STUDIA UBB. PHILOSOPHIA, Vol. 67 (2022), 2, pp. 49-66
(RECOMMENDED CITATION)
DOI:10.24193/subbphil.2022.2.03

RETHINKING (DIS)FLUENCY WITHIN THE SCOPE OF INTERACTIONAL LINGUISTICS AND GESTURE STUDIES

LOULOU KOSMALA*

ABSTRACT. The study of so-called 'disfluency' phenomena (*uh* and *um*, filled and unfilled pauses, self-repairs and the like) has gained a lot of attention in various fields in linguistics in the past few decades, but a majority of studies tend to be production-oriented and often disregard fundamental aspects of face-to-face communication such as interactional dynamics and gesture. This paper presents a multimodal and multilevel model of "inter-fluency", considering different levels of analysis, mainly, talk, gesture, and interaction, by combining different theoretical frameworks and methodologies in gesture studies and interactional linguistics in order to bridge this gap and go beyond previous cognitive-oriented models.

Keywords: Interaction, fluency, gesture, multimodality, interactive model

Introduction

The study of so-called 'disfluency' phenomena (i.e. the study of self-repairs, repetitions, "uh" and "um", pauses etc.) has received a lot of attention in the past sixty years within a variety of research fields, from psycholinguistics (Levelt, 1983; Shriberg, 1994) to conversation analysis (Sacks et al., 1974), but there is a lack of consensus regarding the definition and use of terms. In the fields of psycholinguistics and phonetics, the term 'disfluency' is commonly associated with speech disturbances, disruptions, or errors (Ferreira & Bailey, 2004; Mahl, 1956), while in conversation analysis, most authors use the term "repair" (Sacks et al., 1974) to refer to systematic

* Université Paris Nanterre, EA 4398 PRISMES / EA 370 CREA, e-mail: loulou.kosmala@gmail.com

practices in interaction. The issue, I will argue, is mostly theoretical. Most studies conducted in the field of disfluency in psycholinguistics tend to be largely production-oriented, mostly based on speech production models at the utterance level (Levelt, 1989). Studies in conversation analysis and interactional linguistics, however, focus exclusively on features of talk-in-interaction such as preference structure (Stivers & Robinson, 2006; Yule, 1996), stance-taking (Kärkkäinen, 2006), or participation framework (Goodwin, 2007), among others. The aim of this paper is to offer an integrated model of fluency which goes beyond production-oriented views of 'disfluency' phenomena by combining different approaches and theoretical frameworks, such as gesture studies, multimodality, and interactional linguistics. This paper further presents a unified view of fluency and disfluency phenomena, hence *(dis)fluency* (Crible et al., 2019) as highly complex constructs revolving around multiple processes at the same time which are in constant *interaction* with one another. As this paper will show, the concept of fluency should not be restricted to one view or one model, which invites us to consider it from different dimensions of language (speech, gesture, and interaction) as deeply embedded within social structures, hence going beyond production-oriented approaches isolated from larger interactional contexts.

This paper is structured as follows: I first provide a brief overview of production-oriented models of disfluency as largely defined in the fields of psycholinguistics and phonetics, then review other theoretical frameworks relevant to the study of (dis)fluency but which offer additional dimensions; I then present an integrated framework of *inter-fluency* which combines these different approaches, and illustrate its application with an example taken from a corpus of audio-visual data analyzed with specific tools. Lastly, I conclude this paper with a discussion around the concept of interaction and its definition, as well as its major role in the study of (dis)fluency. The present work is largely based on Kosmala (2021b) but this paper focuses more specifically on the concept of *interaction*, whether it is to describe the role of intersubjectivity in *talk-in-interaction*, or to highlight the *interaction* and *interrelation* between the different modalities and dimensions of language within the study of fluency.

