



Disfluency markers and their facial and gestural correlates. Preliminary observations on a dialogue in French

Elgar-Paul Magro

Paris III – Sorbonne Nouvelle University, Paris, France

Abstract

The aim of this article is to try to establish any observable regularities between the vocal and the visual expression of disfluency markers in a French spontaneous dialogue. The data show different configurations for different types of disfluency markers. Thus “*euh*”s are typically accompanied by mutual eye contact and no gesture; interrupted eye contact takes place less frequently, on occasions where speech planning is more seriously impaired (syntactical disruption and combination of “*euh*” with other disfluency markers). False starts seem to be typically accompanied by gesture production whereas eye contact can be maintained if the speaker relies or not on the listener to resolve the speech production problem. The article takes up the idea that disfluency markers can be classified along a continuum throughout the speech formulation process, going from the most discreet to the most prominent. It suggests that the more prominent the disfluency, the more likely is the visual channel to play a role (interrupted eye contact and gesture production).

1. Introduction

Few researchers so far have considered hesitation phenomena from a multimodal perspective. When they do, they either focus on facial expressions, particularly on eye contact, or else they focus on hand movement (or gesture), rarely on both.

This pilot study mainly tries to address the following two questions:

- a) which regularities can be observed between the vocal and visual channels during disfluencies in spontaneous speech?
- b) what light do these regularities throw upon the speech formulation process itself?

At this stage, it is important to specify and define the variables of the visual mode that have been taken into account for the purpose of this study and to define them for the sake of clarity. They amount to two: eye contact and gesture. Eye contact is what normally takes place between the speaker and the listener during a conversation (Kendon [6], Bouvet & Morel [1]). Although inter-individual differences have been observed, the speaker typically looks less often in the direction of the listener than the latter looks at him. By gesture, one must understand here “the movements of the hands and arms that we see when people talk” and that “are closely synchronized with the flow of speech” (McNeill [9] : 1, 11).

2. Literature review

Published research (Exline & Winters [5], Kendon [6], Brossard [2]) states that, during hesitation, the speaker typically looks away from the listener. The main reason put forward to account for this behaviour is that, during hesitation, the speech formulation process requires extra cognitive efforts on the speaker’s part; thus the speaker prefers to look away from the listener so as to avoid being distracted by his gaze. In

this article, the term “interrupted eye contact” will be used when the speaker looks away from the listener during disfluent speech; the term “mutual eye contact” will be used to design instances when the speaker keeps looking at the listener despite disfluency markers in his speech.

Research taking into account hesitation and gesture points out differing views according to whether the disfluencies arise from a speaker expressing himself in his mother tongue or else in a foreign language. As regards native speakers, McNeill [9] and Seydeffinipur & Kita [12] stress the absence of gesture during hesitant speech. In the case of non-native speakers, various authors (e.g. Proceedings of ORAGE 1998 [7], 2001 [3]) point out the abundance of gestures accompanying the speech formulation process.

On the basis of the above assumptions, we set out this study with the following starting hypothesis, or expectation: namely, that the disfluencies of the corpus under observation would be mainly accompanied by interrupted eye contact and rather limited gesture production.

3. Corpus

The corpus used for the purpose of this study is a spontaneous dialogue between two female French friends, Anne and Soline, aged 25, both coming from the region around Paris and having pursued tertiary education. What is meant by “spontaneous” here is speech that is not linked to any prior written text and not recorded in a speech laboratory – in fact, the recording took place at the house of a common friend of the two speakers.

The general theme of discussion was not imposed and the conversation went from reminiscing holidays to discussing boyfriends. The excerpt which was analyzed is a 90-second narrative by Soline, particularly rich in disfluencies as the speaker encounters difficulties, amongst others, in remembering the proper names of the places she visited in Italy with her parents some years ago.

An audio and video recording (on MiniDisk and digital cameras respectively) was first carried out. As illustrated in Figure 1 below, the two speakers were seated at an angle, at an approximate distance of half a metre from each other. Each video camera was positioned on either side of the speakers outside the interactive space of the dialogue, at an approximate distance of one and a half metres and at an angle of 45° from the speakers. The cameras were neither totally visible, nor totally hidden, but both speakers stated at the end of the recording that they had not been upset by the presence of the cameras.

