question are built through:

> the simultaneous use of structurally different kinds of semiotic practices (language, gesture, and the structure of the page being worked with) in different media which mutually elaborate each other (Goodwin, 2007, p. 55).

Similarly, Paul presents a number of examples from the book illustrating instances of assonance in the poem through speech and gesture, by deploying a series of manual actions directly on his book and notes. The latter, which can be considered as deictic-anaphoric gestures in this context[159], further offer discourse cohesion, as the same gestural patterns (beat movements and flicks of the wrist) were initiated in the same specific spatial area (the desk) and were repeatedly associated with specific referential expressions found in the text (Levy & McNeill, 1992). Paul also draws a parallel between the paired lexical terms, by making use of space; for instance, "enflammer" is produced on the left side of his notes, while "entreprise" is produced on the right side, with the pen raised in the air. In addition, he is also seen writing something on his notes (l.7), during which he momentarily suspends the course of his multimodal discourse activity, marked by a complex fluenceme sequence following the *VOC+MS* pattern, and a retraction of his manual gesture.

In sum, this excerpt has shown the different vocal and manual actions Paul had to mobilize in order to deal with the task at hand, alternating between (1) presenting his assignment through talk and gesture, (2) engaging with the audience, (3) reading his notes, and (4) writing on his sheets of paper. Paul managed to handle all these activities simultaneously without much disrupting the course of his presentation, but this cannot always be achieved so easily, as presentations have to be carried out quite continuously in an organized manner. This leads us to the following example, taken from Linda's (F1) presentation.

[159] Note that these gestures were coded as *deictic-anaphoric* in our quantitative analysis (cf section II. 2.1.3.) because of their use of space to place referents in discourse (cf Chap. 2, section II. 2.2.3), but they could also be regarded as *discursive* gestures to a certain extent, since they were also used to mark emphasis. Thus, both functions may overlap here, which further motivated our need to use intercoder reliability on 15% of the data (cf Chap. 2, section II.2.2.3).

Excerpt 1.4. Linda's presentation

1 hh alors (0.728) mmm pardon (1.132) je cherche la page (smiles)
(7.512)
(looks through her notes; uses her pen to look for the right page))

2 hh donc euh l'auteur nous partage (0.404) tout le long de:e [/] de cette première partie hhh. eum les [/] ses souvenirs de premier émois.
((gazes towards her notes))
3 (0.304) ses randonnées en montagne e:et ses pêches à la mer.
(1)**********(2)*******
((1. holds out her hand, palm facing up with slightly bended fingers; 2. extends her palm and fingers with a slight wrist motion))

As this brief example shows, Linda's use of her space is not as carefully organized as Paul's, given the multiplicity of objects found on her table, with about 8 different sheets of paper, her laptop, and her book. In fact, she seems to be experiencing difficulties managing these various media simultaneously, as marked by the significantly long silence found in her discourse, lasting up to seven seconds. The latter is often considered a *lapse* in the CA literature, and refers to actions whereby speakers refrain from speaking and selecting a next-to-speak (Sacks et al., 1974). It is often defined by perceived length, as they tend to be considerably long (3 seconds or more, McLaughlin & Cody, 1982). In this case, however, the setting is entirely different, as Linda cannot invite another speaker to speak, so she needs to deal with her difficulties alone. Therefore, the term "lapse" is not exactly appropriate, as its identification is strongly determined by turn-taking. In her study on academic lectures, Rendle-Short (2005) identified periods of "non-talk" i.e, when presenters stop talking and do not engage with their audience, as they attend to other *presentation-relevant* activities, such as looking through one's notes or interacting with the computer. She argued that periods of long talk were not necessarily viewed as problematic, as they marked a period during which presenters transitioned from "topic-talk" to non-talk, and from

engagement to disengagement with the audience, and this shift can be achieved through visible bodily behavior. In Linda's case, we can see her gazing towards her different sheets of paper while trying to find the right one, and she also uses her pen at some point to help her. She is thus fully oriented to this activity, and this is signaled by her bodily behavior. Note that this period of nontalk is preceded by a complex fluenceme sequence made of two explicit editing phrases ("pardon" and "je cherche la page"), a non-lexical sound ("mm"), and two unfilled pauses (of 728 and 1132 milliseconds) during which she is smiling, perhaps to save face. Here the fluenceme sequence functions both as a time-buying and a signaling tool, projecting to the audience that more time is required for Linda, who cannot find the right page from her notes. Therefore, this cessation of talk may be treated as relevant here, as it signals to the audience that the talk is momentarily being put on hold in order for Linda to pursue the delivery of her presentation, which gives her more time to accomplish her actions. After some time, she eventually manages to resume the delivery of her presentation, but she still seems absorbed in her notes, as she still does not gaze towards her audience, despite carrying out a gestural activity (l. 3). In fact, she spent 90% of her time gazing towards her notes during her presentation overall (cf Appendix 4, Table 77), which is considerably higher than the average of her group (about 70%, see Fig. 62, section 2.1.3.). This shows that, even in periods of "fluent" talk, which tend to be characterized by mutual gaze (cf Fig. 63, section 2.1.3.) she very rarely engaged with her audience, as she was too focused on her own performance.

To conclude, these two examples from the presentation-sessions have shown that the emergence of fluencemes is intricately embedded within the continuous activity of giving an oral assignment, during which speakers must find ways to deal with the content and temporality of their presentation flow, as well as their spatial and material environment. In the first excerpt, Paul successfully shifted between different presentation-relevant activities through talk, gaze, and gesture, by making use of his gesture space to place referents within his discourse in a way that was visible for his audience; in the second excerpt, however, Linda experienced difficulties managing multiple media simultaneously, and had to delay quite significantly the course of her presentation in order to attend to other actions and focus on her own performance. Once more, we can find different degrees of fluency and disfluency at several levels of analysis, mainly speech, content, gesture, and interaction. Despite being a "monologic" task, where audience participation is not expected, the presenters are still delivering a

presentation *to* them, and must therefore display mutual orientation and engagement, while still being able to talk quite continuously without too many interruptions. This may present a number of challenges for the students, which could potentially explain the high rate of fluencemes found in their presentations. This is further discussed in the next section.

III. Discussion

The present section addresses our research questions formulated earlier (section I.1.4), by drawing on a selection of findings obtained from our statistical treatments (section II.2.1.) and our multimodal qualitative analyses (section II.2.2.). The section is structured as follows: we first report on differences in distribution between class presentations and conversations, and discuss the effect of style and setting on fluency and gesture (3.1.); then stress the importance of *audience design* across the two situations (3.2.), and conclude on the situatedness of discourse (3.3.) and its implications for the multifunctionality and multimodality of inter-(dis)fluency.

3.1. Effect of style and setting on fluency and gesture

One of the main questions this chapter sought to answer was whether style and setting had an effect on (dis)fluency and gesture production (RQ1), and whether significant differences would be found across the two situations (RQ2). As we have seen (cf section I), differences in (dis)fluency or gesture behavior cannot solely be measured by isolating one factor, such as task complexity, mode of delivery, or degree of preparation, which has often been done in the literature (except for Bortfeld et al., 2001, among others). The present study does not intend to be restrictive, and thus presents style as multidimensional, in line with Eskénazi (1993).

3.1.1. Beyond the degree of preparation or mode of delivery: multi-dimensional analysis of language style

As presented earlier, a number of studies in the field of psycholinguistics and (dis)fluency research have insisted on the *spontaneous* nature of fluencemes, and the fact that they are systematically produced in spontaneous productions, as opposed to carefully read speech (e.g. Goldman-Eisler, 1958; Silverman et al., 1992 ; Shriberg, 1994). However, most of these studies carried out elicitation experiments to obtain such findings, which offers little ecological validity, and hence does not truly reflect

the use of fluencemes in situated discourse (see section 3.3.). It would appear that in the case of DisReg, the role of preparation did not affect fluencemes positively, since students produced significantly more fluencemes in class presentations ("prepared" speech) than conversations ("spontaneous" speech). Even though the participants were extensively reading their notes, and had prepared their assignment at home, it did not stop them from producing a high number of fluencemes, in the exception of one participant, F2, which will be further discussed in section 3.2.2. A multiplicity of other factors thus needs to be taken into account to interpret such differences in fluency behavior.

Another body of research in corpus-based linguistics and Second Language Acquisition have compared (dis)fluency rates in monologic versus dialogic situations, sampling from different types of speakers and data types, resulting in contradictory results. For instance, Schatcher et al., (1991) found higher rates of fluencemes in interviews than in lectures, and Duez (1982) found that pauses were more frequent in political and casual interviews than political speeches. By contrast, Tottie (2016) found more pauses during narrative turns than in conversations between relatives and friends, and Michel et al., (2007) found fewer filled pauses in dyadic phone conversations than in messages left on a recording machine. As we have seen, other variables, such as topic familiarity and utterance complexity, may also come into play. For example, Bortfeld et al., (2001) showed that (dis)fluency rates were associated with heavier planning demands, in line with Beattie (1979), Oviatt (1995), and Shriberg (1994) who found more (dis)fluencies in longer utterances. Bortfeld et al., (2001) and Merlo & Mansur (2004) also found an increase in (dis)fluency rates when speakers discussed unfamiliar and abstract topics.

The present study is situated within this existing body of research, but aims to offer a more comprehensive approach to inter-(dis)fluency by including the notion of language style and multimodal communication setting. In line with Eskénazi (1993), we take on a multi-dimensional approach to style, further grounded in sociolinguistics, gesture studies, and Conversation Analysis, by taking into account a number of inter-related factors, such as audience design (cf section 3.3.), register, turn-taking mechanisms, and the overall embodied material environment. We also address a gap in the literature regarding the study of gesture with respect to setting, which has received little attention, except for Bavelas et al's (2008) study that compared the rates of gestures in dialogues and monologues. In sum, we believe that mode of delivery or

task complexity alone are not sufficient to interpret differences in fluency and gesturing behavior across situations. Our corpus study has revealed quantitative and qualitative differences in fluenceme use, thanks to the multiple formal and functional variables included in our analysis which, we believe, has shed some light on the interplay of factors impacting inter-(dis)fluency.

3.1.2. Fluenceme rate, distribution, and patterns of co-occurrence across the two situations

Overall, significant differences were found in the distribution of fluencemes in the two situations. In sum, a higher rate of fluencemes was found in class presentations, with significantly longer unfilled pauses, and a higher proportion of non-lexical sounds, as well as filled pauses. The latter were also more often realized with the nasal variant (*eum*) in class presentations than conversations, suggesting a longer delay (Clark & Fox Tree, 2002). In addition, a slightly higher proportion of complex fluencemes was found in class, but without differences in length. So far, these findings suggest that class presentations require more time for planning and monitoring, which is consistent with Tottie (2016) who found that filled pauses were closely associated with planning demands, as they tended to occur more frequently in utterance-initial position in long narratives or thoughtful presentations of evidence, which require more planning than casual conversations. This is also confirmed by the positive correlation found between unfilled pause duration and utterance length, with longer pauses associated with longer utterances. The latter were significantly longer in class presentations than conversations. A tendency for fluenceme sequences to occur in utterance-initial position was also found in presentations, which could reflect a possible rhythmic and stylistic style, in line with Duez (1982). Indeed, class presentations require students to produce clear intelligible utterances and pay close attention to their speech, in order to present structured arguments with a careful choice of words. However, we also mentioned the role of anxiety and self-consciousness in formal situations with higher stakes (graded assignments in the case of DisReg), which resonates with several previous studies (Broen & Siegel, 1972; Christenfeld & Creager, 1996; Tottie, 2014). Although it is not possible to find a direct connection between anxiety and (dis)fluency production, the effect of stress, or at least self-consciousness could be a possibility, given the amount of time students spent gazing towards their notes and not engaging with their audience, suggesting that they

were paying more attention to their own production than to their group of interlocutors. This is further discussed in section 3.2.

In addition, let us not overlook the weight of individual differences in the data, as many were found across the two situations. For instance, F2 is the only speaker who actually produced fewer fluencemes during his presentation (16.5 phw) than in the conversation (21.4 phw), and D1 produced about equally the same amount in the two situations (20.2 and 19.9). Others, like B2, produced considerably more fluencemes during the presentation (25.6) than when she was engaged in the interaction (9.7). In fact, she produced relatively few fluencemes during the conversation in comparison to the average of her group (20.4). Similarly, a group of speakers produced on average filled pauses of longer duration during their presentation than in the conversation (e.g. A1, D1, E1, E2, see Table 69, Appendix 4), while others showed the opposite tendency (e.g. D2 and F2). These findings are consistent with our study of the SITAF Corpus, where many individual differences were found across speakers within the two language groups, which further confirms that fluency is in part dependent on personal speaking style, regardless of setting or proficiency, in line with De Jong (2016a). These individual differences were also illustrated in the gesturing behavior of speakers, which leads us to the next section.

3.1.3. Gestural distribution and gaze behavior

Contrary to what we expected (cf H5), a higher rate of gestures was found in class presentations than in the conversations, which is not consistent with Bavelas et al's., (2008) findings that gestures were more frequent in dialogues than monologues. But once more, these differences can be explained by the type of experimental procedure used in their study, which relied on a picture elicitation experiment to create the monologue and dialogue conditions. This is very different from our study, which is based on semi-naturalistic settings (cf Chap. 2, section I.1.1.), so specific attention needs to be paid to the ecology of these gestures. In addition, our supposedly "dialogue" situation is also quite dialogic in a sense, since the students were presenting their assignment to the group of students facing them. As our quantitative findings revealed, a majority of the gestures produced during the presentations served discursive functions, i.e. they were used to mark emphasis, present an idea, or structure aspects of multimodal discourse, and the latter were used significantly more during class presentations than conversations. Conversations, on the other hand,

comprised a higher proportion of interactional and representational gestures. This was also found in Bavelas et al's (2008) study, who reported a similar result. In the conversations, representational gestures were also found to occur more frequently during fluencemes (28% of all gestures during fluencemes) than without them (14% of all gestures outside fluencemes), while interactive gestures occurred more frequently outside fluencemes (43%) than during them (27%). Thinking gestures, on the other hand, almost occurred only exclusively during fluencemes in the two situations, which is consistent with our findings from the SITAF Corpus (Chap. 3). A more detailed analysis of thinking gestures is provided in Chapter 5.

As our qualitative analyses further demonstrated, speakers often made use of interactional gestures to perform a series of actions, such as establishing common ground, displaying a stance, or addressing their interlocutor in the course of their interactive practices. During their oral presentations, however, students almost never addressed their audience, except for a few exceptions (cf *Excerpt 1.1.* in section 2.2.1.), but mostly made use of gestures to segment discourse and mark information structure. In particular, we described one example (excerpt 1.3.) in which the presenter made use of his gesture space to emphasize a selection of words, and each change of gesture coincided with a key word. Once again, this further reflects the effect of situation on the functions of the gestures. In addition, a considerable proportion of gazing towards the piece of paper was found in the presentation-sessions, amounting to 70% on average, as opposed to only 27% of gazing towards the interlocutor, which is a significant difference with the conversations. This finding seems somewhat at odds with the considerable number of gestures found in class presentations (10.2 phw as opposed to 7.6 in the conversations), which means that, even though the students were engaged in a relatively dense gestural activity, they did not truly engage with their audience, as they were too engrossed in their notes. In addition, findings further showed that speakers were more likely *not* to establish eye contact when they produced fluencemes in both situations, which is consistent with our findings from the SITAF Corpus where the two language groups were found to withdraw their gaze more frequently during fluencemes both in their L1 and L2 (see Chap. 3, section II. 2.1.3.). This further emphasizes the fact that gazing away is a very common practice of (dis)fluency, regardless of language or setting, as it enables speakers to momentarily retreat from the current activity to attend to other relevant ones, such as retrieving an item from memory, looking for a specific word, checking for a sentence in a book, etc.

However, we also showed several instances of mutual gaze coordinated with fluencemes in the conversations (see sections II. 2.1.3. and 2.2.3.), during which speakers were engaged in interactive practices, which further reveals the potential for fluencemes to embody interactive processes as well, and not only intrapersonal ones. In sum, fluencemes enable speakers to transition between periods of planning and monitoring with periods of engaging and conversing with co-participants, and this transition can further be manifested in accompanying visual-gestural behavior; in the case of class presentations however, it would appear that students were primarily focused on their own performance, and this was further confirmed by the overwhelming proportion of fluencemes performing *own communication management* (99%), as well as the high instances of gazing towards the notes.

3.2. The importance of audience design

This section discusses the importance of *audience* or *recipient design* with respect to language style and setting, in line with the frameworks of Conversation Analysis and sociolinguistics. Unlike previous studies from Bavelas et al., (2008) or Michel et al., (2007) the "monologic" productions from the DisReg Corpus were elicited in front of an actual *real* audience (not an experimenter alone), and this has a number of consequences on the participants' behavior. As Bell further (1984) claimed, style is essentially the reflection of speakers' response to their audience, and "non-audience" factors, such as setting or topic also derive their effect with respect to the type of addressee(s) found in a certain situation (Bell, 1984, p. 145). This notion is truly relevant to the present study of DisReg, as the two settings comprise an entirely different set of co-participants.

3.2.1. Discourse identities within complex participation frameworks

On the one hand, class presentations involve one presenter (or two) and an audience of passive hearers, whose relationship is highly asymmetric, but also fixed and predetermined. Indeed, throughout their presentation, students are acting as *presenters*, and their status is not expected to change until the end of their talk. The audience, too, is bound to remain "passive"[160] at all times, and is not invited to speak,

[160] However, it should be noted that the audience probably displays tokens of participation as well, without necessarily speaking, through head nods, mutual gaze, etc. Unfortunately, this type of evidence cannot be verified, since they could not be videorecorded as part of the study.

except eventually at the end of the student's presentation[161]. In conversations, on the other hand, the participants' status and discourse identities are continuously being (re)shaped in the course of the interaction, which invites them to take on multiple identities, overlapping with one another. This was especially illustrated in Example 2.2 where Paul (B1) skillfully alternated between different roles in the course of his storytelling activity, from the role of entertaining storyteller to television host, in order to build a humorous narrative. His interlocutor, Paula (B2) was shown to attentively attend to his talk without interrupting him, and by displaying tokens of appreciation and understanding. In another example, however, (Excerpt 2.3) the hearer, Alex (D1), played a bigger role in the storytelling activity performed by Jenny (D2), the speaker, by initiating another topic of conversation involving their common experience. The latter resulted in a misunderstanding, which momentarily disrupted the progressivity of the storytelling sequence, but which also provided an opportunity for participants to share common ground and display mutual understanding.

3.2.2. Class presentations and the presenters' orientation to their talk

In institutional settings, the main discourse objectives and orientations of the speakers differ radically from casual conversations (cf section I. 1.1.4) where not much is at stake, excerpt for the need to maintain continuity and alignment between co-participants. In comparison, during class presentations, the main goal of the presenter is to deliver a successful assignment for which they will later receive a grade, and this success is mostly conditioned by their ability to provide well-constructed analyses in a given topic. While most of this work is carefully prepared at home, the challenge remains for students to give a real time performance in front of the whole classroom, and this compels them to constantly work on their own production as the presentation unfolds. As we have seen, a majority of students were fully oriented to their own talk, as indicated by the high rate of fluencemes used to manage their own communication (OCM) and the number of gestures oriented towards their own discourse. We may wonder whether fluencemes were used so frequently in this context as a result from their overburdened utterances as well as their inability to manage their online planning and engage with their audience at the same time, following the *Cognitive Burden* view of (dis)fluency (cf Chap. 1, section II.2.2.1.). However, as maintained

[161] Based on our experience, this is true for most French presentations at an undergraduate level in French universities, but this claim is not verified by actual evidence.

multiple times throughout this thesis, we do not believe that (dis)fluency should be restricted to episodes of difficulty or trouble, as they can also function as relevant *collateral* signals (Clark & Fox Tree, 2002). Note for instance the fluency behavior of a particular speaker in the following excerpt, taken from the presentation delivered by F2, pseudonymized as Matt:

Excerpt 1.5.a

1 (0.670) par cette réforme Cléon est accusé de démagogie (0.400) par ces detracteurs.
2 hhh. puisqu'il instaure un système de corruption en prétendant soutenir le peuple.
3 hhh. et en espérant surtout le soutien de celui-ci.
4 (0.876) il faut dire que Cléon étant le successeur de Periclès
5 (0.463) très populaire et très considéré.
6 (0.404) Périclès pas Cléon.
7 il est notamment celui qui voulut le Parthénon.
8 (0.760) temple en hommage d'Athenas patronne de la cité.

As pointed out earlier, Matt is the only speaker in the data who actually produced more fluencemes in conversation than in class, as he only produced 16.5 fluencemes per hundred words during his presentation, which is significantly lower than the average of his group (28.5 phw). This is clearly illustrated in this brief excerpt of 24 seconds, during which he only produced simple fluencemes (inbreaths and unfilled pauses) mostly in utterance-initial position. His speech is loud and clear, very articulated, and he skillfully makes use of pauses to mark discourse boundaries, reflecting a stylistic function. In sum, Matt represents the ideal *fluent* speaker, who seems very good at public speaking, and whose voice is very pleasant to hear. However, over the course of these 24 seconds, Matt does not produce a single gesture, and his eyes remain fixed on the sheets of paper he is holding with both hands, giving little room for any kind of gestural activity, as illustrated in the figure below[162].

[162] In fact, he only produced 11 gestures during his presentation.

Figure 66. *Matt's (F2) visible bodily behavior during his presentation (Excerpt 1.5.a)*

Even though Matt sounds perfectly fluent *to the ear*, his visible bodily behavior conveys a total lack of communicativeness, as he looks like he is rehearsing his speech alone, or recording himself on the microphone, without paying any attention to his surroundings. Compare now with this second brief excerpt, also taken from Matt's presentation, a minute or so later:

Excerpt 1.5.b

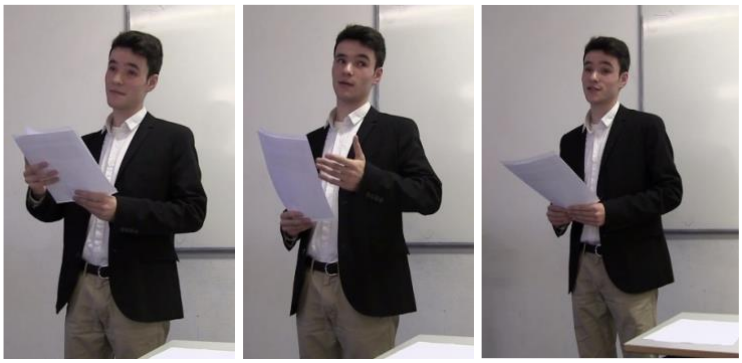
Matt is analyzing a quote from the book he is presenting[163], which includes a list of several terms conveying the notion of cupidity (*Grigou*, *radin*, *lésine* and *Harpagon*, not included in the transcription), but he is having doubts about the use of the term "Harpagon". This excerpt starts from there.

1 **(0.380)** avec un doute sur Harpagon parce-que:e (0.554)
((smiles, gazes towards audience, A.))
Victor-Henri Debidour connait le personnage de Molière.
2 mai:ais j'ai pas de connaissances d'un personnage **euh** Harpago:on
((quickly extends his right open hand sideways, B.))
euh du cinquième siècle avant notre ère.
3 donc eum (0.497) peut-être que:e (0.503) dans la version grecque
euh il existe une euh (0.662) [!] (0.361) [/] une euh +//.
5 Aristophane a:a aussi procédé à une création d'une telle ampleur
ma:ais là j'ai pas de version grecque pour m'en rendre compte.
((raised eyebrows + shoulder shrug C.))

It is interesting to note a total change in his behavior in comparison to the previous excerpt. Here Matt deploys a high number of fluencemes to spontaneously elaborate

[163] His presentation is about Sophocle's *Œdipe Roi* and its translation by Victor-Henri Debidour.

on his interpretation of the use of the term *Harpagon* in the text. He is not reading from his notes anymore, and acting as the skilled oral performer, but rather speaks with *his own voice*, giving his own personal opinion, as he displays a series of communicative and expressive behaviors (smile, shoulder shrug, gaze towards the audience, manual gesture etc.,) as illustrated in the figure below.



A. B. C.

Figure 67. *Matt's display of visible expressive behaviors (Excerpt 1.5.b)*

His utterances are filled with a number of complex fluenceme sequences occurring in medial and final positions, momentarily delaying and interrupting the course of his verbal and vocal flow, as opposed to the previous excerpt, where his pauses were carefully used for discourse marking and emphasis. But this time, Matt is actively oriented towards his audience, and only relies occasionally on his notes. His style is now much more dialogic, as it looks like he is expecting some kind of validation from his audience through his gaze and gestures.

In sum, these brief excerpts have shown that the degree of (dis)fluency should not be restricted to examples of perfectly well-formed fluent utterances as opposed to cases of highly disfluent ones, but should include the full scope of semiotic behaviors participants have at their disposal, which further reveals how they may use them to (dis)engage with their interlocutor(s). While a particular delivery produced in the speech signal may *sound* "disfluent", it may in fact *look* rather "fluent" in the visual-gestural channel. This further reflects our multi-level approach to inter-(dis)fluency (cf Chap. 1, section III), which relies on three dimensions (speech, gesture, and interaction) in order to provide a more comprehensive picture of fluency, without restricting it to one dimension or the other.

3.3. The multifunctionality and multimodality of inter-(dis)fluency in situated discourse

To conclude, our analysis of the DisReg Corpus has enabled us to identify specific multimodal social practices reflecting differences in fluency and gesturing behavior across the two settings (RQ3). In sum, fluencemes are highly sensitive to the contingencies of talk-in-interaction, and reflect different degrees of interactional fluency, depending on whether they occur at transition relevant places to yield a turn, or display disaffiliative actions in turn sequence openings (see the case of tongue clicks in examples 1.2 and 2.1). It is thus essential to examine these dynamic markers in local situated activities and explore their relation to visuo-gestural practices, whether they co-occur with a gestural activity, or are produced while withdrawing one's gaze. These visual gestural resources, as already shown in our previous study on the SITAF Corpus (cf Chap. 3), play an essential role in determining whether fluencemes are more interaction-oriented (ICM) or production-oriented (OCM), further reflecting the multifunctionality and multimodality of inter-(dis)fluency phenomena. This multimodal view of language, whereby our human abilities are projected onto the world through practical uses of our hands and bodies (Boutet, 2018; Morgenstern & Boutet, forth.; Streeck, 2009b; Streeck et al., 2011) plays an essential role in the understanding of inter-(dis)fluency, as shown in our qualitative analyses which revealed a relationship between fluencemes, gestures, actions (i.e. writing on one's notes) and manipulation of objects (fiddling with the pen). Taken together, the deployment of these different semiotic behaviors was shown to contribute the building of multimodal utterances, exploring aspects of fluency that were not visible in quantitative findings only. Once more, we strongly maintain that quantitative statistical treatments conducted on different formal and functional variables should be complemented with specific multimodal qualitative analyses of situated discourse, as to shed light onto the interplay of multimodal features affecting fluency across language styles and settings.

Conclusion to the chapter

In conclusion, the general aim of this chapter was to describe potential differences in fluency and visual-gestural behavior across two distinct styles and settings, i.e. institutional class presentations versus casual face-to-face interactions. In this chapter, we presented the notions of style and setting as *multidimensional*, encompassing a wide array of inter-related factors such as audience design, multimodal setting, turn-taking mechanisms, or register, thus going beyond differences in type of delivery (i.e. read between spontaneous) or mode (i.e. monologue versus dialogue). While a lot of research has been conducted on the different effects affecting (dis)fluency (topic, genre, task type, anxiety, register etc.,) many of these studies were designed very differently from one another, relying on different experimental procedures, sampling from different types of speakers, eliciting different types of production, hence resulting in contrasting results. Gestures, on the other hand, have received very little attention with respect to language style in corpus-based studies. The aim of our study on the DisReg Corpus was to provide an overview of fluency and gesture within their situated ecology, based on semi-naturalistic data and on the same pairs of speakers across the two situations, which allows for efficient quantitative treatments, as well as micro analyses of the data. The quantitative findings are summarized in the table below, taken from section II. 2.1.3.

As our findings revealed, a number of significant differences were found across the two situations, mainly a higher rate of fluencemes, longer unfilled pauses, and longer utterances in class presentations than the conversations, as well as more instances of utterance-initial fluencemes in class as opposed to more utterance-final ones in conversation. More gestures were also found during the class presentations, but a majority of them served discursive functions, as opposed to conversations which contained more interactive and representational gestures. In addition, a significant proportion of gazing towards paper was found during presentations, which reveals a lack of engagement towards the audience, despite dense gestural activity. The lack of visible engagement during the presentations (marked by the absence of mutual gaze) was further illustrated in our multimodal qualitative analyses, where students were shown to be mostly focused on their own performance and the material objects around

them (their notes, their book, their pen, etc.) reflecting more intrapersonal processes and pertaining to *own communication management* (OCM).

In conversations, on the other hand, several interactional practices involving embodied displays of intersubjectivity were exemplified, thus further reflecting the interactional dimension of fluencemes (*interactive communication management, ICM*).

	Class presentations	**Conversations**
VOCAL-VERBAL FEATURES FEATURES (FLUENCEMES)		
Fluenceme rate	Higher rate in class presentations than conversations	
Distribution	more NL sounds, filled pauses and longer unfilled pauses (correlated with utterance lenght)	more repetitions, interruptions and prolongations
Filled pause type	more eum-type filled pauses	more euh-type filled pauses
Combination type	more complex sequences	more simple sequences
Utterance position	more instances of utterance-initial fluencemes	More utterance-final fluencemes
VISUAL-GESTURAL FEATURES (WITHIN AND OUTSIDE FLUENCEMES)		
Gesture rate	Higher rate in class presentations than conversations	
Gesture distribution	more pragmatic gestures overall, and more discursive gestures (with and without fluencemes)	more referential gestures overall, and more interactional gestures (with and without fluencemes)
	no significant differences for thinking gestures, but occurred almost exclusively during fluencemes	
Gaze behavior	more instances of gazing towards piece of paper (with and without fluencemes)	more instances of gazing away and towards the interlocutor (with and without fluencemes)

A number of limitations in this study should also be noted. First, as explained earlier (Chap. 2, section I.1.4) we worked on a selected sample of DisReg to approximately match the size of SITAF, but our sample is still relatively small and only representative of a selection of the students' productions, so our findings need to be taken with caution. The other limitation to our data sample is that it only represents French students from a French university, which does not exactly reflect our contrastive and crosslinguistic approach to inter-(dis)fluency. But as explained earlier (Chap. 2, section I.1.4), we had initially intended to build another similar corpus of American students at an American university using the same collection procedures, but our project was severely impacted due to the outbreak of COVID-19 in early 2020.

If the situation allows, we will resume our data collection project to conduct the same analyses on American English data to offer crosslinguistic comparisons. Lastly, as pointed out before in our conclusion to Chapter 3, we are aware that the statistical methods used in our quantitative analyses are rather simple and resulting in binary outcomes, so more complex data analysis techniques, such as multiple correspondence analysis, or mixed linear regression models, containing fixed and random effects, should be used in the future to explore possible systematic relationships between fluency and other variables. However, we still believe that our quantitative analyses yielded a number of significant results, which can be of interest to researchers in (dis)fluency and gesture studies.

In addition, while the present study does not primarily intend to be pedagogical-oriented, it may still open perspectives onto future work in class pedagogy, offering tools for students to better perform their assignments in class, by making simultaneous use of their voice, body, and eye contact to increase "eloquence" (e.g., Papanas et al., 2011). The issue with a majority of oral presentations found in French universities is that they tend to be too content-oriented, which may explain why students spend most of their time dealing with the written content on their notes, instead of engaging with their audience. However, this type of hypothesis would require more data from a larger sample of speakers in different universities in order to be supported. In addition, fluencemes should not be systematically stigmatized as performance errors or "hallmarks of youth" (Fox Tree, 2007), representing poor communication skills, as we have shown them to be dynamically ambivalent systems, relying on a multiplicity of resources.

Highlights of Chapter 4:

- Fluencemes and gestures are highly sensitive to situational features, as this chapter has demonstrated the effect of setting on different variables (fluenceme rate, fluenceme type, gesture rate, gesture function, gaze direction, etc.)
- We take on a multidimensional view of language style to characterize differences between institutional and more casual settings.
- Drawing from CA, sociolinguistics, and gesture studies, several inter-related factors need to be taken into account to examine differences in fluency and gesture behavior across situations, mainly audience design, register, turn-taking mechanisms, and the overall material environment.
- Class presentations and face-to-face interactions globally differ on a number of dimensions, mainly at the level of speech production, visual-gestural behavior, and interactional dynamics, and these differences should not be restricted to one level or another.
- Speakers may *sound* disfluent in the speech signal but they may *look* fluent in the visual-gestural channel, or the other way around, so (dis)fluency should not be restricted to a performance error or poor communication skills at the level of speech production.
- Quantitative and qualitative analyses need to be combined in order to measure the degree of inter-(dis)fluency at different levels (speech, gesture, and interaction) found across situations to show different but complementary aspects of our investigation.

Chapter 5. On the relationship between Inter-(Dis)fluency and Gesture

Introduction to the chapter

In the previous chapters, we stressed the importance of multimodality within the study of (dis)fluency through the analysis of visual-gestural behavior during "fluent" and "disfluent" stretches of speech (following Graziano & Gullberg, 2018) in other words, within and outside fluencemes. In line with previous work (e.g. Christenfeld et al., 1991; Graziano & Gullberg, 2018; Yasinnik et al., 2005, see Chap. 1, section III.3.3.4) our studies conducted on the SITAF Corpus and the DisReg Corpus have shown that gestures tend not to occur extensively during fluencemes (only about 20-25% of the time, see Chapters 3 and 4, sections II. 2.1.3.), and that (dis)fluency tends to be associated with little gesture activity and gaze withdrawal. However, as our multimodal qualitative analyses have further revealed, several instances of *embodied inter-(dis)fluency* may emerge in the course of the participants' multimodal utterances, that is, when speakers deploy a combination of fluencemes and gestures to (co-)build the fluency of discourse. As pointed out before, *speech* (dis)fluency has typically been characterized by a suspension of the vocal and verbal modality (i.e. interruption of a syntactic structure, or suspension of the voice in the acoustic channel), but without paying much attention to the surrounding interactional environment and the other types of modalities or resources available to speakers. As argued throughout this thesis, the aim of the present work is to go beyond narrow definitions of fluency and disfluency, and offer a more integrated multilevel and multimodal approach, grounded in different interdisciplinary frameworks (cf Chap. 1). To that aim, we analyzed the distribution of fluencemes with regard to gesture phrasing and phases of gestural action, as well as gesture types (Kendon, 2004), based on a consistent functional classification (in line with Cienki, 2004; Kendon, 2004; Müller, 1998, 2015 see Chap. 2, section II.2.2.3.). This annotation system was applied systematically to our data as part of our studies on the SITAF and DisReg Corpus in order to compare general patterns of distribution across languages and settings.