Production-oriented models of disfluency

In the field of psycholinguistics, the study of disfluency, or *speech errors* (see Levelt, 1983, 1989; Menn & Dronkers, 2016) has mainly been concerned with the analysis of utterance surface structures according to speech production models which identify different mental operations at different stages of execution or

production[1]. When speaking, the primary goal for speakers is to produce maximally acceptable speech in both content and form (Hieke, 1981, p. 150) which compels them to monitor their own speech by following different steps, such as message construction, formulation, and articulation (Levelt, 1983). A number of researchers have been interested in sudden breaks or changes in the monitoring process, in other words, when "speech breaks down" (Lickley, 2015, p. 12): whenever the speaker detects an error in the speech apparatus. This process is commonly identified as a departure, or shift from an *ideal fluent delivery* (Ferreira & Bailey, 2004) known as *disfluency*. Researchers in psycholinguistics and phonetics have thus been interested in the structure of disfluent speech events, and have identified several parts, or regions, illustrated in the figure below, taken from Shriberg (1994, p. 8).

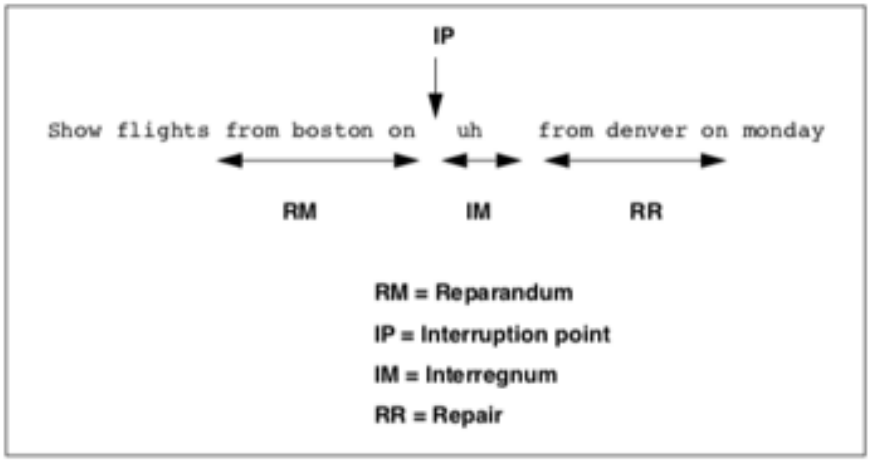


Figure 1. Disfluency Regions (Shriberg, 1994)

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

[1] Extensive work has also been conducted on L2 fluency (e.g. De Jong, 2018; Gilquin, 2008; Götz, 2013, among others), which is not the primary focus of this paper, although it has also been included in the present inter-fluency model. Read Kosmala 2021a for more information.

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).

Let us now illustrate this model with an utterance taken from the SITAF Corpus (Horgues & Scheuer, 2015) which has been analyzed in detail in previous work on disfluency (Betz & Kosmala, 2019; Kosmala, 2021b, 2021a; Kosmala et al., 2019).

but I I'm not sure (be)cause here um (0.768) [!]
here (0.898) if you:u uh I ain't got the w word

This utterance is taken from a native French speaker talking in her second language (more details will be provided in the following sections) who appears to be experiencing difficulties in her speech production. An expert in disfluency research would commonly make the following observations regarding the number and types of disfluencies in this segment (two repetitions, two filled pauses, two silent pauses, one tongue click, one syllable prolongation etc) and where they are located in this utterance (between the *Reparandum* and *Repair* at the *interruption point* within the *Interregnum*). Drawing from this type of analysis, we can make the preliminary assumption that this particular speaker is highly *disfluent*, given the number of disfluencies found in her speech. This could be related to many cognitive processes, such as difficulties in grammatical encoding or lexical access in her second language (Hartsuiker & Notebaert, 2009; Hilton, 2008), or it could reflect stress and anxiety (Christenfeld & Creager, 1996). This type of analysis has been conducted for years within the fields of psycholinguistics, phonetics, and computational linguistics to understand how spoken speech production systems work, with applications in speech modelling and human-machine dialogue (Betz et al., 2018; Eklund, 2004; Eklund & Shriberg, 1998).

