



HAL
open science

A multimodal approach to markedness in spoken French

Gaëlle Ferré

► **To cite this version:**

Gaëlle Ferré. A multimodal approach to markedness in spoken French. *Speech Communication*, 2014, 57, pp.268-282. 10.1016/j.specom.2013.06.002 . hal-01422556

HAL Id: hal-01422556

<https://hal.science/hal-01422556>

Submitted on 26 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Multimodal Approach to Markedness in Spoken French

Gaëlle Ferré¹

School of Languages & Linguistics Laboratory (LLING), University of Nantes, France

Abstract

This study aims at examining the links between marked structures in the syntactic and prosodic domains (fronting and focal accent), and the way the two types of contrast can be reinforced by gestures. It was conducted on a corpus of 1h30 of spoken French, involving three pairs of speakers in dialogues. Results show that although the tendency is for marked constructions both in syntax and prosody not to be reinforced by gestures, there is still a higher proportion of gesture reinforcing with prosodic marking than with syntactic fronting. The paper describes which eyebrow and head movements as well as hand gestures are more liable to accompany the two operations. Beyond these findings, the study gives an insight into the current models proposed in the literature for gesture-speech production.

Keywords: contrast, fronted syntactic constructions, prosodic emphasis, gesture reinforcement, gesture-speech production models

1. Introduction

Since the introduction of the concept by Halliday (1967), quite a large body of research described *Information Structure* in discourse, which refers to “the partitioning of sentences into categories such as focus, background, topic, comment etc...” (Büring, 2007:445). Information focus has been described by Halliday as “one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative” (1967:204). In this paper, we are especially interested in the expression of focus on parts of message blocks in what can generally be termed *marked structures*. Dooley & Levinson (2000:38) describe unmarked utterances as “configurations [which] are susceptible to a variety of interpretations, [and] hence become general-purpose, default constructions”, whereas “marked constructions are used only for specific discourse-pragmatic purposes”. Lambrecht, in a more specific way, describes the difference between marked structures and unmarked ones saying that “while the marked member [of a sentence] is positively specified for some pragmatic feature, the unmarked member is neutral with respect to this feature” (1994:14) and that “given a pair of allosentences, one member is pragmatically unmarked if it serves two discourse functions while the other member serves only one of them” (op. cit.). He explains that in syntax for instance, whereas a sentence showing regular SVO constituent order can be interpreted as an all-new sentence type (in broad focus), it may also be the case that only one constituent of the sentence is informative and therefore under (narrow) focus. Such a sentence is unmarked in terms of its interpretation. On the contrary, a clefted sentence for instance only permits the narrow-focus reading and is therefore marked insofar as it establishes some sort of contrast which has been defined as the way “each entity, predication, etc. in an utterance refers back to, alters and/or updates the existing discourse model” (Calhoun, 2009:56).

In speech, contrast can be operated in different modes. In the verbal mode, one way of establishing contrast is by using fronted constructions – also termed elsewhere *thematic* or

¹ Corresponding author. Tel: +33 2 40 14 11 34
E-mail address: Gaelle.Ferre@univ-nantes.fr

topicalized constructions (Combettes, 1999) – such as clefted sentences, for instance. Lambrecht (1994) distinguishes three focus-structure types in the syntactic domain which all express some sort of contrast: *predicate focus*, *argument focus* and *sentence focus*. The three types of contrast will be described in section 2.1 of this paper.

In the vocal mode, contrast is operated by the production of accents that “stand out as perceptually more prominent than newness accents” (Krahmer & Swerts, 2001:391). By *newness accents*, the authors mean accents produced in unmarked utterances with a broad focus interpretation, i.e. in which all the information is new and no particular item is discursively salient. We will see in section 2.2 of this paper that, similarly to what is found in the syntactic domain, different types of contrasts can also be expressed through prosodic means that play different functions in discourse.

In the visual mode, some co-speech gestures are not necessarily linked to the semantic content of the speech they accompany. For instance hand beats do not carry semantic features, like it is the case of iconic gestures. They are produced as a way of drawing the listener’s attention to some element of speech, i.e. contrast this item with surrounding information, although it remains to be known what type of gesture exactly may establish such a contrast. One way of enquiring into this question is to look at the type of gesture produced together with marked utterances in other modes. In previous work (Ferré et al., 2007), it was observed that reinforcing gestures tended to be produced together with degree adverbs and connectors (discourse markers which introduce a conversational sequence). This is not surprising if we consider that degree adverbs are more emphatic in nature than other types of adverbs and that connectors are used to distinguish the sequence they introduce from previous discourse and therefore establish a contrast. We are not aware of other studies in which the relationship between gesture and (morpho-)syntactic units are looked into. However, quite a number of studies enquired into the relationships between gestural and prosodic marking and they will be described in section 2.3.

Much in the same line of research, the working hypothesis developed in this paper is that utterances may be – but are not necessarily – marked in several modes simultaneously, and that this potential combination of marking strategies enables to “distinguish different degrees of focus in utterances which show variable discursive saliency”² (Lacheret, 2003:159). In a corpus of conversational French described in section 3, we therefore examine the links – or absence of links – between utterances which are marked in the syntactic domain – namely fronted constructions – and utterances marked in the prosodic domain believing that “kontrast can be predicted by syntactic and prosodic features and their interaction” (Calhoun, 2009:53). We also look at the gestures which systematically accompany contrast in the two modes. In the final discussion, we show how the findings of this study are in (dis)agreement with current gesture-speech production models proposed in the literature.

² Our translation.