The goal of the present chapter is to further explore the multimodal dimension of inter-(dis)fluency by documenting the different forms and functions of gestures co-

occuring with fluencemes, or occurring within their vicinity, in their situated, embodied, and multimodal environment, thus taking a step further from our initial functional classification of gestures used in our quantitative analyses. The present chapter will thus only present detailed qualitative analyses of the data across the two corpora, and pay specific attention to the temporal relationship between (dis)fluency and gesture and their synchronicity in terms of gesture phases, as well as the deployment of different articulators (i.e. hand, face, eyes, shoulders, and trunk) and the shape, configuration, orientation, movement, and position of gestural sequences in the gesture space, following a more *form-based approach* to gesture (Bressem & Müller, 2014; Ladewig & Bressem, 2013; Müller et al., 2013). Several references will also be made to the gestures analyzed in previous chapters, as to establish a typology of gestural variants in relation to inter-(dis)fluency. The present chapter is structured as follows: we first illustrate the temporal relationship between (dis)fluency and gesture phrasing through several examples on the synchronization of speech and gesture production (section I), then document several visual-gestural practices embodying inter-(dis)fluency (section II), and conclude on the multimodality of inter-(dis)fluency in embodied situated language use, further questioning the notion of language (section III).

I. Synchronization of Speech and Gesture

In Chapter 1 we reviewed a number of studies conducted on (dis)fluency and gestures which targeted aspects of gestural and speech suspension (cf Chap. 1, section III. 3.3.4). While some showed that gestures tended to be suspended prior to the production of (dis)fluencies (e.g., Seyfeddnipur, 2006; Seyfedinnipur & Kita, 2001) others have stated that they were more likely to begin during them (e.g., Beattie & Butterworth, 1979). As Graziano & Gullberg (2018) pointed out, most of the studies that investigated the timing of gestures relative to (dis)fluency have presented contradictory findings, mostly due to methodological and theoretical differences. Following the notion that speech and gesture form a tightly integrated, orchestrated, and unified system (e.g. Kendon, 2004 ; McNeill, 1992) Graziano & Gullberg (2018), along with other researchers (e.g. Chui, 2004 ; Esposito & Marinaro, 2007 ; Yasinik et al., 2005) provided evidence that gesture suspension tended to synchronize with speech suspension, hence suggesting that gesture production is an integral component of utterance construction (Kendon, 2004). Similarly, our studies conducted on the

SITAF and DisReg Corpus have shown a similar tendency, with a proportion of about 25% to 35% of fluencemes[164] which occurred during phases of gestural actions, such as *preparation*, *hold*, or *retraction*, regardless of language or setting[165] (cf Chapters 3 and 4), further giving support to the temporal patterning of speech and gesture. The following analyses, taken from our two corpora under study, will focus specifically on this relationship by showing the deployment of gestures within different gesture phases and their relationship to (dis)fluency. The excerpts chosen are purposefully very brief as to zoom in on particular instances of gesture phases (i.e., gesture hold, retraction, and preparation) and forward gesturing.

1.1. Hold and retraction: suspension and interruption in the two modalities

In this section, we focus on two specific gesture phases deployed after the stroke, mainly the (post-stroke) hold, and the retraction of the gesture. As explained earlier, (cf Chap. 1, section III. 3.3.4), gesture *hold* refers to hand gestures that are temporarily frozen in a static position, and *retraction* refers to a moment of relaxation, when hands return to the initial rest position, following a gesture stroke, or a previously held one. The two following examples illustrate the way speakers momentarily suspend or interrupt the course of their multimodal utterance by either holding their hands in the same position in the gesture space, or by returning them to their initial rest position. Gesture holds have already been illustrated in our previous qualitative analyses (e.g. Excerpts 1.A, 2.A, and 2B in Chap. 3), but here we will focus more precisely on their timely coordination with vocal suspensions in the acoustic channel. The first example is taken from Pair 3 in French in the SITAF Corpus. All examples use the gestural notation system created by Kendon (2004) described in Chapter 2, (see Chap. 2, section I.1.5.3), and the fluenceme sequences that accompany these gesture phases are marked in brackets and in bold in the transcription.

[164] However, one methodological limitation should be noted. Gesture phases were only coded during fluencemes, while gesture strokes were coded in the whole data (i.e. both within and outside fluencemes, cf Chap. 2, section II.2.2.3.), but it would have been interesting to compare the proportion of gesture holds within fluencemes with those outside fluencemes to get a more precise idea of their temporal relationship.

[165] There was still a tendency, however, for the American and French groups to hold their gestures more frequently during fluencemes in their L2 than in their L1 in SITAF. No significant differences were found across the two situations in DisReg.

Ex. Hold (A)

In this excerpt, the two tandem partners are talking about differences between travelers and tourists (cf Excerpt 2A in Chap. 3), and here the American non-native speaker (Julia, A03) argues that the traveler, unlike the tourist, settles down more permanently in a foreign country.

1 JUL: mai:is (0.460) peut-être euh il s'installe **(1.080) mm** mieux
[FLUENCEME SEQ.]
1.**********
((both lax open hands, palm facing up, spread bent fingers + hold))

1.

2 MAR: +< d'accord.
3 JUL: ou il s'installe **euh (1.280) [!]** plus
[FLUENCEME SEQ.]
2.********** *****
((both hands coming down, same configuration as 1. + hold))

2.

forteme**:ent (0.680)** dans [/] dans le pays.
[FLUENCEME SEQ.]
******3.********** ~~~~**********
((hold + Palm Down Open Hand))

3.

In this example, Julia is producing a series of formally similar manual gestures, where both her lax flat hands are first slightly raised with spread bent fingers to the center of her trunk, then coming down to the lower part of her body, with both palms facing up (1 and 2); the initial stroke of the gesture first coincides with the verb "installer" (settle down), and is then repeated with several beat movements, as she repeats the verb in line 2. What is interesting to note is that every time she held her hands in the same position (as shown in the transcription with the underlined asteriks), she also momentarily suspended the course of her vocal utterance. In line 1, she first delays the course of her vocal flow by pausing for a significant amount of time (about one second), and producing a nasal vocalization ("mm"); then in line 3, she produces another series of vocal and non-lexical fluencemes (a filled pause, an unfilled pause and a tongue click), and this time she holds her gesture during them. She also lengthens the final syllable of the adverb "fortement" and produces another 680 ms pause near the end of her utterance, during which she also holds her hands in the same position. Each time, the same type of gesture is held, i.e., both lax flat open hands facing up with spread bent fingers, and each of them is also accompanied by specific facial expressions; she is either looking away (1), closing her eyes (2) or slightly frowning (3). The first two times she withdraws her gaze, and she seems to be engaged in a word searching activity; but the third time, after she retrieves the word that she has potentially been looking for ("fortement"), she gazes back at her interlocutor and frowns, perhaps to signal uncertainty, and check with her partner that she has used the correct lexical word. She in fact uses a similar multimodal repair strategy as the one described in *Excerpt 2.A* from Chapter 3 (section II. 2.2.2.), where she momentarily held her hands in the same position, gazed back at her interlocutor, and shook her head as to request assistance from her partner. This time, however, the native speaker did not intervene, as it was not considered a relevant next-action in this context. In sum, Julia suspended various parts of her multimodal utterances as she went along with her discourse, and she did so with the help of vocal and gestural markers of suspension, which illustrates how both modalities were momentarily suspended in the multimodal communication channel, in order to build meaning.

Ex. Hold (B)

In this second excerpt from the conversation-sessions in the DisReg Corpus, Lea (E1) is talking about one of the classes on Greek mythology she is attending at university, and explains to her partner Tina (E2) that she does not really enjoy the class overall,

for various reasons, (not in the transcription) but that she really enjoys studying ancient Greek.

1 LEA: mais par contre c:ce que j'aime bien dans ce cours c'est le grec ancien.
((places stretched fingers from LH on the table))
2 LEA: (0.413) fin parce-que:e +//.
3 TINA: +< ah oui.
4 LEA: ouais bah parce-que ils mettent beaucoup **de:e [/] de** mots
[FLUENCEME SEQ.]
~~~~~ **********
((brings RH to the table, produces two circles with her wrist + hold))

par rapport au:u (0.404) [/] par rapport au rythme et au grec ancien.
[FLUENCEME SEQ.]
~~~~~**********
((moves her right hand to the side + hold + moves back and forth))

Here Lea is using the desk next to her as a medium to metaphorically map out certain aspects of the structure and rhythm of ancient Greek, as she performs a series of manual actions with her fingers directly on the table, as if drawing something on a piece of paper. The table seems to embody an imaginary piece of paper, and her extended fingers seem to represent a pen that Lea is holding and using for writing, as she makes a series of circles with her wrist. Performing these gestures may enable Lea to make sense of her appreciation for ancient Greek, through embodied sensory experience (Müller et al., 2013). This is further shown in her second manual action by which she moves her right hand back and forth to the side, with the same hand configuration, as if drawing a line. Two cases of gesture holds are found here, both in

line 4 in the transcript. Each hold is followed by a lexical affiliate ("mots" and "rythme") whose content seems to be related to the gesture strokes initiated before the holds: the series of circles performed with her wrist may conceptualize the representation of different words ("beaucoup de mots") in the text, while her back-and-forth motion may evoke the notion of rhythm in ancient Greek ("au rythme et au grec ancien"). It would appear that these holds enabled Lea to transition from gestural to linguistic units, from one semiotic field to the other, in order to build lexical meaning. In addition, just like the previous example, the gesture holds perfectly coincided with the production of fluenceme sequences, both following the *VOC+MS* pattern as they comprise at least one vocal marker and one morpho-syntactic marker. The linguistic process of delaying the production of upcoming constituents in the verbal modality (through prolongations, pauses, and word or phrase repetitions) is hence also manifested in the gestural modality, where the production of her gestures is temporarily held, and then resumed during the target words.

As Kendon (2004) noted, when a speaker employs a post-stroke hold (Kita 1993), the expression conveyed by the gesture stroke seems to be prolonged, as to extend its meaning, and it would appear to be the case in these two examples as well. In the first example, the non-native speaker first deployed her hands in the gesture space to introduce the concept of settling down in a foreign country, and she did so by moving her hands down, with palm-up open hands and bent fingers. We may wonder whether the downward direction of her gesture may have conjured up a metaphorical mapping with the phrasal verb in English "settle DOWN", reflecting a specific pattern of thinking in her L1 (Stam, 2006), since the single-word verb in French "installer" in her L2 does not contain any particle indicating path and ground. While this is only an unfounded speculation, it is still interesting to note that this downward movement was repeated several times whenever the speaker was talking about settling down more permanently in a foreign country. Her hands were also held in this very same position, as if she was *holding on* to the idea she was conveying to her partner. Similarly, in the second example, Lea shared with her interlocutor her appreciation of her class on Greek mythology by making sense of the multiplicity of words and the type of rhythm found in Greek, through touch and sensory imagery, i.e., enacting a drawing or writing action directly on the table. These multimodal actions were also held, as she transitioned from one gesture to the next, one semiotic field to another, reflecting a smooth, and *fluent* adjustment in her multimodal narrative flow. This further

demonstrates that the notions of suspension or delay, that are inherent to (dis)fluency, are not necessarily associated with disruptive or *disfluent* actions despite the fact that they may be conveyed with a priori verbal disfluent forms in the verbal channel. It is thus crucial to consider all channels of communication to further our understanding of inter-(dis)fluency in multimodal discourse, and the synchronization between gesture and speech suspension illustrated in these two examples further gives support to the view that gesture and speech work together as a unified system. We shall now move to the analysis of two examples which focus on the retraction phase.

Ex. Retraction (A)

In this French excerpt from the SITAF Corpus (Pair 17), the two tandem partners are talking about social media and whether it makes people lonely.

1 RUTH: eum (0.475) [!] mais en même temps c'e:est [/] c'est difficile
****
parce-**que:e hh. eum** je veu:ux être dans les deux (1.362) eum +/.
[FLUENCEME SEQ.]
********-.-.-.-.-.-.-.-.-.-~ ~ ~~ ~ ~~ ~ ~**********.-.-.-.-.-
((palm-up open hands move together in alternate motion + retraction))

2 LOLA: aller aux fêtes de tes amis mai:is en même temps.
3 RUTH: mm mm.

Here the non-native American speaker, Ruth (A17) is talking more specifically about her own experience of staying abroad (in France), and explains that it is difficult to say whether she feels isolated or not, because on the one hand she feels connected to her friends through social media (not in the transcription), but on the other she wishes she could share these moments with them (l.1). It appears that she is considering the pros and cons of using social media, how it may help her staying in touch with her friends, but how it may also further isolate her from them. She expresses this idea by metaphorically representing the action of decision making through her gestural behavior, as she is moving her palm-up open hands up and down in an alternate

motion in her center space, as is weighing one object against another on a scale, and she reiterates this motion several times throughout the excerpt. Note that this gesture is also held during a prolongation ("c'e:est", line 1), showing a similar instance of suspension in the two modalities. While she seems very engaged in this decision-making activity, she suddenly interrupts the course of her gesturing flow in mid-utterance as both her hands fall back into her lap, and she gazes away. This shift in her gesturing activity also coincides with a complex fluenceme sequence (*PR+NL+FP*) of rather long duration (1625 ms in total) which also suspends the course of her speaking activity. She immediately resumes both her speaking and gesturing activity, in other words, her *languaging* activity (cf section III.3.1.), at the end of the fluenceme sequence, marking a resumption of her decision-making practice, as she reiterates her *weighing up* gesture (Müller et al., 2013). This gesture seems to take over the verbal channel, as she pursues its production with several repeated and alternated up-and-down motions during a rather long pause of 1.362 milliseconds. This time, her vocal suspension in the acoustic channel (marked by a pause) does not coincide with a gestural suspension (unlike previous examples), but is in fact taken up by a gestural activity, giving room to another relevant semiotic field. It is interesting to note that as soon as she fills the acoustic space with sound (with a nasalized filled pause "eum" at the end of her utterance, following the silence) her gestural activity is once more brought to an end, as her hands fall back to rest position.

This second shift in Ruth's languaging activity seems to be recognized by her partner as a relevant opportunity for speaker change, as she takes the floor to *verbalize* the meaning conveyed through Ruth's weighing up gesture ("aller aux fêtes de tes amis mais en même temps" line 3), which is perhaps reflecting her role as an expert native speaker who wants to help her tandem partner by providing additional verbal content, or it may as well reflect her role as a co-participant of the conversation who wishes to display shared understanding. As we have seen earlier, these two discourse identities may be existing simultaneously (cf Chap. 3, section II. 2.2.3, *Excerpt 4a*). This further shows that instances of multimodal suspension, which reflect a point of interruption in the two modalities, may also promote the activity of seeking understanding, which echoes Sikveland & Ogden's (2012) work on held gestures across turns to generate shared understandings. Alternatively, Lola's action to take the turn may also be interpreted as a sign that Ruth failed to keep the floor, as she was experiencing difficulties with her production, so she was interrupted, thus marking a disruption in

the progressivity of her current turn. This would be a more negative, DISfluent interpretation. Once more, the line between fluency and disfluency, or communication flow versus breakdown, is not straightforward, and needs to comprise multiple dimensions (speech, gesture, and interaction).

Ex. Retraction (B)

In this second excerpt from the DisReg Corpus, Linda (F1) and Matt (F2) are talking about TV shows and the fact that whenever a new one comes out and the first season is over, they have to wait for another year before they can watch the second season, which makes it difficult to remember the plot and characters (not in the transcription). Here Linda draws a parallel with new episodes that come out every week on TV.

1 LINDA: euh comme regarder **euh (0.425)** les [//] épisode après episode
[FLUENCEME SEQ.]
**********-.-.-.-.-.-.-.-.-.-.-~~~~**********
((moves her right hand forward away from her, vertical palm facing her body with extended thumb, looks up + retraction))

2 LINDA: (en)fin chaque semaine tu dois te remettre dans le truc euh
~~~~~**********
((moves her right hand with a rotating motion; gazes towards MATT))
3 MATT: bah ouais.
4 LINDA: moi je trouve ça:a en fait finalement je:e [//] j'aime pas trop quoi bon.

Here Linda first initiates a gesture that seems to convey a continuous process through time by moving her right hand further away from her body in the central space with several beat motions, her vertical palm facing her trunk and her thumb extended. This gesture is synchronized with the verb "regarder", and is then followed by a fluenceme sequence comprised of a filled pause, a silence and a syntactic repair, during which her hand moves back to rest position. Just like the previous example, Linda interrupts her languaging activity as she is trying to reformulate parts of her utterance to describe the
~~~~~

process of watching one episode after the other every week, and this interruption is embodied in the two modalities. She then repeats the same gesture initiated at the beginning of her turn, and this time it coincides with the noun phrase “episode après episode”, which offers additional verbal lexical content to the gesture’s intended meaning. She is also gazing towards Matt at this moment, and the latter demonstrates his alignment through verbal backchanneling (“bah ouais”). It thus seems that the first initiation of her manual gesture served as a *foreshadowing* of her next action (Streeck, 2009, cf Chap. 2, section I.1.5.1. on forward gesturing), as she later elaborates on the process of watching an episode each week. This is further conveyed by her cyclic gesture, marked by a rotating movement of her right wrist (see section 2.1.2. on cyclic gestures). In addition, her multimodal interruption marked a transition from a self-oriented activity (i.e. reformulating a verbal expression while looking up) to an other-oriented one (i.e. sharing this piece information with her partner while looking towards him).

To conclude, the four examples analyzed above have illustrated instances of suspension or interruption in the verbal/vocal and visual-gestural modality, reflecting a unified and co-orchestrated process in the two modalities. These suspending/interrupting activities may further embody relevant interactional actions in the course of the multimodal exchange, marking a transition from a self-oriented practice to the next communicative move. This leads us to the analysis of the preparation phase, which further sheds light on the timely coordination between speech and gesture.

1.2. Preparation: preparing speech and gesture in tandem

When a manual action enfolds, it is usually marked by a preparatory movement phase where the hand(s) move from a resting position to prepare the execution of the gesture stroke; this is known as the *preparation phase* (Kendon, 1972; Kita et al., 1997; McNeill, 1992). Unlike the post-stroke hold and the preparation phase, this one occurs before the stroke, and therefore does not mark a suspension or an interruption, but rather an initiation, or a projection. In the three following examples, we show how this phase also coincides with the emergence of vocal and verbal fluencemes.

Ex. Preparation (A)

The first excerpt is taken from Pair 18 in the SITAF Corpus. The two tandem partners, Rosie (A18) and Sophie (F18) are comparing teenage years with adulthood, and here the American native speaker (Rosie) is presenting a number of arguments in favor of adulthood, saying that as an adult she believes that she gets more opportunities to experience life more, as opposed to when she was a teenager and things always seemed to be the same[166].

1 ROS: **(0.549) um (0.330) [!] I think that (1.296)**
[/] I just think (0.451) that (1.175) the life +//.
[FLUENCEME SEQ.]
~~~~~~~~**********
((left hand moves forward in preparation in a slow motion + beat movement with her vertical open palm))

2 ROS:(**0.338)** like when I was a teenager my life was very um (1.429)
[FLUENCEME SEQ.]
~~~~~****************************-.

((left hand moves in preparation to her left then produces a beat movement and brings her palm down + return to rest position))

3 ROS: it was all the same.
~~~~~~~~~***********
((performs a cyclic rotation with her left hand with her vertical open palm))

4 SOP: (chuckles)

5 ROS: it was (0.374) the same day after day.
~~~~~*************************************
((performs a similar cyclic gesture with her left vertical palm open hand))

[166] Note that this is the same example used to illustrate our multi-level analysis of (dis)fluency in Chapter 2 (Chap. 2, section II, Fig. 26).

Rosie begins her first utterance with a very long and complex fluenceme sequence, comprised of multiple fluencemes (five unfilled pauses, one filled pause, one tongue click, and a repetition of the segment "I just think that"). During most of this fluenceme sequence, Rosie keeps her hands in rest position, but then she initiates a gesture with her left hand, as it slowly moves forward in preparation during her fifth pause of 1.175 milliseconds. Then, as she introduces a noun phrase ("the life") she produces a beat movement with her vertical left open hand. This multimodal phase of preparation, characterized by a preparation of a gestural and a linguistic unit, also marks the beginning of her discourse unit as she launches a new topic ("the life") that is later re-introduced in her next utterance, following her self-interruption. In her second utterance, she produces another unfilled pause of much shorter duration (338 ms) in utterance-initial position, during which her left hand moves once again in preparation, but this time to the left periphery of her gesture space, her palm facing down, with slightly bent fingers. The left side of her gesture space thus seems to metaphorically embody a specific region of time (her teenage years), and she performs a series of cyclic gestures in the same location, as to represent a habitual and routinely process that captures her life as a teenager ("it was all the same", "the same day after day"). A similar instance of multimodal preparation is found in the second example.

Ex. Preparation (B)

This is the same extract analyzed in Chapter 4 from Pair D (*Louis Garrel* example, section II. 2.2.3), but here we would like to focus more specifically on the first two lines of the excerpt, before Jenny went on with the telling of her humorous anecdote involving French actor Louis Garrel.

1 *JEN: hhh. ma:ais euh la dernière fois j'étais au café (il) y'a une pote (laughs) qui me raconte qu'elle avait une pote (laughs) complètement bourrée qui était dans un bar.
2 *JEN: et elle était en fait **euh** assise en face de Louis Garrel.
[FLUENCEME SEQ.]
~ ~~ **********
((moves her vertical left palm open hand in preparation + extends her arm towards ALX))

In line 2, Jenny slowly extends her left arm and hand forward, with her flat open hand, palm facing right, as to place Louis Garrel's referent in the narrative space (the bar), adopting the character's viewpoint (McNeill, 1992), i.e., reflecting the protagonist's body onto her own, and orienting attention to this particular space. Her deictic gesture, which makes use of her whole palm to point towards the target location in the gestural narrative space, also provides some information regarding its distance, suggesting that the two people from the story are sitting relatively close to each other. This information is further confirmed when she informs Alex that the protagonist of the story was sitting opposite Louis Garrel (at the end of line 2); therefore, the initiation of this gesture paved the way for her upcoming verbal spatial description. In addition, this gesture preparation introduced the character's viewpoint, functioning as a cohesive device which created visual coreference in space. As Gullberg (2006) showed, speakers consistently make use of space to introduce different discourse entities, which is known as "spatial anaphoricity" (Debreslioska et al., 2013, p. 435) i.e., when referents are introduced into multimodal discourse with gestures, as to signal their accessibility. Similarly, in the previous example, Rosie placed a specific referent in time (her life as a teenager, placed in the left periphery of her gesture space) perhaps to disambiguate between her adulthood and her teenage years, and she later made use of this specific location in space to express the monotonous regularity of this time period. What is interesting to find is that in both examples deictic-anaphoric gestures were initiated **during fluenceme sequences**, and that the gesture preparation was timely coordinated with speech preparation, showing once more a harmonious cooperation between the two modalities. This further shows that (dis)fluency processes are not only associated with discourse suspension or interruption, but may also index new sequences of multimodal talk and (re)introduce discourse topics. While we have only shown examples from conversational data so far, instances of gesture preparation co-occuring with fluencemes can also be found in institutional discourse.

Ex. Preparation (C)

This excerpt is taken from Laura's (C2) presentation in which she is analyzing the role of female servants in Molière's play *Les Fourberies de Scapin*.

1 donc hhh. euh elles intéragissent dès le début avec les personnages principaux

2 hhh. (0.400) e:et eum (1.313)

[FLUENCEME SEQ.]

~~~~~~~~~******

((moves both her hands in preparation then brings them together, palms pressed against one another))

par cette idée de personnage exposante.

~~~~**********.-.-.-.-.-.-.-.-.-.~~~~**********

((brings both her lax flat hands forward, palm up, bent fingers + retraction, then repeats the same gesture))

3 eum déjà je voudrais établir un point.

-.-..-.-.~~~~~~~~~~~~~**************************

((hands return to rest position + initiates a similar gesture))

In line 2, Laura produces a complex fluenceme sequence in utterance-initial position, comprised of an inbreath, two unfilled pauses, one prolongation, and one filled pause, following the *VOC+NL* pattern. As she projects a new stretch of talk, her hands also move together from her desk in preparation, and she brings them together, with her palms pressed against one another, while looking up. This gesture is then held during the pause, and is followed by a different gesture in which she extends her lax flat open hands forward, away from her trunk, with her palms facing up and bent fingers. The first gesture seems to embody a posture of "getting ready" for the next stretch of discourse, as it is followed by a series of *discursive* gestures (see Chap. 2, section II.2.2.3. and Chap. 4, section II. 2.1.3) used for presentation or emphasis (three instances of these gestures occur during target phrases "cette idée", "exposante", and

"établir un point"). In an entirely different context, this *palms-pressed-together* gesture may be considered as an emblem, i.e., a symbolic gesture conventionalized within a specific community of speakers. In this case, both palms pressed together may refer to the act of praying (Hunsicker & Goldin-Meadow, 2013, Kendon, 1995), and is in fact recognized as such in late 14th century religious medieval paintings (Schleif, 1993). In other contexts, it is known as the "Namaste" gesture, often found in yoga practices, or the "Mudra" gesture which has been practiced for centuries throughout South Asia to symbolize respect or good will in Hindu or Buddhist culture (Rajput, 2016). However, in this specific institutional context, this gesture has entirely lost this religious meaning. Kendon (1995) also documented this gesture in one of his papers on Southern Italy gesture use, and identified it as the *Mani giunte* gesture ("joined hands"). In his video-recorded data of spontaneous conversations between Italian speakers, he noticed the use of this gesture when speakers wanted to "make visible certain implications of what is being said that is not made explicit verbally" (Kendon, 1995, p. 259). Kendon specified that this type of gesture was commonly used by speakers as an "appeal" to the listeners to accept the consequences of what the speakers have been saying. But because this is not conversational data, it is difficult to say whether this *joined-hands* gesture serves the same function as the one described in Kendon's. We may in fact also wonder whether this type of gesture may solely represent the speaker's gestural home position. We later see her regularly place back her hands in this very same position in space, but with her fists closed, which may be a handshape variation of the initial *palms-pressed-together* or *hands-joined* gesture. For further research, it would be interesting to explore the data more thoroughly, and see if we could document similar types of gestures that are typical of class presentations, with recurrent formational and semantic features. While this type of investigation does not fall within the scope of this chapter, it is still relevant to note the temporal coordination between the initiation and hold of this specific gesture and the emergence of fluencemes, again reflecting the notion of preparation and projection in the two modalities.

In addition, it is important to note that when she produced this gesture, her gaze was not directed towards her notes (as students very frequently did throughout their presentations, see Chapter 4, section II.2.1.3), but she was looking up. This may be an indication that she changed her attention towards "a world of thought" in which she needed to make up her mind about what to say next (Heller, 2021). Heller (2021)

refers to this specific type of gaze behavior as *imaginative gaze*, by which speakers are stepping out of their current activity to focus on their own thoughts. Similarly, Goodwin & Goodwin (1986) spoke of a "middle-distance" look, reflecting a change of orientation, and perhaps to a larger extent a display of "doing thinking" (Heller, 2021). This leads us to the following section, where we focus most specifically on embodied practices of inter-(dis)fluency and discuss the notions of doing thinking.

II. On the visual-gestural practices embodying inter-(dis)fluency

In this section, we explore more specifically the notions of (dis)fluency and hesitation in visible bodily behavior, and document recurrent visual affiliates of (dis)fluency, mainly thinking postures and word searching manual actions (2.1.), embodied displays of intersubjectivity (2.2.), and gestural modes of representation (2.3.). Analyses are also based on the SITAF and DisReg Corpus, with a number of excerpts that have already been introduced in the previous chapters, but with a slightly different take.

2.1. Doing thinking as an interactional practice

In Chapter 1, we mentioned the display of a specific facial expression in social interaction, known as the "thinking face", a term coined by Goodwin & Goodwin (1986) in their paper on joint word searches (cf Chap. 1, section III. 3.2.3.). Despite what the term "thinking" entails, the authors do not view this practice as a reflection of inner cognitive processes, but rather as an interactive display of the speakers' continued engagement in a joint activity. Such facial displays may also reveal a state of uncertainty, and do not necessarily indicate a cognitive process, but rather invoke particular types of social organization, as Goodwin (1987, p. 116) put it:

> First, a speaker can both bring the material being looked for into a position of salience that it would not otherwise have and make the task of searching for that material the primary activity that the participants to the conversation are then engaged in. This shift in activity changes the participation framework of the moment and with it the ways in which those present are aligned towards each other, as well as the behavior they are engaged in. Second, through the way in which a speaker performs the display of uncertainty, he or she can make a variety of proposals about the social position of others present. Thus a speaker

> can signal that others present share with him or her access to the material marked as problematic and invite them to aid in the search for it.

More recently, Heller (2021) conducted a study on joint decision making, following the work of Goodwin & Goodwin (1986). Based on a corpus of monolingual and multilingual children recorded within a school setting in which they were asked to perform a series of argumentative tasks, Heller (2021) showed how speakers and recipients combined various semiotic resources to create complex multimodal gestalts (Mondada, 2014) which embodied practices of "doing thinking". Such practices include a combination of body postures, particular gaze practices, and linguistic resources. In her analysis of thinking displays, she documented a series of multimodal gestalts including "imaginative gaze", wandering of the eyes, and thinking postures. The latter is characterized by an "inflexible" body posture where the face and hands are fixed, as a signal that neither gesture nor hand movement are expected during the display. In addition, thinking postures were shown to index an embodied change of mind, a transition into the display of doing thinking. This is also marked by gaze withdrawals, where the speaker embodies a change of orientation, and displays to his or her partner that he or she is no longer oriented towards their external surroundings, but rather directs attention "inwards, toward a world of thought, in which she first needs to make up her mind before she can share her ideas with her co-participants (Heller, 2021, p. 8). However, Heller still argues that this display of *doing thinking* should not be conceptualized as an external manifestation of inner cognitive processes, but should rather be regarded as a relevant public practice and multi-party activity performed for the co-participants in order to mobilize their visual attention. She concluded that these displays worked as relevant resources for shaping the different emerging participation frameworks within the activity-in-progress, performing both interactional and epistemic functions. Following Goodwin & Goodwin (1986) and Heller (2021) the present section aims to document different uses of the thinking face along with other bodily manifestations, mainly body orientation and manual gestures, specifically in relation to inter-(dis)fluency and the notion of hesitation.

2.1.1. Multimodal gestalts of doing thinking: embodied markers of hesitation?

In Chapter 1, we described the different uses of the term *hesitation* in the (dis)fluency literature, and argued that this term was not truly representative of the phenomena under study. Indeed, not every fluenceme reflects an act of choice or indecision, and the term "hesitation" remains too restrictive to instances of trouble or uncertainty (cf Chap. 1, section II. 2.2.2), which does not cover the whole range of functions and uses of fluencemes. However, it may be argued that in some specific contexts the core notion of hesitation could be found within embodied displays of thinking, which may reveal the emergence of *embodied hesitation*, and therefore instances of *doing hesitation*. We may thus wonder whether the act of hesitating, as in, choosing among different options and feeling uncertain or hesitant about this choice, may also be embodied within this specific social practice. The following instances, taken from previous examples in Chapters 3 and 4, and summarized in the table below, further explore this idea.

These multimodal gestalts of doing thinking, or doing hesitation, largely echo the work of Heller (2021), showing similar facial and body displays documented in her work, mainly self-touch, "imaginative" gaze, "inflexible" posture, wandering of the eyes, and gaze withdrawals. Similar facial features were also found in Bavelas & Chovil (2018), where they regarded thinking faces as *collateral signals* (cf Chap. 1, section II.2.2.1), further describing them as "somewhat stylized gestures in which the speaker pauses, turns his or her head or looks away, often with a blank, puzzled, or thoughtful face" (Bavelas & Chovil 2018, p. 111). Similarly, in the examples reported above, these displays were deployed at relevant transition points during which speakers resorted to temporary solitary practices. These practices were further made visibly available to the recipients, who did not interrupt the speakers while they were delaying the course of their multimodal talk. As Bavelas & Chovil (2018, p. 111) further stated: "it is a temporary hesitation that requires nothing of the addressee but to wait". Once more, the notion of delaying or suspending speech is further embodied in recurrent, recognizable, and salient facial displays which evoke epistemic stance, and to a larger extent the notion of *hesitation*.

Table 39. *Summary of embodied displays of thinking in previous examples (Chap. 3 and 4)*

Illustration	Example	Description	Context	Participant's status
	Excerpt 1.b (Chap. 3)	Squinting of the eyes, eyebrow frown, gazes away, hands in rest position	Speaker is looking for her words, marked by a series of vocal fluencemes in the acoustic channel	Non-native French speaker in tandem interaction
	Excerpt 1.b (Chap. 3)	Smiles, looks up and brings her head up, hands in rest position	Speaker explicitely signals her word-searching problems with an explicit editing phrase ("I ain't got the word here")	Non-native French speaker in tandem interaction
	Excerpt 1.b (Chap. 3)	Squinting of the eyes, winces, gazes away, hands held in the same position	Produces a series of unintelligible words before reformulating	Non-native French speaker in tandem interaction
	Excerpt 2.3. (Chap. 4)	Gazes away, hand resting on chin	Speaker is thinking about what she was talking about before, marked by a pause and a click	French native speaker in the course of her story retelling

This act of hesitating, marked by vocal and gestural markers of suspension in the multimodal flow, may also mark a *fluent* and *smooth* transition towards a change of participation and of pace. In the case of the non-native speaker during the tandem interaction (*Excerpt 1.b*., Chap. 3, section II. 2.2.2.), she interrupted and delayed the course of her utterance multiple times to look for the right words in her target language, while signaling her continued engagement towards the activity-in-progress. In the case of the French native speaker (*Excerpt 2.3*., *Chap. 4, section II. 2.2.3*), who was engaged in her storytelling activity, her delay marked a relaunching of her story preface, which was previously interrupted by a series of insertion-sequences.

Similar practices are reported in the following examples. In particular, we will focus on instances of *self-touch*. As Heller (2021) claimed, touching a part of one's body implies that "the individual gets entangled in the haptic-kinetic perception of her own body and shields herself from other stimuli" (p. 8). Ekman & Friesen (1969) spoke of "self-adaptors" i.e., a series of manipulations performed on the body, such as wiping around the corners of an eye, squeezing one's leg, touching one's hair, etc. Even though self-touch, or self-adaptors are often either dismissed from gesture analysis, or

associated with interactional disengagement and a form of self-involvement, Streeck (2020) argued that they could also be viewed as an engagement display. In a series of examples during which co-participants touched themselves at the same time, Streeck further put forward the social and cooperative nature of self-touch, and the relevant use of the human body at meaningful transitional moments in social interaction (i.e., coordinated facial touch, collaborative drinking, and corrective interchange; see Streeck, 2020, pp- 11-16). Note that not every instance of self-adaptor or self-touch embodies a display of thinking or hesitation, but it may be one component contributing to the multimodal gestalt of doing thinking (see section 2.1.). Additionally, thinking postures which involve self-touch may be also regarded as *stylized*, as Heller (2021) and Bavelas & Chovil (2018) noted, and echo similar famous representations found in Art, such as "The Thinker" by Rodin, or even "The Scream" by Edvard Munch (cf Excerpt D further below). The following examples further illustrate this point. Embodied displays of thinking are carefully described in the transcriptions with regard to their synchronicity with fluenceme sequences, and only gesture strokes occurring outside these postures are mentioned in the main texts.