However, this type of analysis does not give us the full picture. Most analyses conducted in disfluency research are based on decontextualized utterances and focus exclusively on processing and planning processes associated with their production, but we rarely get to see their pragmatic and interpersonal dimension in larger interactional contexts (except for a few, see Allwood et al., 1990; McCarthy, 2009; Tottie, 2014, among others). There are even fewer studies which address the role of gesture and gaze with regards to disfluency (except for a few, Seyfeddinipur, 2006; Graziano & Gullberg, 2013; Tellier et al., 2013; read Kosmala 2021a for review). In addition, there are several underlying problems with the term "disfluency" (read Kosmala, 2021b for a full review) which presupposes a

disruption, or a problem to be repaired; yet so-called disfluencies are a highly natural aspect of spontaneous talk, as they 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). The term "disfluency" thus stems from the monolithic and mythical assumption that a speaker is either "fluent" or "disfluent" or that a structure either reflects "fluency" or "disfluency". But language represents so much more than a binary opposition or a series of words in decontextualized utterances; it is an embodied experience, grounded in our overall environment comprised of our own bodies, our movement in space, and our *interaction* with the people and objects around us. The present work thus stresses the need to situate (dis)fluency phenomena within a larger interactional and multimodal framework, going beyond previous production-oriented models of disfluency.

Beyond the production model: the interplay of speech, gesture, and interaction

In this view, (dis)fluency should not be solely regarded as a mental process, isolated from other visible cues in interaction, but as a multimodal process which includes all semiotic features of language (the stream of speech, hand gestures, body posture and orientation, gaze behavior), following the frameworks of interactional linguistics, gesture studies and multimodality.

Interactional linguistics is an interdisciplinary framework which brings together a growing community of linguists who are interested in studying many aspects of grammar and prosody from a specific interactional approach. One of its major theoretical influences (among two others, read Couper-Kuhlen & Selting, 2001) 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, 1996). 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. 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. Another major contribution is found in the field of social interaction and linguistic anthropology, and more specifically in the work of C. Goodwin and M.H. Goodwin (Goodwin, 1981, 2003, 2017; Goodwin & Goodwin, 1986, 1996, 2004) 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). In addition, the participation framework is 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, which leads us to the field of gesture studies and multimodality.

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 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).

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). Stivers & Sidnell (2005) further distinguished between the vocal-aural and the visual-spatial modalities of face-to-face multimodal communication: the vocal modality 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), and the visuo-spatial modality includes the study of visible behavior, such as hand gestures, gaze, and body orientation within the spatial environment. Extensive work has also been done on the classification of hand gestures (read Kendon, 2004; Ferré, 2019; and Kosmala, 2021b for review) which can be analyzed on the basis of their form (handshape, orientation, execution etc.) or function in discourse (pragmatic versus referential). More recently, a closely related framework has emerged, known as *Analyse de Discours Multimodale* (Analysis of Multimodal Discourse, Ferré, 2019) which integrates visible, verbal and vocal resources specifically within the area of discourse analysis. Speech is further subdivided into the verbal mode (discourse at the segmental level), the vocal mode (at the suprasegmental level) and the gestural mode.

In this view, speech cannot be separated from its context of occurrence, which is constantly being (re)-shaped by speakers' actions within the ongoing interaction, and interaction is not exclusively built by speech, but by a combination of semiotic features which are harmoniously coordinated within the materiality of the exchange. This further invites us to reconsider the concept of fluency, based on these different frameworks: while studies in so-called "disfluency" phenomena seem in contradiction with the tenants of interactionist studies, I believe that they could supplement one another as to provide a richer and more complex picture of the phenomena under study. It is not surprising to note that the term "disfluency" is virtually excluded from all researchers' analyses in interactional linguistics, given what the term entails. Interactional linguists have opted for terms such as "repair" (Goodwin & Goodwin, 2004; Schegloff et al., 1977), although homonymous with the one used by psycholinguist Levelt (1983), has entirely different implications, as it is said to be "neither contingent upon error, nor limited to replacement" (Schegloff et al., 1977, p. 363). The aim of this paper is therefore to bridge the gap between production-based psycholinguistic studies conducted on "disfluency" and interactional, multimodal approaches to social interaction.