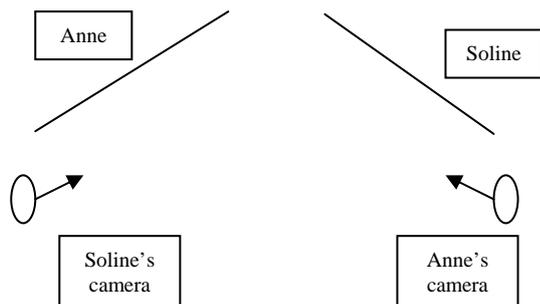


Figure 1: Diagram showing the recording set-up for the corpus.

At a later stage, a synchronized editing of the video was obtained on computer format (AVI file). This made it possible for Soundforge to be used to transcribe and observe minutely the eye, hand and arm movements of the speakers (25 images per second). The transcription of gestures was based on the identification of their movement phases : preparation, stroke and retraction (cf. McNeill [9], Bouvet & Morel [1]). On the other hand, a rigorous observation of the eye and head movements of Soline (the main speaker) over a long sequence of the recording made it possible to pinpoint the direction of Soline's gaze, that is whether it was oriented towards Anne (the listener) or elsewhere. Although a small error margin is not to be totally excluded, the observation work carried out as well as the observer's intuitive experience of speaker and conversation participant made it possible to infer whether a disfluency was accompanied by mutual or interrupted eye contact.

The examples illustrated in the plates below show the end result of the edited video: Soline, the main speaker in this excerpt, and the one whose disfluencies have been analyzed, is on the right frame of each photo; on the left frame is Anne, the listener. As can be observed in these illustrations, each frame not only shows the head, trunk, hands and knees of one particular speaker but also includes the hands of the other speaker. This detail, which initially came across as an undesirable result of the recording, was actually beneficial for the transcription of the gestures produced, as the hands of each speaker could be observed from two different angles (one on each frame).

4. Results

The data show that not all disfluencies are accompanied by interrupted eye contact and lack of gesture. On the contrary, a detailed analysis suggests that different disfluency markers tend to display different types of facial and gestural behaviours.

For the purpose of this study, only two types of disfluency markers were taken into account, namely "euh"s (the French equivalent of the English "uh"s and "um"s) as well as false starts. One of the main reasons behind this choice lies in the high frequency with which they occur in the dialogue excerpt analyzed in this paper. Besides, it is generally admitted that "euh"s are considered to be the most typical markers of hesitation in French, whereas false starts are easily identifiable and enable to consider larger stretches of disfluent speech than "euh"s do.

4.1. The case of "euh"s

4.1.1. General remarks

Soline's speech, whose disfluencies have been analyzed, contains fifteen occurrences of disfluent "euh"s. Four of these occurred at overlaps in her speech (at moments when Anne, the listener, gives some sort of backchannel signals). Since it is not clear whether these four "euh"s are a result of a real disfluency in Soline's speech or rather a way to assert her wish to keep her turn, they have not been taken into account for this study.

Out of the eleven remaining "euh"s, only one was accompanied by gesture production, which seems to comfort McNeill's [9] claim. On the other hand, only three out of eleven were accompanied by interrupted eye contact, whereas during the remaining eight the speaker keeps looking at the listener. This seems quite surprising in the light of the common point of view adopted by the doxa.

Let us first give an example of each category of "euh", one with interrupted and another with mutual eye contact. The numbers within brackets indicate the duration of silent pauses in milliseconds, the colons shows the amount of lengthening (each colon marks an additional lengthening of 200ms).

E.g. (1): *c'est en (550) euh::: (600) ben près de Florence c'est quoi la région?*

(it's in (550) euh (600) well next to Florence what's the region?)

Soline cannot remember the name of the region around Florence in Italy and after producing an 800ms euh, she decides to ask Anne for help. The eye contact is interrupted at the beginning of the utterance and is resumed on the word "Florence". Soline's averted gaze does not fix one particular point but is very mobile. Cf. Plate 1 below.



Plate 1: Illustration of Example (1). Disfluent "euh" - interrupted eye contact and absence of gesture: "*c'est en (550) euh::: (600) ben près de Florence c'est quoi la région?*".

E.g. (2) *et euh::: donc euh: eux ils ont l'habitude de louer une maison euh::: là-bas*

(and euh so euh they have the habit of renting a house euh over there)

Soline is trying to explain the geographical context for the voyage she is narrating. During these three "euh"s of varying length (350 to 800ms) which she produces within the same utterance, she does not look away from Anne. Cf Plate 2 below.