Thinking Posture (A)

The first excerpt is taken from Pair 3 in English in the SITAF Corpus, where the participants are discussing whether prisoners should have the right to vote. Marina, the non-native speaker spoke first and gave her opinion, mainly that prisoners should still keep their rights as citizens even though they are in jail, since they are going to go back to society eventually (omitted from the transcription). It is now Julia's turn to speak, and she feels somewhat conflicted about this topic.

```
      1     *JUL: yeah and you know what you said
                        **********
      ((both PUOH oriented towards MAR + gazes towards MAR))
            if they get out that's [//] they're eventually gonna l [//] be
                      *****                                          **
((brings her arms and hands to her left away from her center+ gazes towards
MAR))
            back in society and they're gonna live there so.
            ****                          ********
((brings her flat PUOH to her left with a rotating motion + gazes away))
      2     *MAR:     +< yeah.
```

((nods gazes towards NS + gazes up))
3 *JUL: hhh. um [!] yes a [//] and hhh. you know it's obviously
***
((brings her hands to her chest))
(0.600) a little bit um (0.720) conflicting.
******
((deploys lax flat open hands, fingers extended, palms facing her trunk))
4 *JUL: (be)caus:se you know I think of (0.570) you know like oh
****
((brings lax flat open hands forward + looks up))
what did they do to get in jail?
5 and now they're gonna choose you know
*** ****
((performs rotating motions with one wrist then two + looks up))
like their future or something.
******
((moves lax flat open hands upward, palms facing forward))
→ 6 *JUL: but I don't actually like **um (0.878) [!]**
[FLUENCEME SEQ.]
((scratches her cheek with index finger)) ((thinking posture))

I don't [/] I do:on't (0.909) [/] I don't feel like that's valid
****
((extends right PUOH with rotating motion))
(0.400) feeling like that.

Julia's thinking posture in line 6, characterized by several recurrent features identified earlier, mainly imaginative gaze, self-touch, and inflexible posture, closely resembles the one from the DisReg Corpus described in Excerpt 2.3., (Chap. 4), where the index finger is resting on the chin. This facial display is in fact a widely recognized and conventionalized embodiment of thinking or skeptical stance, and is also found in digital communication in the form of an *emoji* i.e., a small image encoded in text messages and other forms of digital communication, known as the *Thinking Face*

Emoji (🤔 see Gawne & McCulloch (2019). By performing this thinking posture, Julia manages to simultaneously mark epistemic stance towards her utterance, as well as her personal involvement towards the argumentative task at hand, with her use of the emotion verb "feel" in line 6. This *doing thinking* practice is also recognized as such by her conversational partner, who displays her understanding through head nods and shared gaze. As further marked in the multimodal transcription, the display of this thinking posture is perfectly synchronized with a complex fluenceme sequence comprised of a filled pause, an unfilled pause, and a tongue click. In this context, the tongue click may function as a stance marker, illustrating another slightly different function and turn position from the ones described in Chapter 4 (see Chap. 4, section. II.2.1.) It is interesting to note that a tongue click was also found during a similar thinking posture in Excerpt 2.3. from Chapter 4, which may reveal another recurrent vocal feature of *doing thinking*[167].

Thinking Posture (B)

The second excerpt is also taken from the same pair in the SITAF Corpus (Pair 3) but this time in French, where the roles are reversed, and they are talking about differences between being a traveler and a tourist (cf *Excerpt 2.A* from Chap. 3 and excerpt *Hold (A)* from this Chapter). Here the non-native speaker (Julia) is once more seeking confirmation from the native speaker regarding the choice of a word, just like she did in previous examples (cf Excerpt *Hold (A)* from section I.1.1., and *Excerpt 2A* from section II. 2.2.3 in Chap. 3).

```
      1     *JUL: parce-que:e (0.630) il est peut-être p [//] euh plus
            attentive euh à:à ce qui [//] à ses entourages?
            ********          ********          ************
         ((lax PUOH with bent fingers + moves her hands and arms in a circle
      away from her trunk + PUOH gesture))
      2     *MAR:             +< (nods) attentif
      2     *MAR:                          à son entourage (nods)
      3     *JUL:                                +< son entourage (nods)
→     4     *MAR: son enviro(nement) [//] ouais [/] son (0.248) [/] son
                                          [FLUENCEME SEQ.]
```

[167] However, let us not forget that fluencemes are multifunctional. While tongue clicks may function as emotional or interactional displays in certain interactional settings, they were also found to occur very frequently in contexts of lengthy talk presentations, where they mostly indexed new sequences of talk. This further reflects the effect of setting on the use of fluencemes (see Chapter 4).

((series of thinking postures))

5 *JUL: (laughs)

entou(rage) [//] son entourage oui non ça va (laughs)

****

((finger from left hand extended pointed towards NNS+ gazes towards NNS))

6 *JUL: +< oui son entourage (smiles)

Here Marina performs a series of thinking postures, characterized by eyebrow frowns, gaze withdrawals and alternated head movements, following the repair sequence initiated by Julia in line 2. Once again, the non-native speaker made use of her voice, tone, and gesture to seek confirmation from her interlocutor, regarding the choice of the word "entourage". The latter is thus positioned as a native "expert" fluent speaker within the overall tandem setting. She first corrects Julia's use of the wrong possessive pronoun in the plural form "ses entourage" (l. 1) to a singular form "son entourage" (l.2), followed by Julia's repetition of the target noun phrase in line 3, marking the end of the repair sequence. However, in this case, Julia does not resume her argumentative activity right away, as Marina seems to be experiencing doubts about the use of this specific lexical word as well, and begins to search for other relevant options (with the truncated word "enviro" in line 4 which probably meant "environement"). We can thus note a shift from Marina's initial other-repair sequence towards a self-oriented one, characterized by a series of truncated words and word repetitions, during which she also looks up, frowns, and moves her head to the left then to the right, as if she was weighing up the different options. This was also found in Excerpt *Retraction (A)* (cf section I.1.2.) where the non-native speaker performed a *weighing-up* gesture, further embodying her decision-making activity. This display is once again deeply representative of the act of hesitating, whereby speakers need to make a decision before pursuing the next move. Marina then brings her hand and index finger to her chin, but this time her index finger is raised, and is brought to her mouth. This raised stretched index finger, documented in Bressem & Müller (2014) and observed in *Excerpt 2.2*. in Chapter 4 (section II. 2.2.3.) can be used to draw attention to other

participants, as we have seen earlier, but they may also signal thematic shifts, which may be the case here, as Marina shifts back to a joint activity. Indeed, she then extends her left arm towards Julia, with the same index finger pointed towards her, as she finally retrieves the lexical item initiated by her partner "son entourage".

Thinking Posture (C)

The third excerpt is taken from Pair 09 in the SITAF Corpus introduced in Chapter 3 (see Excerpt 3A, Chap. 3, section II. 2.2.2.) in the exchange in French where they are also discussing whether prisoners should have the right to vote. The native speaker (Emilie, A09) speaks first and seems a bit at loss for words as she has never thought about this type of issue before, but then provides a series of arguments in favor of voting[168].

→ 1 ***EMI:** mais **euh (0.840) euh wow (laughs) je sais pas quoi dire eum (2.120)**

((EMI gazes away, touches her ear, her body oriented to the side))

((ART gazes towards paper, hand resting on chin, body crouched))

2 bah c'est pas parce-qu'ils sont prisonniers que:e (0.410) i:ils pourraient pas voter

((Emilie slightly moves her body towards ART, gazes towards him))

((ART remains in the same inflexible thinking posture))

3 *EMI: ils font quand même parti d'une nation.

[168] This example was also briefly shown to members of a research group on tandem interactions within *SeSyLiA* at Sorbonne Nouvelle University in March 2021.

4 *EMI: et euh (0.490) c'est pas parce-qu'ils sont euh [//] ils ont fait u:un [/] un crime que:e (0.470) [//] fin qu'ils ont pas c [//] qu'ils ont quand même pas leur [/] leur mot à dire sur euh (0.430) [/] sur un président sur la politique puisque ça les concerne aussi.

5 *EMI: indirectement même s'ils sont un peu en autarcie quand ils sont en prison.

6 EMI: euh fin ils sont concernés par (0.480) [/] par le gouvernement par la politique et +/.

→ **7 *ART: +< hhh. ou:ii ma:ais** je sais pas parce-que quand euh (0.410)

[FLUENCEME SEQ.]

((readjusts his glasses; gazes towards EMI; winces + looks up))

quand euh (0.410) être a empri(sonné) [//] euh emprisonné.

8 *ART: c'e:est [/] (1.030) c'est déjà euh

****

((extends his left hand to his upper lower right periphery))

parce-que normalement on est libre

9 *ART: si on est emprisonné c'est [/] c'est parce-qu'o:on [/] on a fait un crime et euh [/] et donc euh o:on euh abandonne euh le droit d'être libre.

10 *ART: alors pourquoi pas le droit de vote?

Three instances of thinking postures can be observed in this excerpt. The first two are deployed together almost simultaneously by the two different speakers at the beginning of the excerpt, when Emilie takes the turn with utterance-initial fluencemes (*FP*+*UP*+*FP*) as she rotates her body to the side, gazes away, and touches her right ear. She remains in this position until the end of her second fluenceme sequence comprised of an explicit editing phrase "je sais pas quoi dire", a nasalized filled pause ("eum") and an unfilled pause of significantly long duration (2220 ms), indexing her next piece of talk. Similarly, her partner Arthur also displays a very characteristic thinking posture, which closely resembles the posture depicted by Rodin in *The Thinker* (cf Fig. 69 at the end of this section), a famous sculpture representing a male individual in deep

thought; his body is crouched, his hand resting on his chin, and he displays a salient thinking face. Arthur's posture is almost identical, and it is further relevant to note that his whole body remains in this fixed position until line 7 when he takes his turn. Even though these two thinking postures were initiated at the same time by the two speakers at the beginning of the sequence, as they were engaged in collaborative joint thinking, the exchange takes a whole different turn when Emilie shifts from her solitary thinking activity and turns her body towards Arthur in line 2, as she begins to provide a number of arguments in favor of the topic. However, Arthur remains in the same inflexible posture, and completely withdraws himself from the interaction, as he shows utter disengagement from the ongoing activity, marked by his lack of vocal and visual participation, i.e., he does not respond to any of what Emilie is saying with either vocal or visual backchanneling (e.g. head nods) at *backchannel relevant spaces* (Heldner et al., 2013) i.e., intervals where it is relevant for other speakers to provide backchannel. Note that this type of asymmetrical positioning was also found during their exchange in English where the roles were reversed (see *Excerpt 3a*, section II. 2.2.2., Chap. 3) and when the French non-native speaker constantly gazed towards the piece of paper and showed little participation in the ongoing exchange. As Heller (2021) noted, recipients' displays of *doing thinking* function slightly differently from the ones initiated by speakers, as they may project an intention to take the turn, or serve as a signal of *upcoming disagreement* in the next turn. In this case, Arthur's change of thinking posture in line 7 indexed a change of orientation as well, within the trajectory of his upcoming turn, as he interrupted Emilie's prior turn to display his disaffiliation (as cued by the discourse marker "mais" and a non-response "je sais pas"; he later argues that prisoners should *not* have the right to vote). We can see his body shifting back towards the interactional space, his head raised with a wince, as he readjusts his glasses. This display is also perfectly synchronized with a fluenceme sequence initiated at the beginning of his turn ("hhh. ou:ii ma:ais") following the VOC+NL pattern.

Thinking Posture (D)

The final excerpt is taken from the conversational exchange between Dan (D1) and Laura (D2) from the DisReg Corpus. They are talking about the assignments they are

preparing for their class presentations, and more specifically about the relationship between servants and masters in Molière's play *Le Malade Imaginaire*[169].

1 *LAU: hhh. mais Toinette du coup n'est pas du côté de sa maîtresse pour le coup
2 *LAU: hhh. elle est plutôt du côté bon de son autre maîtresse la fille
****
((performs a cyclic rotation with both hands then brings them down))
→ **3 *LAU: mais en l'occurrence (0.571) mmm +/.**
[FLUENCEME SEQ.]
((thinking posture + leans her head to her left and right))

3 *DAN: bah moi j'ai plus l'impression que (0.572) es [//] euh fin que Argan c'est vraiment son maître.
******
((brings extended fingers to the table+ gazes towards LAU))

Laura displays another type of thinking posture in line 3 in turn-final position, which, to our knowledge, has not yet been documented in previous studies[170], and presents slightly different characteristics. We find instances of frowning, looking up, and wincing, which are, so far, recurrent features of doing thinking, but the speaker also brings both her palms to her cheeks, her fingers almost touching her ears. Just like the native speaker Marina (F03) from the *Thinking Posture (B)* example, she also leans her head to her left and right in alternated motions, further embodying the action of decision-making, which may be another recurrent feature of doing thinking and doing hesitation (summarized in Table 40 at the end of this subsection). Once more, this posture is perfectly synchronized with the emergence of a fluenceme sequence

[169] This example was also shown to members of the *Co-Op Lab* in UCLA in February 2020.
[170] In fact, Ellinor Ochs thought that this type of posture was very peculiar when she watched the excerpt during the data session.

(following the VOC+NL pattern), which also becomes a transition relevant place for Dan, who takes his turn and gives his opinion on the current matter.

Note that a similar facial display was also performed by Paul during his class presentation (cf Figure below, from *Excerpt 1.3*, Chap. 4, section II. 2.2.4), so we may wonder the extent in which it may be considered a thinking posture as well. Even though he is not frowning nor wincing, he still brings his palms to his cheeks, and looks slightly dismayed. More work needs to be done on the potential significance of this gesture, which can be considered a self-adaptor in some contexts, but which may also bear other forms of symbolic or social meaning.

***Figure 68**. Paul (B1) bringing his palms to chis cheeks during his presentation (taken from Excerpt 1.3., Chap. 4, section II. 2.2.3.)*

This posture may also echo another famous art work *The Scream*, a painting by Expressionist artist Edvard Munch, depicting a man standing by a bridge, screaming, with a horrified look on his face, his palms placed on his cheeks, covering his ears. Even though the facial expressions are radically different, our two speakers displayed a similar posture, hinting once more to the fact that embodied displays of thinking in social interaction bear stylized or iconic properties, and are in permanent interaction with popular culture and art (Boutet, 2018).

As Müller (2015, p. 453) rightly put it: "Gesturing hands are so intricately bound to the act of speaking that they function as an icon for speaking in the visual arts". This was shown across our four excerpts, where several references to works of art or digital illustrations were mentioned, as summarized in Figure 69.

***Figure 69**. Thinking displays as stylized and iconic postures imbricated into art and popular culture*

To conclude, our different analyses of embodied displays of thinking or hesitation documented in this section have illustrated several recurrent features of this social practice, emerging within complex multimodal gestalts (Heller, 2021, Mondada, 2014) or compound enactments (Debras, 2017, Streeck, 2009). Following the work of Heller (2021), we introduced several embodied practices of doing thinking, which were shown to function as relevant displays in specific interactional sequences (repair, disaffiliation, or turn projection) involving multiple modalities (vocal fluencemes and a series of facial and body displays). This further acknowledges the embodied nature of hesitation, and to a larger extent of inter-(dis)fluency, characterized by a number of multimodal markers of suspension (speech, hand gesture, and body suspension) to manage the fluency of multimodal discourse. Table 40 summarizes the different visual and vocal resources identified in this section.

***Table 40**. Summary of features embodying practices of doing thinking*

FACIAL AND BODY DISPLAYS
- Imaginative gaze
- Squinting and wandering of the eyes
- Eyebrow frown
- Raising one's head up and looking up
- Self-touch (hand resting on chin, or touching one's ear or face)
- Body crouched, or oriented to the side in an inflexible posture
- Head leaning back and forth to the left and right
VERBAL AND VOCAL DISPLAYS
- Nasalized vocalizations and filled pauses (*mm, eum, um*)
- Tongue clicks (*tsk, ttut*)
- Silent pauses
- Explicit editing phrases (*je sais pas quoi dire, I ain't got the word*)

2.1.2. Cyclic gestures and the searching activity

So far, we have explored the display of salient thinking postures across languages and settings in particular interactional sequences, and they were all characterized by a number of recurrent visible features, such as inflexible posture, imaginative gaze, frown, or self-touch, suggesting a change of orientation towards a "world of thought". This was also conveyed by a suspension of the speakers' visible bodily activities, which echoes our analyses of gesture holds in section 1.1. However, embodied displays of thinking or hesitation can also be more dynamic, and involve joint collaborative word searches marked by specific gestural actions. While Heller (2021) or Bavelas & Chovil (2018) focused more specifically on stylized inflexible postures, we may also find other instances of *doing thinking* that involve a richer gesturing activity. In Chapter 2, we mentioned Gullberg's use of the term *thinking gestures*, which are metapragmatic gestures that do not relate to referential content[171] but rather comment on "the breakdown itself" (Graziano & Gullberg, 2018), and the latter are also known as

[171] However, it should be noted that Ladewig (2011 , 2014) claimed that some cyclic gestures could be used with a *referential* function to illustrate a mental process such as scooping or thinking. She also further noted that gestures can perform a referential and a pragmatic function simultaneously, so the line is not always so clear. In our data, we systematically coded "thinking gestures" (which includes cyclic gestures, but also other types of gestures such as finger snaps) as a subtype of "pragmatic gestures" but we are aware that this type of annotation is limited to a certain extent. We further discuss our limitations in the Conclusion.

Butterworth or "conduit" gestures (in McNeill, 1985 and Tellier & Stam, 2012), and refer to gestures used for word searches specifically (cf Chap. 2, section II. 2.2.3). As Graziano & Gullberg (2018) pointed out, many of these word-searching gestures involve a rotation of the wrist, and they have in fact been extensively documented in Ladewig's work on recurrent gestures (Ladewig, 2011, 2014; Ladewig & Bressem, 2013) under the label "cyclic gesture". As explained earlier in our methodology (cf Chap. 2, section II. 2.2.3) the category "thinking gesture" was used in our functional classification to refer to a large class of metapragmatic gestures used to enact word searching and thinking activities, in line with Gullberg (2011) and Graziano & Gullberg (2018), but without annotating their formal features, to remain consistent with our functional criteria. As we shall see in this chapter, some of these gestures involve cyclic movements, but we may find other types of gestural activities, such as finger snaps, or finger tapping, explored in section 2.1.3. The aim of the present subsection is to analyze the forms and functions of cyclic gestures more specifically, during and outside fluencemes, and see how they may further relate to the multimodal practice of doing thinking within embodied word-searching activities. We first summarize the different cyclic gestures illustrated in our previous analyses and report on their forms and functions, then move on to the ones that were used for word searches more specifically, to analyze their coordination with fluencemes within situated practices.

It should first be noted that cyclic gestures perform a wide range of functions other than lexical search, which have been consistently documented by Silva Ladewig and members of the ToGoG group (Müller et al., 2013 in Chap. 1, section III. 3.3.3.). As explained earlier in Chapter 1, this body of research largely focused on a form-based approach to gesture, starting with the assumption that the articulation of hand shapes, movements, positions, and orientations of bodily behavior played a central role in the formation of meaningful units of visible action. They further view gesture as a form of "embodied conceptualization" (Müller et al., 2013) whereby gestures make sense of meaning through embodied sensory experience. They introduced the term "recurrent" gestures (also known as *pragmatic* gestures, see Table 3 in Chap. 1, section III. 3.3.3.[172]) to refer to conventionalized gestures with a stable form-meaning relation that

[172] As explained throughout Chapters 1 and 2, we do not use the terms "recurrent gesture" nor "cyclic gesture" in our quantitative annotation model because we want to remain consistent with functional criteria (while "cyclic" refers to a rotating movement, "recurrent" refers to a conventional character, and none of these terms refer to a function). The aim of the present section is to provide more qualitative form-based analyses as a supplement to our previous quantitative annotations.

is used repeatedly across different contexts of use. Ladewig (2011; 2014) focused more specifically on the cyclic gesture, which, as introduced earlier, is characterized by "a continuous circular movement of the hand, performed away from the body" (Ladewig, 2001, p. 3). The formal and semantic core of this gesture, Ladewig (2011) added, is based on the image schema[173] CYCLE (Johnson, 1987), and is involved in several metaphorical mappings such as TIME-IS-MOTION-THROUGH-SPACE or BODY-IS-MACHINE. The basic meaning of the cycle (i.e., continuity, process, or duration) is hence reflected in all instances of the gesture, but it also varies according to its context of use. Ladewig (2011) further identified 4 different contexts in which cyclic gestures were found to occur most frequently, in her data of German every-day conversations between relatives and friends, listed as the following:

- a word or concept search,
- descriptions,
- a request,
- an enumeration.

Cyclic gestures were found to occur most frequently in contexts of lexical search, which already suggests a close relationship between this specific type of recurrent gesture and the notion of inter-(dis)fluency. In fact, Ladewig (2014, p. 1563) made an interesting comment regarding the production of cyclic gestures during pauses:

> By representing the ongoing searching process, these gestures fulfill the same function as verbal disfluency markers, mainly indicating that the speaker is engaged in a searching process. As such, when used without speech, these variants of the cyclic gesture replace verbal markers of hesitation and work as a turn-holding device.

However, as argued several times before, we do not exactly agree with the idea that these gestures "*replace* verbal markers of hesitation", as we believe that it is rather a matter of *contextual configuration* (Goodwin, 2000), i.e., a selection among the most relevant semiotic resources available to the speaker in a specific context (also following the *SrB*, Cienki, 2012, 2015). In addition, the claim that such gestures replace verbal markers systematically assumes that hesitation and (dis)fluency phenomena are *only*

[173] An image schema can be defined as the following: "a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience" (Johnson, 1987, p. 14)

produced in the verbal modality, which, as argued multiple times before, is too restrictive. We will return to this point at the end of this subsection. Table 41 summarizes the different cyclic gestures found in our previous analyses of the SITAF and DisReg Corpus in Chapters 3 and 5. It includes a description of the gestures and the contexts in which they occurred, as well as information regarding their temporal coordination with fluencemes, and their functions. As Ladewig (2011, 2014) noted, cyclic gestures can perform both pragmatic and referential functions. They can be used to depict an ongoing action, process, or abstract entity, reflecting the semantic core of *cyclic continuity* (marked by the image schema CYCLE as explained earlier), and the continuous circular movement of the gesture.

***Table 41**. Summary of cyclic gestures found in previous analyses (Chap. 3 and 5)*

Illustration	Example	Context	Within/outside fluencemes	Gesture description	Potential function
	Excerpt 1B (Chap. 3)	French non-native speaker reformulates her previous utterance	Outside, co-occurs with the utterance "I repeat"	Both hands raised to the upper center space, index fingers extended; perform a series of rotating movements in alternating motion	Referential; refers to the process of starting over
	Excerpt 1B (Chap. 3)	French native speaker completes NNS' previous utterance and provides a list of lexical items	Outside; co-occurs with the noun phrases "religieuses politiques"	Right flat open hand,palm facing her trunk; performing a series of rotating movements in the center space; oriented towards interlocutor	Referential; refers to an abstract enumerating process
	Excerpt 4a (Chap. 3)	American native speaker invites NNS to speak by initiating an utterance "social media is like..."	within (complex fluenceme sequence UP+IR)	Right flat open hand palm facing left, moving in a rotating motion in the right periphery of the gesture space; oriented towards interlocutor	Performative; used to encourage the interlocutor to speak
	Excerpt Retraction (B) (Chap. 5)	French native speaker is talking about episodes that come out every week on TV	Outside, occurs during the utterance: "chaque semaine tu dois te remettre dans le truc"	Right flat open hand, palm facing her trunk, moving in a rotating motion in her center space	Referential; refers to the weekly routine of watching an episode every week
	Excerpt Preparation (A) Chap. 5)	American native speaker is talking about the monotonous aspects of her life as a teenager	Outside; co-occurs with the noun phrase "the same", then "the same day after day"	Right flat open hand, palm facing out, moving in a rotating motion in the right periphery of her gesture space	Referential; refers to a daily routine

In a majority of the examples above (except for *Excerpt 4.a*), the cyclic gestures performed referential functions[174], as they referred to continuous and habitual

[174] However, as noted before, referential and pragmatic functions may also overlap, which is why the term "*potential* function" is used in the table.

processes, such as a weekly or daily routine, or the process of enumerating or starting over. This semantic core of cyclic continuity is also found in excerpt 4.a, but not in a referential way, as the gesture does not exactly refer to a process per se, but rather fulfills a *performative* function (Müller, 1998), enacting an interactional move (Kendon, 1995), by inviting the interlocutor to take the turn. This type of function was also illustrated in Ladewig (2011) in her data of German conversations. In addition, these five gesture variants show similar formational patterns with respect to hand shape and orientation (flat open hand moving away from the body and oriented towards interlocutor) except for *Excerpt 1B* where the speaker used both her hands and held out her index fingers, and Excerpt *Preparation (A)* where the palm is facing out. In sum, the examples documented above share close formational and functional characteristics with previous cyclic gestures documented in previous studies (e.g. Bressem & Ladewig, 2013; Ladewig, 2011, 2014; Müller et al., 2013), thus giving more support to their conventional nature. While the present section does not intend to dwell on the analysis of recurrent gestures, the aim of this short introduction on cyclic gestures is to explore its relation with inter-(dis)fluency and embodied displays of thinking. Out of the five examples summarized above, only one of them occurred during fluencemes, and served a performative function. Further in line with our view of fluency and hesitation as an embodied practice, we shall now turn to the analysis of two examples from the data where cyclic gestures co-occurred with fluencemes in contexts of word search. As explained before, cyclic gestures that are executed during fluencemes typically perform word-searching functions, as Ladewig & Bressem (2013, p. 218) put it:

> In its primary use, the context of a word/concept search, the cyclic gesture presents the searching activity as a continuous activity, thereby reflecting the semantic core of cyclic continuity. It fulfills a performative function, more precisely a meta-communicative function, showing that the speech activity of searching for a word/concept is in progress.

Once again, this semantic core of cyclic continuity is conveyed in the process of searching for the next word, phrase, or concept, which is another inherent feature of inter-(dis)fluency e.g., Tottie (2014) who suggested the term *planner* to talk about filled pauses, which were shown to serve planning and word-searching functions (cf Chap. I, section II. 2.2.1). The two following analyses further illustrate this point.

Ex. Cyclic Gesture (A)

The first excerpt is taken from Pair 10 of the SITAF Corpus during their exchange in English. Here the two tandem partners are discussing whether tuition fees at university guarantee a better quality of teaching. The non-native speaker (Juliette, F10) spoke first and argued against the topic, saying that tuition fees may treat some students unfavorably, especially the ones who cannot afford it.

```
    1  JULT: um (0.800) that's m:make uh (sighs) an [/] an handicap
                          ~~~~~~~*********-.-.-.-.-.-.-.-.-          *****
((holds out her hands in the lower center space, with bent fingers, palms
facing up+ produces a beat motion downwards + repeats a similar gesture with
a rotation of the wrists))
    2  BET:  oh yeah yeah yeah yea:ah.
                                 *****
                          ((extends her left PUOH towards Juliette))
    3  JULT: hhh. uh (0.370) because uh they have to work maybe
→   4  JULT: uh o:or [/] or um they [!] hhh. [/] they have to be priv(ate)
                 *****       ****
((raised shoulders, rotation of the wrists with fists closed in the lower
center space))
       uh (0.400) yeah (laughs)
  [FLUENCEME SEQ.]
       ***********-.-.-.-.-.-.-.-.-
((lax flat palm-up open hands moving in a rapid sequence of rotating movements
+ winces + hands return to rest position))
```

```
    5  BET: yeah (nods)
    6  JULT: so private or about something or yeah.
    7  BET:                              +< yeah.
```

Juliette produces a cyclic gesture in line 4 in the transcription. She uses both her lax flat open hands, palms facing upwards in her center space, and moves them away from

her trunk in a series of quick rotating motions. Her gesture is synchronized with a complex fluenceme sequence produced in utterance-medial position, which is comprised of a filled pause and an unfilled pause. We can also see her wincing and gazing towards her interlocutor, who is fully oriented to Juliette's talk, as cued by her series of vocal and visual backchannels (*yeah* and head nods). The facial expression displayed during the cyclic gesture shares similar features with the ones found in *thinking postures* described earlier (cf section 2.1.2), although this time it is accompanied by a gestural action. In this case, the gesture occurs in a context of word search, and Juliette first initiates a lexical item in a truncated form ("private"), abandons the search (marked by the utterance final "yeah"), then re-introduces the target word in her subsequent utterance (l. 6). As Ladewig & Bressem (2013) and Graziano & Gullberg (2019) showed, such variants of the cyclic gesture perform a meta-pragmatic function, signaling that a search is currently in progress. The word search is hence not only manifested in the vocal channel, but in the visual-gestural channel as well, giving more support to the multimodal nature of inter-(dis)fluency phenomena. In this case, it was deemed more relevant to combine vocal and visual-gestural resources to embody a word search than to use speech alone, perhaps because of Juliette's expressive difficulties (note in line 1 her wincing and the tension in her downward beat movement before retrieving the word "handicap"). Indeed, the non-native speaker seems to be struggling with her current word search, suggesting that the cyclic gesture may further convey the effort and endeavor associated with the word-finding activity, marked by the series of rapid circular movements. However, while certain multimodal word searches may reflect instances of difficulties in contexts of L2, they are not necessarily associated with trouble, as shown in the next example.

Ex. Cyclic Gesture (B)

This is taken from Laura's (C1) presentation in which she is introducing the first scene of the play (*Le Malade Imaginaire),* known as *exposition* – when information about the characters and their background is revealed, as to set the scene.

```
      1     on parle au théâtre d'une scène d'exposition
                          ******
((brings her two PUOH forward on the table with bent fingers))
      2     on pourrait dire que ces personnages sont en quelque sorte
                         *****                          ***************
```

((brings her two PUOH forward higher in the gesture space with bent fingers; gesture executed with a rotation of the wrists))

hhh. euh les eum [!] [/] les exposantes
*****************1. ***********
[FLUENCEME SEQ.]

1.

((1.series of circular rotations)) ((PUOH brought forward with bent fingers))

eum hhhh. du début de la pièce (**0.965**) **[!]** xxx
***********2. ******3.**
[FLUENCEME SEQ.] [FLUENCEME SEQ.]

2. 3.

((2.series of circular rotations + PUOH brought forward with bent fingers))
((3. right hand moving with a single rotation of the wrist))

Here the speaker is producing a series of *palm-presenting* gestures, where the Open Hand Supine is "presented" or "displayed" into the frontal space to introduce elements in discourse (Kendon, 2004, p.265). Note that the latter were identified as *discursive* gestures in our quantitative analyses, as they are mostly used for discourse structuring and emphasis. Indeed, the gestures in this excerpt are executed with a rotation of the wrists, and the strokes coincide with target words in Laura's discourse (e.g., "théâtre", "exposantes") as she is also gazing towards her audience as to seek their attention and maintain visual contact[175]. Between these palm-presentation gestures, she produces three different cyclic gestures in line 2 (see illustrations, 1, 2, and 3). The first two take

[175] Note that, Laura spent slightly more time gazing towards her audience during class presentations (about 30% of the time, as opposed to 17% for the average of her group).

up most of her gesture space and involve a series of repetitive circular rotations with her lax flat palm open hands, her palms facing her trunk, as she either raises her index finger from her right hand in the air (see 1), or brings her bent fingers upward (see 2). The gestures also co-occur with fluenceme sequences produced in utterance-medial position; the first one is comprised of an inbreath, a nasalized filled pause and a tongue click, and a repetition of the pronoun "les" ("hhh. euh les eum [!] les"), and the second one comprises another nasalized filled pause and inbreath ("eum hhh".). These two sequences are produced within a complex noun phrase, before the head noun ("les exposantes") and a prepositional phrase ("du début de la pièce") which functions as a post-modifier of the head. Laura is thus elaborating on the role of the characters in this exposition scene, and re-uses the noun "exposition" to use it as an adjective "exposante" to qualify the characters from the play. Once more, this searching, or rather *elaborating* process involves vocal and gestural actions. The third cyclic gesture, which is only executed with the right hand in the lower center space with the index finger extended, and involves a single circular rotation of the wrist, is also produced during a fluenceme sequence made of an unfilled pause of 965 milliseconds, and a tongue click. Note that the speaker performed a similar swallowing activity during the production of the tongue click, which is consistent with a previous analysis provided in Chapter 4 (e.g. excerpt 1.2). This time the fluenceme sequence is produced near the end of Laura's utterance before an unintelligible word. In addition, these repetitive circular rotations (also found outside fluencemes preceding the palm-presenting gestures) mark a continuity with what is being said in her multimodal discourse; her repetition of the same cyclic movement during the vocal fluencemes thus further contribute to the *fluency* of her talk. Unlike the previous example, these variants of the cyclic gesture are not exactly associated with word-finding difficulties but rather embody a thinking and elaborating process as cued by the *repetitive* cyclic motions.

To conclude, the notion of cyclic continuity, process, or duration, which is inherent to the cyclic gesture, may also apply to the notion of fluency. As we have seen, cyclic gestures are conventional, recognizable recurrent gestures that are very frequent in spoken discourse and serve a number of referential and pragmatic functions. When the latter co-occur with fluencemes, they further serve meta-communicative functions as a comment on the "breakdown" (Graziano & Gullberg, 2018) and the word searching process. While we do not believe that such gestures "replace" fluencemes

(Bressem & Ladewig, 2013), or that fluencemes are necessarily associated with a "breakdown", we argued that they further took part in the multimodal activity of looking for a word or a concept, or to a larger extent of doing thinking, which involves a combination of vocal, verbal, and visual-gestural activities. The image schema CYCLE, which, as Ladewig showed, is inherent to cyclic gestures, also strongly echoes the notion of continuity, which can largely be attributed to a continuous flow embodying *fluency*. Unlike our previous analyses illustrating the temporal coordination between fluencemes and gestural phases of suspension or interruption (sections 1.1.), illustrating a "breakdown" in the two modalities, this section has further illustrated the way fluency may also embody the notions of progression and continuity.

2.1.3. Other gestural practices of doing thinking

So far, our examples have looked into the embodied practice of *doing thinking* based on the analysis of thinking postures (Heller, 2021) and cyclic gestures (Ladewig, 2014), which have been thoroughly documented in the literature. As we have seen, thinking postures may function as relevant displays of doing thinking, signaling that a speaker is currently oriented towards a "world of thought", and cyclic gestures, when they are produced during fluencemes, may signal that a word search is currently in progress. While thinking postures tend to be characterized by an inflexible posture and gaze withdrawal (self-oriented), cyclic gestures involve dynamic repetitive circular motions and mutual gaze (other-oriented). Both of these practices, however, as argued, may embody the practice of doing thinking, doing hesitation, or doing *searching*, which involve a temporary suspension or interruption of the current activity to either retreat into a more solitary practice (DISfluency), or maintain its ongoing flow (INTER-FLUENCY), reflecting once more different degrees of inter-(dis)fluency.