Towards an integrated framework of inter-fluency

The present model invites us to reconsider fluency from multiple dimensions (also see Candea, 2000; Grosman, 2018; Segalowitz, 2016) situated within a larger integrative framework. The concept of *fluency* is not restricted to the notions of ideal delivery or second language proficiency, but is understood in broader terms such as "communicativeness" "smoothness" "fluidity" "progressivity" 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). Following McCarthy's (2009) notion of *confluence*, which focuses on the co-creation of fluency, I have proposed the term *inter-fluency*[2], with the prefix "inter" to draw a parallel to the notions of intersubjectivity, interpersonal relations, and interaction. In addition, the prefix *inter* further symbolizes the constant ongoing interaction between the different dimensions of fluency, exemplified in the figure below.

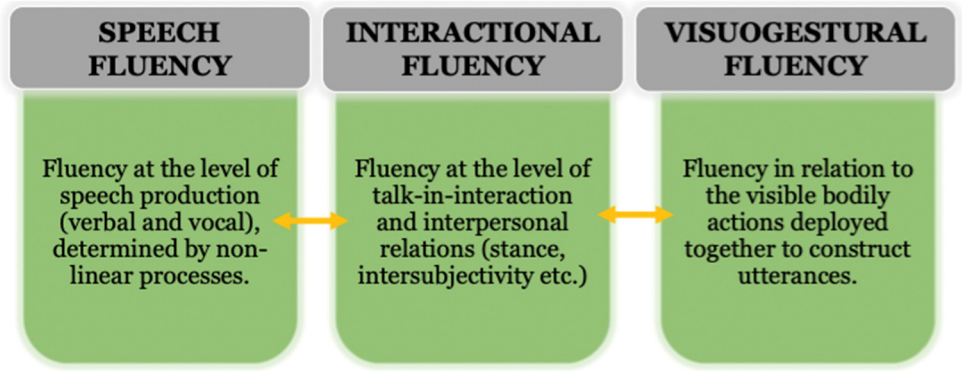


Figure 2. Multidimensional model of inter-fluency (Kosmala, 2021b, p. 96)

[2] In my PhD dissertation (Kosmala, 2021b) I discuss the terminology in detail regarding the use of terms 'disfluency' versus 'fluency' or '(dis)fluency'. The term '(dis)fluency' is also sometimes used in this paper as a reference to Crible et al.'s (2019) work who focused on the dynamic, flexible and ambivalent nature of these processes, also used extensively in Kosmala (2021b). This paper aims to gradually remove the 'dis' from 'disfluency' which remains too closely associated with disruptive features of speech that are negatively connotated.

This model comprises three different dimensions. First, the *speech* dimension (equivalent to Segalowitz's (2006) *utterance fluency*) is restricted to the level of speech production (verbal and vocal) which takes into account morphosyntactic and temporal features of speech, in line with previous psycholinguistic models. The second dimension, the *interactional* dimension (similar to Grosman's (2008) *socio-interpersonal dimension*) further includes the situated conversation languages practices at play, and the *visuogestural* dimension (which echoes Götz's (2013) *nonverbal fluency*) considers visible bodily behavior during utterance construction. The term "utterance" here is not only restricted to the speech level, as it follows Kendon's acceptation, defined as the following: "the ensemble of actions, whether composed of speech alone, of visible action alone, or a combination of the two" (Kendon, 2004, p. 111). In addition, the term "fluenceme" is adopted in this model, following Götz (2013) to replace the phrase "disfluency marker".

In sum, the present definition of inter-fluency involves multiple dimensions that are not mutually exclusive, but interactively complementing one another in the course of the interaction. In some contexts, a verbal utterance that is considerably 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 may potentially disrupt the progressivity of an interactional sequence (e.g. with turn-initial uhms displaying a dispreferred action, see Hoey, 2014; Yule, 1996). This is exemplified in the following section.

Methodology and tools

Let us now illustrate this model with an excerpt from a videotaped corpus, initially presented in the first section. This excerpt, as explained earlier, is taken from the SITAF Corpus (Horgues & Scheuer, 2015) which includes face-to-face interactions between tandem partners at Sorbonne Nouvelle University alternatively speaking in their L1 and their L2, in French and English. The following excerpt shows an entire interactional sequence, using multimodal transcription conventions (cf. Appendix) between a French speaker (Elena) and an American speaker (Francis), who are talking about the prices of tuition fees at university (also analyzed in Kosmala, 2021a).