2. Theoretical background

2.1. Syntactic fronting

Considering marked constructions in the syntactic domain, Lambrecht (1994) distinguishes three types of focus structures that show similarity in form – fronting – but assume different functions: *predicate focus*, *argument focus* and *sentence focus*. As summarized in Lacheret (2003:137-138), the three types of focus have the following functions:

Predicate focus corresponds to dislocations and topicalizations and its role is to introduce a new topical referent in the mental representation of the listener (left dislocations and topicalizations) or maintain a topical referent in the mental representation of the listener (right dislocations – which were not considered in the present paper). Dislocated elements, according to Dooley & Levinson (2000:37), have “their own intonation contours, are intermediate in phonological prominence, and their task of relocating a cohesive tie is intermediate in newsworthiness between focus and presupposition”. An example of left dislocation in French³ is “**Les ânes**, c’est vraiment insupportable” (*The donkeys, they are terrible*). In this example, the dislocation is of a full Noun Phrase, which is co-referent with the pronoun in the matrix clause. However, dislocation may also be made with a pronoun as in “**nous**, ça nous barrait” (lit. *Us, we were bored*). One must note that pronominal dislocation is much more frequent in spoken French than it is in spoken English, and that all personal pronouns may be detached. This type of dislocation is generally used in its contrastive value in written French. Gregory & Michaelis (2001) made the same observation for English and even observed that this type of dislocation is also dependent on genre and familiarity. Topicalization is very close in form to dislocation, the difference between them being that “TOP contains a gap in the clause which corresponds to an argument position that the pre-clausal NP can be construed as filling, whereas LD [left dislocation] contains an argument-position pronoun which is coreferential with the pre-clausal NP” (Gregory & Michaelis, 2001:1667). In speech, topicalizations are almost always introduced by a topic marker (like *concerning, about...*). An example of topicalization is “**Au niveau animaux**, c’est tout ce qu’ils ont” (lit. *Concerning animals, it is all they have = These are the only animals they have*).

Argument focus corresponds to cleft and pseudo-cleft constructions. It aims at “identifying an argument for a given proposition” (Lambrecht, 1994:336). Simple focus establishes a contrast between the argument and any other type of information that may occupy the same position, whereas contrastive focus establishes a contrast between the argument and another one (Creissels, 2004:11). An example of pseudo-cleft construction is “**Ce qui me gênait**, ouais, c’était ça” (*Yes, what I didn’t like was this*) and of cleft construction, “**C’est un truc qui** me dit rien du tout” (*It is something that I don’t feel like doing*).

³ The examples provided in this section are from the CID corpus of spoken French described in section 3.

At last, *Sentence focus* corresponds to presentative constructions and aim at introducing a new discourse referent or reporting an event. In French, there are two types of presentative constructions (“il y a ... qui” *There is/are ... who/that* and “j’ai/on a ... qui” *I/we have ... who/that*). An example of a presentative construction in the corpus is “**Y avait ma sœur et des amis qui** étaient venus me rejoindre” (lit. *There were my sister and some friends who came to visit me*).

This typology is the most complete one described in the literature to the best of our knowledge. It accounts for all the examples met in the corpus and was therefore adopted as it suits our needs. Adverbials, as well as spatial and temporal clauses may also be fronted, and although they constitute a “point of departure” (Dooley & Levinson, 2000:35) in information structure which is also the case of the constructions described above, they are not considered as topics on which a comment can be made, and are therefore not considered in the scope of this paper.

2.2. *Prosodic focus*

There has been a vast body of studies in the literature on prosodic focus. To give a general definition to the term, focus has been described as “the general phenomenon of attributing prominence to a constituent” (Féry, 2001:161). Prosodic prominence indicates “a pragmatically construed semantic domain (...) marking different activation states of discourse referents” (Lambrecht, 1994:208). Prosodic prominence can indicate that the whole utterance is considered as new in the activation state, and therefore in broad focus, “often indicated only by the presence of a final falling tone” (Féry, 2001:176) but also by the lengthening of the final syllable. But prosodic prominence can also “indicate certain topic discontinuities in the discourse” (Lambrecht, op.cit.), in which case only part of the utterance is new in the activation state and therefore in narrow focus, “often signaled by an initial boundary tone” (Féry, op.cit.). The boundary tone is produced with higher pitch and intensity, according to Lacheret (2003:138), but we will see below that duration also plays a role in certain types of focused constituents.

We won’t discuss the presence or absence of lexical stress in French here, as it is out of the scope of this paper, but as summarized in Féry (2001), there are two different opinions on the question. The traditional view holds that French has a regular lexical stress on the last full syllable of lexical words (a syllable in which the vowel is not schwa). Lexical stress in French is then fixed as opposed to other languages like English and does not contribute to the meaning of words. Lexical stress on the last syllable affords a position for the assignment of accent at higher levels of the intonational structure, namely the Accentual Phrase (AP) and the Intonational Phrase (IP) as defined by Selkirk (1978). In both types of phrases, accent is also assigned to the last full syllable of the phrase, and is described as a “final, grammatical or logical accent” (Féry, 2001:157). Its role is to indicate phrasing in unmarked utterances, which is the reason why the French language is often considered as differing from Germanic languages in terms of prosody: “In French, the primary correlate of information structure is not accentuation, but rather prosodic

phrasing” (Féry, 2003). In order to highlight a syntactic constituent, it is possible in French (although not in all contexts) to make it a separate prosodic phrase in sentence initial position (also called thematic position or theme). However, a second strategy may be adopted by speakers in narrow focus utterances with the presence of an “insistence or emphatic accent” (Féry, op.cit.) in any position in the utterance (which may also be the last syllable of course). A more recent view holds that because accent in French plays a grammatical role in broad focus sentences and a discursive-pragmatic role in narrow focus sentences, it is not derived from lexical stress and that lexical stress does not exist in French. Here, we are especially interested in accents at higher levels of the prosodic hierarchy and especially emphatic accents and are therefore not so much concerned with the existence of lexical stress.