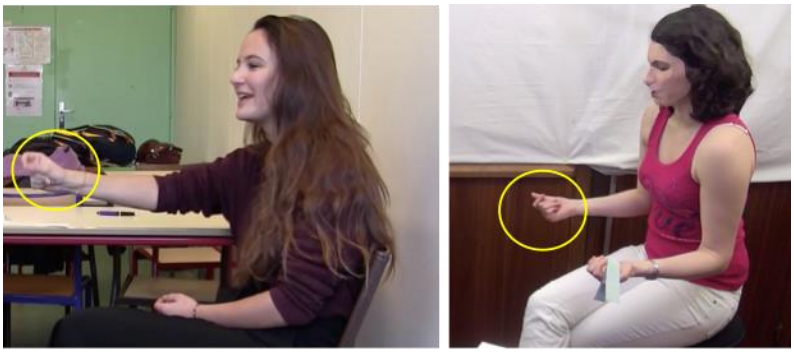
We shall now turn to other gestural practices of doing thinking, which, unlike cyclic gestures, have not received as much attention in the literature. The first one is the finger snap. This involves a forceful and rapid movement of the thumb towards the index finger, usually performed with a single hand, creating a snapping or clicking sound. Unlike the cyclic gesture, the finger-snap gesture has not been much documented in previous papers, to our knowledge. This may be explained by the fact that it is rarely found in video-recorded data of conversations, and is used in entirely

different contexts (e.g. at a musical performance[176]). A quick reference to this gesture is found in Poggi's (2001) paper on her typology of gestures. In one section, she focuses on the semantic content of gestural actions, whether they communicate about abstract or concrete objects, or convey speakers' mental states (i.e., beliefs, goals, and emotions). In particular, she mentions the types of gestures that display metacognitive information towards a speaker's utterance, which is very similar to the types of metacommunicative functions described earlier in section 2.1.2. She claims (Poggi, 2011, p. 3):

> we provide metacognitive information as we inform about the source of what we are saying: we may be trying to retrieve information from our long-term memory, as we imply when we "snap thumb and middle finger" (=I am trying to remember); or we may try to concentrate, that is, to draw particular attention to our own thinking, as we imply by "leaning chin on fist" (Rodin's Thinker posture)

Interestingly, Poggi also makes a reference to the thinking posture described earlier in section 2.1, but she does not provide any example or illustration of these gestures in context. In addition, while she focuses on the externalization of "mental states" through gesture (in line with the *cognitive-psychological* approach to gesture, (cf Chap. 1, section III. 3.3.2), we maintain that such manifestations should be regarded as relevant social and interactional displays, which emerge from specific participation frameworks in particular multimodal settings, following Heller (2001) and Goodwin & Goodwin (1986). In addition, as Streeck (2020) further emphasized, manual gestures should not be viewed as *mere reflections* of the inner workings of the mind, but rather as dynamic actions abstracted from haptic acts, i.e., touch, physical actions, or manipulation of objects. It is still relevant to note, however, that finger snaps, just like cyclic gestures, perform meta-communicative functions, as they embody the process of thinking, searching, or "remembering", according to Poggi. In Chapters 2 and 3, we briefly mentioned the emergence of finger-snap gestures in two excerpts, which are illustrated in the figure below.

[176] For more information, read https://www.nytimes.com/2015/11/22/fashion/snapping-new-clapping.html (last retrieved on June 12th 2021)



Excerpt 2.3 (Chap. 4) Excerpt 1a (Chap. 3)

Figure 70. *Occurrences of finger snaps in previous chapters*

In *Excerpt 2.3.*, the finger-snap gesture occurred near the end of the excerpt, when Jenny resumed her storytelling activity, following the digression sequence initiated by Alex. The gesture was not performed on its own, but was coupled with other visible behaviors, as it was performed right after Jenny's thinking posture (cf section 2.1.1., Table 39), and immediately before her interactive gesture by which she extended her left arm and PUOH towards Alex to display mutual understanding (see Fig. 65 in section II. 2.2.3, Chap. 4). The finger-snap gesture was also produced at the end of Alex's pause. In this case, the gesture seems to index a change of participation from a solitary thinking activity to an intersubjective one, which reflects the same type of function served by tongue clicks, which in fact create a similar clicking sound performed with the tongue (cf Chap. 4, section III 2.2.2. on tongue clicks).

In *Excerpt 1a*, the finger-snap gesture emerged in a word-searching context in L2 during a complex fluenceme sequence within Sally's discourse, as she explicitly signaled her word-finding difficulties ("I don't know how the word") thus directing her attention (and her interlocutor's) towards the current search-in-progress, presenting it as a relevant activity to pursue the ongoing exchange. Here the snapping motion was performed twice with the left hand, which was previously held, as Sally expressed her expressive difficulties. This occurrence of the finger-snap gesture functioned very differently from the one described in Excerpt 2.3. as it rather offered a meta-communicative comment on the current search, a practice also found in cyclic gestures. A very similar pattern is also found in the following example.

Ex. Thinking Gest - Finger Snap

This excerpt is taken from Pair C in the conversation-session (DisReg) where Laura (C2) and Dan (C1) are still talking about the role of Toinette in Molière's *Le Malade Imaginaire (cf* Excerpt *Thinking Posture (D)*. Dan suggests that Toinette (the servant)

is kind of acting like a second wife towards her master (omitted from the transcription), and he further elaborates on this idea.

1 DAN: ba:ah elle est très maternelle avec lui +/.
2 LAU: +< ouais c'est vrai qu'elle le gronde plus au début
3 DAN: genre euh +//.
4 DAN: ouais
→ 5 DAN: pis même à un moment elle hhh. fin **mmm comment dire euh**
[FLUENCEME SEQ.]
****
((head oriented to the right; snaps his fingers twice))

un moment elle lui demande de prendre son traitement
****
((right hand held + gazes towards LAU))
fi:in je trouve ça presque ironique
****
((extends his left PUOH towards LAU and performs a series of rotating motions))
6 LAU: oui après c'est pour un peu se moquer d'elle je pense (laughs)
7 DAN: +< voilà

Similarly, the finger-snap gesture performed by Dan also emerges in a searching context. Here Dan is trying to find an appropriate and relevant illustration of Toinette's role as a mother figure in the play. He does so by first qualifying her as "maternal", which Laura agrees with, and then begins to search for a specific moment from the play ("à un moment" l. 5). This is conveyed by his complex fluenceme sequence, comprised of a nasal vocalization, an explicit editing phrase and a filled pause ("mmm comment dire euh") during which he initiates the finger-snap gesture. Just like the non-native speaker from Excerpt 1a, the snapping motion is performed twice with the left hand in the upper lower gesture space. However, unlike Sally the L2 speaker, Dan is not experiencing language-related difficulties, but is rather concerned

with the relevance of what he is about to say[177]. The snapping motion may further conceptualize the dynamic and embodied act of thinking, derived from the motor activity and clicking sound of the gesture, which further creates a kinesthetic and tactile experience for the speaker (Streeck, 2021) to seek or grasp his next piece of discourse to present to his interlocutor. This is also visible in his gaze, as he is gazing away while performing the snapping motion, further reflecting a temporary disengagement from the interactive task at hand, which echoes our previous analyses of thinking postures characterized by gaze withdrawal (cf section 2.1.1.). In addition, this example shows that the practice of *doing thinking* is not only characterized by temporarily "frozen" thinking postures, but may involve dynamic gestural actions as well, such as cyclic or finger-snap gestures.

This leads us to a third type of thinking gesture variant found in the data, characterized by a tapping of the fingers. This is illustrated in the two examples below, taken from DisReg.

Ex. Thinking Gest – Finger tapping (A)

This is taken from the same exchange between Laura and Dan, and Laura is talking about whether servants are truly loyal to their masters.

1 LAU: hhh. mais c'est peut-être ça qui serait intéressant parce-que euh (0.400) du coup (0.557) est-ce que les serviteurs aussi ils assurent la loyauté
****
((taps fingers from her left hand on the table twice))

→ 2 LAU: mais **eum (1.042) [!] d'un autre côté (1.265)**
****
((taps her fingers four times by raising her fingers higher + looks up))

et j'ai oublié que j'allais dire

[177] In fact, when Dan first suggested the idea that Toinette acted like a mother figure towards her master, Laura did not seem very convinced and reacted with surprise, which may explain Dan's motivation to look for relevant arguments in favor of his idea in order to convince her.

3 DAN: (laughs)

Ex. Thinking Gest – Finger tapping (B)

This is taken from Pair E of the DisReg Corpus where Tina (E1) and Lea (E2) are talking about the books they had to read for the semester.

1 LEA: et toi du coup euh à part euh la littérature euh comparée et Agrippat t'as quoi d'autre comme euh bouquin?
2 TINA: euh olala (sighs) c'est une grande question ça hhh. ba:ah en (f)ait euh attends à part Agripa:at bah j'ai pas d'autre livre
3 LEA: ah ouais
4 LEA: ah t'as de la chance (laughs)
→ 5 TINA: ah si j'en ai t(u) sais c'est [/] c'est un cours là **eum hhh.**
****
((eyes closed + taps her fingers four times on the table))

c'est une UE libre!
****
((gazes towards LEA, extends her left arm towards LEA with a PUOH))

In the two examples above, the speakers performed a similar gesture with their left hand that involves several beat motions of their fingers against the table next to them. In both cases this finger-tap gesture was produced during fluencemes (although for Laura the gesture was also initiated prior to the fluenceme sequence); and they also emerged in contexts of deep thinking, or deep search, while they were withdrawing her gaze. In Excerpt A, Laura is wondering whether servants are truly loyal to their masters, and is looking for something to add that may be insightful, which requires additional time for thinking, so she gazes up and inserts a silent pause of rather long duration (1.265 milliseconds) while tapping her fingers against the table, signaling a change of participation towards a self-oriented search. Similarly, in Excerpt B, Tina is thinking about the other books she may have read during the semester and remembers one that she had in a specific class, called an "UE libre" a sort of optional class that is not part of the student's major. She does not remember the name of this class right

away so she delays her utterance for some time with a nasalized filled pause and an inbreath, and also retreats into a solitary word search, during which she closes her eyes and taps her fingers against the table. In both cases, the two speakers withdrew their gaze while performing the tapping motion, just like the previous excerpts introduced above with the finger snaps. And just like the finger-snap gesture, this gesture is performed with several motions of the fingers which create a soft noise. This further reveals the haptic dimension of gestures, which are truly sensitive to their environment and the different objects and artifacts placed around them (Morgenstern & Boutet, forth.; Streeck, 2021, 2020). Similarly in Excerpt Hold B (cf section I. 1.1.), Tina also interacted with the table next to her to make sense of her appreciation of Ancient Greek by enacting a drawing action directly on the table. We can thus view these gestures as abstractions from actual physical actions to make sense of the situation in highly specific sequential contexts, following Streeck (2020, 2021). It is further relevant to note that the finger-tap gestures could have easily been produced on the participants' bodies (i.e. on their lap) and not necessarily on the desk, which further gives support to the close relationship between language, the body and the material environment. Gaze plays a key role here; unlike the excerpts from section 2.1.2. with cyclic gestures, where the gaze was oriented *towards* the interlocutor as the speakers were searching or elaborating a piece of discourse, suggesting a more communicative intention, here the speakers shifted back their gaze towards their partner only after the search was over, suggesting that the search was more self-oriented and did not require the partner's participation.

To conclude, the aim of this section was to demonstrate the ways our vocal, gestural, and physical actions all interact to jointly perform specific social practices in discourse. More specifically, we focused on the practice of "doing thinking", a term coined by Heller (2021) who followed the work of Goodwin & Goodwin (1986) on specific facial displays known as "thinking" faces, and who identified multimodal gestalts of doing thinking, characterized by imaginative gaze, frowning, and inflexible posture (section 2.1.1.). Similarly, we presented a number of qualitative analyses from our data which showed similar instances of doing thinking, and examined them more specifically through the lens of inter-(dis)fluency and hesitation. We showed how the process of hesitating, or *doing hesitation*, may also be viewed as an embodied act, making use of the voice, the face, and the body, thus going beyond vocal or acoustic parameters. In addition, we included the deployment of other relevant gestural actions

into the practice of doing thinking, mainly cyclic gestures, finger snaps, and finger-tapping gestures. The latter were all characterized by a series of repetitive dynamic motions (either a circular motion performed with the wrist, a snapping or a beat motion performed with the fingers) and were also sometimes accompanied by specific facial displays (e.g., frown, gaze withdrawal). This has further shown that the act of doing thinking can also invoke more dynamic gestural actions which make use of the material space around them. In addition, we paid specific attention to the role of gaze during these thinking practices, and could further distinguish between practices of *doing self-centered thinking,* marked by gaze withdrawals indexing a change of participation into a temporary solitary practice, and *interactional thinking*, i.e., communicative displays oriented towards the interlocutor (cf our examples of mutual gaze during cyclic gestures in section 2.1.2). Our analyses have also further given support to the multimodal dimension of inter-(dis)fluency, which is manifested by a number of vocal and gestural resources that are altogether shaped by contextual and situational features. While this section has only focused on rather solitary self-oriented practices (except for cyclic gestures which were accompanied by mutual gaze), we shall now turn to the analysis of intersubjective displays.

2.2. Embodied displays of stance and intersubjectivity

As explained in Chapters 3 and 4, fluencemes are multifunctional, and may relate to planning, or monitoring, production-oriented processes (*Own Communication Management*), or may as well embody interactional and communicative actions (*Interactive Communication Management*); this ambivalence can further be determined by their accompanying visual-gestural features, depending on the type of gaze direction, facial display, or gesture deployed. In the previous section on thinking postures, we focused on temporary solitary practices during which participants typically "collected their thoughts" while signaling their continued engagement towards the activity-in-progress, but without necessarily orienting towards the interlocutor (as marked by instances of gaze withdrawal, frowning or body oriented to the side). However, we did find instances of mutual gaze during cyclic gestures, suggesting more intersubjective displays. Similarly, in this section, we focus on a selection of visible and bodily displays which deal with interpersonal mechanisms (i.e., turn-taking management, displays of understanding,), and perform or enact communicative actions in the interaction. The latter are typically known as *interactive*

(Bavelas et al., 1992, 1995) or *performative* gestures (e.g., Cienki, 2004, Kendon, 2004, Müller, 2015), and very often take the form of palm-up open hands oriented towards the interlocutor as to offer, present, give, request, or handle over a piece of discourse, an argument, or a turn (Bavelas et al., 1992; Kendon, 1995, 2004; Müller, 2017; Müller et al., 2013; Streeck & Hartge, 1992). They may also involve instances of pointing towards the interlocutor, as to include them in the interactional task, or cite their previous contribution (Bavelas et al., 1992). However, performative/interactive gestures do not only deal with turn-taking matters, but may also further convey speakers' attitudes towards what they are saying (e.g., indifference, indignation, obviousness etc.,), and this is often conveyed by shoulder shrugs, or head shifts (Debras & Cienki, 2012; Kendon, 2004; Streeck, 2009). While all these types of gestural actions have been thoroughly documented and carefully analyzed in different papers within the field of gesture studies over the past 30 years, little is known about their coordination with fluencemes. Such instances have already been illustrated in previous chapters, and are summarized in the table below.

In Excerpt *Teenage Years* from the SITAF Corpus (cf Chap. 3, section II. 2.2.1.), we described how the non-native speaker displayed epistemic stance, evaluating her teenage years as a confusing time period. The multi-component shrug (shoulder shrug, head tilt, palm-up open hands, cf Chap. 1, section III. 3.3.3.) occurred during a pause and was followed by a kind of rhetorical question "What am I doing here?" during which she looked at her interlocutor and smiled, further conveying *epistemic indetermination* (see Debras, 2017; this could be glossed by "I don't know what I'm doing here").

In Excerpt 2b (Chap. 3, section 2.2.2.), the non-native speaker extended his left palm-up open hand towards his interlocutor at the end of this turn to invite her to take the floor, as he was talking about the types of sensitive topics that friends may share, further including her in the current topic of conversation. This gesture also coincided with a fluenceme sequence at a transition relevant place, when his partner took the turn.

Table 42. *Summary of communicative displays deployed during fluencemes in previous chapters*

Illustration	Example	Context	Visual-gestural description
	Excerpt Pair-7-teenage years (Chap. 3)	French NNS is displaying a stance towards her utterance "What am I doing here?"	Shoulder shrug, palm-up open hands in the lower gesture space, head shake and body oriented towards interlocutor
	Excerpt 2B (Chap. 3)	American NNS is inviting his interlocutor to take the turn	left Palm-Up Open Hand extended towards the interlocutor
	Excerpt 2.2. (Chap. 4)	French native speaker is adressing his interlocutor and an imaginary audience	Stretched index finger from left hand is raised and slightly oriented towards the interlocutor

Lastly, in Excerpt 2.2. (Chap. 4, section II, 2.2.3.), the speaker raised his index finger from his right hand and slightly oriented it towards his interlocutor as to acknowledge her presence (and the presence of an imaginary audience as he was telling a humorous anecdote), and hand other relevant discourse material. This gesture was also produced during a pause following the production of a funny catch phrase “le saviez-vous”, further contributing to the humorous dimension of his story. While all the gestures documented above served comparatively different functions (i.e., turn-taking, stance-taking, attention-seeking, etc.,) and occurred at different turn-positions (turn-medial and turn-final) they all performed communicative illocutionary acts and embodied an interactional move. In fact, Kendon (1995) referred to this class of gestures as “illocutionary marker gestures” (as opposed to “discourse unit marker gestures” which mark discourse structure[178]). In addition, all these gestures occurred without speech, in other words, during fluencemes. This further reveals that fluencemes produced in the speech channel may provide relevant opportunities for *languagers* (Kendon, 2014) to perform multimodal communicative

[178] Note that a similar distinction was made in our quantitative gesture annotation, as we distinguished between *interactive/performative* gestures and *discursive/parsing* gestures (within the *pragmatic gesture* category, cf Chap. 2, section II, 2.2.3).

actions within the interaction; even though the flow of speech is momentarily suspended or interrupted, it *opens up*[179] to another semiotic field in a different modality to build *deliberate expressiveness* (Kendon, 2004, p. 13-14). This is further illustrated in the following examples.

Ex. Interactive Gest. (A)

This very brief extract is taken from Pair B in the DisReg Corpus, right before *Excerpt 2.2* (Chap. 4, section 2.2.2.) when Paul shared a funny anecdote with Paula. They just read one of the topics written on the piece of paper, and decided that they would talk about funny anecdotes at university, but they are actually struggling to find one. This excerpt was also presented at the IPRA Conference in June 2021 (Kosmala et al., 2021).

1 *PAULA: eum une anecdote t'en as pas une à dire avant que je t'en
****
((extends her left open-palm towards PAUL))
trouve une?
2 *PAULA +< sur l'université

→ **3 *PAUL: eu:um:m:m (0.400) [!] mmm:m:mm**
((gazes away, shakes his head then extends his left PUOH towards Paula who then averts her gaze and looks puzzled))

((1.358))
4 *PAUL: ah si j'en ai une elle est très très précise.
*** *********
((extends his index finger towards Paula)) ((finger bunch))
5 *PAULA: ok (smiles)

[179] The term "open up" was suggested by Simon Harrison after hearing our talk at the LSPPC6 Conference (Kosmala & Morgenstern, 2021) to suggest that fluencemes embodied an "opening up" rather than a "breakdown" in discourse.

As Yule (1996, p. 78) noted on his book on pragmatics, "delay represents distance between what is expected and what is provided. Delay is always interpreted as meaningful". Here Paul initiates a significant delay in the course of this small exchange, marked by a series of lengthened filled and unfilled pauses and vocalizations. This fluenceme sequence occurs in standalone position and takes up most of Paul's turn, which marks a considerable suspension in both the vocal acoustic channel and the interactional flow, as the progressivity of the exchange is momentarily disturbed, as further cued by the lapse of 1.358 milliseconds following the fluenceme sequence. This long period of "nonspeech" constitutes a relevant round of *possible self-selection* (Sacks et al., 1974) for Paula, who is strongly encouraged to take the turn, but who refrains from doing so, thus projecting a dispreferred next-action. Paul's invitation to Paula to take the turn is not expressed verbally, but is manifested in his visible bodily behavior, as he mobilizes a number of relevant resources to perform this interactional move. He is first seen looking up and tilting his head to the right, his body and head not moving, hence displaying an inflexible thinking posture, then shifting his attention back to Paula by gazing towards her, and quickly extending his left Palm-Up-Open Hand towards her to invite her to take the turn, while shaking his head as he begins to *hum* the nasal sound ("mmm") as if singing a tune. Paula then quickly averts her gaze and looks puzzled, suggesting that she is not ready to take the turn, since she has not found an anecdote yet; she in fact invited him to do so in the first place (cf line 1), so it may be a way for her to deny any responsibility towards the current turn-at-talk. After some time, following the lapse, Paul then prefaces a change of state with the token "ah" (similar to "oh" see Heritage, 1984), as he extends his index finger upwards towards Paula to indicate that he has finally found an anecdote (which is very similar to the interactive gesture from the same exchange listed in Table 42 above). The issue at hand is then resolved at this sequence-final turn, further indexing the initiation of his story (when Excerpt 2.2. begins).

In sum, what this excerpt illustrates is the way vocal, a priori *non-lexical* fluencemes, devoid of propositional meaning, can in fact bear highly meaningful and lexical properties in social interaction. From a strictly verbal or vocal point of view, Paul's utterance in line 3 is highly "disfluent", and bears no semantic weight; as some psycholinguists would say, it is simply "noise" in the signal (see Linell, 2004, p. 146). However, from a *multimodal interactional* perspective, this sequence of fluencemes embodies a series of relevant actions working simultaneously with one another; Paul

is both deeply engaged in the act of thinking and *hesitating* (and this is clearly understood by Paula who does not interrupt him and may perhaps also be "hesitating" in silence) as well as the act of conversing with his partner. This is further shown in his interactive gesture, coupled with a shift in his gazing behavior, which once again illustrates the intersubjective dimension of fluencemes, as cued by their accompanying visual-gestural actions. Moving now to the analysis of a second example of intersubjective displays during fluencemes from the SITAF Corpus.

Ex. Interactive Gest (B)

This excerpt is taken from Pair 16 of the SITAF Corpus during the tandem exchange in English. Here the two partners are talking about whether prisoners should get the right to vote, and after agreeing that prisoners are still citizens and must definitely keep this right (omitted from the transcription) the French non-native speaker (Elisa, F16) is wondering about the types of serious crimes that actually prevent prisoners from voting.