Excerpt: Tandem interaction in French

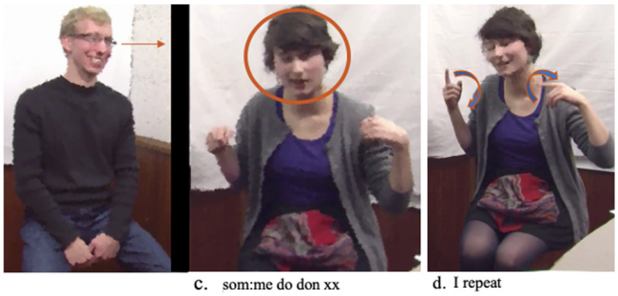
→ **1 *ELEN: but I [/] I'm not sure (be)cause here um (0.768) [!] [/] here (0.889) if you:u uh I ain't got the w word here ((thinking face a.)) ((looks up; smiles b.))**

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.))
4 *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.

As stated in the first section of the paper, Elena is experiencing a number of lexical and grammatical difficulties in her second language, which makes her verbal utterances highly "disfluent" from a strictly verbal perspective. However, unlike the isolated utterance presented earlier, this excerpt shows us that fluency mechanisms do not operate on a single level but on several interrelated ones, which are not restricted to the speech dimension. Here Elena is enacting a lexical search activity by coordinating vocal fluencemes and bodily actions which enable her to project the current progressivity of her search. As she is trying to make a point (that students don't have to pay a lot of tuition fees to get good education) she first displays a state of uncertainty with a *thinking face* (Goodwin & Goodwin, 1986; picture a.), while suspending the course of her utterance, 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, in which both hands are rotating as to convey the process of starting over (picture d.). Her tandem partner, Francis, 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). In sum, while Elena's utterances are highly "disfluent" from a strictly verbal perspective, it doesn't stop her from pursuing her word search activity without her partner's assistance. 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).

This example highlights the interactional dimension of fluency, which does not solely reflect online cognitive processes, but also relies on participation and cooperation, in line with the frameworks of Interactional Linguistics. This novel approach to inter-fluency can be analyzed with different tools, adopting previous annotation systems conducted on (dis)fluency (Crible et al., 2019; Pallaud et al., 2019) combined with conversation-analytic methods to social interaction (Sacks et al., 1974) as well as gestural notation systems (Kendon, 2004). All these analyses were conducted using the annotation software ELAN (Sloetjes & Wittenburg, 2008) which is a multilayer and multipurpose annotation tool developed at the Max Planck Institute for Psycholinguistics to provide a technological basis for the annotation of multi-media recordings. Therefore, this type of analysis favors audio-visual data in spontaneous, naturalistic and ecological settings within situated and social activities, known as *talk-in-interaction*. Three essential analytic orientations emerge from this conversation-analytic approach to interaction (Atkinson et al., 2002, p. 204): 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', which are accomplished through the deployment of multiple semiotic modalities. In a similar 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 a previous corpus-based study (Kosmala, 2021b), the present methodology was applied to two videotaped corpora, the SITAF Corpus (mentioned earlier) and the DisReg Corpus (Kosmala, 2020) which includes recordings of French students engaged in two different communication settings, during individual class presentations, and in pairs during a conversational talk. These two datasets were chosen for their multimodal quality as well as their ecology; they share a similar set of features as they both include semi-realistic situations of students interacting within a shared institutional and social environment, the university. Therefore, this allows for an efficient and reliable quantitative treatment of the corpus sample (following corpus-based analyses, e.g. Crible et al., 2019) as well as micro-qualitative analyses of the data (following conversation-analytic methods, e.g. Sacks et al., 1974). These two corpora also capture different interactive situations (tandem exchange, conversation between friends, and individual oral presentations) during which students are

engaged in different tasks across different settings and languages. Several recurrent interactive multimodal practices were identified in the corpus study, which further provides an interactive frame for the analysis of fluency.