Much of the work on narrow focus is based on the analysis of a certain type of prosodic focus, namely contrastive focus. Dooley & Levinson (2000:36) state that “a contrastive statement (...) differs in one or more particulars from an already activated propositional framework (...)”. This contrastive statement may either “replace the existing filler of a plot” or “select between alternatives to fill an empty slot”. Contrastive focus is marked by an accent that is perceptually more prominent than the default accent in unmarked utterances (Krahmer & Swerts, 2001; Katz & Selkirk, 2011) and contrast involves an increase of pitch protrusion, duration and/or intensity (Katz & Selkirk, 2011:81). To this, one must add that contrastive focus is realized on the first syllable of the focused constituent that possesses a consonantal onset and its characteristics also involve a rising/falling pitch pattern whose peak is aligned with the first syllable, together with a significant lengthening of the onset consonant (Astésano et al., 2004:1). The latter authors observe that these tendencies in the production of contrastive focus have also been reported for other types of focus like intensification focus in non contrastive utterances. As to post-focus constituents, different positions have been held. Most authors (among which Dohen et al., 2004; Dohen & Loevenbruck, 2004, 2005; Kohler, 2006) observe a “global F0 and intensity compression in the post-focus sequence either through a low plateau, a late but steady fall or a constant fall until the end of the utterance” (Dohen et al., 2004:156). Yet, Féry (2001:167) states that post-focus constituents are “realized with a low and flat intonation, or alternatively with a high and flat intonation until the end of the sentence”, which is also observed in our corpus and is worth mentioning here.

In terms of perception, some studies mention the fact that “focus is very well perceived through the auditory modality” (Dohen et al., 2004:156). The parameters which are used by the listeners to perceive contrastive focus are pitch and duration, although not in equal proportion, as mentioned by Herment-Dujardin & Hirst who note that “duration alone is not sufficient to express emphasis but it nevertheless plays an important part in the perception of emphasis when it is associated with other parameters, F0 and semantics. (...) Unlike duration, F0 alone can express emphasis” (2002:382). At last, Dahan & Bernard (1994:504) mention that although intensity varies with the production of contrastive stress, it plays a small role in its identification.

Another important cue in the perception of contrastive focus is that the focused constituent is perceived better if it is embedded in an utterance, whereas “the difference in salience tends to

disappear if the accented word is heard in isolation” (Kraemer & Swerts, 2001:391). Context is also mentioned as an important factor in the identification of contrastive stress by Dahan & Bernard (1994).

Although we may state that prosodic prominence in narrow-focus utterances always establishes some kind of contrast, the contrast does not necessarily oppose two discourse constituents at the level of the utterance. It has a much broader array of discourse functions which are listed in Calhoun (2009:64-65), one of the most complete existing pragmatic typologies to the best of our knowledge. She distinguishes true ‘contrastive accent’ on a word which is “directly differentiated from a mentioned, semantically related word” from other categories like what she terms as:

- ‘Subset’, i.e. “a member of a more general set mentioned in the context”,
- ‘Adverbial’ focus when “a focus-sensitive adverb such as *only* or *even* is associated with the word”,
- Focus in ‘answer’ that “completes a question (or open proposition) by the other speaker”,
- Focus in ‘correction’ that “corrects or clarifies a previous word or phrase”,
- Focus in ‘other category’ when the emphatic accent does not fall into the previously mentioned types.

It is true that the different types of prosodic emphases that we analyze in this paper may not be distinguished by any prosodic differences in the signal, but since they play different functions in discourse, they may well be accompanied by different types of gesture, knowing that gesture is very sensitive to discourse function. However, in order to match more closely the types of emphasis met in our corpus, we propose a reorganization of Calhoun’s typology as:

- Discourse contrast when an item of discourse stands in contrast with another as in the following example from our corpus⁴ “ils sont déterminés à les tuer assez rapidement parce que **eux** non plus ne peuvent pas dormir” (*they mean to kill them quickly [the roosters] because **they** can’t sleep either*).
- Retrieval of a word or of an idea may also generate a prosodic emphasis on the retrieved item as in “le volet s’ouvre vers la euh vers la **droite**” (*the shutter opens on the uh on the **right***).
- Another possible context for prosodic emphasis is self-correction. The emphasis may then fall on the corrected item, but may also fall on a corrective adverb as in “Myriam, elle était pas spécialement accrochée au prénom quoi **enfin** au nom” (*Myriam didn’t really mind about the [baby’s] first name, its last name **rather***).
- Other-correction may also generate an emphatic stress when a speaker contradicts another participant as in “mais **si** il est convoqué” (*of **course** he will play [the match]*)⁵.

⁴ The emphasis falls on the syllable noted in boldface.

⁵ ‘Si’ is difficult to translate having no direct equivalent in English. In this particular context, it means *contrary to what you think, he will play the match*.

- Prosodic emphasis may also serve to express a high degree or reinforce the degree of something as in the following examples: “elle a accouché vers Noël, elle a mis **vingt** heures si tu veux” (*she gave birth around Christmas, it took **twenty** hours you see*). Here the focal accent enables the speaker to express a judgment on how long the labor lasted, ‘twenty hours’ being presented as a very long time. In the next example, “ouais quelle **gamelle** elle a pu se prendre pour éclater ses chaussures à ce point” (*yes what a **cropper** she must have come to smash her [ski] boots in this way*), the speaker emphasizes the fact that the fall she is talking about must have been particularly violent for the ski boots to be in such a state. These two examples can be considered as instances of focalization.

Scholars very often associate prosodic focus with fronted syntactic constructions since one of their common ‘phonological motivations’ is to “enhance the phrasing” (Féry, 2001:153). Rialland et al. contest the fact that “cleft sentences [which] are generally viewed as focus-related sentence types (...) are prosodically marked for focus” (2002:595). They find that “the relation between focus and clefting is far less direct than is usually assumed and that clefts may have various types of informational organization besides the prototypical focus/post-focus one” (op.cit., p. 598). They found that there exist different types of cleft utterances, some of which are all new utterances in broad focus. Besides, it is generally assumed that cleft utterances are linked to prosodic focus because of their rising contour which enables the speaker “to establish a link with another element later in the Intonation Phrase” (Féry, 2001:162). Yet, this rising contour is not necessarily related to the accent type presented above and is not always a marker of emphatic stress. Therefore, one may wonder if any kind of emphatic stress is needed to reinforce a syntactic construction which is already marked and this will be one of the objects of study in the present paper.