1 *ELI: so what are the crimes (0.360) you know that (0.420) made
**** **********
((extends her left PUOH with index finger extended towards Beth + brings her flat right palm-up open hand forward))
p(eople) [/] that made people (0.370) uh unable to vote?
****
((brings both her palms-up open hands forward in the lower gesture space))
2 *ELI: (0.650) like (0.410) crimes o:or (0.510) I don't know?
***** ********
((flips both her palms upwards, hands resting on her lap + moves her left palm higher to the side with a swaying motion))
3 *BET: +< oh I believe that you can still vote in the US if you're a prisoner.
4 *ELI: +< oh (raises her eyebrows).
5 *ELI: (0.350) really? (lifts her shoulders)
6 *BET: yes [/] yes [/] yes (nods)
7 *ELI: and even if you I don't know killed someone?
******* ********
((right hand: points to the right)) ((left hand: performs a waving motion towards her shoulder))
→ **8 *BET: +< ex(cept) [//] unless hhhh. [/] unless**
~~~~~~~~~~~********

((looks up, extends her index finger towards ELI from her right hand and performs two beat motions + hold))

you've committed um (0.620) a certain level of felony.
**********+.+.+.+.+.+. **********
((bends her fingers from her right hand and sways back and forth))

Here the two participants are talking about the United States more specifically, so Beth, the American speaker, does not only position herself as a native speaker of American English, but also as someone who has a certain knowledge of her country's body of law and its rules and regulations. In fact, Elisa further seeks knowledge from Beth, as marked by her series of WH-questions on the matter, and her palm-up open hand gestures oriented towards her partner (lines 1-2), which further invite her to take the turn, which she does in line 3. Beth first claims that prisoners do have the right to vote in the United States, which comes as a surprise to Elisa, as signaled by her exclamations ("oh" and "really") and facial displays (raised eyebrows and lifted shoulders). Elisa thus re-initiates a third question (l. 7), which further prompts Beth to re-shape her previous assessment, perhaps to align with Elisa's expectations. A complex fluenceme sequence emerges at the beginning of Beth's subsequent turn in line 8, during which she reconsiders the matter at hand. She first initiates a preposition ("except") but in truncated form, replaces it with another one ("unless" in its complete form), followed by an audible inbreath, and a repetition of the initial word. As shown in the multimodal transcription, Beth also extends her index finger towards Elisa in her lower gesture space during the fluenceme sequence, and holds it the same position until her production of a second fluenceme sequence in mid-utterance (*FP+UP*) when her gesture is retracted. The gesture initiated during the first sequence shares close formal and functional properties with the one performed by Paul in *Excerpt 2.2.* (cf Table 42 above) except for the hand configuration and position in the gesture space. In Paul's case, the index finger was raised in the center space with the palm vertical,

while here it is only extended in the lower gesture space with the palm up. In both cases, nonetheless, the two gestures are oriented towards the partner and can thus be considered *interactive* and be recognized more specifically as *delivery* gestures (cf Bavelas et al., (1995) from Chap. 4, section II. 2.2.3) i.e. gestures used to hand over information relevant to the speakers. Even though Beth is pointing towards Elisa, it is not to indicate an object, person, or location in this case (as pointing gestures often do e.g. Kendon & Versante, 2003), but rather to acknowledge her partner's presence and present an upcoming piece of discourse as potentially relevant material to the activity-in-progress. In addition, Beth is also seen gazing up, then down, and she only shifts her gaze back towards her partner when she finishes her utterance and introduces the noun phrase "a certain level of felony". This shift in gaze behavior may be interpreted as a sign that Beth is dealing with multiple orientations at the same time: on the one hand she is re-elaborating her previous assessment, which prompts her to modify parts of her talk and think about what to say next (self-oriented); and on the other she may also be wishing to capture her partner's attention and react to what they have been discussing in the prior turn (other-oriented). Once more, this example reveals the multidimensionality and multifunctionality of inter-(dis)fluency phenomena, which can further be determined by paying close attention to their co-occurring visual-gestural behavior and gaze within the turns-at-talk.

To conclude, these two examples have further demonstrated that we cannot strictly separate the speech channel from the gestural channel, nor claim that one modality replaces the other, as the different articulators used in discourse to build meaningful language (i.e. the lips, the tongue, the mouth, as well as the limbs) do not operate independently from one another, but are rather co-deployed harmoniously in ways that are relevant to the *fluent* achievement of the interactional task at hand. In addition, these examples have further put forward the interactional dimension of *inter-fluency*, which should not only be regarded as a cognitive internal and mental process underlying planning or thinking processes, but as a dynamic interactive practice which is very sensitive to the affordances of the situation.

2.3. Gestural modes of representation: beyond lexical retrieval

Across Chapters 1 and 2, we explained a common distinction found in previous classifications between gestures that relate to the referential content of discourse (often known as referential gestures, topic gestures, depictive gestures, iconic gestures,

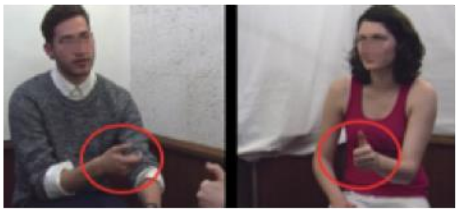
among other terms, cf Chap. 1, section III. 3.3.3.) and those that relate to interaction or discourse itself, but not propositional content (mostly known as pragmatic gestures, cf Chap. 1, section, III. 3.3.3. and Chap. 2, section II. 2.2.3.). The previous subsections largely focused on pragmatic gestures (more specifically thinking gestures and interactive gestures), capturing the pragmatic dimension of fluencemes in situated discourse. The present section aims to discuss the use of referential gestures in multimodal discourse with regard to inter-(dis)fluency behavior, and most specifically *representational* ones. As explained earlier, Müller et al., (2013) and Müller (2014) described several techniques developed by gesturers' moving hands in interaction, who make use of different *gestural modes of representation*[180] (molding, drawing, representing, and acting), and which all emerge from practical manual actions, such as holding, grabbing, brushing, etc., (cf Chap. 1, section III. 3.3.3)[181].

Table 43 summarizes the representational gestures deployed during fluencemes and analyzed in previous excerpts, and describes more specifically their mode of representation. As it shows, several speakers deployed representational gestures across languages and settings in our two datasets to build discourse, by making use of slightly different techniques. In the SITAF Corpus, we presented two excerpts from Pair 11 where meaning was co-constructed and cooperated between the two parties, who deployed similar gestures in tandem. In Excerpt 4b, the speakers relied on the *acting mode*, where the hands re-enacted an actual manual activity (i.e., typing on one's phone) to build meaning around social media; in Excerpt 4c, on the other hand, they relied on the *drawing mode*, as they outlined the contour of a wave in the gesture space to metaphorically represent the notion of ups and downs.

[180] Note that this distinction is also present in Moro et al., (2020, p. 233) in which they further subdivided representational gestures into "modelling gestures" when the gesturing body is used as a model for an object, "enactment", when the gesturing body enacts a specific action, and "depiction", when the speaker outlines the shape of an object and traces it in the air.

[181] Note that Streeck (2009) made a subtle distinction between depicting and *ceiving* gestures, and the latter are said to conceptualize a thematic object , but without necessarily depicting them. However, as Streeck said so himself, this distinction can be a difficult one, so we preferred to analyze the gestures of this section using Müller et al., (2013)'s method.

Table 43. *Different gestural modes of representation performed during fluencemes in previous chapters*

Illustration	Example	Context	Gestural mode of representation
	Excerpt 4b. (Chap. 3)	The two tandem partners are talking about social media in English, and more specifically about twitter and facebook	**Hands acting out** the action of typing on one's phone to represent twitter and facebook
	Excerpt 4c. (Chap. 3)	The two tandem partners are talking in French about teenage years and the fact that life very often goes "up and down"	**Hands drawing** a wave in the air as to metaphorically represent the meanings of ups and downs
	Excerpt 1.1. (Chap. 4)	The class presenter is talking about the beeping sound the phone makes when one makes a call	**Hand tracing** the path and movement of a line in the air to represent the beeping sound of the telephone
	Excerpt 2.2. (Chap. 4)	The native French speaker is talking about funny-looking shoes ("gants de pied")	**Hand representing** the shoes by embodying the object as an whole (the hand represents the shoes which shows individual fingers)
	Excerpt 2.2. (Chap. 4)	The native French speaker is talking about funny-looking shoes ("gants de pied")	**Hand outlining** a series of circles as to metonymically represent a specific attribute (i.e. individual toes)

In the DisReg corpus, we presented similar excerpts from the presentation- and conversation-sessions. More specifically, in Excerpt 1.1, the class presenter also relied on the *drawing mode*, by tracing out the path and movement of a line in space to represent the beeping, staccato-like sound on the telephone. This technique was also used in Excerpt 2.2 (final illustration) to draw a series of circles in the air to metonymically represent specific properties of an object (the parts of the shoes that show individual toes). Finally, in the fourth example from Excerpt 2.2., the speaker relied on the *representing mode* to embody the whole object which became itself a kind of "manual sculpture" (Müller et al., 2013, p. 712).

There is one technique that has not been documented in our analyses so far, known as the *molding mode*, were the hands "create a transient sculpture, such as a frame or a bowl" (Müller et al., 2013, p. 712), which is briefly illustrated in the example below.

Ex. Representational Gest – Molding

This excerpt is taken from Pair A in the DisReg Corpus. Here David (A1) is talking about another board game where players can use little bags. This excerpt is purposefully very short as to focus on the representational gesture used in this context[182]. Here David is talking about a specific board game where players can dig all sorts of rocks that can be put inside a little bag (not in the transcription), but he does not find the target word ("petits sacs") right away, as he first produces a truncated word, lengthens the pronoun ("de:es") and produces a tongue click before producing the lexical item, potentially signaling a word search[183].

*DAV: hhh. et donc ce qui est assez nouveau c'est que t'as
des pie [//] **de:es [!] des p(e)tits sacs [/] des p(e)tits sacs.**
1.*** 2.**********
((brings his palms down with bent fingers; **hands mold an object in the center space i.e., little bags; gazes at his hands**))

A gesture is first initiated during the truncated word, where the palms are brought downwards in the lower center space with bent fingers, but it is very quickly retracted and followed by a second one where David seems to be *molding* the shape of the little bags which contains the rocks. Note that this gesture is first produced during the prolongation, and is then repeated and slightly amplified as he produces the tongue click and repeats the target word, further reflecting a coordination between the verbal and gestural modalities.

[182] This excerpt was also presented at the *Laughter Workshop* in October 2020 (Kosmala, 2020b)
[183] As explained earlier, tongue clicks can serve a word-searching function (see Chap. 4, section II. 2.2.2.)

In Chapter 3, we further argued against cognitive-oriented approaches to representational gestures with theories such as the *Lexical Retrieval Hypothesis* (Krauss et al., 2000), or *Information Packaging Hypothesis* (Kita, 2000) following the work of Graziano & Gullberg (2018). As argued, we do not believe that speakers produce representational gestures during fluencemes to overcome a "lack of skill" in their L2, or to facilitate access to a lexical item, which would be too restrictive, suggesting that speech and gesture work independently from one another. This claim was further corroborated by our quantitative findings, as we found a higher proportion of representational gestures outside fluencemes (about 85%) than during them (about 15%), as well as no significant differences between L1 and L2 (see Chap. 3, section II. 2.1.3.). In Chapter 4, a slightly higher proportion of representational gestures was found during fluencemes in DisReg compared to SITAF (about 28%) but such gestures were also significantly more frequent in face-to-face conversations compared to class presentations (28% as opposed to 5%, see Chap. 4, section II. 2.1.3), which reflects the importance of style and setting on gesture use, which is not restricted to difficulties in lexical access. As the examples listed above have further shown, such manual gestures skillfully make use of several representational techniques that rely on speakers' manual abilities to (co)-build meaning in ways that are relevant to the current activity. In Excerpts 4b and 4c, the two tandem partners produced similar gestures in tandem to co-elaborate on the meaning surrounding social media and teenage years. In Excerpt 1.1., the presenter deployed this gesture to represent a beeping staccato-like sound from a telephone, and make it visibly available for her audience to check that they understood what the author meant in the novel. Lastly, in Excerpt 2.2 the speaker made use of several representational gestures to introduce a referent with iconic properties within his multimodal discourse, and further illustrate the comical aspect of the funny-looking shoes. As Müller et al., (2013, p. 716) further noted:

> Speakers choose between different gestural modes: they may trace, mold, represent an object or they may act with it and each time they will highlight a different dimension of this object.

Similarly, in all our examples, each speaker focused on a specific aspect or dimension of an object (its shape, path or movement, size, or a specific attribute) to build their narrative. Further in line with the *scope of relevant behavior* theory (Cienki, 2015a), and the notion of *contextual configuration* (Goodwin, 2000) we further maintain that it is a matter of selection among the most relevant type of semiotic resources available

to a speaker that will motivate the use of representational gestures in a given context. Müller & Tag (2010) put forward a similar theory known as the *theory of a dynamic focus on attention*, which proposes that specific aspects of meanings are foregrounded by the speakers, depending on the flow or current focus of attention within a particular setting, so highly specific hand shapes will more likely be used in contexts of richly embodied experiences, while reduced gestures depict more "prototypical" situations. (Müller et al., 2013, p. 716).

To conclude, the fact that all these gestures were produced during fluencemes (i.e., without "fluent" speech) does not necessarily mean, in our opinion, that they replaced speech to "compensate" for language difficulties (Krauss et al., 2000) or to "activate" spatio-motoric information to manage cognitive difficulties (Kita et al., 2017), in line with cognitive-oriented approaches. As maintained multiple times in this thesis, we situate inter-(dis)fluency and gesture in a larger interactional framework where we regard the accomplishment of multimodal discourse as *inherently interactional*, rather than solely resulting from mental or cognitive states. Both functions (interactive and cognitive) may thus co-exist simultaneously depending on the type of gesture produced, and the direction of gaze. It is thus essential to consider the emergence of fluencemes and gestures in their multimodal context of use and within their situated ecology. In addition, the present work is grounded in a *dynamic* view of language, whereby the construction of meaning is the result of a highly dynamic and interactive process, which is continuously changing its focus, zooming in or zooming out on particular features, selecting among the most relevant semiotic field in a particular context, treating one or the other as no longer relevant, and which are all in turn the product of interactive demands, following the *SrB* (Cienki, 2015), the notion of *contextual configuration* (Goodwin, 2000), and Müller & Tag's (2010) *dynamic focus on attention* theory. We suggest that fluencemes, along with gestures, are integral components of this interactive and dynamic process of doing language, or *languaging*. This leads us to our last section.

III. The multimodality of inter-(dis)fluency in situated language use

The aim of this chapter was to offer a comprehensive review of the different types of visible and gestural behaviors that typically occur during fluencemes or in their close vicinity, in order to give support to the multimodal nature of inter-(dis)fluency. The quantitative analyses conducted on our dataset of conversations and presentations in French and English described in Chapters 3 and 4 initially consisted in identifying and annotating all the verbal and vocal fluencemes present in the speech channel, following previous work in (dis)fluency research. However, as pointed out throughout this thesis, inter-(dis)fluency should not be restricted to the analysis of verbal and vocal markers, but should include the diverse range of semiotic resources surrounding them (i.e., the body, the face, the hands, gaze, the material environment and its artifacts). Indeed, living *languagers* (Kendon, 2014) do not build the fluency of their multimodal discourse by strictly using their voice, but by mobilizing a wide array of relevant resources in multiple media in an orchestrated relationship in order to construct multimodal *languaging*. The term *languaging*, briefly introduced in Chapter 1 (cf section IV.4.1.) is further discussed in the next section, along with the concept of language.

3.1. Inter-(dis)fluency in multimodal languaging

As mentioned earlier (cf Chap. 1, section IV, 4.1.) the term *languaging* has been used by a number of researchers across different fields of linguistics over the past 15 years, mainly Linell (2009), Kendon (2004), Morgenstern (2020), and Swain & Watanabe (2012), among others. This term was initially defined as "a process of making meaning and shaping knowledge and experience through language" (Swain, 2006, p. 98) and as "linguistic actions and activities in actual communication" (Linell, 2009, p. 274). More recently, this term has been adopted more specifically to refer *to multimodal language use*, deeply rooted in multimodal analyses of motivated and conventionalized language forms (sounds, words, tones, gestures). This began with Kendon (2014) who criticized the concept of "language" developed in structural linguistic research in the mid 20th and early 21st century as being too narrow, and too centered on speech models of production, which he believes to be inadequate to refer to the way spoken languages work. As emphasized multiple times before, when

speakers engage in discourse, they do more than "utter words" (Kendon, 2014, p. 12) in a linear fashion and in decontextualized situations, as they mobilize different kinds of visible actions that are integrated to the activity of utterance construction at specific moments within the interaction. In sum, they're not just speaking, but *doing* language, or languaging. Similarly, in the work of Morgenstern (2020), and Morgenstern & Boutet (forth.), their notion of *language* is deeply grounded in *embodied action*, and can thus only be truly captured in its ecological and material environment. They also use the term *languaging* to refer to the coordination of available semiotic resources used to construct and give meaning to interactive productions. Similarly, the present study of inter-(dis)fluency is situated within this existing body of research. We believe that the different analyses introduced in this chapter further give support to his new model of language, and that the study of fluencemes should not be overlooked in the analysis of multimodal discourse, for a number of reasons elaborated below.

First, we have shown a timely coordination between speech and gesture, with instances of speech suspension combined with gesture suspension/retraction, and speech preparation with gesture preparation (section I) where the two modalities were combined to suspend, interrupt, restart, initiate, or project an upcoming, or previous, piece of multimodal discourse. This shows that the notions of suspension, interruption, or preparation, which are inherent to *speech* (dis)fluency, are not only marked verbally or vocally, but gesturally as well, which further gives support to this model of multimodal languaging, which operates upon a number of semiotic systems.

Second, we have shown that inter-(dis)fluency may also further be manifested in recurrent embodied interactional practices of doing thinking, doing hesitation, or doing searching (section 2.1.), characterized by a number of specific displays (i.e. frown, imaginative gaze, inflexible postures) and gestural actions (i.e. cyclic rotations, finger snaps, and finger tapping gestures). Once again, a timely coordination was found between the emergence of vocal and verbal fluencemes in the speech channel and particular visual-gestural affiliates, further emphasizing the embodied and multimodal nature of inter-(dis)fluency.

Third, we have shown that the multifunctionality of inter-(dis)fluency phenomena can further be determined by their accompanying visual-gestural behavior, which may signal the current orientation of a speaker in a particular setting, whether they are engaged in self-oriented practices, or other-oriented ones (section 2.2.). As we have seen, speakers make use of their voice, body, gaze, and gesture to

yield a turn, or catch someone's attention, and in such cases fluencemes may function as relevant interactional displays, signaling not only *hesitation* or a solitary word search, but intersubjective actions. Despite being a priori "non-lexical", "disfluent" forms in the verbal and vocal channel, they have the potential to gain symbolic meaning in specific multimodal contexts, which further stresses the need to analyze them with a *multimodal* perspective.

Lastly, we have shown how fluencemes may also be used in contexts of gestural depiction, whereby speakers rely on different manual techniques of representation to (co)-construct meaning in interaction. We further argued in favor of a *dynamic* view of language, which relies on a constantly changing focus or flow of attention, determined by the current context and situation, the task at hand, the interlocutor(s), and the availability of semiotic resources. In this view, speakers do not "replace" one modality with another (e.g. speech with gesture), suggesting that they work independently, but rather select among their scope of available communicative behaviors and contextual fields that are deemed more relevant to a given situation.

To conclude, we believe that the study of inter-(dis)fluency presents a number of relevant implications for the analysis of multimodal *languaging*; while fluencemes tend to be associated with little gestural activity, their coordination with gestures and other types of visible actions should not be overlooked. In this chapter, we have further given support to the concept of fluency as a multimodal, multilevel, and transitional process, not only marking a transition from "fluent" to "disfluent" speech, but embodying a change of orientation, from solitary to joint activities, a change of pace, with gestural and speech suspension, or a change of semiotic field, from the speech flow to the gestural flow. Inter-(dis)fluency should thus be regarded as an *embodied* process, reflecting a number of bodily practices, summarized in the next section.

3.2. Summary of the most recurrent visible features embodying inter-(dis)fluency

In section II, we identified a number of recurrent visual-gestural practices accompanying fluencemes across three specific social practices: doing thinking (section 2.1.), displaying stance and intersubjectivity (section 2.1.) and (co)-constructing meaning (section 2.2.), summarized in the table below. In sum, inter-(dis)fluency relates to a multiplicity of cognitive and interactive processes, indexing a potential change of orientation, focus, or participation, which can altogether be

manifested with the voice, the face, the body, the eyes, and the hands, or all of them combined, depending on a number of contextual and situational factors. A number of recurrent visible features were identified, which were all found to occur during fluencemes across our various examples from the SITAF and DisReg Corpus, which further gives support to the multimodal and embodied nature of inter-(dis)fluency. This does not mean, however, that the different types of visible actions documented above are exclusively tied to one of the three practices (e.g. cyclic gestures perform a wide range of functions other than word search, see section 2.1.2).

Table 44. *Summary of the most recurrent visible features found during fluencemes*

Recurrent social practice embodying inter-(dis)fluency	Different types of visible action
Doing thinking, hesitation, or searching	Thinking posture (thinking face, inflexible posture, frown, gaze withdrawal)
	Self-touch (hand resting on chin, touching one's face)
	Body crouched, oriented to the side
	Head leaning back and forth to the left and right
	Continuous cyclic orientations of the wrist
	Repetitive finger-snap motions
	Repetitive finger-tap motions
Display of stance and intersubjectivity	Mutual gaze
	Body oriented towards interlocutor
	Shrug
	Palm-Up Open Hand gesture
	index finger extended and raised towards interlocutor
(Co)-construction of meaning	Hands re-enacting a specific action
	Hands drawing, or tracing an object, a path or movement in the gesture space
	Hands representing an object as a whole
	Hands molding the shape of an object

This also does not imply that inter-(dis)fluency solely covers these three practices, as fluencemes may also be used to perform other actions, such as structuring or emphasizing certain aspects of discourse, which has not been thoroughly investigated in this section. It should further be noted that this analysis was based on a selection of excerpts and qualitative analyses extracted from the data, but more work needs to be done on the rest of the data to document other types of recurrent visible behaviors tied to specific actions during inter-(dis)fluency. We further discuss our methodological limitations in the conclusion.

Conclusion to the chapter

The aim of this chapter was to further explore the multimodal dimension of inter-(dis)fluency by drawing on a number of qualitative analyses extracted from the data illustrating the different visible practices occurring during fluencemes and in their close vicinity. In this chapter we focused on a more *form-based* approach to gesture, documenting formally similar gestural patterns, thus taking a step further from our initial gesture functional typology used in the quantitative annotations. We first illustrated the tight relationship between speech and gesture production through several examples of gestural holds, retraction, and preparation coordinated with fluencemes, reflecting a unified process performed in the two modalities. We then documented a series of recurrent visible practices found during fluencemes, with several facial and body displays of doing thinking, as well as manual actions (cyclic gestures, finger snaps, finger-tapping gestures), intersubjective displays (with palm-up open hand gestures, shrugs, raised index fingers), and other gestural modes of representation (molding, acting, drawing, etc.). These analyses led us to a closer understanding of the notion of language, and to a larger extent, *languaging*, which captures multimodal language use grounded in situated and embodied discourse. In this view, fluency should not solely be regarded as a vocal or temporal phenomenon, restricted to mental states or proficiency levels, but as a fully multimodal process, relying on a multiplicity of resources in situated discourse. Hence the practice of *doing hesitation,* or *doing fluency*, can further be recognized as an embodied social act, rather than as a mere by-product of verbal processes.

However, this chapter also raises a few questions regarding methodology in gesture research. For the sake of clarity and consistency, we chose to work on a functional classification of gestures with a finite set of categories in our quantitative analysis, in order to get a clear idea of the gestural distribution and tendencies found in our two datasets. However, it could be argued that some gesture functions may overlap in certain situations (e.g. pragmatic and referential functions), which makes the coding process quite difficult, and ultimately requires inter-coder reliability. While we did use inter-coder reliability (cf Chap. 2), this was only performed on 15% of the data, which remains limited. In addition, we did not annotate the different forms, handshapes, or configurations of the gestures in the data at a quantitative level, which would have been highly time consuming, but which would also have resulted in a high

number of different categories, making statistical analysis difficult. Our solution was thus to combine quantitative annotations with qualitative analyses of the data in order to reflect different aspects of the gestures at different levels (form, function, shape, context of use etc.). However, the ideal solution would be to integrate a multi-level annotation system based on *both* the forms and functions of gestures, see for example *The MultiModal MultiDimensional* (M3D) labelling scheme for the annotation of audiovisual corpora (Rohrer et al., 2020). But such models do not incorporate the analysis of inter-(dis)fluency, which is fundamental to the present study, and which, as maintained throughout this chapter, needs to find a place within the field of gesture studies.

Highlights of Chapter 5:

- Inter-(dis)fluency is not only marked vocally or verbally, but relies on a multiplicity of modes and semiotic resources.
- Even though the vocal channel remains the "default" mode, inter-(dis)fluency is the result of a dynamic and interactive process which relies on a constantly changing flow or focus on attention which chooses among the most relevant types of communicative behavior at a given time in the interaction.
- Several recurrent visible bodily practices were identified during fluencemes, mainly thinking posture, self-touch, gaze withdrawal, but also manual actions such as cyclic gestures, finger snaps, finger-tapping gestures.
- Although inter-(dis)fluency tends to be associated with hesitation and disengagement, it may also embody communicative actions, as further cued by their accompanying visual-gestural features (mutual gaze, body oriented towards interlocutor, palm-up open hands or raised index finger pointed towards the other), reflecting once more different degrees of communicative and interactive fluency.
- Gestures produced during fluencemes do not merely "replace" vocal fluencemes or "compensate" for language difficulties, but rather work together and alongside fluencemes to jointly perform a number of relevant practices, such as signaling that a current word search is in progress, displaying intersubjectivity, or (co-) constructing meaning.
- More work needs to be done on the multimodality of inter-(dis)fluency and its relation to multimodal languaging to further our understanding of "language".

General conclusion

I. Theoretical and methodological contribution

The present study is situated within a large existing body of research in fluency and disfluency phenomena and aimed to offer a more comprehensive approach by integrating different theoretical frameworks, mainly usage-based and corpus-based linguistics, interactional linguistics, and gesture studies. In Chapter I, we stressed the need to situate inter-(dis)fluency in a larger integrated framework in order to bridge the gap between "traditional" production-based psycholinguistic studies conducted in disfluency research, and interactional, multimodal approaches to social interaction. This also invited us to consider (dis)fluency from multiple dimensions, following Lickley (2015), Segalowitz (2016) Götz (2013), and Grosman (2018). We will return to this point in section III of this conclusion.

In Chapter I, we reviewed a number of theoretical frameworks which all had different but interrelated perspectives on (dis)fluency phenomena. We started with psycholinguistics, which is one of the first major field of research that systematically investigated these phenomena, and which introduced major classifications on (dis)fluency types and offered systematic analyses regarding their distribution in speech production (e.g., Levelt, 1983, 1989, Maclay & Osgood, 1959; Shriberg, 1994, cf Chap. 1, section I). This later paved the way for other seminal studies on (dis)fluency in other lines of research, such as Second Language Acquisition and Corpus-based linguistics. However, we also pointed out a number of issues underlying the term "disfluency" in the literature, which presupposes a problem, or a *deviation* from *ideally fluent* speech. The term "fluency" on the other hand, is traditionally found in Second Language Acquisition research, and refers to aspects of L2 performance and native-like proficiency levels, based on several temporal variables (speech rate, duration of pauses, rate of "disfluencies", etc.). The notions of fluency versus disfluency have thus been consistently distinguished from one another, despite the constant overlap of terms found across theoretical research fields (Chap. 1, section II). Some researchers argued that disfluencies should be called markers of *fluency* instead of disfluency, as they have been shown to serve many positive planning functions in discourse other than just signaling an interruption in the speech signal (e.g. Tottie,

2014). This constant opposition between fluency and disfluency is reflected in two contrasting views of these phenomena, one that regards them as a cognitive burden, and another one which considers them as a communicative signal (cf Chap. 1, section II. 2.2.1.). Throughout this thesis, it has come to our understanding that the real conflicting issue regarding the notions of fluency versus disfluency is not only terminological, but theoretical and methodological as well. The fact that these phenomena have been systematically analyzed from a strictly verbal and formal perspective (except for a few notable exceptions, e.g. Tottie, 2014; Allwood et al., 2015; McCarthy, 2009, Clark & Fox Tree, 2002 etc, see Chap. 1, section II. 2.2.1. and III.3.2.2) has hindered their evaluation on the basis of discourse, interaction, or even gesture. This led us to a review of other theoretical fields which took on a new perspective to these phenomena and which were altogether relevant to our study. In particular, we focused on the frameworks of cognitive grammar and usage-based linguistics, which accounted for a more dynamic approach, considering fluencemes (Götz, 2013) as fluid categories, whose degree of symbolic meaning, lexicality, conventionalization, and to a larger extent *fluency*, were altogether shaped by repeated instances of specific patterns in different contexts of use. In light of this approach emerged innovative cognitive and usage-based frameworks on (dis)fluency (e.g. Crible et al. 2019; Grosman, 2018; Götz, 2013; Segalowitz, 2016) which served as a basis for our multidimensional definition of inter-(dis)fluency (cf section III of this chapter). In addition, we also took into account the framework of interactional linguistics, which provided essential conversation-analytic tools for our study of inter-(dis)fluency based on their position within the turns-at-talk and their sequential development in the exchange, captured in situated talk-in-interaction, thus going beyond their analysis in the verbal production channel (e.g., Mondada, 2007, Sacks et al., 1974). Lastly, we integrated the frameworks of *gesture studies* and multimodal interaction, further vouching for a view of (dis)fluency as an embodied and multimodal phenomenon, tightly related to the deployment of visible bodily actions (manual gestures, facial displays, gaze direction, body movement). In this respect, we regarded language as an embodied mode of action, reflecting the notion of *languaging* (Kendon, 2014, Linell, 2009, Morgenstern, 2020, see Chap. 5, section III), as inherently situated within its ecological and multimodal environment, comprised of the material, interactional and gestural space. Additionally, the field of gesture studies provided relevant gesture classifications, grounded in a functional-communicative approach, which was applied

to our study of embodied inter-(dis)fluency through the analysis of its temporal coordination with gesture (cf Chapters 3, 4, and 5).

Our integrated and interdisciplinary approach to inter-(dis)fluency, drawing on multiple research orientations and perspectives from previous research, was further reflected in our *mixed-method* methodology which relied on quantitative and qualitative analyses, described in Chapter 2 (cf Chap. 2, sections II and III). Following previous corpus-based approaches to (dis)fluency, we implemented a (dis)fluency annotation model, targeting different levels of analysis, i.e., level of individual fluencemes (form, duration, type), fluenceme sequence (patterns of co-occurrence, sequence type, length, utterance position, communication management), and visual-gestural level (gesture phase, gesture type, gaze direction). This model was applied systematically to our two datasets (The SITAF and DisReg Corpus) using different annotation and statistical tools (CLAN, ELAN, Excel, and statistical tests). In addition, in line with interactionist approaches to social interaction, we relied on *data-driven methods* with detailed micro qualitative analyses of a selection of excerpts from the data. These types of analyses, as pointed out in Chapter 2, are rarely addressed in psycholinguistics and cognitive linguistics, as the latter rather focus on isolated decontextualized utterances, or quantitative findings alone (cf Cienki, 2013). We thus argued that the combination of usage-based, interactional, and multimodal approaches could enable us to bridge the gap between large corpus-based quantitative studies and data-driven single case analyses or collection studies (cf Mazeland, 2006). It was argued that, while quantitative methods give a robust and statistically valid overview of the data, they fail to illuminate particular instances in a specific interactional sequence, whose complex information can never be truly conveyed in quantitative findings. On the other hand, micro qualitative analyses, although truly illuminating all the ongoing relevant interactional and social processes shaping the course of a specific exchange, only rely on a small selection of instances, thus disregarding all other instances of the same phenomena in the whole dataset. This was particularly relevant to our study of fluencemes, which were highly frequent in speech overall across the two datasets (amounting to 6042 tokens in total), and which consequently did not systematically exhibit essential features of talk-in-interaction. Reciprocally, their use was not systematically restricted to contexts of speech error or internal mental operations. This further emphasizes their dynamic and fluid nature,

whose degree of *fluency,* understood here as communicativeness, flow, or continuity, depends on a number of contextual and language features.

II. Inter-(dis)fluency across languages and settings: summary of the main findings

In the present thesis, we investigated the role of inter-(dis)fluency across languages and settings in order to measure the effect of language proficiency and language style on fluenceme use, and observe general differences in their pattern of distribution. Our first study was conducted on the SITAF Corpus (Chap. 3) which comprises video recordings of French and American students interacting in their first and second language, while engaged in an argumentative task. We compared the distribution of verbal fluencemes and gestures across the two speaker groups (American Group and French group) respectively in their first language and second language. One of the main research questions this study sought to answer was whether L2 fluency differed significantly from L1 fluency, and we showed how such differences could be measured by looking at temporal variables and fluency rates (following previous work in Second Language Acquisition), as well as sequential, positional, and visual-gestural features, leading to more subtle differences.

2.1. Study on the SITAF Corpus: native versus non-native productions

Overall, our results showed a higher rate of fluencemes in L2 than L1 (for both groups), with a number of differences in distribution: for the American Group, more repetitions and filled pauses were found in L2, but more self-interruptions and unfilled pauses in L1. For the French group, however, more self-interruptions and prolongations were found in L1, but more non-lexical sounds in L2, exhibiting a different pattern of distribution. In addition, the American group produced more complex fluenceme sequences in their L2 than in their L1, comprising a higher number of markers combined on average. However, no significant differences were found for the French group. Differences were also found in the sequence configurations: American speakers showed a tendency to produce sequences which mainly consisted in the *VOC+MS* configuration in their L2, and the *VOC+VOC* configuration in their L1, suggesting preferences for stalling strategies in the L1, as opposed to a mixture of stalling and