Plate 2: Illustration of Example (2). Disfluent “euh » - mutual eye contact and absence of gesture. “*et euh::: donc euh: eux ils ont l’habitude de louer une maison euh::: là-bas*”

4.1.2. “Euh”s and their distribution

At this stage we tried to find ways to possibly account for the distribution of these two categories of “euh”s (with mutual or interrupted eye contact). So we tried to look at the impact that the following three factors may have on their distribution: namely (a) the duration of the “euh”s, (b) their combination with a silent pause and (c) the degree of syntactical disruption they bring about.

4.1.2.1 The duration of “euh”s and eye contact

Out of the three “euh”s accompanied by interrupted eye contact, one is lengthened (350ms) and two are very long (750, 800ms). On the other hand, the duration of the eight remaining “euh”s accompanied by mutual eye contact varies from brief (200ms) to very long (800ms). Although one would expect duration of “euh”s to be of some significant influence on the behaviour of the speaker’s gaze direction, the small numbers of “euh”s under observation here do not make it possible to make any statement in this sense.

4.1.2.2 The combination of “euh”s with silent pauses and eye contact

Since no correlation could be found between eye contact and the combination of “euh”s with vowel lengthening, we tried to see if the data suggested any possible correlation between eye contact and the combination of “euh”s with other disfluency markers.

The three “euh”s during which the speaker looks away from the listener are immediately followed by a silent pause. On the contrary, none of the eight other “euh”s during which the speaker does not turn away her look from the listener is immediately followed by a silent pause. This suggests that the combination of an “euh” with a silent pause might have a direct relation with eye contact during speech.

4.1.2.3 The level of syntactical disruption and eye contact

Another interesting observation to point out is that two of the three “euh”s that are accompanied by interrupted eye contact are also accompanied by an interrupted syntactical structure. They are the same two “euh”s that are immediately followed by a silent pause (cf. 4.1.2.2. above). In other words, the speaker does not finish off the syntactical structure that precedes the “euh” and finds the need to start afresh (either a repetition or a new start).

On the contrary, this syntactical disruption does not happen in any of the eight “euh”s that are accompanied by mutual eye contact. In all of these instances, if the “euh”s were to be

deleted from the speaker’s discourse, the resulting syntax would not bear the traces of any disfluency.

This might suggest a possible correlation between the extent of syntactical disruption followed by an “euh” and the interruption or otherwise of eye contact.

4.2. The case of false starts

The discourse of Soline in the dialogue excerpt under observation contains seven false starts, that is seven syntactical structures that are left abandoned in the favour of new ones in the immediate linguistic context. Six of these are not isolated but appear in sequences. These figures do not include, for obvious reasons, any incomplete utterances located at turn-taking frontiers, in which case they are clearly interrupted by the listener and not by the speaker herself.

4.2.1. Gesture

False starts present a major difference compared to “euh”s in the visual channel. Contrary to “euh”s, they are predominantly accompanied by gestures (six out of seven). Moreover, when false starts appear one after the other in a sequence, as is twice the case in our corpus, gesture seems to illustrate and throw light on the formulation process that is under way.

Let us have a closer look at a sequence of three false starts in our corpus.

E.g. (3): Soline is trying to give a list of the places she and her parents visited in Tuscany. After mentioning Florence and Siena, she takes up the same syntactical structure to add a third place name, but apparently she does not manage to recall it immediately. This gives rise to the following sequence of structures, of which (c), (d), (e) (in italics) are false starts:

- (a) on a fait Flo^{rence}
- (b) on a fait Si^{enne} (1300)
- (c) *on a fait^{ait} ::*
- (d) *parce qu’on é^{ait} dans le: dans la ré^{gion}:*
- (e) *on avait lou^é:: une:*
- (f) à Grosse^{to} on était (600)

(we visited Florence / we visited Siena / *we visited / cos we were in the region / we had rented a / Grosseto that’s where we were)*

It is interesting to note that no gesture accompanies the first of the three successive false starts: (c). Presumably, the speaker at this stage is not yet conscious of the amplitude which her disfluency is going to cover.

The second false start (i.e. (d)) is accompanied by a gesture which clearly shows the cognitive effort of word-searching going on. The fingers of the speaker’s right hand, oriented at an angle towards the listener, rub against the thumb six times. This gesture accompanies the whole of the utterance. Incidentally, this utterance introduces a change of strategy : following the unproductive attempt in (c) to find the desired place name, the speaker moves out of the narrative per se and shifts to an explicative mode (cf. the connector “parce que” and the change in tenses from narrative (passé composé) to extranarrative ones (imparfait, plus-que-parfait). Plates 3A and 3B try to illustrate this finger-rubbing gesture.