2.3. Gesture reinforcement of marked constructions

Bolinger (1986:199) claims that “gesture and speech/intonation are a single form in two guises, one visible and the other audible”, a view also shared by McNeill (1992, 2005) who proposes in the Growth Point Theory that gesture and speech stem from the same mental imagery and constitute one integrated system. On the contrary, de Ruiter (2000) proposes in the Sketch Model that gesture and speech constitute two separate systems which interact at a very early stage in the conceptualization of the message before grammatical and phonological encoding of the message. Therefore, linguistic factors do not have any impact on the gestural expression. Krauss (2000) also proposes a model (Lexical Gesture Process Model) in which gestures are produced before speech encoding since they facilitate lexical retrieval. However, Kita & Ozyürek (2003), as well as Kita et al. (2007) showed that the types of syntactic constructions valid in a language do have an influence on the gestures produced by the speakers. The Interface Model they propose therefore posits that gesture and speech constitute two systems which interact throughout the different stages of formulation and that grammatical encoding also has an impact on the type of gesture produced and its synchronization with speech.

In order to determine whether the gesture and speech systems interact at an early or later stage of the formulation, it is interesting to enquire into the possible interaction between gesture and marked constructions: if no distinction can be found between gesture accompaniment of unmarked and marked constructions, this means that neither the grammatical encoding nor the phonological one have an influence on the type of gesture produced and therefore that gesture production must be decided on before the encoding of speech. If on the contrary different gesture types are met with marked and unmarked constructions, this could well mean that either gesture can be planned at different stages of the formulation or that there exist different gesture planners.

Several studies examined the links between gesture and the expression of focus, considering either the temporal organization of gesture and speech, or the gestures which accompany different types of foci or morphosyntactic and discourse constituents.

In terms of the temporal organization of gesture and speech, Wilmes looked at hand/arm movements produced together with speech and “found differences between so-called new-information foci and contrastive foci in their relation to gestures. Whereas the former align with the onset of an associated gesture, the latter align with the stroke of that gesture” (2009:8).

Dohen and her colleagues found that the “presence and salience of the visual cues enhances perception” (Dohen & Loevenbruck, 2005:2413). They studied the links between prosodic contrastive focus and some lower-face parameters. Their working hypothesis was “that the main articulatory consequence of contrastive focus is hyper-articulation. Hyper-articulation can be achieved in various ways, including increase in the amplitude of lip and/or jaw opening and closing movements, increase in lip spreading or narrowing” (op.cit., p. 2414). Their results in both the 2004 and 2005 papers showed that contrastive focus is regularly marked with “a large jaw opening associated with a high opening velocity, a long phrase-initial lip closure and a post-focal hypo-articulation”. The phrase-initial lip closure is consistent with what has been observed in studies in prosody as it corresponds to the lengthening of the initial consonant [m] in the syllable they tested which was [ma].

Krahmer et al. (2002), Krahmer & Swerts (2007) and Swerts & Krahmer (2008) studied the links, both from the production and perception point of view, between some gestures (eyebrow raises, hand beats and head nods) and acoustic prominence, especially contrastive focus. They showed in experimental studies that these gestures facilitate the perception of prominence, but also that, when produced together with speech, they influence some acoustic parameters of speech.

In a perception experiment involving talking heads, Al Moubayed et al. (2010a, 2010b), showed that “when head-nods and eyebrow raise gestures are visualized during prominent syllables, they can aid speech perception. On the other hand, the results do not present a strong evidence on whether visualizing them during non-prominent syllables may hinder or aid perception” (2010b:67). They also stated that “head-nods have been shown to be a stronger cue in the perception of prominence than eyebrows” (2010a:303).

At last, Ferré et al. (2007) studied the relationships between hand/head gestures and morphosyntactic constituents as well as prosodic accents. They found a link between the presence of metaphoric hand gestures and connectors. Connectors are conjunctions or discourse markers used at the beginning of Turn Constructional Units and therefore play a very important role in the information structure of the message. We know as well that metaphoric gestures participate in the topical organization of discourse (Ferré, 2011); it is therefore not surprising to find an association of these gestures with connectors. Ferré et al. (op.cit.) found another link between adverbs and head shakes, which they attributed to the evaluative value of the type of adverb accompanied by a head shake, namely negation particles and degree adverbs like 'super'. In this study, no relationship could however be established between any accent type in all-new utterances or narrow-focus ones and gesture reinforcement, a failure we attributed at the time to the small size of the corpus, although we will see in the present paper that this relationship is far from being established.

The aim of the present study is to enquire deeper into the organization of information structure in spontaneous French. To this purpose, we propose to look at the possible correlations between fronted constructions and prosodic emphasis, as well as the possible production of gesture with both types of constructions.

3. Data and methodology

The original character of the present work is that it is based on a corpus of conversational French, whereas most of the studies mentioned in the previous section based their analysis on elicited speech. The corpus consists of a video recording of spontaneous conversations in French. The whole corpus is 3 hours long, but only a subset was annotated for multimodal phenomena and was considered here. This lasts 1h30 and involves 3 pairs of speakers. The videos were recorded at Aix en Provence by R. Bertrand and B. Priego-Valverde, and the corpus is freely distributed online⁶. Speech was transcribed orthographically into words and phonetically into phonemes using Praat (Boersma & Weenink, 2005), and has been used since in the nationally-funded project OTIM for multimodal treatment. The details of the recording conditions were reported in Blache et al. (2009).