Discussion

The aim of the present model is to offer a multidimensional account of fluency phenomena, going beyond previous cognitive-oriented models of speech production and integrating other theoretical frameworks which consider talk from a multimodal and interactional perspective. Several assumptions or questions emerge from this interactional approach:

- Disfluency phenomena should not solely be regarded in terms of a binary opposition between "fluency" and "disfluency, but rather as a multi-level embodiment of the notion fluidity and flow.
- Fluency may result from two systems of communication, interactive communication management, and own communication management (following Allwood, 2017). In this sense, inter-fluency does not only reflect internal cognitive processes, but also exhibit essential features of talk-in-interaction. The present model suggests that these different dimensions should work together and include visible bodily behavior to capture the complexity of human interaction.
- Fluencemes are merely "disfluency markers" indexing a suspension point in the speech flow, they are highly flexible and dynamic categories which are shaped by their context of use. Context is understood here in terms of (1) the immediate neighboring environment of the fluencemes, (2) the syntactic position of fluencemes within the verbal utterance, (3) their sequential position within a turn; (4) their co-occurrence with bodily actions, (5) the situated language activity speakers are currently engaged in, and (6) the overall material environment, i.e. the objects they are manipulating.
- Speakers continuously adjust their body and talk for the co-participants of the exchange and rely on a multiplicity of semiotic resources and diverse media to build meaning in interaction.

Appendix

Transcription conventions for fluencemes and gestures
based on CHAT and CA conventions.

CHAT conventions (MacWhinney, 2000)	
+/	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)	
[!]	tongue click
.hhh	inbreath
hhh	outbreath
creaky	creaky voice
(( ))	description of events, or analyst's comment
Gesture annotation (Kendon, 2004)	
~ ~ ~	preparation of gesture stroke
***	gesture stroke
<u>***</u>	hold
-.-.-	return to rest position

REFERENCES

Allwood, J. (2017). Fluency or disfluency? *Proceedings of DiSS 2017, the 8th Workshop on Disfluency in Spontaneous Speech*, 1.

Allwood, J., Nivre, J., & Ahlsén, E. (1990). Speech Management—On the Non-written Life of Speech. *Nordic Journal of Linguistics*, *13*(1), 3–48.
https://doi.org/10.1017/S0332586500002092

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.

Betz, S., Carlmeyer, B., Wagner, P., & Wrede, B. (2018). Interactive Hesitation Synthesis: Modelling and Evaluation. *Multimodal Technologies and Interaction*, *2*(1), 9. https://doi.org/10.3390/mti2010009

Betz, S., & Kosmala, L. (2019). Fill the silence! Basics for modeling hesitation. *The 9th Workshop on Disfluency in Spontaneous Speech*, 11.

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.
https://doi.org/10.1177/00238309010440020101

Boutet, D. (2018). *Pour une approche kinésiologique de la gestualité* [Habilitation à diriger des recherches]. Université de Rouen-Normandie.

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.* [PhD Thesis, Université Paris III- Sorbonne Nouvelle].

Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology*, *70*(3), 451. https://doi.org/10.1037/0022-3514.70.3.451

Clark, H.H. (2006). Pauses and hesitations: Psycholinguistic approach. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 244–248). Oxford: Elsevier.

Couper-Kuhlen, E., & Selting, M. (2001). Introducing interactional linguistics. *Studies in Interactional Linguistics*, *122*.

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.

De Jong, N.H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*(3), 237–254.

Disfluency: Interrupting speech and gesture. (2006). [Unpublished doctoral dissertation, Radboud University]. http://eprints.soas.ac.uk/21385/

Dollaghan, C.A., & Campbell, T.F. (1992). A procedure for classifying disruptions in spontaneous language samples. *Topics in Language Disorders*.

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.

Ferré, G. (2019). *Analyse de discours multimodale. Gestualité et prosodie en discours*. UGA éditions.

Ferreira, F., & Bailey, K.G.D. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences*, *8*(5), 231–237.

https://doi.org/10.1016/j.tics.2004.03.011

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.

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.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers*. Academic Press.

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. (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., & 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*.

Grosman, I. (2018). *Évaluation contextuelle de la (dis) fluence en production et perception: Pratiques communicatives et formes prosodico-syntaxiques en français* [PhD Thesis]. UCL-Université Catholique de Louvain.