Plate 3A: Illustration of Example (3d). False start – mutual eye contact and word-searching gesture (repeated finger-rubbing). “*parce qu’on é^{ait} dans le: dans la ré^{gion}.*”



Plate 3B: Illustration of Example (3d). False start – mutual eye contact and word-searching gesture (repeated finger-rubbing). “*parce qu’on é^{ait} dans le: dans la ré^{gion}.*”

Since the word-searching in (d) above was not fruitful, the speaker gives a new twist to her speech planning strategy and prefers to start afresh by evoking a concrete referent this time, namely the lodging they had rented (later on in the dialogue, we learn it was a house). The gesture that is produced now is no longer a gesture of word-searching, but one which shows the grasp of an object. In fact, the speaker raises her right hand, then palm facing downwards she lowers it whilst cupping it in the form of the claws of a bird of prey. Plausibly enough, the object being grasped is the house she has in mind or possibly the place whose name she is after. The gesture of prehension is illustrated in Plate 4 below.



Plate 4: Illustration of Example (3e). False start – mutual eye contact and gesture of prehension (grasping). “*on avait lou^é une.*”

This strategy pays off as before the speaker finishes her utterance in (e), she has already found the name she was looking for, namely the town of Grosseto. Interestingly enough, this final syntactical structure which puts an end to the series of false starts analyzed above is also accompanied by a gesture. This time it is an abstract pointing gesture,

whereby the forefinger of the speaker’s right hand points out to an imaginary point within the conversation space. The wider space circumscribed during the previous false start thanks to the gesture of prehension, is now clear-cut, precise: the proper name has at last been identified. Plate 5 illustrates this pointing gesture.



Plate 5: Illustration of Example (3f). End of false start series – mutual eye contact and pointing gesture. “*à Grosse^{to} on était.*”

4.2.2. Eye contact

As far as eye contact is concerned, no clear-cut tendency arises in the case of false starts. In three out of seven occurrences the eye contact is mutual, in the remaining four it is interrupted. It is not very clear to us yet at this stage which are the main factors that may account for this distribution. One hypothesis is that it may significantly depend on whether the speaker decides or not to rely on the listener or not to overcome the source of his disfluency.

It may be of interest to note that when false starts appear in a cluster (which is often the case in our corpus), the first false start of the series behaves differently to the following false starts. Thus in the series analyzed in 4.2.1. above, the first false start is accompanied by interrupted eye contact, following which the speaker looks back at the listener till the final completed utterance arises. This possibly bears a link with the speaker’s change of narrative strategy pinpointed above. In the second series (equally made up of three false starts) it is the other way round. On the other hand, the only occurrence of a false start in isolation is accompanied by interrupted eye contact. Thus at a certain point in all the false starts observed, there is interrupted eye contact.

5. Discussion

It may be argued that disfluency in speech takes place along a sort of continuum throughout the speech formulation process going from the more discreet to the more prominent markers. The results observed above suggest that the more prominent the disfluency in the vocal channel, the more prominent it is likely to be in the visual channel.

As a matter of fact, only the more prominent “euh”s seem to be accompanied by an interrupted eye contact; that is, those “euh”s which are combined to a silent pause and/or the need for a syntactical restart. Gestures do not intervene at this point. As regards false starts, they cover longer stretches of disfluency than “euh”s, often contain other disfluency markers (vowel lengthening, internal repetition as in (3d) above or even combination with an “euh” as in (1)), and their incompleteness clearly brings about a higher level of syntactical disruption than “euh”s. Which means they come off as more prominent disfluency markers than “euh”s. Interestingly, one can note that this is where gestures take place.

Thus one may argue that the more the speech formulation seems to be seriously impaired, the more this will be marked on both the vocal and the visual level.

It seems interesting at this stage to mention that different researchers have shown interest in the role of gesture within the speech production process, proposing different speech production models which place gesture at either the conceptualization level or the formulation level (e.g. de Ruiter [4], Kita [7] and Krauss, Chen & Gottesman [8]). Our preliminary findings do not offer much support to either of the two major models but we understand that, at a later stage, our research may contribute to this debate, which is still open.