3.1. Marked syntactic structures

Marked syntactic structures were coded by the author using Anvil (Kipp, 2001) on the basis of the written transcription only. Yet, the sound was turned on for ambiguous cases (arising mostly with pronominal dislocation described in section 2.1, which can be confused in writing with repetition of the pronoun due to word search). Below is a list of the operations taken into account

⁶ <http://crdo.up.univ-aix.fr/index.php?lang=en&nbre=12&offset=0>

in this paper, which are the ones treated as marked syntactic constructions in the vast body of literature and which are summarized in Prévost (2003):

- Left dislocation: we noted this type of construction *Full NP dislocation* (FULL NP DS) when the dislocated item contained a noun possibly followed by a relative clause, and *Pronominal dislocation* (PRO DS) when the dislocated item was a pronoun. Because of its frequency in spoken French, pronominal dislocation is sometimes not considered as contributing much to the syntactic marking of utterances, but we decided to consider it nonetheless in this study.
- We also noted the so-called pseudo-cleft (PSEUDO-C) and cleft constructions, knowing that we included structures of the type “It is true that...” among cleft sentences although these structures constitute fossilized phrases.
- We noted presentative constructions (PRES) as well, which are always of the following type in the corpus: “There was/were X that/who...”.
- At last, topicalizations (TOP) are not very frequent in spoken French but were noted nonetheless.

Reliability was assessed by having two inexperienced coders determine the type of highlighting met in a subset of marked constructions. 30 scripted sentences from the corpus, each containing one type of syntactic fronting, were presented to the two coders independently, in the form of an online survey. The subset contained a balanced number of fronting types which were presented in a random order. Agreement between coders was 95 % ($n = 60$).

The distribution of the different types of fronting in the corpus is shown in Table 1. The most common structures are dislocations (with a high number of pronominal dislocations as mentioned above) and presentative constructions, far above (pseudo-)cleft utterances and topicalizations. Among the 6 speakers involved in the study, variation in the number of marked constructions used was found for two participants: one participant produced globally less (-7 %) marked syntactic constructions in proportion to her total amount of speech (to the exclusion of pauses) than the others, whereas a second participant produced slightly more marked syntactic constructions in proportion to his total speaking time (6 % more syntactic highlighting), and especially more cleft constructions and pronominal dislocations than the others. The remaining four participants produced marked syntactic constructions of all types in equal proportions.

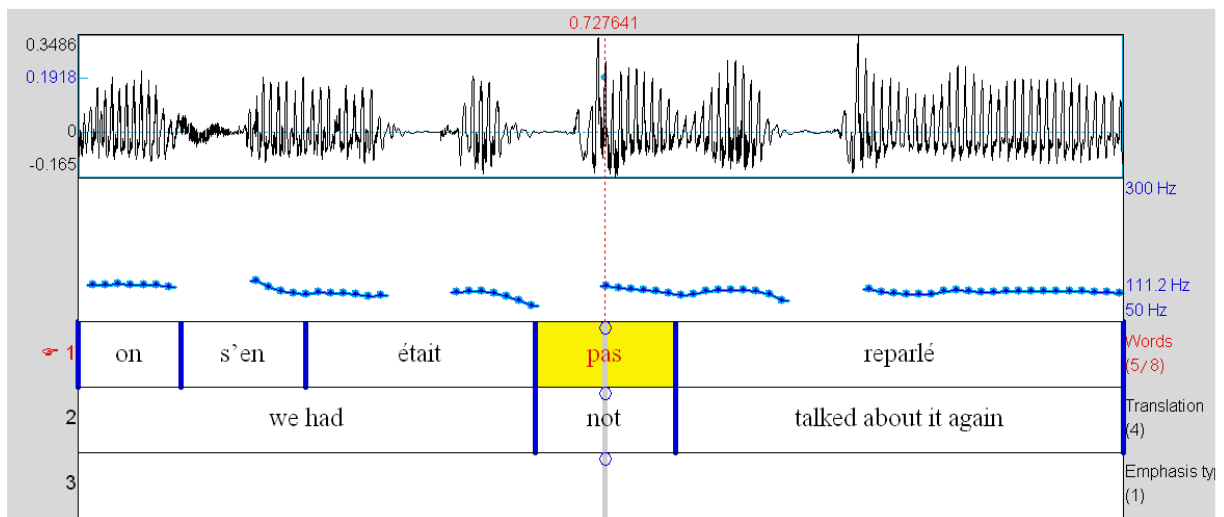
Table 1. Number of fronted constructions studied in the corpus.

Fronting type	Nb
PRO DS	116
FULL NP DS	81
PRES	58
CLEFT	36

TOP	14
PSEUDO-C	9
TOTAL	314

3.2. Prosodic emphasis

Perceived prosodic emphasis involving the parameters described in section 2.2 was noted on the sound only using Praat and then imported in Anvil by the author. What was noted in Praat was the entire Intonational Phrase that carried the emphasis and in Anvil, multilinks were created between the notation of the emphasis and the word that carried it to be able to retrieve the information. The reason for this annotation choice is that there would otherwise be a large difference in granularity between syntactic constructions, prosodic accents and gesture. *Figure 1* below shows an example of two utterances which are almost identical semantically speaking (*We didn't speak about it later*). They were produced by the same speaker and whereas the first utterance shows the unmarked prosodic contour for this type of statement, the second curve shows a strong emphatic stress on “pas” (*not*) characterized by a strong initial plosive and a step up in pitch that is then forming a plateau up to the end of the utterance. The function of the high plateau is pragmatic insofar as it adds an evaluative comment of unexpectedness to the utterance, whereas the strong emphatic stress on “pas” reinforces the negation particle which can then be understood as meaning *not at all* and which is an example of focalization as will be described below. Whereas no emphasis type has been noted for the first utterance which bears no emphatic stress, FOC has been noted below the second utterance on the whole Intonational Phrase.