repair mechanisms in the L2. For the French group, however, the *VOC+MS* pattern was used more frequently in their L1 than in their L2, showing once again an opposite tendency. Overall, these results seemed to suggest that some specific patterns were more prominent than others, especially *VOC+MS* and *VOC+VOC,* but their use was not systematically determined by levels of proficiency, but rather by language preferences. Slight differences were also found in the positions of the fluenceme sequences in the two language groups: American speakers produced a slightly higher proportion of sequences in medial position in L2 than in L1, while the French group produced slightly more utterance-final fluencemes in their L1 than in the L2.

Most importantly, one of the major contributions of this study was to analyze inter-(dis)fluency with regards to gaze and gestural behavior, in order to go beyond a traditional view of L2 fluency which has too often been restricted to temporal variables in SLA research. Following previous work (e.g. Gullberg, 1998, Kita, 1993, Stam 2006), we expected a higher rate of gestures in L2 than in L1, and our findings confirmed this prediction, as the two speaker groups produced significantly more gestures in their L2 than in their L1, both in sequences with and without fluencemes (even though gestures did not frequently occur during vocal fluencemes), which demonstrated a higher gestural activity in the L2 overall. In addition, we argued against cognitive-psychological approaches to gesture with theories such as the *Lexical Retrieval Hypothesis* (LRH; Krauss et al., 2000), or *Information Packaging Hypothesis* (Kita, 2000) following the work of Graziano & Gullberg (2018). As argued, we did not believe that speakers produced more gestures in their L2 because of lexical problems (Beattie & Butterworth, 1979; Krauss & Hadar, 1999), or that they produced referential gestures during fluencemes to overcome a "lack of skill" in their L2, which would be too restrictive, suggesting that speech and gesture worked independently from one another. Indeed, as our quantitative results showed, the two speaker groups produced a higher rate of referential gestures in their L1 than in their L2 overall, which challenged the idea that speakers produced more referential gestures in their L2 to deal with lexical difficulties (Stam, 2001). In addition, the two speaker groups were found to produce a higher proportion of referential gestures in sequences *outside* fluencemes than during them, while a large majority of the gestures produced during fluencemes were pragmatic ones, contrary to what the LRH suggested. We also noted a higher rate of thinking gestures in the L2 than in the L1 overall (for both groups, and almost exclusively during fluencemes) which may reflect one prominent feature of L2

fluency as a display of *doing thinking* (Heller, 2021). Such gestures, along with thinking faces, were also examined in the qualitative analyses, and we showed how they were used as an interactional practice to display the progressivity of a word search. These displays were then further documented in Chapter 5 (further discussed in section III of this chapter). When it comes to gazing behavior, the two groups showed a tendency to withdraw their gaze more often during sequences of fluencemes than without them, and this was the case both in L1 and in L2, which demonstrates a notable feature of (dis)fluency in general, regardless of language proficiency. We will return to this point in the following section, and in section III. In sum, this study has revealed a number of characteristics differentiating native and non-native inter-(dis)fluency and visual-gestural behavior, summarized below:

- **Differences in frequency**, with a higher rate of fluencemes in L2, more complex sequences containing a higher number of markers combined (only for the American group).
- **Differences in marker and form distribution**, with a preference for certain markers in a specific language to build overall fluency (e.g. higher proportion of NL sounds in the L2 than in the L1 for the French group, or more self-interruptions in the L1 than in the L2 for the American group; more um-type filled pauses in the L2 than in the L1 for the French group, etc.,).
- **Different patterns of combination in the two languages** (e.g. *VOC+MS* in L2 versus *VOC+VOC* in L1 for the American group).
- **Slight differences in utterance position** (e.g., more utterance-medial fluencemes in the L2 compared to the L1 for the American group).
- **Differences in gestural behavior**, with a higher rate of gestures in L2 than in L1 (for both groups), as well as a higher proportion of held gestures during fluencemes in L2 than in the L1 (for both groups) as well as more thinking gestures in the L2 (for both groups).

However, it should be noted that a high degree of variability and dispersion was found in the data, with a number of crosslinguistic and individual differences, as well as inter-group variation, leading to more nuanced findings, and further suggesting that L2 fluency is highly speaker- and language- specific (De Jong, 2018). In addition, our qualitative analyses further shed light on the individual multimodal communication

strategies yielded by the L2 speakers, who made use of a variety of semiotic features (voice, face, gaze, manual gestures, the body, and material objects around them) to deal with language difficulties in the course of interaction. We further stressed the need to study L2 fluency within situated language use, following interactional frameworks such as *CA-SLA* (Pekarek-Doehler, 2006) and the notion of *interactional competence* (Galaczi & Taylor, 2018). We claimed that L2 competence, and to a larger extent, L2 fluency, was not solely the result of internal cognitive processes related to encoding difficulties (e.g. Hilton, 2009) but a relevant interactional tool for maintaining the *confluence* (McCarthy, 2009), i.e., interactional fluency, of the exchange. As our qualitative analyses further revealed, the types of gestures that were produced during fluencemes were not necessarily used to overcome language difficulties or to deal with intrapersonal problems, but rather to display forms of engagement in the interactional task at hand, leading to multimodal joint productions. We concluded that visible bodily features, such as gaze and manual gestures, played a major role in understanding the ambivalence of inter-(dis)fluency phenomena: while some speakers momentarily retreated from the current activity by withdrawing their gaze and orienting towards the piece of paper (Excerpts 3a and 3b), others relied on mutual gaze and other communicative gestural activities (interactive and representational gestures) to co-construct meaning with their partner (excerpts 1a and 2b). Once again, this study on the SITAF Corpus underlined the interactional ambivalence of fluencemes in situated tandem interactions, and the fact that their usage is highly contextual and depends on a number of multimodal and situational features (cf section III below). This multi-level ambivalence was further explored in our second study on the DisReg Corpus.

2.2. Study on the DisReg Corpus: individual class presentations versus dyadic conversations

In our second study, we targeted differences in setting and style (Chap. 4). The second dataset under study comprises recordings of French students engaged in two different speaking tasks in entirely different settings, mainly graded class presentations in front of the classroom, versus casual dyadic face-to-face interactions between friends or classmates. One of the main questions this study sought to answer was whether style and setting had an effect on (dis)fluency and gesture production, and whether significant differences would be found across the two situations. We further presented

the notions of style and setting as multidimensional, encompassing a wide array of inter-related factors, such as audience design, multimodal environment, turn-taking mechanisms, or register, in order to identify the different types of variables characterizing the two situations. As explained, (Chap. 4, section I) the two situations differ quite significantly on a number of levels: class presentations are characterized by a number of institutional constraints (no interaction with the audience, the student needs to present a specific assignment prepared at home that will later be graded), as well as spatial ones (the student is alone, sitting or standing in front of a desk, facing a group of students and the teacher). In addition, the style of the student must be formal, intelligible, and to do so students may rely extensively on their notes to deliver a successful presentation. On the contrary, face-to-face interactions are more informal and do not require the participants to talk for a limited amount of time on a restricted topic. Moreover, the two partners know each other fairly well, which gives them opportunities to share common ground and take part in joint activities.

All these features were shown to have an effect on the use of (dis)fluency and visual-gestural behavior: a higher rate of fluencemes was found in class presentations, with significantly longer unfilled pauses, and a higher proportion of non-lexical sounds as well as filled pauses. The latter were also more often realized with the nasal variant ("eum") in class presentations than conversations, suggesting a longer delay (Clark & Fox Tree, 2002). In addition, a slightly higher proportion of complex fluencemes was found in class, but without differences in length. A tendency for fluenceme sequences to occur in utterance-initial position was also found in presentations, which potentially reflected a rhythmic and stylistic style, in line with Duez (1982). As to the sequence configurations, the *VOC+NL* pattern was more prevalent in class presentations than conversations, which further reflected the recurrence of non-lexical sounds in this specific institutional context. Overall, the distributional differences found across the two situations (i.e., longer pauses, slightly more complex sequences, and more instances in utterance-initial position) suggested that class presentations required more time for planning and monitoring processes than conversations. This was further confirmed by the mean length of utterances which were much longer in class than in conversation, and the positive correlation found between unfilled pause duration and utterance length, with longer pauses associated with longer utterances, as well as the overwhelming proportion of

fluenceme sequences which pertained to *Own Communication Management* (almost a majority of instances) in presentations.

When it comes to gesturing and gazing behavior, a number of differences were also found across the two situations. A higher rate of gestures was found in class presentations, with a higher proportion of discursive gestures. Conversations, on the other hand, comprised a higher proportion of interactional and representational gestures, in line with Bavelas et al.'s (2008). No significant differences were found for the proportion of thinking gestures however, but the latter almost occurred only exclusively during fluencemes in the two situations, which is consistent with findings from the SITAF Corpus. As the qualitative analyses further showed, speakers often made use of interactional gestures in the conversations to perform a series of communicative actions, such as establishing common ground, displaying a stance, or addressing their interlocutor in the course of their interactive practices (cf Excerpts 2.2 and 2.3). During their oral presentations, however, students almost never addressed their audience, except for a few exceptions (cf Excerpt 1.1.), but mostly made use of gestures to segment discourse and mark information structure. In addition, a considerable proportion of gazing towards the piece of paper was found in the presentation-sessions, with a relatively very small proportion of gazing towards the interlocutor, which was a significant difference with the conversations. This finding seemed somewhat at odds with the high gesturing activity found in the presentations; we thus concluded that even though the students produced more gestures in class than in conversation, they did not use them to truly engage with their audience (as they would more often do in conversations), since they were too engrossed in their notes. In addition, findings further showed that speakers were more likely *not* to establish eye contact when they produced fluencemes in both situations, which was consistent with SITAF where the two language groups were found to withdraw their gaze more frequently during fluencemes than outside them, both in their L1 and L2. This further emphasized the fact that gazing away is a very common practice of (dis)fluency, regardless of language or setting, as it enables speakers to momentarily retreat from the current activity to attend to other relevant ones, such as retrieving an item from memory, looking for a specific word, checking for a sentence in a book, etc. However, we also showed several instances of mutual gaze coordinated with fluencemes in the conversations (see excerpts 2.2 and 2.3 during which speakers were engaged in interactive practices), which further revealed the potential for fluencemes to embody

interactive processes, and not only intrapersonal ones. This ambivalence was further explored in Chapter 5, where we documented a series of practices involving gaze withdrawal (i.e. with thinking gestures) and mutual gaze (i.e. with interactive gestures). We will return to this point in section III. In sum, just like SITAF, we found a number of characteristics differentiating inter-(dis)fluency and visual-gestural behavior across the two situations, summarized below:

- **Differences in frequency**, with a higher rate of fluencemes in presentations, and slightly more complex sequences than in the conversations.
- **Differences in marker and form distribution**, with a preference for certain markers in a specific setting to build overall fluency (e.g. higher rate of NL sounds, filled pauses and unfilled pauses of longer duration in class, as opposed to more repetitions, interruptions, and prolongations in conversation).
- **Different patterns of combination in the two settings** (e.g. *VOC+MS* in the conversations versus *VOC+NL* in the presentations).
- **Differences in utterance position** (e.g., more utterance-initial fluencemes in the presentations, as opposed to a higher proportion of utterance-final ones in the conversations).
- **Differences in gesture distribution**, with a higher rate of gestures in class than in conversation, as well as a higher proportion of discursive gestures in class, as opposed to a higher proportion of interactive and representational gestures in conversation.
- **Differences in gaze behavior,** with an overwhelming proportion of gazing towards paper in class, and more instances of gazing towards interlocutor and away in the conversations compared to the presentations.
- **Differences in communication management**, with a majority of fluencemes performing own communication management and a quasi-absent proportion of interactive communication management in class presentations, as opposed to the conversations.

It should also be noted that just like SITAF, a great number of individual differences were found in the data, displaying different tendencies and patterns of distribution across speakers, which further confirmed that fluency is in part dependent on personal speaking style, regardless of setting or proficiency. This further gives support to the

key role of qualitative analyses in linguistic research, which enable us to capture differences that are not otherwise visible in quantitative findings only. In particular, we showed how participants adjusted their body and talk to perform the task at hand: in class presentations, the presenters relied on the different objects they had within reach (their notes, their pen, their book, etc.,) to maintain the continuity of their presentation, and attend to several presentation-relevant activities (i.e. look for the right passage from the book, organize their notes, etc.). In conversations, on the other hand, the participants relied much more on their partner to perform a series of interactive actions (i.e. display a stance, yield a turn, etc.). Our analyses further revealed the interplay between fluencemes, gestures, actions, and manipulation of objects, which can all be deployed together to build the fluency of multimodal discourse. We concluded that fluencemes and gestures were highly sensitive to a number of multimodal situational features, other than language proficiency, which further gives support to the claim that they should not be considered in decontextualized utterances but in situated language use.

2.3. Synthesis

In this thesis, we sought to provide a complex picture of inter-(dis)fluency phenomena and the several potential underlying factors and variables affecting their use in two specialized datasets of French and English. Although they were compiled using quite different procedures, the two data samples are comparable to a certain extent (i.e. use of video, similar corpus size, similar speaker profiles, see Chap. 2, section I.1.4) which enabled us to get an overall idea of the different patterns of distribution of fluencemes and gestures found across languages and settings. As specified in the previous section, our two studies have shown that (dis)fluency and gesture are highly sensitive to language and contextual multimodal factors, but are also in part dependent on individual style. We stressed the need to combine quantitative and qualitative analyses to capture these differences, and consider inter-(dis)fluency from a multidimensional perspective (see section III of this chapter). Table 45 summarizes the main quantitative findings described in Chapters 3 and 4 from the SITAF and DisReg Corpus.

Table 45. *Summary of the main findings in SITAF and DisReg*

	SITAF			DisReg	
VOCAL-VERBAL FEATURES FEATURES (FLUENCEMES)					
	L1	**L2**		**class**	**conversation**
Fluenceme rate	Higher rate in L2 than L1 (both groups)			Higher rate in class than conversation	
Distribution	more self-interruptions and unfilled pauses (American group) more self-interruptions and prolongations (French group)	more identical repetitions and filled pauses (American group) more non-lexical sounds (French group)		more NL sounds, filled pauses, and longer unfilled pauses	more repetitions and prolongations
Filled pause type	more um-types (American group) more euh-types (French group)	more um-types (French group); more euh-types (American group)		more eum-types	more euh-types
Combination type	higher proportion of complex sequences (only American group, no significant differences for the French group)	higher proportion of simple sequences (only American group, no significant differences for the French group)		sligthly higher proportion of complex sequences	slightly higher proportion of simple sequences
Combination pattern	**VOC+MS (American) VOC+MS (French)**	higher proportion of VOC+MS (American)		higher proportion of VOC+NL	higher proportion of VOC+MS
Utterance position	**more utterance-final (French group)**	more utterance-medial (American group)		more utterance-initial	more utterance-final
Communication Management	no significant differences between L1 and L2 (both groups)			overwhelming proportion of own communication management	slightly higher proportion of interactive communication management
VISUAL-GESTURAL FEATURES (WITHIN AND OUTSIDE FLUENCEMES)					
	L1	**L2**		**class**	**conversation**
Gesture rate	Higher rate of in L2 than in L1			Higher rate in class than conversations	
Gesture distribution	higher proportion of held gestures during fluencemes in L2 than L1			no significant differences between the two settings	
	more referential gestures (both groups)	more thinking gestures (both groups) which almost never occured outside fluencemes		no significant differenes for the thinking gestures, but they occured almost exclusively during fluencemes	
				more discursive gestures	more interactive and referential gestures
Gaze behavior	no significant differences between L1 and L2 (both groups)			majority of gazing towards paper	higher proportion of gazing towards interlocutor

As explained in Chapter 2, the same annotation system was applied to the two data samples, using the same numerical and categorical variables in both studies in order to triangulate evidence from the two corpora and further give support to the ambivalent status of (dis)fluency. As the table shows, inter-(dis)fluency can be characterized by a number of recurrent features, or traits, such as form, duration, type, pattern of co-occurrence, etc., and some of them were found to be more prominent in certain situations. For example, *combination type* was shown to be affected by language proficiency specifically in the American group from the SITAF Corpus, with a higher proportion of complex sequences containing more fluencemes in the L2 than in the L1, suggesting a more complex pattern of co-occurrence in L2 fluency. However, these differences did not reach statistical significance within the French group, which may further reveal language characteristics, or individual differences.

In addition, some aspects of inter-(dis)fluency and gesture were found to differ significantly in SITAF, but not DisReg, or the other way around. For instance, no significant differences were found in the proportion of fluencemes performing *Own Communication Management* (OCM) versus *Interactive Communication Management* (ICM) in SITAF, as a majority of them served intrapersonal functions, regardless of language proficiency or language group. In DisReg, however, significant differences were found between the two situations, with a quasi-absent proportion of fluencemes serving interpersonal functions in class presentations, contrary to the conversations. This is not surprising, given the number of differences characterizing the two situations (level of interactivity, type of audience, task type etc.) which inevitably had an effect on the function of fluencemes. In addition, the speakers from the SITAF Corpus were shown to produce a higher proportion of held gestures during fluencemes in their L2 than in their L1 (both groups), while no significant differences were found in DisReg between the two situations. This may reveal another feature of L2 fluency marked by a higher degree of suspension in the two modalities. Similarly, no differences were found in gaze behavior between L1 and L2 in SITAF both within and outside fluencemes, while it exhibited radically different patterns across the two situations in DisReg, with an overwhelming proportion of gazing towards the piece(s) of paper in class presentations (both within and outside fluencemes) compared to the conversations which relied more on mutual gaze. This is one distinct feature of class presentations, noted in Chapter 4, which is marked by a quasi-total absence of mutual gaze and engagement with the audience. Lastly, both language groups from the SITAF

Corpus showed a tendency to produce more thinking gestures in their L2 than in their L1, while no differences were found in DisReg, hinting once more to the fact that the association between the activity of doing thinking and the emergence of fluencemes may be more prominent in L2 than in L1 at a quantitative level. In Chapter 5, however, we also showed how such displays could also be manifested in the L1, at a qualitative level.

Lastly, some of the findings described in the two corpora also converge in some aspects: all speakers were shown to keep their hands in rest position and gaze away more frequently during fluencemes than outside them, which reveals a recurrent feature of inter-(dis)fluency overall, marked by little gestural activity and gaze aversion. Other notable features can be noted: in SITAF, the French group produced a higher proportion of euh-type filled pauses in their L1 compared to the L2, and in DisReg, the French students produced more *euhs* in the conversations than in the presentations. These findings may suggest that *euh* is a common conversational marker in French, often found in spontaneous face-to-face conversations, as opposed to *eum* which is rarer overall and tends to occur in different contexts (L2 English discourse and L1 French class presentations). Similarly, the *VOC+MS* pattern was found more frequently in French L1 (SITAF) and in French conversation (DisReg), and more utterance-final fluencemes were found in these two situations as well (French L1 and French conversation), which could further suggest another feature of conversational fluency in L1 French. Although these are simply preliminary observations which go beyond the scope of this thesis, these findings may still be of interest to researchers in (dis)fluency, and may call for further work in the field (cf section IV).

III. Beyond Disfluency: Towards a multidimensional scale of inter-(dis)fluency

Following the work of Crible et al., (2019) the present thesis aimed to explore the *ambivalent status* of fluencemes, emphasizing the fact that they should not be restricted to one label (“cognitive burden”) or another (“communicative signal”) hence vouching for a more flexible and dynamic approach. Although other researchers have previously noted the ambivalence of fluencemes in prior research (e.g. Götz, 2013, Tottie, 2014, Allwood, 2017, among others), Crible et al., were the first to systematically explore this idea with the application of a reliable annotation system.

One of the most valuable contributions of Crible et al.,'s work within the field of (dis)fluency research was to go beyond the common distinction made between "fluency" and "disfluency" and provide methodological tools to uncover the complexity and duality of these phenomena across languages and settings, thus placing them in a functional continuum, with a range of potentially (dis)fluent functions, depending on a number of contextual and situational factors (e.g., genre, position, co-occurrence, etc.). Our annotation scheme was largely adapted from their work, and some variables, such as *sequence type* (simple versus complex) and *sequence pattern* (pattern of combination) were particularly useful in our own work, and further gave support to the fact that (dis)fluency is better understood in terms of constructions, rather than isolated tokens, exhibiting recurrent patterns of combination (e.g. *VOC+VOC*, *VOC+MS*). This dynamic view of fluencemes was also put forward by Cienki (2012, 2015b) who acknowledged the dynamic and complex nature of fluencemes which, along with gestures and intonation, have the potential to gain symbolic status, and form an integral part of multimodal talk and usage events.

In addition, the present study stems from the valuable work of Götz (2013), Segalowitz (2016), and Grosman (2018) who introduced *multidimensional* models of fluency, distinguishing between different levels of analysis, such as utterance, cognitive, and perceived fluency for Segalowitz (2016), or productive, perceptive, and nonverbal fluency for Götz (2013). However, none of their conceptual models have been effectively implemented into an annotation scheme, in the exception of Grosman's (2018), which was used for experimental purposes (cf Grosman et al., 2019). As specified in Chapter 1, the view of fluency as *multidimensional* is the central contribution of this thesis. The conflating views on "fluency" versus "disfluency", or the total absence of the term "disfluency" in the field of interactional linguistics can largely be explained by differences in perspective and choice of dimension. In psycholinguistics, Second Language Acquisition, and even Corpus Linguistics, the main focus remains the level of *speech production*, thus zooming in on (dis)fluency rates, their position in the utterance, their combination with other markers, and the different ways speech can be suspended and interrupted. In Interactional Linguistics however, the main focus is on the sequential development of the exchange, and the timely ordering of turns in interaction, hence almost systematically regarding fluencemes as conversational displays. And in Gesture Studies, very little attention is paid to fluencemes since the latter rarely co-occur with gestures, which is the main

topic of interest in this field. As pointed out in Chapter 1, the main issue with the existing literature is that even though inter-(dis)fluency has been analyzed across different research disciplines within a variety of theoretical and methodological frameworks, very few of them communicate with one another. This further led us to the development of an *integrated* approach to inter-(dis)fluency, drawing from different theoretical frameworks, and adopting a mixed-methods methodology (cf section I of this chapter, and Chapters 1 and 2). In this view, the ambivalence of fluencemes, put forward in the work of Crible et al.,'s (2019), can further be examined by considering three different dimensions of inter-(dis)fluency (speech, gesture, and interaction, cf Chap. 1, section IV), as illustrated in the figure below:

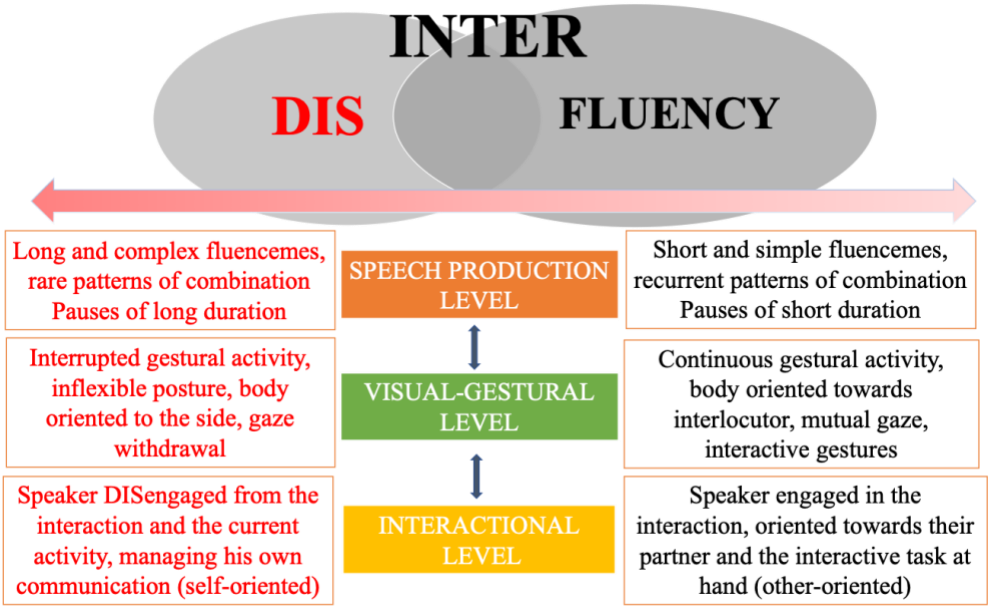


Figure 71. *Multidimensional scale of inter-(dis)fluency*

The core notion of *fluency*, understood here as flow, continuity, progressivity, or communicativeness, is a flexible and dynamic process which is constantly reshaping its focus, and which can potentially be interrupted, suspended, or disrupted, at different levels of analysis. In the verbal and vocal level, a short and simple vocal fluenceme is not often deemed disruptive and "disfluent", since it is barely perceptive in the vocal channel, but a very long sequence containing a dozen of different markers can mark a significant delay in the speech signal, hence disrupting its initial delivery. This is the traditional view of disfluency, as cued by several temporal variables, as well as others, such as syntactic position, or co-occurrence with other markers. The visual-gestural and interactional dimensions offer entirely different views on these phenomena, since a priori "disfluent" forms in the speech signal can still perform

"fluent" communicative actions in the interactional flow. For instance, in Chapter 3, we showed a series of examples from the SITAF Corpus in which non-native speakers produced a number of fluencemes in the verbal channel, hence temporarily interrupting the flow of speech, but who also made use of their hands and gaze to co-construct meaning with their interlocutor, hence contributing to the interactional fluency of the exchange (Excerpts 4a, 4b, and 4c). Conversely, in Chapter 4, we showed how one particular student, who was ideally "fluent" from a speech production perspective, as he produced very few fluencemes compared to the rest of the group, was in fact quite "disfluent" from a visual-gestural and interactional perspective, since he seemed very *DIS*engaged from his audience, only focusing on his own production (cf Excerpt 1.5.a.). In Chapter 5, we further put forward the role of gaze direction to index changes of participation, from self-oriented cognitive practices to other-oriented interactive ones. In some cases, these dimensions were shown to converge (i.e. gaze withdrawal associated with a self-oriented practice marked by a very long fluenceme sequence), and in others they were divergent (e.g. a short fluenceme disrupting the progressivity of the exchange).

Lastly, it was claimed multiple times in this thesis that the integration of bodily visible behavior was fundamental to the conceptualization of this multidimensional model. Even though the verbal or vocal mode remains the default focus of (dis)fluency (given the limited number of occurrences accompanying gestures overall), speakers may still build the fluency of their discourse by choosing among different semiotic resources, following the *scope of relevant behavior* theory (Cienki, 2012), the notion of *contextual configuration* (Goodwin, 2000), and Müller & Tag's (2010) *dynamic focus on attention* theory (cf Chap. 5). Across chapters 3, 4, and 5, we illustrated a timely coordination between speech and gesture, with instances of speech suspension (vocal and verbal fluencemes) coordinated with gesture suspension or retraction, and speech preparation with gesture preparation. This showed that the notions of suspension, interruption, or preparation, which are inherent to *speech* (dis)fluency, were not only marked verbally or vocally, but gesturally as well, which further gave support to the view of language, or *languaging* as a unified system operating upon a diverse range of semiotic systems. In addition, we showed how inter-(dis)fluency could also be manifested in recurrent embodied interactional practices (Chap. 5). In particular, we focused on three specific ones (i.e., doing thinking, displaying intersubjectivity, and co-constructing meaning) which were characterized by specific

features in gaze behavior, body orientation, posture, gesture use, fluenceme type, and a coordination was further found between the emergence of vocal and verbal fluencemes in the speech channel and particular visual-gestural affiliates in the gestural channel. We concluded that fluencemes, along with gestures, gaze, body orientation, and object manipulation, were integral components of the interactive and dynamic process of doing language. Despite the fact that fluencemes rarely co-occur with gestures, their coordination with bodily visible behavior should not be overlooked, as we maintain that these phenomena should be regarded as complex, ambivalent, multimodal systems, altogether shaped by a myriad of interrelated factors across different contexts of use.

IV. Perspectives for future work

As noted in Chapters 3, 4, and 5, the present study presents a number of limitations, which calls for further research in the field. First, our data sample remains quite small and limited, especially for DisReg which is based on a selective sample used to match the size of SITAF. While we justified our choice to work on a small corpus for the present study (cf Chap. 2), we intend to extend our analysis to the whole dataset for further research, in order to see if the results presented here could be generalized to a larger sample. In addition, due to the health crisis of Covid 19, we could not achieve our data collection project and compare the American group from the SITAF Corpus with American students at Berkeley University as part of the DisReg Corpus, which would have allowed for more crosslinguistic comparisons. As explained at the end of Chapter 4, we will resume this project to complete our existing dataset of French students in DisReg, if the situation allows.

In addition, we also mentioned the limits of our statistical analyses, which were chosen to remain quite simple, as statistical, purely quantitative descriptions were not the core approach of this study. Indeed, we chose to mostly use a binary categorization ("L1" versus "L2", "class" versus "presentations", or "gesture" versus "no gesture", etc.); given the multiplicity of factors potentially affecting (dis)fluency use, more complex statistical models, random and fixed effects, such as mixed-model regression models, or multiple correspondence analysis, should be used in the future, in order to uncover which aspects of discourse may be more affected by (dis)fluency across different contexts of use. More progress regarding our methodology could already be considered for further investigation. While we started with the analysis of "traditional"

fluencemes, following previous typologies in (dis)fluency research, we also included the analysis of non-lexical sounds in our typology, since they were found to very often cluster with fluencemes, and should thus not be overlooked (cf Chap. 2), but we mainly focused on two types (inbreaths and tongue clicks), because they were found to occur most frequently with other fluencemes, while we paid less attention to other types of laughing, breathing and sniffing phenomena in our quantitative analyses. In a more recent model of (dis)fluency annotation built for typical and atypical speech, Dirdirkovà et al. (in press) suggested to include all types of non-lexical sounds in the typology, without an a priori hierarchy. Similarly, we used a duration threshold for the annotation of silent pauses, which is questionable to a certain extent, given the number of pauses that were left out from the analyses. In future research, we may wish to follow Campione & Véronis (2002) and Betz (2021) who did not use a cut-off point for their identification of pauses. In a similar vein, the phonetic, acoustic, and prosodic dimensions of inter-(dis)fluency remain quite underexplored in our study, since we mostly focused on visual-gestural features, which have received less attention in the (dis)fluency literature. This opens up perspectives of collaboration with experts in the fields of phonetics and prosody (e.g., Dodane, 2020; Exare, 2017; Ferré, 2008; Horgues & Scheuer, 2015; Lelandais, 2019; Ogden, 2020) to further our understanding of inter-(dis)fluency as a truly multimodal process. In addition, several questions regarding methodology in gesture research were raised in Chapter 5. While we focused on a functional typology in our quantitative model in order to work on a finite set of categories that would be better suited to statistical treatments, other aspects of gestures, such as form, handshape, orientation, configuration, quality of movement, flow, segments involved, and the like, were completely left out from the quantitative annotations, and were only observed in the qualitative analyses. It was thus suggested to use a multi-level gestural annotation system for future work, following previous ones such as *The MultiModal MultiDimensional* (M3D) labelling scheme for the annotation of audiovisual corpora (Rohrer et al., 2020).

Lastly, the aim of the present thesis was to go beyond traditional production-oriented approaches to (dis)fluency phenomena in order to capture the multimodality of these processes and integrate several dimensions (speech, gesture, interaction), but it should be noted that our initial analysis was still based on the annotation of verbal and vocal fluencemes in the *speech channel*, which was the starting point of this study, following previous work in (dis)fluency research. One outcome of this study would be

to include other forms that also embody the notion of (dis)fluency in the *visual-gestural channel*, such as thinking postures, held gestures, palm-up gestures, among other visible practices which have been documented in this study.

So far, the present study has made two central theoretical and methodological contributions to the current field of (dis)fluency research, which are, in our opinion, also relevant to the field of linguistics and the study of *language* in general. First, this study has provided several tools for the analysis of fluency on a multidimensional scale, combining several methods from different theoretical frameworks, such as Conversation Analysis, gesture studies, and psycholinguistics. What the quantitative findings have suggested so far is that the complexity of inter-(dis)fluency phenomena cannot easily be broken down into a finite number of categories, and that despite general tendencies in the data, it was deemed necessary to integrate all aspects of human communication to capture the intricacies of these processes. The complexity and multifunctionality of these processes were in fact highlighted in the qualitative analyses, as they shed light on the importance of individual variation and contextual features. Second, this study has introduced a fresh and innovative approach to fluency as a multifaceted, multimodal, and dynamic phenomenon, without restricting it to temporal variables, language proficiency, repair processes, or speech error, which, we hope, may further help us unravel some of the most fascinating issues surrounding these phenomena.

References

Abrahamson, D., & Bakker, A. (2016). Making sense of movement in embodied design for mathematics learning. *Cognitive Research: Principles and Implications*, *1*(1), 1–13.

Adams, T. W. (1998). *Gestures in foreigner talk*. [Unplublished PhD thesis]. University of Pennsylvania.

Akhavan, N., Göksun, T., & Nozari, N. (2016). Disfluency production in speech and gesture. *CogSci*. Cognitive Science Society.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, *5*(4), 307–331.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, *15*(6), 593–613.

Alibali, M. W., & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. *Video Research in the Learning Sciences*, 349–365.

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*.

Allwood, J., Nivre, J., & Ahlsén, E. (1990). Speech Management—On the Non-written Life of Speech. *Nordic Journal of Linguistics*, *13*(1), 3–48.

Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics*, *31*(3), 397–411.

Atkinson, J. M., & Heritage, J. (1984). *Structures of Social Action*. Cambridge University Press.

Atkinson, P., Becker, H., Bergmann, J. R., Blumer, H., Davis, F., Garfinkel, H., Glaser, B., & Strauss, A. (2002). Analysing Interaction: Video, Ethnography and Situated Conduct. In T. May (Ed.), *Qualitative research in action*. SAGE Publications.

Ayaß, R. (2015). Doing data: The status of transcripts in Conversation Analysis. *Discourse Studies*, *17*(5), 505–528.

Azaoui, B. (2015). Fonctions pédagogiques et implications énonciatives de ressources professorales multimodales. Le cas de la bimanualité et de l'ubiquité coénonciative. *Recherches En Didactique Des Langues et Des Cultures. Les Cahiers de l'Acedle*, *12*(12–2).

Azi, Y. A. (2018). Fillers, Repairs and Repetitions in the Conversations of Saudi English Speakers: Conversational Device or Disfluency Markers. *International Journal of Linguistics*, *10*(6), 193–205.

Bailey, K. G., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*(2), 183–200.

Barlow, M. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics*, *18*(4), 443–478.

Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2001). Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication*, *33*(1), 5–22.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*(4), 469–489.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and social psychology bulletin*, *21*(4), 394—405.

Bavelas, J., & Chovil, N. (2018). Some pragmatic functions of conversational facial gestures. *Gesture*, *17*(1), 98–127.

Bavelas, J., Gerwing, J., & Healing, S. (2014). Effect of dialogue on demonstrations: Direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Processes*, *51*(8), 619–655.

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., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*(1/2), 001–030.

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.

Beaupoil-Hourdel, P. (2015). *Acquisition et expression multimodale de la négation. Etude d'un corpus vidéo et longitudinal de dyades mère-enfant francophone et anglophone* [Unpublished PhD Thesis]. Sorbonne Nouvelle-Paris 3.

Belhiah, H. (2013). Gesture as a resource for intersubjectivity in second-language learning situations. *Classroom Discourse*, *4*(2), 111–129.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Bell, A. (2006). Speech Accommodation Theory and Audience Design. In J. L. Mey & R. E. Asher (Eds.), *Concise encyclopaedia of pragmatics (second edition)* (pp. 992–994). Elsevier Oxford.

Benus, S., Enos, F., Hirschberg, J. B., & Shriberg, E. (2006). *Pauses in Deceptive Speech*. *Proceedings of ISCA, 3rd International Conference on Speech Prosody*.

Betz, S. (2020). *Hesitations in Spoken Dialogue Systems* [Unpublished PhD Thesis]. Bielefeld University.

Betz, S., Carlmeyer, B., Wagner, P., & Wrede, B. (2018). Interactive Hesitation Synthesis: Modelling and Evaluation. *Multimodal Technologies and Interaction*, *2*(1), 9.

Betz, S., & Gambino, S. L. (2016). Are we all disfluent in our own special way and should dialogue systems also be? *Elektronische Sprachsignalverarbeitung (ESSV) 2016*, *81*.

Betz, S., & Kosmala, L. (2019). Fill the silence! Basics for modeling hesitation. *The 9th Workshop on Disfluency in Spontaneous Speech*, 11.

Betz, S., & Wagner, P. (2016). Disfluent Lengthening in Spontaneous Speech. *Elektronische Sprachsignalverarbeitung (ESSV) 2016*.

Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: A social semiotic perspective. *International Journal of Social Research Methodology*, *14*(3), 191–206.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

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.

Boulis, C., Kahn, J. G., & Ostendorf, M. (2005). The role of disfluencies in topic classification of human-human conversations. *AAAI Workshop on Spoken Language Understanding*.

Boutet, D. (2008). Une morphologie de la gestualité: Structuration articulaire. *Cahiers de Linguistique Analogique*, *5*.

Boutet, D. (2010). Structuration physiologique de la gestuelle: Modèle et tests. *Lidil. Revue de Linguistique et de Didactique Des Langues*, *42*, 77–96.

Boutet, D. (2018). *Pour une approche kinésiologique de la gestualité* [Habilitation à diriger des recherches]. Université de Rouen-Normandie.

Boutet, D., Morgenstern, A., & Cienki, A. (2016) Grammatical Aspect and Gesture in French : a kinesiological approach. *Vestnik RUDN*, *20*(3), 131—150.

Bowers, J., Pycock, J., & O'brien, J. (1996). Talk and embodiment in collaborative virtual environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 58–65.

Brand, C., & Götz, S. (2013). Fluency versus accuracy in advanced spoken learner language. *Errors and Disfluencies in Spoken Corpora*, *52*, 117–137.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*(2), 274–296.

Bressem, J., & Müller, C. (2014). A repertoire of German recurrent gestures with pragmatic functions. In *Handbücher zur Sprach-und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK) 38/2* (Vol. 2, pp. 1575–1591). De Gruyter Mouton.

Brouwer, C. E. (2004). Doing pronunciation: A specific type of repair sequence. *Second Language Conversations*, 93–113.

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, *32*(10), 1439–1465.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language* (pp. 347–360). Springer.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, *96*(1), 168–174.

Bybee, J. (2008). *Usage-based grammar and second language acquisition*. Routledge.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press.

Calvert, M., & Brammerts, H. (2003). Learning by communication in tandem. In T. Lewis & L. Walker (Eds.), *Autonomous Language Learning in Tandem*. Academy Electronic Publication.

Campbell, J., Hill, D., & Driscoll, M. (1991). Systematic Disfluency Analysis: Using SDA to determine stuttering severity. *Annual Convention of the American Speech-Language-Hearing Association, Anaheim, CA*.

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *Speech Prosody 2002, International Conference*.

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*. [Unpublished PhD Thesis]. Université Sorbonne Nouvelle – Paris III.

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.

Candea, M., Vasilescu, I., & Adda-Decker, M. (2005). Inter-and intra-language acoustic analysis of autonomous fillers. *DISS 05, Disfluency in Spontaneous Speech Workshop*, 47–52.

Cenoz, J. (1998). *Pauses and Communication Strategies in Second Language Speech*. (ERIC Document ED 426630).

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT press.

Christenfeld, N., Schachter, S., & Bilous, F. (1991). Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research*, *20*(1), 1–10.

Christodoulides, G., Avanzi, M., & Goldman, J.-P. (2018). DisMo: A morphosyntactic, disfluency and multi-word unit annotator. An evaluation on a corpus of French spontaneous and read speech. *ArXiv Preprint*

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, *37*(6), 871–887.

Cienki, A. (2004). Bush's and Gore's language and gestures in the 2000 US presidential debates: A test case for two models of metaphors. *Journal of Language and Politics*, *3*(3), 409–440.

Cienki, A. (2005). Image schemas and gesture. *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, *29*, 421–442.

Cienki, A. (2012). Usage events of spoken language and the symbolic units we (may) abstract from them. *Cognitive Processes in Language*, 149–158.

Cienki, A. (2015a). 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. (2015b). Spoken language usage events. *Language and Cognition*, *7*, 499–514.

Cienki, A. (2016). Cognitive Linguistics, gesture studies, and multimodal communication. *Cognitive Linguistics*, *27*(4), 603–618.

Cienki, A. (2017a). Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard*, *3*(s1).

Cienki, A & Irishkhanova, O.K. (2018) *Aspectuality across Languages*. John Benjamins.

Ciolek, T. M., & Kendon, A. (1980). Environment and the spatial arrangement of conversational encounters. *Sociological Inquiry*, *50*(3–4), 237–271.

Clark, H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111.

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

Clark, H. H. (2002). Speaking in time. *Speech Communication*, *36*(1), 5–13.

Clark, H. H. (2006). Pauses and hesitations: Psycholinguistic approach. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 244–248). Oxford: Elsevier.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242.

Cochet, H., & Vauclair, J. (2014). Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(3), 278–301.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass*, *2*(4), 589–602.

Couper-Kuhlen, E., & Selting, M. (2001). Introducing interactional linguistics. *Studies in Interactional Linguistics*, *122*.

Coupland, N. (2001). Dialect stylization in radio talk. *Language in Society*, *30*, 345–375.

Crible, L. (2017). *Discourse markers and (dis)fluency across registers: A contrastive usage-based study in English and French* [PhD Thesis]. UCL - Université Catholique de Louvain.

Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. John Benjamins Publishing.

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses. *Languages in Contrast*, *17*(1), 69–95.

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* (Corpora and Language in Use-Proceedings 4). Presses universitaires de Louvain.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson Education.

Crookes, G., & Rulon, K. A. (1985). *Incorporation of corrective feedback in native speaker/non-native speaker conversation* (Technical Report No. 3). Center for Second Language Classroom Research. Social Science Research Institute. University of Hawaii.

Cucchiarini, 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.

Cutting, J. (2001). The speech acts of the in-group. *Journal of Pragmatics*, *33*(8), 1207–1233.

Cutting, J. (2002). The in-group code lexis. *HERMES-Journal of Language and Communication in Business*, *28*, 59–80.

Danino, C. (2018). Les petits corpus—Introduction du numéro thématique. *Corpus*, *18*.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

De Jong, N. H.,(2016a). Fluency in second language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment*. Mouton de Gruyter.

De Jong, N. H. (2016b). 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.

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. *The 6th Workshop on Disfluency in Spontaneous Speech (DiSS)*, 17–20.

De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, *19*(2), 85–114.

De Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture*, *7*(1), 21–38.

Debras, C. (2013). *L'expression multimodale du positionnement interactionnel (multimodal stance-taking): Étude d'un corpus oral vidéo de discussions sur l'environnement en anglais britannique* [Unpublished PhD Thesis]. Université Sorbonne Nouvelle -Paris III.

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.), *La correction en langue(s) – Linguistic Correction/Correctness*. 91—115.

Debras, C., & Cienki, A. (2012). Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking. *International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 932–937.

Debras, C., Horgues, C., & Scheuer, S. (2015). The multimodality of corrective feedback in tandem interactions. *Procedia-Social and Behavioral Sciences*, *212*, 16–22.

Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, *50*(7), 431–456.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 533–557.

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.

Didirkovà, I., Dodane, C., & Diwersy, S. (2019). The role of disfluencies in language acquisition and development of syntactic complexity in children. *The 9th Workshop on Disfluency in Spontaneous Speech*, 85.

Didirkovà, I., Crible, L., Dodane, C., Kosmala, L., Morgesntern, A., Monfrais-Pfaudwel, M. C., & Hirsh, F., (in press) Towards an inclusive system for the annotation of (dis)fluency in typical and atypical speech. *Proceedings of DiSS 2021 – Disfluency in Spontaneous Speech*.

Dingemanse, M. (2020). Between sound and speech: Liminal signs in interaction. *Research on Language and Social Interaction*, *53*(1), 188–196.

Dodane, C. (2020). *Au commencement était la prosodie: Du langage en émergence à l'histoire de la description de la parole* [Habilité à Diriger des Recherches]. Université de Toulouse Jean Jaurès.

Dodane, C., & Hirsch, F. (2018). L'organisation spatiale et temporelle de la pause en parole et en discours. *Langages*, *3*, 5–12.

Dodane, C., Nunes de Vasconcelos, A., Scarpa, E., & Barkatdefradas, M. (2016). Disfluences dans le langage de l'enfant: Une perspective trans-linguistique (français et portugais brésilien). *Glossa*, *121*, 15–37.

Dollaghan, C. A., & Campbell, T. F. (1992). A procedure for classifying disruptions in spontaneous language samples. *Topics in Language Disorders*.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A Psycholinguistic Perspective. *Studies in Second Language Acquisition*, *20*(3), 349–385.

Drew, P., & Heritage, J. (1992). Analyzing talk at work: An introduction. In P. Drew & J. Heritage (Eds.), *Talk at work: Interaction in institutional settings: Vol. Studies in interactional sociolinguistics*. Cambridge University Press.

Du Bois, J. W. (1985). Competing motivations. In J. Haiman (Ed.), *Iconicity in syntax* (Vol. 6, pp. 343–365). John Benjamins Publishing.

Du Bois, J. W. (2007). The stance triangle. *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, *164*(3), 139–182.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. *Talking Data: Transcription and Coding in Discourse Research*, *45*, 89.

Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, *25*(1), 11–28.

Duez, D. (1991). *La pause dans la parole de l'homme politique*. Editions du Centre national de la recherche scientifique.

Duez, D. (1997). Acoustic markers of political power. *Journal of Psycholinguistic Research*, *26*(6), 641–654.

Duez, D. (2001a). Caractéristiques acoustiques et phonétiques des pauses remplies dans la conversation en français. *Travaux Interdisciplinaires Du Laboratoire Parole et Langage d'Aix-En-Provence (TIPA)*, *20*, 31–48.

Duez, D. (2001b). Signification des hésitations dans la parole spontanée. *Revue Parole*, 17–18.

Dumont, A. (2018). *Fluency and disfluency: A corpus study of non-native and native speaker (dis) fluency profiles* [PhD Thesis]. UCL-Université Catholique de Louvain.

Duranti, A. (2006). Transcripts, like shadows on a wall. *Mind, Culture, and Activity*, *13*(4), 301–310.

Duranti, A. (2011). Linguistic anthropology: Language as a non-neutral medium. *The Cambridge Handbook of Sociolinguistics*, 28–46.

Dwijayanti, I., Budayasa, I. K., & Siswono, T. Y. E. (2019). Students' gestures in understanding algebraic concepts. *Beta: Jurnal Tadris Matematika*, *12*(2), 133–143.

Edlund, J., Hirschberg, J. B., & Heldner, M. (2009). *Pause and gap length in face-to-face interaction*.

Eguchi, M. (2016). Investigating the Relationships Between Vocabulary and Clause-internal Pauses and its Development in L2 Speech. *Proceedings of the Pacific Second Language Research Forum (PacSLRF2016). Hiroshima: Japan Second Language Association*.

Eisenbeiss, S. (2010). Production methods in language acquisition research. *Experimental Methods in Language Acquisition Research*, 11–34.

Eitel, A., & Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, *11*(1), 107–121.

Eklund, R. (2001). Prolongations: A dark horse in the disfluency stable. *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*.

Eklund, R. (2004). *Disfluency in Swedish human–human and human–machine travel booking dialogues* [PhD Thesis]. Linköping University Electronic Press.

Eklund, R., & Shriberg, E. (1998). Crosslinguistic disfluency modelling: A comparative analysis of Swedish and American English human–human and human–machine dialogues. *5th International Conference on Spoken Language Processing, 30th November-4th December, 1998, Sydney, Australia*, *6*, 2627–2630.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*.

Elo, J., & Pörn, M. (2018). Challenges of implementing authenticity of tandem learning in formal language education. *International Journal of Bilingual Education and Bilingualism*.

Enfield, N. J. (2005). The body as a cognitive artifact in kinship representations: Hand gesture diagrams by speakers of Lao. *Current Anthropology*, *46*(1), 51–81.

Eskénazi, M. (1993). Trends in speaking styles research. *Third European Conference on Speech Communication and Technology*.

Esposito, A., & Marinaro, M. (2007). What pauses can tell us about speech and gesture partnership. *Nato security through science series human and societal dynamics*, *18*, 45.

Esposito, A., McCullough, K. E., & Quek, F. (2001). Disfluencies in gesture: Gestural correlates to filled and unfilled speech pauses. *Proceedings of IEEE Workshop on Cues in Communication*, 1–6.

Exare, C. (2017). *Les aspirations intrusives dans l'anglais des apprenants francophones* [Unpublished PhD Thesis]. Université Paris 3 Sorbonne Nouvelle.

Fehringer, 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.

Ferguson, C. A. (1975). Toward a characterization of English foreigner talk. *Anthropological Linguistics*, 1–14.

Ferré, G. (2008). Récits de femmes Analyse multimodale du récit conversationnel en français: Une étude de cas. *Congrès Mondial de Linguistique Française*, 081.

Ferreira, F., & Bailey, K. G. D. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences*, *8*(5), 231–237.

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Origins and Evolution of Language and Speech*, *280*, 20–32.

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, *19*(5), 921–928.

Fischer, K. (2000). *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Markers*. Mouton de Gruyter.

Ford, C. E., & Thompson, S. A. (1996). Intonational, and pragmatic resources for the management of turns. *Interaction and Grammar*, *13*, 134.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Fox, B. A., Thompson, S. A., Ford, C. E., & Couper-Kuhlen, E. (2013). Conversation Analysis and Linguistics. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (p. 726). Blackwell Publishing.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738.

Fox Tree, J. E. & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151–167.

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, *65*(2), 161–175.

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.

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, *15*(3), 219–236.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs.

Gawne, L., & McCulloch, G. (2019). Emoji as digital gestures. *Language@ Internet*, *17*(2).

Geeraerts, D. (2006). A rough guide to cognitive linguistics. In *Cognitive linguistics: Basic readings* (pp. 1–28). De Gruyter Mouton.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. Academic Press.

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.

Ginzburg, J., Fernández, R., & Schlangen, D. (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, *7*, 9–1.

Ginzburg, J., & Poesio, M. (2016). Grammar is a system that characterizes talk in interaction. *Frontiers in Psychology*, *7*, 1938.

Goffman, E. (1967). On face-work, interaction ritual: Essays on face-to-face behavior. *Pantheon, New York*, 5–46.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279.

Goldman, J.-P., Avanzi, M., & Auchlin, A. (2010). Hesitations in read vs. Spontaneous French in a multi-genre corpus. *DiSS-LPSS Joint Workshop 2010*.

Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. *Language and Speech*, *1*(3), 226–231.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*(2), 103–113.

Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers*. Academic Press.

Goodwin, C. (1987). Forgetfulness as an interactive resource. *Social Psychology Quarterly*, 115–130.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10), 1489–1522.

Goodwin, C. (2003). The body in action. In *Discourse, the body, and identity* (pp. 19–42). Springer.

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).

Goodwin, C. (2017). *Co-operative action*. Cambridge University Press.

Goodwin, C., & Goodwin, M. H. (1996). Seeing as a situated activity: Formulating planes. In D. Middleton & Y. Engestrom (Eds.), *Cognition and Communication at Work*. Cambridge University Press.

Goodwin, C., & Goodwin, M. H. (2004). Participation. *A Companion to Linguistic Anthropology*, 222–224.

Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology*, *19*(1), 283–307.

Goodwin, M.H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*(1–2), 51–76.

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.

Green, J., Franquiz, M., & Dixon, C. (1997). The myth of the objective transcript: Transcribing as a situated act. *Tesol Quarterly*, *31*(1), 172–176.

Grosjean, F., & Deschamps, A. (1972). Analyse des variables temporelles du français spontané. *Phonetica*, *26*(3), 129–156.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, *31*(3–4), 144–184.

Grosman, I. (2018). *Évaluation contextuelle de la (dis) fluence en production et perception: Pratiques communicatives et formes prosodico-syntaxiques en français* [Unpublished PhD Thesis]. UCL-Université Catholique de Louvain.

Grosman, I., Simon, A. C., & Degand, L. (2019). Empathetic hearers perceive repetitions as less disfluent, especially in non-broadcast situations. *Proceedings of DiSS, 2019, the 9th Workshop on Disfluency in Spontaneous Speech*.

Guaïtella, I. (1993). Functional, acoustical and perceptual analysis of vocal hesitations in spontaneous speech. *ESCA Workshop on Prosody*.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish* (Vol. 35). Lund University.

Gullberg, M. (2011). Multilingual multimodality: Communicative difficulties and their solutions in second-language use. *Embodied Interaction: Language and Body in the Material World*, 137–151.

Gullberg, M. (2014). Gestures and second language acquisition. In C. Müller, A. Cienki, S. H. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body- Language- Communication: An International Handbook on Multimodsality in Human Interaction*. Mouton de Gruyter.

Gullberg, M., & McCafferty, S. G. (2008). Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition*, *30*(2), 133–146.

Gumperz, J. J., & Berenz, N. (1993). Transcribing conversational exchanges. *Talking Data: Transcription and Coding in Discourse Research*, *91122*.

Gürbüz, N. (2017). Understanding fluency and disfluency in non-native speakers' conversational English. *Kuram ve Uygulamada Egitim Bilimleri*, *17*(6), 1853–1874.

Hai, T. (2017). Hesitation phenomena expressed in native Russian and Chinese and accented Russian* spontaneous speech. *Phonetics Without Borders*, 86–92.

Halliday, M. A. K. (1973). *Explorations in the functions of language*. Arnold.

Harrison, S. (2009). Grammar, gesture, and cognition: The case of negation in English. *Unpublished Doctoral Dissertation, University of Bordeaux*.

Hartsuiker, R. J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*.

Hashemi, M. R., & Babaii, E. (2013). Mixed methods research: Toward new research designs in applied linguistics. *The Modern Language Journal*, *97*(4), 828–852.

Hayashi, M. (2003). Language and the body as resources for collaborative action: A study of word searches in Japanese conversation. *Research on Language and Social Interaction*, *36*(2), 109–141.

Heath, C., & Luff, P. (2011). Gesture and institutional interaction. *Embodied Interaction: Language and Body in the Material World*, 276–288.

Heldner, M., Hjalmarsson, A., & Edlund, J. (2013). Backchannel relevance spaces. *Nordic Prosody XI, Tartu, Estonia, 15-17 August, 2012*, 137–146.

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.

Hepburn, A., & Bolden, G. B. (2013). The conversation analytic approach to transcription. *The Handbook of Conversation Analysis*, *57*, 76.

Heritage, J. (1997). Conversation analysis and institutional talk. *Handbook of Language and Social Interaction*, *103*, 47.

Heritage, J. (2009). Conversation analysis as social theory. *The New Blackwell Companion to Social Theory*, 300–320.

Hieke, A. E. (1981). A Content-Processing View of Hesitation Phenomena. *Language and Speech*, *24*(2), 147–160.

Hilton, H. (2009). Annotation and analyses of temporal aspects of spoken fluency. *Calico Journal*, *26*(3), 644–661.

Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In *Prosody: Theory and experiment* (pp. 335–350). Springer.

Hoetjes, M., & Van Maastricht, L. (2020). Using Gesture to Facilitate L2 Phoneme Acquisition: The Importance of Gesture and Phoneme Complexity. *Frontiers in Psychology*, *11*.

Hoey, E. M. (2014). Sighing in interaction: Somatic, semiotic, and social. *Research on Language and Social Interaction*, *47*(2), 175–200.

Hoey, E. M. (2015). Lapses: How people arrive at, and deal with, discontinuities in talk. *Research on Language and Social Interaction*, *48*(4), 430–453.

Hoey, E. M. (2020). Waiting to inhale: On sniffing in conversation. *Research on Language and Social Interaction*, *53*(1), 118–139.

Holmes, V. M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes*, *3*(4), 323–361.

Holt, B., Tellier, M., & Guichon, N. (2015, September). The use of teaching gestures in an online multimodal environment: The case of incomprehension sequences. *Gesture and Speech in Interaction 4th Edition*. https://hal.archives-ouvertes.fr/hal-01215770

Horgues, C., & Scheuer, S. (2015). Why some things are better done in tandem. In *Investigating English Pronunciation* (pp. 47–82). Springer.

Horgues, C., & Scheuer, S. (2017). Misunderstanding as a two-way street: Communication breakdowns in native/non-native English/French tandem interactions. *International Symposium on Monolingual and Bilingual Speech 2017*, 148.

Hough, J., Tian, Y., De Ruiter, L., Betz, S., Kousidis, S., Schlangen, D., & Ginzburg, J. (2016). Duel: A multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter. *10th Edition of the Language Resources and Evaluation Conference*.

Hunsicker, D., & Goldin-Meadow, S. (2013). How handshape type can distinguish between nouns and verbs in homesign. *Gesture*, *13*(3), 354–376.

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, *81*, 72–90.

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology*, *4*, 255.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11–12), 19–40.

Irishkhanova, O.K., & Cienki, A. (2018) The semiotics of gestures in Cognitive Linguistics: Contribution and challenges. *Voprosy Kognitivnoy Lingvistiki*, *4*, 25-36.

Jakobson, R. (1960). Linguistics and poetics. In *Style in language* (pp. 350–377). MA: MIT Press.

Jefferson, G. (1974). Error correction as an interactional resource. *Language in Society*, 181–199.

Jefferson, G. (1996). A case of transcriptional stereotyping. *Journal of Pragmatics*, *26*(2), 159–170.

Jefferson, G. (2004). Glossary of transcript symbols. *Conversation Analysis: Studies from the First Generation. Amsterdam: John Benjamins*, 13–31.

Jehoul, A. (2019). *A multimodal study of filled pauses. On the interplay of speech and eye gaze*. Leuven University.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133.

Johnson, W. (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. *Journal of Speech & Hearing Disorders. Monograph Supplement*.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Kärkkäinen, E. (2006). Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies*, *26*(6), 699–731.

Kasper, G., & Færch, C. (1983). *Strategies in interlanguage communication*. Longman Publishing Group.

Keevallik, L., & Ogden, R. (2020). Sounds on the Margins of Language at the Heart of Interaction. *Research on Language and Social Interaction*, *53*(1), 1–18.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63.

Kendon, A. (1972). Some relationships between body motion and speech. In A. W. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–210). Pergamon Press.

Kendon, A. (1976). The F-formation system: The spatial organization of social encounters. *Man-Environment Systems*, *6*(01), 1976.

Kendon, A. (1980). A description of a deaf-mute sign language from the Enga Province of Papua New Guinea with some comparative discussion. *Semiotica*, *32*(1/2), 81–117.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters* (Vol. 7). CUP Archive.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, *23*(3), 247–279.

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

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language.' *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 2–13.

Kendon, A. (2017). Languages as semiotically heterogenous systems. *Behavioral and Brain Sciences*, *40*.

Kendon, A., & Versante, L. (2003). Pointing by hand in Neapolitan. *Pointing: Where Language, Culture, and Cognition Meet*, 109–137.

Kita, S. (1993). *Language and thought interface: A study of spontaneous gestures and Japanese mimetics*. University of Chicago.

Kita, S. (2000). How representational gestures help speaking. *Language and Gesture*, *1*, 162–185.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Lawrence Erlbaum.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32.

Kita, S., Van Gijn, I., & Van der Hulst, H. (1997). Movement phases in signs and co-speech gestures, and their transcription by human coders. *International Gesture Workshop*, 23–35.

Kjellmer, G. (2003). Hesitation. In defence of er and erm. *English Studies*, *84*(2), 170–198.

Koester, A. (2010). Building small specialised corpora. *The Routledge Handbook of Corpus Linguistics*, *1*, 66–79.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164.

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). On the distribution of clicks and inbreaths in class presentations and spontaneous conversations: blending vocal and kinetic activities. *Proceedings of Laughter and Other Non-Verbal Vocalisations Workshop*.

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., & Crible, L. (2021). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech*, 00238309211010862.

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.

Kosmala, L., & Morgenstern, A., (2021). Multimodal languaging in tandem collaborative interactions: Focus on embodied inter-(dis)fluencies. *Asia-Pacific Languages for Specific Purposes & Professional Communication Association*. June 4th. 2021. Hong Kong, China.

Kosmala, L., Candea, M., & Morgenstern, A., (2021). The deployment of inter-(dis)fluencies in the course of multimodal communication: a French case study in two interactional settings. *17th International Pragmatics Conference*. June 29th 2021. Winterthur, Switzerland.

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. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*(2), 2–19.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process. *Language and Gesture*, *2*, 261.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, *31*(6), 533–552.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, speech, and sign* (pp. 93–116). Oxford University Press.

Kurhila, S. (2001). Correction in talk between native and non-native speaker. *Journal of Pragmatics*, *33*(7), 1083–1110.

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Ladewig, S. H. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, *Volume 6*.

Ladewig, S. H. (2014). Recurrent gestures. In C. Müller, A. Cienki, S. H. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body- Language- Communication: An International Handbook on Multimodsality in Human Interaction*. Mouton de Gruyter.

Ladewig, S. H., & Bressem, J. (2013). New insights into the medium hand: Discovering recurrent structures in gestures. *Semiotica*, *2013*(197), 203–231.

Landis, J. R., & Koch, G. G. (1977a). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.

Landis, J. R., & Koch, G. G. (1977b). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Langacker, R. W. (1987). *Foundations of Cognitive grammar, vol. 1 theoretical prerequisites*. Stanford University press.

Langacker, R. W. (1995). Cognitive grammar. In *Concise History of the Language Sciences* (pp. 364–368). Elsevier.

Langacker, R. W. (1998). Conceptualization, symbolization, and grammar. In M. Tomasello (Ed.), *The new psychology of language* (pp. 1–40). Lawrence Erlbaum.

Langacker, R. W. (1999). *Grammar and conceptualization*. Mouton de Gruyter.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.

Lapadat, J. C. (2000). Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology*, *3*(3), 203–219.

LeBaron, C. D., Mandelbaum, J., & Glenn, P. J. (2003). An overview of language and social interaction research. *Studies in Language and Social Interaction*, 12–42.

Lelandais, M. (2019). *Expression multimodale de la subordination en anglais* [Unpublished PhD Thesis]. Université de Nantes.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.

Lerner, G. H. (1992). Assisted storytelling: Deploying shared knowledge as a practical matter. *Qualitative Sociology*, *15*(3), 247–271.

Levelt, W. J. (1999). Producing spoken language. *The Neurocognition of Language*, 83–122.

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. (1989). *Speaking. From intention to articulation*. MIT Press.

Levelt, W. J., & Schriefers, H. (1987). Stages of lexical access. In *Natural language generation* (pp. 395–404). Springer.

Levy, E. T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Processes*, *15*(3), 277–301.

Lickley, R. J. (2001). Dialogue moves and disfluency rates. *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*.

Lickley, R. J. (2015). Fluency and Disfluency. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 445–474). John Wiley.

Linell, P. (2009). *Rethinking language, mind, and world dialogically*. IAP.

Lopez-Ozieblo, R. (2019). Cut-offs and co-occurring gestures: Similarities between speakers' first and second languages. *International Review of Applied Linguistics in Language Teaching*, *1*(ahead-of-print).

Lopez-Ozieblo, R. (2020). Proposing a revised functional classification of pragmatic gestures. *Lingua*, *247*, 1–47.

Ma, W., & Winke, P. (2019). Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Language Annals*, *52*(1), 66–86.

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.

MacWhinney, B. (2001). *From CHILDES to TalkBank*.

MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung: Online-Zeitschrift Zur Verbalen Interaktion*, *11*, 154.

Matzinger, T., Ritt, N., & Fitch, W. T. (2020). Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. *PloS One*, *15*(4), e0230710.

Mazeland, H. (2006). Conversation analysis. *Encyclopedia of language and linguistics*, *3*, 153-162.

McCafferty, S. G. (1998). Nonverbal expression and L2 private speech. *Applied Linguistics*, *19*(1), 73–96.

McCafferty, S. G. (2004). Space for cognition: Gesture and second language learning. *International Journal of Applied Linguistics*, *14*(1), 148–165.

McCarthy, M. (2009). Rethinking spoken fluency. *ELIA, 9, 11-29*.

McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. *Human Communication Research*, *8*(4), 299–316.

McLaughlin, S. F., & Cullinan, W. L. (1989). Disfluencies, utterance length, and linguistic complexity in nonstuttering children. *Journal of Fluency Disorders*, *14*(1), 17–36.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Meisel, J. (1987). A note on second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (Vol. 83, pp. 83–90). Ablex Publishing Corporation.

Menn, L., & Dronkers, N. F. (2016). *Psycholinguistics: Introduction and applications*. Plural Publishing.

Merlo, S., & Barbosa, P. A. (2010). Hesitation phenomena: A dynamical perspective. *Cognitive Processing*, *11*(3), 251–261.

Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: Topic familiarity and disfluencies. *Journal of Communication Disorders*, *37*(6), 489–503.

Meteer, 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.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). Effects of task complexity and task condition on Dutch L2. *International Review of Applied Linguistics*, *45*(3), 241–259.

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. (2001). Pour une linguistique interactionnelle. *Marges Linguistiques*, *1*, 142–162.

Mondada, L. (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, *9*(2), 194–225.

Mondada, L. (2013). Interactional space and the study of embodied talk-in-interaction. *Space in Language and Linguistics: Geographical, Interactional and Cognitive Perspectives*, 247–275.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, *20*(3), 336–366.

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.

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, *145*, 47–62.

Mondada, L., & Pekarek Doehler, S. (2004). Second language acquisition as situated practice: Task accomplishment in the French second language classroom. *Canadian Modern Language Review*, *61*(4), 461–490.

Mondada, L., & Traverso, V. (2005). (Dés) alignements en clôture. Une étude interactionnelle de corpus de français parlé en interaction. *Lidil. Revue de Linguistique et de Didactique Des Langues*, *31*, 35–59.

Moniz, H. (2019). Processing disfluencies in distinct speaking styles: Idiosyncrasies and transversality. *The 9th Workshop on Disfluency in Spontaneous Speech*, 1–2.

Moniz, H., Batista, F., Mata, A. I., & Trancoso, I. (2014). Speaking style effects in the production of disfluencies. *Speech Communication*, *65*, 20–35.

Moniz, H. (2013). *Processing disfluencies in European Portuguese* [Unpublished PhD Thesis] Universidade de Lisboa.

Moniz, H., Trancoso, I., & Mata, A. I. (2009). Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts. *Tenth Annual Conference of the International Speech Communication Association*.

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: Who talks to whom about what in family dinners. In Letizia Caronia (ed.) *Language and Social Interaction at Home and in School*. Amsterdam: John Benjamins, 46-85.

Morgenstern, A. (2020). The Other's Voice in the Co-Construction of Self-Reference in the Dialogic Child. *Bakhtiniana: Revista de Estudos Do Discurso*, *16*, 63–87.

Morgenstern, A. (2021). The Other's Voice in the Co-Construction of Self-Reference in the Dialogic Child. *Bakhtiniana: Revista de Estudos Do Discurso*, *16*(1), 63–87.

Morgenstern, A., & Boutet, D. (forth.). *The orchestration of bodies and artifacts in French family dinners*.

Morgenstern, A., & Goldin-Meadow, S. (in press). *Gesture in language: Development across the lifespan*. De Gruyter Mouton.

Morgenstern, A., & Parisse, C. (2007). Codage et interprétation du langage spontané d'enfants de 1 à 3 ans. *Corpus*, *6*, 55–78.

Morgenstern, A., & Parisse, C. (2012). The Paris Corpus. *Journal of French Language Studies*, *22*(Special issue 1), 7–12.

Mori, J., & Hayashi, M. (2006). The achievement of intersubjectivity through embodied completions: A study of interactions between first and second language speakers. *Applied Linguistics*, *27*(2), 195–219.

Morita, E., & Takagi, T. (2018). Marking "commitment to undertaking of the task at hand": Initiating responses with eeto in Japanese conversation. *Journal of Pragmatics*, *124*, 31–49.

Moro, L., Mortimer, E. F., & Tiberghien, A. (2020). The use of social semiotic multimodality and joint action theory to describe teaching practices: Two cases studies with experienced teachers. *Classroom Discourse*, *11*(3), 229–251.

Müller, C. (1998). *Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich* (Vol. 1). Spitz.

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. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, *16*(2), 277–304.

Müller, C., Bressem, 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* [Unpublished PhD thesis]. University of Edinburgh.

Notarrigo, I. (2017). *Marqueurs de (dis) fluence en langue des signes de Belgique francophone* [Unpublished PhD Thesis]. Université de Namur.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43–72). Academic Press.

Ochs, E. (1996). *Linguistic resources for socializing humanity*. Cambridge University Press.

Ochs, E., & Schieffelin, B. B. (2011). The theory of language socialization. *The Handbook of Language Socialization*, *71*(1), 1–11.

O'Connell, D. C., & Kowal, S. (2005). Uh and Um Revisited: Are They Interjections for Signaling Delay? *Journal of Psycholinguistic Research*, *34*(6), 555–576.

Ogden, R. (2013). Clicks and percussives in English conversation. *Journal of the International Phonetic Association*, *43*(3), 299–320.

Ogden, R. (2020). Audibly not saying something with clicks. *Research on Language and Social Interaction*, *53*(1), 66–89.

Ogden, R. (2018). The actions of peripheral linguistic objects: Clicks. *Proceedings of Laughter Workshop 2018*, 2–5.

O'Shaughnessy, D. (1992). *Analysis of False Starts in Spontaneous Speech*. The International Conference on Spoken Language Processing, Banff, Alberta, Canada.

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*.

Papanas, N., Maltezos, E., & Lazarides, M. K. (2011). Delivering a powerful oral presentation: All the world's a stage. *International Angiology: A Journal of the International Union of Angiology*, *30*(2), 185–191.

Parisse, C., & Le Normand, M.-T. (2007). Une méthode pour évaluer la production du langage spontané chez l'enfant de 2 à 4 ans. *Glossa*, *97*, 10–30.

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.

Pekarek Doehler, S., & Pochon-Berger, E. (2011). Developing "methods" for interaction: A cross-sectional study of disagreement sequences in French L2. *L2 Interactional Competence and Development*, *56*, 206.

Peltonen, P. (2017). L2 fluency in spoken interaction: A case study on the use of other-repetitions and collaborative completions. *AFinLA-e: Soveltavan Kielitieteen Tutkimuksia*, *10*, 118–138.

Peltonen, P. (2019). Gestures as Fluency-enhancing Resources in L2 Interaction: A Case Study on Multimodal Fluency. *Fluency in L2 Learning and Use*, *138*, 111.

Peltonen, P. (2020). *Individual and Interactional Speech Fluency in L2 English from a Problem-solving Perspective: A Mixed-methods Approach* [Unpublished PhD Thesis]. University of Turku.

Peters, A. M. (2001). Filler syllables: What is their status in emerging grammar? *Journal of Child Language*, *28*(1), 229.

Pinto, D., & Vigil, D. (2019). Searches and clicks in Peninsular Spanish. *Pragmatics*, *29*(1), 83–106.

Poggi, I. (2001). From a typology of gestures to a procedure for gesture production. *International Gesture Workshop*, 158–168.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shaped. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 57–108). Cambridge University Press.

Pomerantz, A., & Fehr, B. J. (2011). Conversation analysis: An approach to the analysis of social interaction. *Discourse Studies: A Multidisciplinary Introduction*, *2*, 165–190.

Rajput, M. S. (2016). The source, meanings and use of "mudra" across religions. *Int. J. Interdiscip. Res. Arts Humanit*, *1*, 37–42.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, *28*(2007), 41–66.

Ratner, N. B., Rooney, B., & MacWhinney, B. (1996). Analysis of stuttering using CHILDES and CLAN. *Clinical Linguistics & Phonetics*, *10*(3), 169–187.

Reber, E., & Couper-Kuhlen, E. (2020). On "Whistle" Sound Objects in English Everyday Conversation. *Research on Language and Social Interaction*, *53*(1), 164–187.

Reed, B. S. (2009). Units of interaction:"Intonation phrases" or "turn constructional phrases." *Actes/Proceedings from IDP (Interface Discours & Prosodie)*, 351–363.

Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97–105). Ablex Publishing Corporation.

Rendle-Short, J. (2005). Managing the transitions between talk and silence in the academic monologue. *Research on Language and Social Interaction*, *38*(2), 179–218.

Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition*, *23*(04), 497–526.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441.

Rodríguez, L. J., Torres, I., & Varona, A. (2001). Annotation and analysis of disfluencies in a spontaneous speech corpus in Spanish. *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*.