6. Conclusion

The aim of this study was to present the preliminary results obtained upon close observation of 1.5 minutes of a conversation between two young female French friends.

This paper has argued that the generally acclaimed views on facial and gestural behaviour during disfluencies (interrupted eye contact and lack of gesture) give only a partial picture of what really goes on. The initial working hypothesis thus needs to be nuanced.

The data show different configurations for different types of disfluency markers. Thus eye contact shows a tendency to be mutual during “euh”s, except where speech planning seems to be more seriously impaired (combined disfluency markers and syntactical disruptions); the data suggest that false starts may prefer interrupted eye contact although no statement can be made in favour of this. Gestures seem to be typically absent with “euh”s whereas they typically accompany false starts, illustrating thereby the progress in the speech planning process.

Finally, the paper suggests that the more the speech formulation is impaired alongside the disfluency continuum, the more markers arise in the visual mode (going from interruption of eye contact to the production of gestures).

It is clear that no general conclusions can be drawn out of such a small corpus. Indeed, the results obtained have to be validated by using much vaster material, ideally constituted of data that is greater both in size and in nature. A further analysis of longer stretches of the same dialogue is currently under way.

Future analysis must also take into account the rates and configurations of eye contact and gesture production in fluent stretches of speech. Such a comparison of fluent and disfluent speech will enable to establish how typical of disfluent speech are the trends of facial and gestural behaviour that have been observed. The observation of the timing relation (e.g. Seyfeddinipur & Kita [12]) between gesture, gaze and speech during disfluencies may also be of interest in future studies.

7. Acknowledgements

This study is part of an ongoing doctoral research supervised by Prof. Mary-Annick Morel [1], [10]. Our gratitude also goes to Prof. Laurent Danon-Boileau [10] and Dr. Danielle Bouvet [1] for their inspiring comments on certain aspects of the research presented in this paper.

8. References

- [1] Bouvet, Danielle & Mary-Annick Morel. 2002. *Le ballet et la musique de la parole*. Paris-Gap: Ophrys.
- [2] Brossard, Alain. 1992. *La psychologie du regard*. Neuchâtel: Delachaux et Niestlé.
- [3] Cavé, Christian, Isabelle Gauitella & Serge Santi (éds.). 2001. *Oralité et gestualité. Interactions et comportements*

- multimodaux dans la communication*, Actes du colloque Orage 2001, 18-22 juin 2001, Aix-en-Provence, France. Paris: L’Harmattan.
- [4] de Ruiter, Jan Peter. 2000. The production of gesture and speech. In McNeill, David (ed.). *Language and gesture*. Cambridge : Cambridge University Press, pp. 284-311.
- [5] Exline, Ralph V. & Lewis C. Winters. 1966. Affective relations and mutual glances in dyads. In Tomkins, Silvan S. & C. Izard (eds.). *Affect, cognition and personality*. London : Tavistock.
- [6] Kendon, Adam. 1967. Some functions of gaze direction in social interaction, *Acta Psychologica*, vol. 26, pp.1-47.
- [7] Kita, Sotaro. 2000. How Representational Gestures Help Speaking. In McNeill, David (ed.). *Language and gesture*. Cambridge : Cambridge University Press, pp. 162-185.
- [8] Krauss, Robert M., Yihsiu Chen & Rebecca F. Gottesman. 2000. Lexical gestures and lexical access: a process model. In McNeill, David (ed.). *Language and gesture*. Cambridge : Cambridge University Press, pp. 261-283.
- [9] McNeill, David. 1992. *Hands and mind. What gestures reveal about thought*. Chicago and London: The University of Chicago Press.
- [10] Morel, Mary-Annick & Laurent Danon-Boileau. 1998 (rééd. 2001). *La grammaire de l’intonation. L’exemple du français oral*. Paris – Gap : Ophrys.
- [11] Santi, Serge, Isabelle Gauitella, Christian Cavé & Gabrielle Konopczynski (éds.). 1998. *Oralité et gestualité. Communication multimodale, interaction*, Actes du colloque Orage 1998, 9-11 décembre 1998, Besançon, France. Paris: L’Harmattan.
- [12] Seyfeddinipur, Mandana & Sotaro Kita. 2001. Gesture as an Indicator of Early Error Detection in Self-Monitoring of Speech. *Proc. DISS’01*, 29-31 August 2001, Edinburgh, Scotland, pp. 29-32.