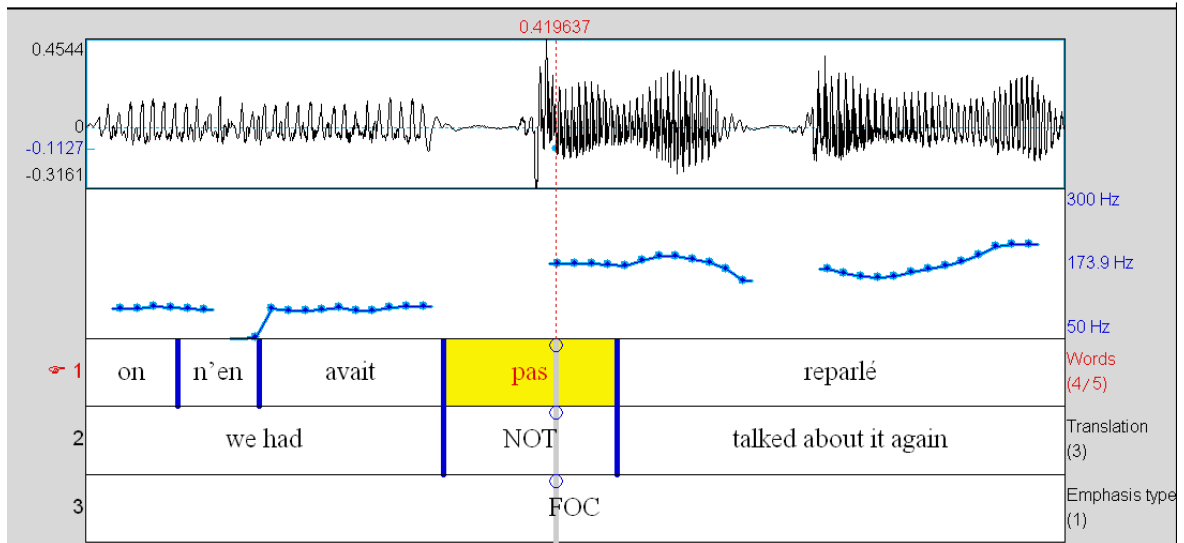


Figure 1: F0 Praat curves in Hz of the utterances “on s’en était pas reparlé” (top) and “on en avait pas reparlé” (bottom) produced by the same speaker.

As shown in section 2.2, prosodic emphasis may assume different functions in discourse organization and an extended typology of the functions listed by Calhoun (2009) was coded as well for each focal accent, based on discourse context, distinguishing the following types:

- Like in Calhoun (2009), when some contrast is explicitly marked in discourse (with phrases like *not... but...*), the emphasis was tagged *Discourse contrast* (DIS CONT).
- When speakers correct a word they uttered or partially uttered just before, the emphasis was tagged *Self-correction* (SC) and when they openly contradict the other participant, the emphasis was tagged *Other contradiction* (OC). These two categories correspond to the ones tagged ‘correction’ and ‘answer’ in Calhoun (op.cit.).
- All the cases of emphasis corresponding to what Calhoun calls ‘subset’ and ‘adverbial focus’ were generally tagged *Focalization* (FOC) as in the sentence “c’était super rapide” (*it was super quick*) in which the speaker emphasizes the first syllable of “super” to mark a high degree of the quality “quick”, without contradicting the addressee in any way.
- At last, *Retrieval* (RET) induces prosodic emphasis when the right word or idea has been found after some hesitation on the part of the speaker. This was not mentioned in Calhoun and would have been tagged as ‘other’ in her scheme.

Table 2 shows the distribution of the different types of prosodic emphasis in the corpus. The data is visibly strongly unbalanced with a much higher proportion of general focalization than other types of emphasis. Although there are slightly more occurrences of discourse contrasts than of other types, they remain in small proportion compared to focalization. Contrary to the use of marked syntactic constructions, no inter-speaker variability was observed for the production by

the 6 participants of prosodic emphasis, the different types of which were used in equal proportion by each speaker in relation to their total speech duration.

Table 2. Number of occurrences in each type of prosodic emphasis.

Prosodic emphases	Nb
FOC	271
DIS CONT	49
RET	37
OC	23
SC	8
<i>TOTAL</i>	388

Reliability was assessed by having two inexperienced coders independently determine the type of highlighting met in a subset of prosodically marked sentences. To this end, in order to avoid confusion due to the use of an unfamiliar computer program, an interactive video was edited. The video showed 30 sequences, each containing one type of focal accent. The different categories of prosodic emphases were represented by an equal number of items all of which were presented in random order. Each sequence of the video played a sound extract from the corpus while showing a still image presenting the scripted version of the sound played. The syllable the coders had to pay attention to in each sound was written down in capital letters in the corresponding script. Coders could replay each sequence as many times as they wanted thanks to the control panel at the bottom of the video. The video automatically stopped after each sound and coders were presented with a still image that showed an interactive quiz. Coders had to select one type of prosodic emphasis in the quiz, after which the video resumed play for the next sound. Results were automatically collected at the end of the video and sent to the author. Agreement between coders was 86.5 % ($n = 60$).

3.3. Co-verbal behavior

The coding scheme and the annotations made for co-verbal behavior by the author have been fully described in Bertrand et al. (2008) and Blache et al. (2009). It is quite complex as it includes a precise description of the gestures produced by speakers, most of which were not used here so we will rather concentrate in this section on the description of the annotations actually used for this particular study.

Co-verbal behavior was annotated by the author using Anvil in a series of different tracks, three of which were actually copied onto the file that included words, prosodic emphasis and syntactic constructions. The three tracks were:

- Eyebrow movements, which has only two values: *raising* and *frowning*.
- Head movements, with the following values: *beat*, *jerk*, *nod*, *shake*, *tilt*, *turn*, *pointing (of the chin)*, *other*. These values encode what is perceived as gesture, not change of posture or direction of the head.

- Hand gestures, with values inspired from McNeill (1992): *beat, deictic, emblem, iconic, metaphoric, butterworth*. The original annotation also included *adaptators* (contact gestures) but these were not taken into account in this study. As noted by McNeill (2005), hand gestures may have several dimensions with an iconic gesture having both a representational and a deictic value for instance, and when this was the case in our corpus, both values were noted in the same tag in Anvil. For the present study, only one of the co-occurring gestures showed two values and both were counted. Each gesture unit considered in this paper started at the beginning of the preparation phase when there was one and ended at the end of the gesture retraction.