Hartsuiker, R.J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*.

Hieke, A.E. (1981). A Content-Processing View of Hesitation Phenomena. *Language and Speech*, *24*(2), 147–160. https://doi.org/10.1177/002383098102400203

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*(2), 153–166.

Hoey, E.M. (2014). Sighing in interaction: Somatic, semiotic, and social. *Research on Language and Social Interaction*, *47*(2), 175–200.

Horgues, C., & Scheuer, S. (2015). Why Some Things Are Better Done in Tandem. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English Pronunciation: Trends and Directions* (pp. 47–82). Palgrave Macmillan UK.

https://doi.org/10.1057/9781137509437_3

Jefferson, G. (2004). Glossary of transcript symbols. *Conversation Analysis: Studies from the First Generation. Amsterdam: John Benjamins*, 13–31.

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.

Kendon, A. (2004). *Gesture: Visible action as utterance* (Cambridge University Press). Cambridge University Press. https://books.google.fr/books?hl=en&lr=&id=hDXnnzmDkOkC&oi=fnd&pg=PR6&dq=Kendon,+A.+(2004).+Gesture:+visible+action+as+utterance&ots=RJ4Tp92VhJ&sig=1lE4boAPssZefmS3uKfCEwBfpCs

Kosmala, L. (2021a). *A multimodal contrastive study of (dis)fluency across languages and settings: Towards a multidimensional scale of inter-(dis)fluency* [Unpublished PhD thesis]. Sorbonne Nouvelle.

Kosmala, L. (2021b). 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.*

Levelt, W.J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W.J. (1989). *Speaking. From intention to articulation*. MIT Press.

Lickley, R.J. (2015). Fluency and Disfluency. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 445–474). John Wiley. https://www.researchgate.net/publication/296707223_Fluency_and_Disfluency

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. transcription format and programs* (Vol. 1). Psychology Press.

Mahl, G.F. (1956). Disturbances and silences in the patient's speech in psychotherapy. *The Journal of Abnormal and Social Psychology*, *53*(1), 1.

McCarthy, M. (2009). Rethinking spoken fluency. *ELIA, 9, 11-29*.

Menn, L., & Dronkers, N.F. (2016). *Psycholinguistics: Introduction and applications*. Plural Publishing.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, *20*(3), 336–366.

Morgenstern, A., & Boutet, D. (forth.). *The orchestration of bodies and artifacts in French family dinners*.

Pallaud, B., Bertrand, R., Prevot, L., Blache, P., & Rauzy, S. (2019). *Suspensive and Disfluent Self Interruptions in French Language Interactions*.

Pomerantz, A., & Fehr, B. J. (2011). Conversation analysis: An approach to the analysis of social interaction. *Discourse Studies: A Multidisciplinary Introduction*, *2*, 165–190.

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

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. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, *102*(1), 161–216.

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.

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.

Seyfeddinipur, M., (2006). *Disfluency: Interrupting speech and gesture* [Unpublished PhD Thesis]. Radboud University.

Shriberg, E.E. (1994). *Preliminaries to a Theory of Speech Disfluencies* [PhD Thesis]. University of California.

Sidnell, J. (2016, March 3). *Conversation Analysis*. Oxford Research Encyclopedia of Linguistics. https://doi.org/10.1093/acrefore/9780199384655.013.40

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

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. (2010). Ecologies of gesture. *New Adventures in Language and Interaction*, 223–242.

Streeck, J., Goodwin, C., & LeBaron, C. (2011). *Embodied interaction: Language and body in the material world*. Cambridge University Press.

Tellier, M., Stam, G., & Bigi, B. (2013). Gesturing while pausing in conversation: Selforiented or partner-oriented?". *The Combined Meeting of the 10th International Gesture Workshop and the 3rd Gesture and Speech in Interaction Conference, Tillburg (The Netherlands)*.

Tottie, G. (2014). On the use of uh and um in American English. *Functions of Language*, *21*(1), 6–29. https://doi.org/10.1075/fol.21.1.02tot

Yule, G. (1996). *Pragmatics*. Oxford University Press.