Rohr, J. L. (2016). *Acoustic and Perceptual Correlates of L2 Fluency: The Role of Prolongations* [PhD Thesis].

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S., & Prieto, P. (2020). The MultiModal MultiDimensional (M3D) labelling scheme for the annotation of audiovisual corpora. *Gesture and Speech in Interaction (GESPIN)*.

Rose, R. L. (1998). *The communicative value of filled pauses in spontaneous speech* [M.A Diss.]. University of Birminghman.

Rossano, F. (2013). Gaze in Conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of Conversation Analysis*. Blackwell Publishing.

Rydell, M. (2019). Negotiating co-participation: Embodied word searching sequences in paired L2 speaking tests. *Journal of Pragmatics*, *149*, 60–77.

Sacks, H. (1992). *Lectures on Conversation* (Vol. 1–2). Basil Blackwell.

Sacks, H., Jefferson, G., & Schegloff, E. A. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/10.1016/B978-0-12-623550-0.50008-2

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. (1982). Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences. *Analyzing Discourse: Text and Talk*, *71*, 93.

Schegloff, E. A. (1991). Conversation analysis and socially shared cognition. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Socially Shared Cognition*. American Psychological Association.

Schegloff, E. A. (1996a). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, *102*(1), 161–216.

Schegloff, E. A. (1996b). Turn organization: One intersection of grammar and interaction. *Studies in Interactional Sociolinguistics*, *13*, 52–133.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, *29*(1), 1–63.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I* (Vol. 1). Cambridge university press.

Schegloff, E. A. (2010). Some other "uh(m)" s. *Discourse Processes*, *47*(2), 130–174.

Schegloff, E. A., Sacks, H., & Jefferson, G. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schettino, L., Di Maro, M., & Cutugno, F. (2020). Silent pauses as clarification trigger. *Laughter and Other Non-Verbal Vocalisations Workshop: Proceedings (2020)*.

Scheuer, S., & Horgues, C. (2020). Potential pitfalls of interpreting data from English-French tandem conversations. *Interpreting Languagelearning Data*, 197.

Schieffelin, B. B., & Ochs, E. (1986). Language socialization. *Annual Review of Anthropology*, *15*(1), 163–191.

Schiffrin, D. (1994). *Approaches to Discourse*. Blackwell Publishers.

Schleif, C. (1993). Hands that appoint, anoint and ally: Late medieval donor strategies for appropriating approbation through painting. *Art History*, *16*, 1–1.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *28*(28).

Schneider, U. (2014). *Frequency, Hesitations and Chunks. A Usage-based Study of Chunking in English*. Albert-Ludwigs-Universität.

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.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Günthner, S., Hartung, M., Kern, F., Mertzlufft, C., Meyer, C., Morek, M., Oberzaucher, F., Peters, J., Quasthoff, U., Schütte, W., … Uhmann, S. (2009). Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforschung : Onlinezeitschrift zur verbalen Interaktion*, *10*, 353–402.

Selting, M., Barth-Weingarten, D., Reber, E., & Selting, M. (2010). Prosody in interaction. *Prosody in Interaction, Amsterdam/Philadelphia, John Benjamins*, 3–40.

Seo, M.-S., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, *42*(8), 2219–2239.

Seyfeddinipur, M., (2006) *Disfluency: Interrupting speech and gesture* [Unpublished PhD Thesis]. Radboud University.

Seyfeddinipur, M., & Kita, S. (2001). Gesture as an indicator of early error detection in self-monitoring of speech. *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*.

Shames, G. H., & Sherrick, C. E. (1963). A discussion of nonfluency and stuttering as operant behavior. *Journal of Speech and Hearing Disorders*, *28*(1), 3–18.

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. (1995). Acoustic properties of disfluent repetitions. *Proceedings of the International Congress of Phonetic Sciences*, *4*, 384–387.

Shriberg, E. (1996). Disfluencies in switchboard. *Proceedings of International Conference on Spoken Language Processing*, *96*(1), 11–14.

Shriberg, E. E. (1994). *Preliminaries to a Theory of Speech Disfluencies* [Unpublished PhD Thesis]. University of California.

Shriberg, E. E. (1999). *Phonetic consequences of speech disfluency*. Sri International.

Sidnell, J. (2016). *Conversation Analysis*. Oxford Research Encyclopedia of Linguistics.

Sikveland, R. O., & Ogden, R. (2012). Holding gestures across turns: Moments to generate shared understanding. *Gesture*, *12*(2), 166–199.

Silverman, K., Blaauw, E., Spitz, J., & Pitrelli, J. F. (1992). A prosodic comparison of spontaneous speech and read speech. *Second International Conference on Spoken Language Processing*.

Simpson, R., Eisenchlas, S., & Haugh, M. (2013). The functions of self-initiated self-repair in the second language Chinese classroom 1. *International Journal of Applied Linguistics*, *23*(2), 144–165.

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.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14.

Slobin, D. I. (1987). Thinking for speaking. *Annual Meeting of the Berkeley Linguistics Society*, *13*, 435–445.

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. https://doi.org/10.1006/jmla.1993.1002

Smotrova, T., & Lantolf, J. P. (2013). The function of gesture in lexically focused L2 instructional conversations. *The Modern Language Journal*, *97*(2), 397–416.

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. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *International Review of Applied Linguistics*, 143.

Stam, G. (2008). What gestures reveal about second language acquisition. *Gesture: Second Language Acquistion and Classroom Research*, *231*.

Stam, G. (2018). Gesture and speaking a second language. *Speaking in a Second Language*, *49*, 67.

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. (2001). Negotiating who presents the problem: Next speaker selection in pediatric encounters. *Journal of Communication*, *51*(2), 252–282.

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41*(1), 31–57.

Stivers, T. (2015). Coding social interaction: A heretical approach in conversation analysis? *Research on Language and Social Interaction*, *48*(1), 1–19.

Stivers, T., & Robinson, J. D. (2006). A preference for progressivity in interaction. *Language in Society*, *35*(3), 367–392.

Stivers, T., & Sidnell, J. (2005). Introduction: Multimodal interaction. *Semiotica*, *2005*, 1–20. https://doi.org/10.1515/semi.2005.2005.156.1

Streeck, J. (2008a). Depicting by gesture. *Gesture*, *8*(3), 285–301.

Streeck, J. (2008b). Gesture in political communication: A case study of the democratic presidential candidates during the 2004 primary campaign. *Research on Language and Social Interaction*, *41*(2), 154–186.

Streeck, J. (2008c). Metaphor and gesture. A view from the microanalysis of interaction. In C. Müller & A. Cienki (Eds.), *Metaphor and gesture* (pp. 259–264). John Benjamins Publishing.

Streeck, J. (2009a). Forward-gesturing. *Discourse Processes*, *46*(2–3), 161–179.

Streeck, J. (2009b). *Gesturecraft: The manu-facture of meaning* (Vol. 2). John Benjamins Publishing.

Streeck, J. (2010). Ecologies of gesture. *New Adventures in Language and Interaction*, 223–242.

Streeck, J. (2014). Mutual gaze and recognition. In M. Seyfeddinipur & M. Gullberg (Eds.), *From Gesture in Conversation to Visible Action as Utterance. Benjamins, Amsterdam* (pp. 35–55). Benjamins.

Streeck, J. (2015). Embodiment in human communication. *Annual Review of Anthropology*, *44*, 419–438.

Streeck, J. (2020). Self-Touch as Sociality. *Social Interaction. Video-Based Studies of Human Sociality*, *3*(2).

Streeck, J. (2021). The emancipation of gestures. *Interactional Linguistics*, *1*(1).

Streeck, J., Goodwin, C., & LeBaron, C. (2011). *Embodied interaction: Language and body in the material world*. Cambridge University Press.

Streeck, J., & Hartge, U. (1992). Gestures at the transition place. In P. Auer & A. Di Luzio (Eds.), *The contextualization of language* (pp. 135–157). John Benjamins Publishing.

Swain, M. (2006). Languaging, Agency and Collaboration in Advanced Second Language Proficiency. *Advanced Language Learning: The Contribution of Halliday and Vygotsky*, 95–108.

Swain, M., & Watanabe, Y. (2012). Languaging: Collaborative dialogue as a source of second language learning. *The Encyclopedia of Applied Linguistics*, 1–8.

Sweetser, E. (2007). Looking at space to study mental spaces. *Methods in Cognitive Linguistics*, *18*, 201–224.

Sweetser, E., & Sizemore, M. (2008). Personal and interpersonal gesture spaces: Functional contrasts in language and gesture. *Language in the Context of Use: Discourse and Cognitive Approaches to Language*, 25–51.

Sweetser, E., & Stec, K. (2016). Maintaining multiple viewpoints with gaze. *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities*, *237*, 257.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*(4), 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9

Szymanski, M. H. (1999). Re-engaging and dis-engaging talk in activity. *Language in Society*, *28*(1), 1–23.

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. https://doi.org/10.1093/elt/ccq020

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In *Planning and task performance in a second language* (pp. 239–273). John Benjamins.

Tellier, M. (2006). *L'impact du geste pédagogique sur l'enseignement/apprentissage des langues étrangères: Etude sur des enfants de 5 ans* [PhD Thesis, Université Paris-Diderot - Paris VII]. https://tel.archives-ouvertes.fr/tel-00371041

Tellier, M. (2008a). Dire avec des gestes. *Le Français Dans Le Monde. Recherches et Applications*, *44*, 40–50.

Tellier, M. (2008b). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235.

Tellier, M., Stam, G., & Bigi, B. (2013). Gesturing while pausing in conversation: Self-oriented or partner-oriented?". *The Combined Meeting of the 10th International Gesture Workshop and the 3rd Gesture and Speech in Interaction Conference, Tillburg (The Netherlands)*.

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.

Ten Have, P. (2007). *Doing conversation analysis*. Sage.

Thompson, P. (2010). Building a specialised audio-visual corpus. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 121–132). Routledge.

Thornbury, S., & Slade, D. (2006). *Conversation: From description to pedagogy*. Cambridge University Press.

Tomasello, M. (1995). Joint attention as social cognition. *Joint Attention: Its Origins and Role in Development*, *103130*, 103–130.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard university press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691.

Torreira, F. J., Bögels, S., & Levinson, S. C. (2016). *Breathing for answering. The time course of response planning in conversation*.

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. R

Tottie, G. (2015). Uh and um in British and American English: Are they words? Evidence from co-occurrence with pauses. In N. Dion, A. Lapierre, & R. T. Cacoullos (Eds.), *Linguistic Variation: Confronting Fact and Theory* (pp. 38–55). NY: Routledge.

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.

Tottie, G. (2019). From pause to word: Uh, um and er in written American English. *English Language & Linguistics*, *23*(1), 105–130. https://doi.org/10.1017/S1360674317000314

Trouvain, J., Möbius, B., & Werner, R. (2019). On Acousitc Features of Inhalation Noises in Read and Spontaneous Speech. *1st International Seminar on the Foundations of Speech: BREATHING, PAUSING, AND VOICE*.

Vasilescu, I., & Adda-Decker, M. (2007). A cross-language study of acoustic and prosodic characteristics of vocalic hesitations. *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, *18*, 140.

Vaughan, E., & Clancy, B. (2013). Small corpora and pragmatics. In *Yearbook of Corpus Linguistics and Pragmatics 2013* (pp. 53–73). Springer.

Vygotsky, L. S. (1934). *Thought and language*. The M.I.T. Press (1985).

Wallbott, H. G. (1995). Congruence, contagion, and motor mimicry: Mutualities in nonverbal exchange. *Mutualities in Dialogue*, 82–98.

Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, *14*(1), 129–182.

Watanabe, M., & Rose, R. (2012). Pausology and hesitation phenomena in second language acquisition. *The Routledge Encyclopedia of Second Language Acquisition*, 480–483.

Witton-Davies, G. (2014). *The study of fluency and its development in monologue and dialogue* [PhD Thesis]. Lancaster University.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review*, *57*(4), 573–589.

Wright, M. (2005). *Studies of the phonetics-interaction interface: Clicks and interactional structures in English conversation* [PhD Thesis]. University of York.

Wright, M. (2011). On clicks in English talk-in-interaction. *Journal of the International Phonetic Association*, *41*(2), 207–229.

Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. *Bilingual Cognition and Language: The State of the Science across Its Subfields*, *54*, 37.

Yasinnik, Y., Shattuck-Hufnagel, S., & Veilleux, N. (2005). Gesture marking of disfluencies in spontaneous speech. *Disfluency in Spontaneous Speech*.

Yule, G. (1996). *Pragmatics*. Oxford University Press.

Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, *24*(5), 666–689.

Zuniga, M., & Simard, D. (2019). Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguistic Research*, *48*(1), 43–59.

Appendices

Appendix 1

Symbol	Explanation	Example	Section
Region-delimiting			
[ ]	onset RM, offset RR	(see all examples below)	4.3.4.1.1
.	IP	(see all examples below)	4.3.4.1.2
Syntactic-word			
r	repeated word	she she liked it [r . r]	4.3.4.2.1
s	word in substituted string	she my wife liked it [s . s s]	4.3.4.2.2
i	inserted word	she liked really liked it [r . i r]	4.3.4.2.3
d	deleted word	it was very she liked it [d d d .]	4.3.4.2.4
Extra-syntactic-word			
f	filled pause	she uh liked it / she uh he liked it [f] [s . f s]	4.3.4.3.1
e	explicit editing term	she sorry he liked it [s . e s]	4.3.4.3.2
p	discourse marker	she liked well she liked it [r r . p r r]	4.3.4.3.3
Inter-sentence-word			
c	coordinating conjunction	she saw it and and she liked it [c . c]	4.3.4.4.1
Diacritics			
-	word fragment	she li- he liked it [s r- . s r]	4.3.4.5.1
~	misarticulated word	shle she liked it [r~ . r]	4.3.4.5.2
^	contracted word	she'd she'll like it [r^s . r^s]	4.3.4.5.3
=	substituted-string fragment	she thought highly she liked it [r s s= . r s]	4.3.4.5.4

Figure 72. *Shriberg's (1994, p. 57) annotation model of disfluencies*

Table 46. *Summary of Crible (2017)'s discourse markers annotation tiers*

Tag	Tier definition	Values
DM	Full-word orthographic transcription of the DM	"and" "but" "or" "donc" "you know" "tu vois" "in fact" "for example" etc.
POS	Source grammatical class of the DM (part-of-speech)	9 values: -Coordinating conjunction (and, but, or) - Adverb (so, actually, now, anyway) - Verbal phrase (you know, I mean) -Pronoun ("quoi" "et tout") -Noun phrase (sort of, "genre") -Adjective ("right" "bon") -Prepositional phrase ("in fact", "for example" "par contre") -Interjection "okay" "yeah"
TYPE DM	Position of the DM on the scale of relationality (more generic filter into the functions of DMs, binary variable)	"non-relational" (no linking function but rather intersubjective purposes such as monitoring) "relational" (grammatical items in the traditional sense of the term, eg "conjunctive" (see Degand & Simon-Vandenbergen)
DOMAIN 1	Functional domain of the DM	"ideational domain" "rhetorical domain" "sequential domain" "interpersonal domain"
FUNCTION 1	Specifies the function of the domain	"cause" "consequence" (ideational domain) "motivation" "conclusion" "relevance" (rhetorical) "topic-shifting" "closing boundary" (sequential) "monitoring" "face-saving" (interpersonal)
DOMAIN 2	Possible second domain of the DM	"ideational domain" "rhetorical domain" etc.
FUNCTION 2	Possible second function of the DM	"quoting" "agreeing" "exception" etc.
POSITION macro	Macro-syntactic position of the DM (dependency structure with all its constituents) Tesnières 1959, Auer 1996, Lindstrom 2001 → strictly linear approach (no functional considerations)	"pre-field" "left-integrated" "middle field" "right-integrated" "post-field" (see macro-syntactic segmentation) "independent" "interrupted"
POSITION micro	Micro-syntactic position of the DM (minimal clause the DM belongs to)	"initial" "medial" "final" "independent" "interrupted"
POSITION turn	Position of the DM in the turn of speech (Bolly et al. 2015)	"turn-initial" "turn-final" "turn-medial" (any other position) "independent turn"
CO-OCC	Whether the DM co-occurs with another & where	"Yleft" (co-occurrence at the left of the DM) "Yright xxx" (co-occurrence at the right where "xxx" stands for the sequence of DMs in context) "Ylr" (co-occurrence at both left & right) "NO" no co-occurrence

Table 47. *Crible's (2017)' annotation of fluencemes (also used in Crible et al., 2019)*

Simple fluencemes (Crible, 2017, p. 107):
1. Unfilled pauses (UP): defined by an interruption of the sound signal lasting more than 200 milliseconds, following Candea (2000); threshold is fixed & does not take account of speaking rate or speaking style variation, due to the very limited potential of the corpus for prosodic analysis
2. Filled pauses (FP): vocalizations characterized by their conventional & neutral phonetic form ("euh" in French) & their function as supporting or maintaining on-going speech (Clark & Fox Tree, 2002)
3. Discourse markers (DM): definition: "grammatically heterogeneous, syntactically optional, multifunctional type of pragmatic markers. Their specificity is to function on a metadiscursive level as procedural cues to constrain the interpretation of the host unit and its context, expliciting the structural sequencing of discourse segments, expressing the speaker's meta-comment on their phrasing, or contributing to the speaker-hearer relationship
4. Explicit editing terms (ET) cover any lexical expression by which the speaker signals some production trouble & which are not identified as DMs or filled pauses; explicit references to lexical access trouble.
5. False-starts (FS) interruptions that leave a segment syntactically and/or semantically incomplete and where no elements from the previous abandoned context are taken up in what follows (Pallaud et al. 2013a)
6. Truncations (TR) interruptions that only apply to words & not segments (as in false starts) if the fragments are repeated and/or completed, the truncation becomes a compound fluenceme

Compound fluencemes (Crible, 2017, p. 108)
→ function with a structure in at least two parts, namely the reparandum & the reparans
7. Identical repetitions (RI): words formally similar to each other & contiguous whether intentionally (because of an overlap) or not; semantic repetitions excluded.
→ repeated in their exact same form and without any semantic addition (Candea 2000)
8. Modified repetitions (RM): words belonging to a segment that is partially repeated but with a change in content, either by a substitution, a truncation, a deletion, or a lexical insertion
9. Morphosyntactic substitutions (SM): any morphological modification in a complete lemma (excluding truncations); can be an addition or deletion of a morpheme
10. Propositional substitutions (SP): any segment replaced by another one which introduces a semantic nuance

Table 48. *Patterns associated with dispreferred responses in English (Yule, 1996, p. 81)*

How to do a dispreferred	**Examples**
Delay/hesitate	Pause, er, em, ah
Preface	Well, oh
Express doubt	I'm not sure/ I don't know
Token Yes	That's great; I'd love to
Apology	I'm sorry; what a pity
Mention obligation	I must do X; I'm expected in Y
Appeal for understanding	You see; you know
Make it non-personal	Everybody else, out there
Give an account	Too much work; no time left
Use mitigators	Really; mostly; sort of; kinda
Hedge the negative	I guess not; not possible

Table 49. *Functional classification of gestures adapted from Müller (1998) in Cienki (2004, p. 439)*

Referential Gestures		Performative Gestures	Discursive Gestures
Concrete reference	Abstract reference	Abstract reference	Abstract reference
– Objects (e.g., a picture frame) – Properties (e.g., the straight edge of a ruler) – Behaviors and Actions (e.g., waving something away) – Events (e.g., the flowing of water) – Relative location (e.g., the fact that a book lies at the bottom of the pile)	– Entities (e.g., the framework of a theory) – Properties (e.g., a "straight" answer) – Behaviors and Actions (e.g., waving away an offer) – Events (e.g., the flow of billions of dollars) – Relative location and relative time (e.g., the obviousness of an argument [it lies on the palm of your hand]; the past is behind us)	– Actions e.g., considering carefully, warding off, dismissing, turning down (requesting, swearing, blessing) [appeal function] (mourning gestures, hand clapping, fist raised in the air) [expressive function]	– Emphasis (e.g., through beats) – Presenting an idea or argument – Linking (by resuming any gesture) – Structuring (e.g., with counting gestures)

Appendix 2

Attestation d'utilisation des données du corpus SITAF
(*Spécificités des Interactions orales en Tandem Anglais Français*)
Projet Innovant Sorbonne Nouvelle-Paris 3, 2012-2014
(Responsables : Céline Horgues et Sylwia Scheuer)

Je soussigné(e) : ……atteste disposer des séquences audio/vidéo du corpus SITAF et /ou des transcriptions qui m'ont été confiées par les responsables du corpus.

J'atteste avoir pris connaissances de la Licence d'utilisation dans laquelle s'inscrit ce corpus et m'engage à en respecter les modalités d'utilisation.
Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 non transposé)
https://creativecommons.org/licenses/by-nc-nd/3.0/deed.fr

J'atteste avoir pris connaissance du Formulaire d'autorisation/Consent form signé par les participant.e.s du corpus SITAF. Je m'engage à respecter l'intégrité des participant.e.s et la confidentialité des données.

Je m'engage à ne pas partager les données du corpus SITAF avec une tierce personne sans l'autorisation des responsables du corpus.

Fait à ……., le ……

Signature :

***Figure 73**. Consent form used for sharing the SITAF Corpus*

UNIVERSITÉ SORBONNE NOUVELLE PARIS 3
Membre fondateur de Sorbonne Paris Cité

NOM : **Prénom :**

Formulaire d'autorisation
Accord pour l'utilisation de données vidéo et audio, Droit à l'image
*P*rojet Innovant Jeunes Chercheurs Paris 3, *SITAF*
S*pécificités des* ***I****nteractions Verbales dans le cadre de* ***T****andems* ***A****nglais-****F****rançais*

Présentation du projet et des conditions de votre participation :

Le projet *SITAF* est piloté par Céline Horgues (celine.horgues@univ-paris3.fr et Sylwia Scheuer (sylwia.scheuer@univ-paris3.fr) et comprend 10 membres enseignants-chercheurs et doctorants en Sciences du Langage et Didactique de l'Université Sorbonne-Nouvelle Paris 3 et d'autres Universités (Paris St Denis-Paris8, Université de Nantes).

Le but de votre participation est de contribuer au recueil de données qui serviront à l'analyse des spécificités des échanges dans le cadre de tandems anglais/français et des atouts que représente ce mode d'apprentissage.

Pour ceci, nous procédons à des enregistrements audio et vidéo de deux rencontres tandems autour de tâches de communication ludiques à réaliser avec votre binôme.

Dans un premier temps, les données audio et vidéo seront analysées par les chercheurs de l'équipe du projet innovant. Votre nom, prénom et autres informations personnelles fournies à l'inscription resteront toujours confidentielles. A ce dessein, aucun nom et prénom réel ne sera mentionné et tous les participants seront désignés sous une appellation générique du type : *Locuteur Francophone # 1.*

Des extraits audio et vidéo des enregistrements et leur transcription pourront être utilisés lors de conférences académiques ou dans des articles de recherche publiés. Les membres du projet s'engagent à ne pas diffuser d'extraits compromettant les personnes filmées. Les membres du projet s'engagent à masquer l'identité des participants (visage) dans les publications écrites.

D'autres chercheurs en sciences du langage et didactique pourront, lors des projets futurs, consulter les données qui seront mises à leur disposition via une plate-forme de mutualisation des ressources en ligne (du type projets sur l'acquisition du langage : *PhonBank, ou Childes, http://childes.psy.cmu.ed/*)

Les données ne seront en aucun cas utilisées à des fins commerciales.

En tant que participant, vous pouvez demander à visionner les enregistrements à tout moment et vous êtes libre de retirer votre autorisation pour une partie ou la totalité d'un enregistrement si cela ne vous convient pas. Vous pourrez par ailleurs avoir accès à toute publication éventuelle si vous en faites la demande.

Votre participation à ce projet est libre et volontaire et elle n'est pas rémunérée.

Les membres du projet vous remercient de votre participation et ils vous assurent qu'ils feront en sorte que les données ne soient utilisées qu'à des fins scientifiques, avec un souci d'éthique et de préservation de l'intégrité et de la vie privée des participants.

Autorisation :

1) J'ai pris connaissance du descriptif du projet ci-dessus et ai obtenu des réponses à toutes mes questions. Dans les conditions exposées ci-dessus, j'accepte de donner mon autorisation pour l'enregistrement, le stockage, la transcription, l'analyse et la présentation (conférences et publications scientifiques) des données audio et vidéo de mes interactions tandems par les membres du *projet S.I.TA.F.*

Oui ☐ Non ☐

2) J'accepte que mes données audio et vidéo soient intégrées à une plateforme de ressources en ligne afin de permettre leur exploitation scientifique par d'autres chercheurs du domaine des sciences du langage et de la didactique.

O ui ☐ Non ☐

Signature du participant, précédée de la mention manuscrite « Lu et approuvé »

Date : **(**mention « lu et approuvé**) :** **S ignature :**

Figure 74*. Consent form used for the collection of the SITAF Corpus*

CONSENTEMENT ÉCLAIRÉ

Je soussigné(e),

Madame, Monsieur……..……..

Né(e) le……., certifiant être majeur(e) et pouvoir donner librement mon consentement à la présente autorisation,

Donne mon accord pour être filmé(e) en classe par Loulou Kosmala dans le cadre de **son projet de recherche,** et autorise l'utilisation de ces données, sous leur forme enregistrée aussi bien que sous leur forme transcrite et anonyme.

Je comprends que ces données seront utilisées seulement dans l**e cadre d'une recherche scientifique à but non lucratif** (thèses, articles scientifiques, exposés à des congrés, séminaires), à des fins d'enseignement universitaire, et pour une diffusion dans la communauté des chercheurs sous la forme d'éventuels échanges et prêts de corpus.

Je prends acte que, pour toutes ces utilisations scientifiques, les données ainsi enregistrées seront anonymes.

J'ai également la possibilité d'obtenir des informations supplémentaires concernant cette étude auprès de l'investigatrice, et de visionner les enregistrements si je le souhaite.

J'autorise l'investigatrice à diffuser mon image au sein de la communauté des chercheurs (thèses, articles scientifiques, séminaires, exposés à des congrès), dans le cas contraire mon visage sera flouté afin de préserver mon anonymat.

Oui ☐ Non ☐

Fait en deux exemplaires pour servir et valoir ce que de droit.

À……, **le**……

Signature du participant

***Figure 75**. Consent form used for the collection of the DisReg Corpus*

Table 50. *Early functional classification system (also found in Kosmala, 2021)*

Category	Description
Speech Management (Default)	Basic feature of (Dis)S. Related to speech processes. When speakers work on their production (planning, repair, lexical search). (See Allwood et al. 1990; choice-related & change-related SM functions). No clear pragmatic intent, seems only to relate to production processes. Basic category which applies to all (dis)fluencies.
ADDITIONAL PRAGMATIC FUNCTIONAL CONTEXTS (PRAGMATIC DIMENSION)	
Discursive Structuring (Swerts, 1998)	Discursive contexts in which (Dis)S are used to structure, mark, punctuate, emphasize parts of speech, introduce a new topic, similar to discourse markers (Crible et al. 2017). ➔ Marking semantic & syntactic coherence Criteria: Co-occur with discourse markers which serve ideational and sequential functions (e.g. *and, but, or* see Crible 2017)
Interactive/ Communicative (Kjellmer, 2003)	Interactional/Communicative contexts in which (Dis)S contribute to the interaction (understood in broad terms: covering sequence organization turn-taking mechanisms, stance, affect) by indexing a stance, reacting to what the other one is saying, turning to the interlocutor, hold/yield a turn etc. Criteria: Occur within dialogic sequences (question/answer; when a speaker initiates a turn, expresses agreement/disagreement to a previous assessment etc.) Co-occur with interpersonal discourse markers (*you know, you know what I mean*), interactional gestures, and/or a gaze towards the interlocutor. E.g.: a lexical search oriented towards the partner is considered communicative
Uncertainty (Smith & Clark 1993)	Contexts in which the speakers overtly display/signal their uncertainty verbally (*I'm not sure, I don't know*) or non-verbally (frown, thinking face) in order to save face (Goffman, 1967, 1971)

Appendix 3

Table 51. *Rate of individual fluencemes (raw and relative frequency) for the SITAF Corpus*

PAR	L1 phw (raw values)		L2 phw (raw values)	
A02	24.5	(43)	61.7	(118)
A03	19.9	(97)	45	(188)
A07	8.9	(28)	32.6	(90)
A09	9.5	(39)	56.8	(110)
A10	30.6	(28)	41.7	(28)
A11	7.9	(34)	32.2	(75)
A13	19.3	(67)	52.6	(122)
A15	14.7	(38)	41.7	(73)
A16	20.5	(53)	25.5	(37)
A17	19.5	(41)	81.6	(275)
A18	22.3	(48)	48.9	(70)
Total	**16.1**	**(516)**	**49.2**	**(1186)**
F02	19.4	(38)	24.8	(30)
F03	17.3	(43)	16.5	(34)
F07	25.6	(127)	23.3	(93)
F09	34.5	(205)	58.7	(181)
F10	35.2	(43)	68.2	(84)
F11	22	(38)	31	(133)
F13	5.9	(19)	28.8	(62)
F15	24.3	(63)	39.4	(93)
F16	6.7	(8)	20.9	(41)
F17	22.5	(51)	26.9	(41)
F18	20.5	(22)	30.4	(21)
Total	**23**	**(657)**	**33.1**	**(813)**

Table 52. *Rate of fluencemes per hundred words (American group, SITAF)*

	L1 (raw)		L2 (raw)		LL score	p value
morpho-syntactic markers						
lexical repair	0.2	9	0.3	9	0.36	< 0.05
morphological repair	0.4	13	1.04	25	8.03	< 0.01
syntactic repair	0.4	14	1.9	48	30.61	< 0.0001
identical repetition	1.8	60	7.5	181	102.80	< 0.0001
self-interruption	0.7	24	0.7	17	0.04	< 0.05
truncated word	0.9	29	2.03	49	12.43	< 0.001
vocal markers						
filled pause	1.53	49	8.01	193	137.3	< 0.0001
prolongation	2.25	72	11.7	283	201.4	< 0.0001
unfilled pause	6.03	193	10.7	260	38.07	< 0.001
peripheral markers						
NL sound	1.50	48	4.73	114	49.68	< 0.001

Table 53. *Rate of fluencemes per hundred words (French group, SITAF)*

	L1 % (raw)		L2 % (raw)		LL score	p value
morpho-syntactic markers						
lexical repair	0.3	11	0.2	5	1.49	< 0.05
morphological repair	0.3	11	1.02	25	7.92	< 0.01
syntactic repair	0.9	27	0.8	20	0.25	< 0.05
identical repetition	3.08	88	3.8	95	2.38	< 0.05
self-interruption	0.84	24	0.6	15	0.95	< 0.05
truncated word	1.02	29	1.4	36	2.19	< 0.05
vocal markers						
filled pause	4.94	141	7.41	182	13.29	< 0.001
prolongation	4.2	121	3.9	98	0.19	< 0.05
unfilled pause	5.25	150	9.57	235	34.02	< 0.001
peripheral markers						
NL sound	1.6	48	3.8	94	22.9	< 0.001

Table 54. *Average duration of filled pauses in SITAF (in ms)*

PAR	L1	Stdev	L2	Stdev
A02	792.29	366.11	512.9	211.86
A03	682.57	224.53	506.81	162.63
A07	689	222.82	524.53	176.84
A09	602.5	116.67	462.76	230.55
A10	N/A	N/A	675	212.13
A11	510	N/A	641.33	389.63
A13	790	257.39	529.67	159.85
A15	330	N/A	422.57	294.37
A16	471.25	63.82	622.75	231.67
A17	552.14	148.21	522.31	206.43
A18	598.57	124.04	578.38	265.46
Total	**658.6**	**238.8**	**514.6**	**192.3**
F02	406.36	175.74	533.25	72.2
F03	372.6	251.45	496.67	192.94
F07	361.82	255.6	567	209.23
F09	373.2	176.19	450.75	169.51
F10	498	170.23	407.4	160.11
F11	491.43	191.7	537.24	248.47
F13	N/A	N/A	616.55	121.8
F15	301.92	181.16	399.07	171.86
F16	534	524.67	586.67	330.81
F17	189.27	83.9	471.7	174.8
F18	850	N/A	309.13	168.45
Total	**371.78**	**214.9**	**465**	**190.5**

Table 55. *Average duration of prolongations in SITAF (in ms)*

PAR	L1	Stdev	L2	Stdev
A02	435	138.14	479.23	227.04
A03	400.07	126.84	475.66	174.95
A07	588.75	170.64	401.82	137.89
A09	575.13	123.28	511.67	295.92
A10	266.67	50.33	371.29	66.93
A11	700	N/A	427.18	190.39
A13	546.43	118.22	497.8	161.64
A15	387.25	98.01	382.25	126.25
A16	312.86	58.87	380.77	113.08
A17	372	135.17	434.98	120.32
A18	396.5	192.29	465.64	154.04
Total	**433.5**	**149.4**	**461.5**	**177.6**
F02	442.83	86.14	576.14	87.43
F03	236.67	90.18	366.67	75.06
F07	450.5	172.32	717.27	280.79
F09	423.84	105.36	430.9	115.95
F10	369.5	134.56	428.09	164.4
F11	328.33	53.45	406.75	100.55
F13	347.25	84.66	623.25	168.79
F15	304.33	121.11	343.33	110.03
F16	334	48.08	337.5	84.6
F17	317.67	69.07	326.22	102.02
F18	480	145.33	366.33	63.89
Total	**383.4**	**122.1**	**459.3**	**187.1**

Table 56. *Average duration of unfilled pauses in SITAF (in ms)*

PAR	L1	Stdev	L2	Stdev
A02	1067.83	1007.55	820.25	344.32
A03	758.42	343.51	629.39	249.7
A07	747	429.34	633.21	238.71
A09	693.42	351.8	785.56	465.96
A10	878.55	426.62	1048.25	584.96
A11	729.6	282.44	695.59	290.67
A13	794.83	423.97	834.69	412.12
A15	717.47	291.76	643.65	288.43
A16	637.11	435.23	579.31	279.21
A17	816.67	282.54	778.2	376.44
A18	692.93	353.3	640.3	238.07
Total	**754.7**	**444.3**	**683.9**	**300.8**
F02	646.5	238.67	889.64	943.56
F03	672.92	395.98	598.23	165.4
F07	591.62	187.72	754.6	300.41
F09	582.44	320.83	820.78	558.62
F10	707.06	376.85	544.27	225.11
F11	876.67	460.9	813.08	410.89
F13	444.57	100.72	757.9	541.57
F15	613.14	279.23	674.9	308.22
F16	625.25	303.51	599	332.44
F17	697.36	237.57	615.87	268.62
F18	748.67	325.37	765.33	202.93
Total	**629.4**	**292.3**	**717.3**	**443.1**

Table 57. *Count of non-lexical sounds for the American speakers (raw values, SITAF))*

NL	L1	L2	Total
click	13	44	57
cough	1	0	1
creaky-voice	4	7	11
hunhun	0	1	1
inbreath	23	47	70
laughter	1	4	5
mm	3	7	10
sigh	3	4	7

Table 58. *Count of non-lexical sounds for the French speakers (raw values, SITAF)*

NL	L1	L2	Total
click	12	30	42
creaky-voice	5	16	21
inbreath	25	35	60
laughter	3	1	4
mm	1	7	8
sigh	2	4	6
unintelligible	0	1	1

Table 59. *Z scores and p values for the distribution of NL sounds (SITAF)*

	American speakers		French speakers	
click	z = -1.401	p = 0.1	z = -0.854	p = 0.3
inbreath	z = 0.785	p = 0.4	z = -0.235	p = 0.8
other	z = 0.681	p = 0.4	z = 0.994	p = 0.3

Table 60. *Average number of markers combined within a sequence (SITAF)*

PAR	L1	Stdev	L2	Stdev
A02	1.9	0.8	2.4	1.4
A03	2.0	1.4	2.6	1.7
A07	1.4	0.6	2.2	1.7
A09	1.5	0.8	2.4	1.9
A10	1.9	1.0	1.9	1.4
A11	1.4	0.5	2.1	1.5
A13	1.6	0.9	2.2	1.6
A15	1.6	0.9	2.0	1.5
A16	1.7	1.1	1.8	1.1
A17	1.6	0.8	2.7	1.9
A18	1.6	0.9	1.9	1.2
Total	**1.7**	**1.0**	**2.4**	**1.6**
F02	2.4	1.7	2.4	1.3
F03	1.7	0.9	1.4	0.7
F07	2.0	1.4	1.9	1.2
F09	2.1	1.4	2.3	1.5
F10	2.3	1.6	2.4	1.4
F11	2.4	1.5	2.1	1.7
F13	1.6	0.8	2.2	2.6
F15	1.7	0.9	1.9	1.2
F16	1.5	0.8	1.6	0.9
F17	1.5	0.8	1.6	1.1
F18	1.7	1.2	1.6	0.9
Total	**1.9**	**1.3**	**2.0**	**1.5**

Table 61. *Raw values and z scores for the proportion of pragmatic and referential gestures in (dis)fluent cycles of speech in L1 and L2*

American group						
	L1			**L2**		
	DIS	**FLUENT**	**Z (*p*)**	**DIS**	**FLUENT**	**Z (*p*)**
Pragmatic	41	200	0.74 (0.4)	87	214	0.84 (0.3)
Referential	13	82	-0.74 (0.4)	18	57	-0.84 (0.3)
French group						
	L1			**L2**		
	DIS	**FLUENT**	**Z (*p*)**	**DIS**	**FLUENT**	**Z (*p*)**
Pragmatic	43	156	0.58 (0.5)	81	218	1.79 (0.07)
Referential	14	62	-0.58 (0.5)	13	63	-1.79 (0.07)

Table 62. *Annotation of gaze direction in SITAF (raw values)*

	American group		French group	
Gaze	**L1**	**L2**	**L1**	**L2**
away	385	419	262	323
in different directions	54	62	62	82
towards interlocutor	434	450	435	459
towards paper	60	129	111	138
Total	**933**	**1060**	**870**	**1002**

Table 63. *Rate of gestures in L1 and L2 in SITAF (raw frequencies and per hundred words)*

	L1		L2	
PAR	**Phw**	**raw frequency**	**phw**	**raw frequency**
A02	6	11	12	23
A03	21	107	15	63
A07	7	21	12	33
A09	5	22	11	22
A10	12	11	37	25
A11	13	55	10	24
A13	2	6	16	37
A15	8	20	3	5
A16	12	30	12	17
A17	13	27	26	89
A18	12	26	27	38
Total	**11**	**336**	**16**	**376**
F02	9	17	9	11
F03	4	11	16	33
F07	11	53	16	65
F09	4	22	13	41
F10	25	30	33	40
F11	24	42	18	77
F13	11	36	20	44
F15	9	24	9	22
F16	13	15	15	29
F17	9	20	6	9
F18	5	5	6	4
Total	**10**	**275**	**15**	**375**

Table 64. *Results on the Z test on gaze direction in SITAF (Z scores and p values)*

	American group	**French group**
away	z = 0.78 ; p = 0.4	z = -0.98 ; p = 0.3
in different directions	z = -0.05 ; p = 0.9	z = -0.85 ; p = 0.3
towards interlocutor	z = 1.82 ; p = 0.06	z = 1.811 ; p = 0.07
towards paper	z = -4.36 ; p < 0.002	z = -0.64 ; p = 0.5

Table 65. *Annotation of gaze in fluent and disfluent stretches of speech (raw values, American group, SITAF)*

	L1		**L2**	
	DIS	**FLUENT**	**DIS**	**FLUENT**
away	164	221	245	174
in different directions	17	37	50	12
towards interlocutor	84	350	138	312
towards paper	14	46	59	70
Grand Total	**279**	**654**	**492**	**568**

Table 66. *Annotation of gaze in fluent and disfluent stretches of speech (raw values, French group, SITAF)*

	L1		**L2**	
	DIS	**FLUENT**	**DIS**	**FLUENT**
away	127	135	181	142
in different directions	27	35	44	38
towards interlocutor	117	318	105	354
towards paper	51	60	54	84
Grand Total	**322**	**548**	**384**	**618**

Appendix 4

Table 67. *Rate of fluencemes in DisReg (raw frequency)*

PAR	class	conversation
A1	117	103
A2	153	93
B1	156	58
B2	103	43
C1	155	154
C2	151	115
D1	83	141
D2	103	75
E1	134	119
E2	103	170
F1	125	198
F2	89	129
Total	**1472**	**1398**

Table 68. *Rate of individual fluencemes (raw values and per hundred words)*

	class (raw)		Conversation (raw)		LL	p value
morpho-syntactic markers						
lexical repair	0.4	19	0.2	16	1.77	<0.05
morphological repair	**0.9**	**45**	**0.4**	**29**	**9.39**	**<0.01**
syntactic repair	0.9	44	0.8	56	0.04	<0.05
identical repetition	2.3	121	2.4	161	0.01	<0.05
self-interruption	**0.1**	**7**	**0.7**	**45**	**21.41**	**<0.0001**
truncated word	1.1	56	0.9	63	0.77	<0.05
vocal markers						
filled pause	**7.5**	**387**	**4.1**	**281**	**59.42**	**<0.0001**
prolongation	**3.1**	**162**	**4.1**	**280**	**7.54**	**< 0.01**
unfilled pause	**6.7**	**345**	**4.7**	**319**	**21.15**	**<0.0001**
peripheral markers						
NL sound	**5.4**	**278**	**2**	**134**	**99.58**	**<0.0001**
explicit editing phrase	0.2	8	0.2	11	0.01	< 0.05

Table 69. *Average duration values of filled pauses in DisReg*

	Class		Conversation	
PAR	**Average**	**Stdev**	**Average**	**Stdev**
A1	429.32	182.13	221.57	114.02
A2	414.11	217.32	241.05	168.63
B1	410.54	180.77	469.12	447.61
B2	521.44	324.45	353.33	118.16
C1	464.43	256.36	390.82	137.4
C2	375.34	272.73	314.65	95.3
D1	413.23	250.43	338.87	221.65
D2	240.95	129.08	353.82	187.6
E1	388.28	191.88	348.39	267.78
E2	467.35	335.01	326.19	161.83
F1	421.69	190.95	356.07	153.91
F2	242.25	65.12	308.87	118.53
Total	**412.09**	**240.18**	**340.05**	**199.77**

Table 70. *Average duration values of unfilled pauses in DisReg*

	Class		Conversation	
	Average	**Stdev**	**Average**	**Stdev**
A1	674.45	268.46	591.81	331.72
A2	687.74	301.06	661.81	433.77
B1	478.84	117.95	410.00	17.32
B2	645.68	228.19	578.85	393.07
C1	646.48	191.42	639.74	352.79
C2	637.52	272.42	764.83	410.32
D1	755.06	731.27	578.93	352.83
D2	668.21	649.90	556.23	260.62
E1	849.79	875.02	516.00	136.34
E2	794.06	435.20	574.07	288.66
F1	827.11	666.58	527.52	222.44
F2	630.34	676.95	539.85	218.26
Total	**695.78**	**543.37**	**594.20**	**323.55**

Table 71. *Average duration values of prolongations in DisReg*

	Class		Conversation	
	Average	**Stdev**	**Average**	**Stdev**
A1	281.78	102.09	382.55	167.54
A2	397.27	106.11	313.06	103.75
B1	325.88	126.69	356.00	47.75
B2	316.14	55.63	400.17	129.99
C1	448.00	213.27	377.83	159.57
C2	259.31	62.37	325.24	105.14
D1	280.67	70.82	341.65	86.29
D2	281.11	120.40	295.75	90.54
E1	401.18	119.58	414.04	341.31
E2	393.90	188.06	346.78	119.99
F1	281.90	96.20	325.02	98.55
F2	348.30	88.62	328.24	102.17
Total	**351.32**	**142.45**	**350.14**	**155.70**

Table 72. *Count of non-lexical sounds in DisReg (raw values)*

NL	prepared	spontaneous	Total
click	63	15	78
creaky-voice	4	4	8
inbreath	206	99	304
laughter	0	4	4
mm	5	8	13
sigh	0	4	4

Table 73. *Average number of markers combined in a sequence (DisReg)*

	Class		Conversation	
	Average	**Stdev**	**Average**	**Stdev**
A1	2.8	1.3	3.7	1.9
A2	3.1	1.2	2.6	0.7
B1	2.9	1.1	2.6	0.9
B2	2.5	0.9	2.7	1
C1	3.3	1.6	2.6	1.2
C2	2.9	1.8	2.6	0.9
D1	2.3	0.6	2.6	0.9
D2	2.3	0.6	2.9	0.9
E1	2.8	1.2	2.9	1.3
E2	3	1.2	2.6	0.9
F1	3.3	1.8	2.7	1
F2	2.7	1.6	2.7	1.3
Total	**2.9**	**1.3**	**2.7**	**1.1**

Table 74. *Count of gestures in class and conversation (raw values)*

PAR	Class	Conversation	Total
A1	41	27	68
A2	38	36	74
B1	92	38	130
B2	43	35	78
C1	14	44	58
C2	92	48	140
D1	36	47	83
D2	61	35	96
E1	68	51	119
E2	9	40	49
F1	23	59	82
F2	11	31	42
Total	**528**	**491**	**1019**

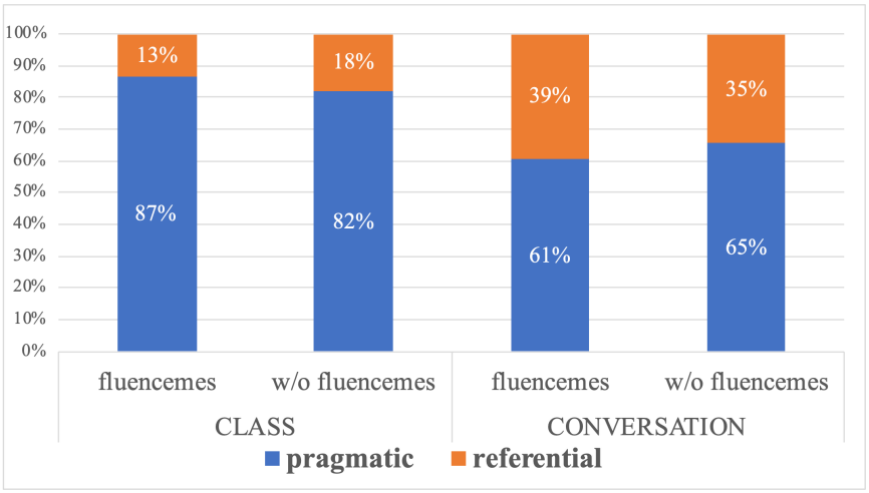


***Figure 76**. Proportion of the two main gesture types in class and conversation with/without fluencemes*

***Table 75.** Annotation of gaze direction in DisReg (raw values)*

	class	**conversation**
away	67	737
in different directions	53	62
towards camera	25	1
towards interlocutor	478	1199
towards paper	1211	23
Total	**1834**	**2022**

***Table 76.** Count and proportion of gaze direction for Participant F1 (Linda)*

Gaze	**Class**	**Conversation**
away	0%	31% (77)
in different directions	0%	4% (9)
towards interlocutor	9% (10)	60% (151)
towards paper	91% (110)	6% (15)

Table 77. *Annotation of gaze direction in DisReg with and without fluencemes (raw values)*

	WITH FLUENCEMES		**W/O FLUENCEMES**	
Gaze	**class**	**conversation**	**class**	**conversation**
away	33	326	34	411
in different directions	35	47	18	15
towards camera	7	0	18	1
towards interlocutor	129	359	349	840
towards paper	531	7	680	16
Total	**735**	**739**	**1099**	**1283**

Index

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, positionelles, fonctionelles, 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 recurrentes et pertinentes aux activités langagières des locuteurs.

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

UNIVERSITÉ SORBONNE NOUVELLE
ED 625 – MAGIIE (Mondes Anglophones, Germanophones, Indiens, Iraniens et Études Européénnes)
EA 4398 PRISMES (Langues, Textes, Arts et Cultures du Monde Anglophone)
Maison de La Recherche
4, rue des Irlandais, 75005 Paris