Co-verbal behavior (which can also be termed “gesture” in the broad sense of the word) was considered as co-occurring with syntactic constructions or prosodic emphasis when it was produced in overlap with either the detached NP in the case of fronted constructions or with the word that bore the focal accent in the case of prosodic emphasis. This was noted in the Anvil file with multilinks for easier retrieval of the information. Whenever two gesture tags co-occurred with syntactic constructions, both were selected as co-occurring gestures if these were head movements, and only the hand gesture whose stroke coincided with the syntactic construction was selected. This distinction was made because of the fact that whereas head movements are generally quite short, hand gestures are much longer and produced in anticipation of the affiliate as shown in Ferré (2010) and therefore the preparation of the second gesture in a sequence containing two gestures may be produced in overlap with the first affiliate. The total number of gestures annotated in the corpus were 578 eyebrow movements, 1289 hand gestures and 2520 head movements, although we will see that only 471 gestures were produced in combination with marked syntactic or prosodic constructions.

Reliability was assessed for hand gestures only by having two inexperienced coders independently determine the type of gesture met in video clips. Again, in order to avoid confusion due to the use of an unfamiliar computer program as well as unfamiliarity with annotation procedures, an interactive video was edited. The video showed 30 sequences, each containing one type of hand gesture. The different categories of hand gestures were represented by an equal number of items all of which were presented in random order. Each sequence of the video played a clip from the corpus (including sound). Coders could replay each sequence as many times as they wanted thanks to the control panel at the bottom of the video. The video automatically stopped after each clip and coders were presented with a still image that showed an interactive quiz. They had to select one type of gesture in the quiz, after which the video resumed play and showed the next clip. Results were automatically collected at the end of the video and sent to the author. Agreement between coders was 71.5 % ($n = 60$).

4. Results

4.1. Syntactic vs. prosodic emphasis

Out of the 314 fronted syntactic constructions, only 51 co-occur with prosodic emphasis and these are distributed evenly among the different types of prosodic emphasis so that no particular type of focal accent is preferred to accompany the syntactic constructions and vice versa. What should be noted is that among the 51 co-occurrences of co-occurring syntactic and prosodic emphasis, the prosodic emphasis occurs on an element of the rheme of the syntactic construction (the matrix clause) in 36 cases and on an element of the theme (the detached constituent) in only 15 cases. Due to the small number of occurrences, no statistical treatment could be done⁷, but the syntactic constructions with the highest percentage of prosodic emphasis on the theme are the PRES, PSEUDO-C and TOP, as illustrated in Table 3. So this is in agreement with the results obtained on clefted constructions by Rialland et al. (2002) mentioned in section 2.2 and applies as well to other types of fronted syntactic constructions so that no direct link can be established between marked constructions in syntax and marked ones in prosody.

Table 3. Percentage of prosodic emphasis occurring on the theme of the different syntactic constructions.

Syntactic constructions	% of prosodic emphasis on the theme
FULL NP DS	2.46
PRO DS	2.58
CLEFT	2.77
PRES	10.34
PSEUDO-C	11.11
TOP	14.28

4.2. Gesture marking

Syntactic constructions and prosodic emphasis may be accompanied with one or more gestures. Before going into the detail of which gestures are preferred in both cases, it is interesting to have a view of the load of gesture reinforcing. As we tested head and eyebrow movements as well as hand gestures, we were interested in knowing if gestures would combine or not when accompanying syntactic constructions or prosodic emphasis. Table 4 gives the number of occurrences and the percentage of gesture combinations and gesture types alone in the syntactic and the prosodic contexts respectively.

⁷ No statistical treatment has been done under 10 occurrences.

Table 4. Number of occurrences and percentage of gesture combinations and gesture types alone in the syntactic and the prosodic contexts⁸.

Gestures	Syntax	%	Prosody	%
head+eyebrows+hands	1	0.38	13	3.35
head + eyebrows	3	1.14	13	3.35
head + hands	3	1.14	25	6.44
eyebrows + hands	6	2.28	19	4.89
head alone	15	5.70	*53	13.65
eyebrows alone	18	6.84	19	4.89
hands alone	72	27.37	114	29.38
TOTAL	118		*256	

A proportion test revealed that there is a higher proportion of total gesture marking with prosodic emphasis than with syntactic constructions (Prop.Test: X-squared = 27.72, df = 1, p-value < .01) and that is probably due to the fact that there is a higher proportion of head movements alone than for syntactic constructions (Prop.Test: X-squared = 9.77, df = 1, p-value = .001). There was no significant difference in the proportion of eyebrow movements or hand gestures alone between both contexts. For the 6 participants in the study, the proportion of gestures produced together with marked syntactic constructions was different compared to the total number of gestures produced by each speaker in the corpus with quite a lot of inter-speaker variability: while 3 participants produced less gestures with marked syntactic constructions than they do otherwise, 1 produced more and 2 produced the same amount. Variability was less important considering the number of gestures produced by the participants in accompaniment with prosodic emphasis: 5 speakers produced more gestures in this context than they do otherwise and 1 of them produced the same amount. However, the differences observed for the two context types were proportionately distributed amongst head/eyebrow movements as well as hand gestures.

4.3. Eyebrow movements

Table 5 shows the number of eyebrow movements that accompany prosodic emphasis and syntactic constructions (which are not themselves linked to emphasis). The proportion test shows that there are slightly more eyebrow movements in the case of prosodic emphasis than in that of syntactic constructions (Prop.Test: X-squared = 3.70, df = 1, p-value = .05). One may think that this might be due to a higher proportion of eyebrow movements in the case of prosodic focalization (with 39 occurrences out of the 65 total number), but the test is not significant. The proportional distribution of frowns and raises co-occurring with prosodic emphasis and syntactic

⁸ Figures preceded by * are statistically significant, but only the grey-highlighted part of the table was tested with a proportion test. Statistics were run with R (<http://www.r-project.org/>).

constructions is of the same order as there is no significant difference between the two contexts in this respect.

Table 5. Number of eyebrow movements accompanying each type of prosodic emphasis.

Prosody	frowns	raises	TOTAL	Syntax	frowns	raises	TOTAL
DIS CONT	1	7	8	PSEUDO-C	0	0	0
FOC	8	31	39	FULL NP DS	1	2	3
OC	3	5	8	PRES	0	3	3
RET	2	5	7	PRO DS	4	11	15
SC	1	2	3	CLEFT	1	4	5
				TOP	1	2	3
TOTAL	15	50	*65	TOTAL	7	22	29

4.4. Head movements

Table 6 below gives the number of head movements (of which only the most frequent types – beats, nods and shakes – are detailed here) that accompany prosodic emphasis. Details are not given for the head movements that accompanied syntactic constructions as only 22 head movements in total were produced with the 263 syntactic constructions not linked to emphasis, evenly distributed among the different head movements and syntactic construction types, and yielding few occurrences each time.

The test first revealed that the proportion of total head movements that accompany prosodic emphasis is higher than the one for syntactic constructions (Prop.Test: X-squared = 29.34, df = 1, p-value < .01). Among the head movements produced with prosodic emphasis, there is a higher proportion of beats (Prop.Test: X-squared = 28.94, df = 1, p-value < .01) and of shakes (Prop.Test: X-squared = 5.26, df = 1, p-value = .02), than in the rest of the corpus. However, the proportion of head nods is not significantly different from the rest of the corpus.

Table 6. Number of head movements accompanying each type of prosodic emphasis.

Prosody	beats	nods	shakes	other	TOTAL
DIS CONT	3	4	2	3	12
FOC	24	10	20	15	69
OC	1	1	2	3	7
RET	2	0	4	3	9
SC	1	0	1	0	2
TOTAL	*31	15	*29	24	*99

We tested the hypothesis that head movements might be more frequent with discourse contrast (DIS CON) and focalization (FOC), but none of the two types of focal accent revealed significant proportions of head movements, which means once again, that head movements are distributed evenly among the different types of prosodic emphasis.

4.5. Hand gestures

Again, only the most productive categories are detailed in Table 7 and 8 below for co-occurring hand gestures.

Table 7. Number of hand gestures accompanying each type of prosodic emphasis.

Prosody	DIS CONT	FOC	OTHER	TOTAL
beats	3	*26	6	*35
deictic	4	17	1	22
emblems	3	25	3	31
iconics	0	25	3	*28
metaphorics	11	37	10	58
TOTAL	22	132	20	*174

Exactly like with eyebrow and head movements, the statistical test showed that the total proportion of hand gestures produced together with prosodic emphasis is higher than the proportion of hand gestures produced with syntactic constructions (Prop.Test: X-squared = 12.33, df = 1, p-value < .01). As far as gesture type is concerned, the proportion of hand beats is higher with prosodic emphasis than in the rest of the corpus (Prop.Test: X-squared = 3.74, df = 1, p-value = .05) and this is mainly due to the fact that the proportion of beats in contexts of focalization is also much higher than in the rest of the corpus (Prop.Test: X-squared = 89.83, df = 1, p-value < .01), although we cannot say that the total proportion of hand gestures in this particular type of prosodic emphasis is higher than in other emphasis types. The test also revealed that the proportion of iconics is lower in contexts of prosodic emphasis than in the rest of the corpus (Prop.Test: X-squared = 6.90, df = 1, p-value < .01). The proportion of all other gesture types that accompany prosodic emphasis was not significantly different from their proportion in the rest of the corpus.

As far as syntactic fronting is concerned, the only significant result is that the proportion of metaphorics is higher in this context than in the rest of the corpus (Prop.Test: X-squared = 15.86, df = 1, p-value < .01). However, this significant increase of the number of metaphorics is evenly distributed among the different types of syntactic constructions so that none of them in particular can be linked to the increase.

Table 8. Number of hand gestures accompanying each type of syntactic fronting.

SYNTAX	PRO DS	FULL NP DS	PRES	OTHER	TOTAL
beats	1	1	0	1	3
deictic	5	1	1	2	9
emblems	0	4	1	2	7
iconics	2	9	5	2	18
metaphorics	13	11	13	8	*45
TOTAL	21	27	20	14	82

5. Discussion

The results presented in the previous section show that gestural reinforcing is higher in the case of prosodic emphasis than in fronted syntactic constructions and this is mainly due to the fact that prosodic emphasis is reinforced by a higher proportion of head movements than syntactic fronting. We have seen also that two or three gestures may combine to reinforce emphasis in other speech modes (verbal mode for syntactic fronting and vocal mode for prosodic emphasis), but the total number of combined gestural reinforcing (13 for syntax and 70 occurrences for prosody) is much lower than the total number of reinforcing made by a single gesture (105 for syntax and 186 occurrences for prosody). This means that speakers generally do not prefer to express emphasis in several modes at the same time, but use emphatic cues in complementary distribution. Al Moubayed, & Beskow (2010) made a similar observation. It explains as well why syntactic reinforcing is accompanied by prosodic emphasis in only 51 occurrences, among which the theme of the construction is highlighted both in the verbal and the vocal mode in only 15 occurrences.

Whenever there is gestural reinforcing, results show that eyebrow and head movements, as well as hand gestures are produced in greater proportion to reinforce prosodic emphasis than marked syntactic constructions, which means that syntactic and prosodic highlighting do not play the same role in discourse, since they are not reinforced by gesture in the same way.

Considering each individual gesture type, results show that eyebrow raises, although in greater proportion with prosodic emphasis than with syntactic fronting, are not produced in greater proportion in the case of prosodic emphasis than in the rest of the corpus. Results also showed that among the different types of prosodic emphasis, eyebrow raising is distributed evenly. This doesn't mean however that eyebrow raising is not linked with emphasis in any way, but rather that it may be a strong enough marker of emphasis to be used on its own instead of being used in combination with markers of emphasis in other modes.

As far as head movements are concerned, results show that whereas head beats and shakes occur in greater proportion with prosodic emphasis than in the rest of the corpus, this is not the case of head nods. The absence of statistical significance for head nods can be explained in two ways: first, it means that the proportion of head nods is not higher in contexts of prosodic

emphasis than in the rest of the corpus. This is probably due to the polysemy of the gesture which is used both as a means of reinforcing speech and as backchannel, i.e. minimal response made by the co-participant who is not presently holding the speech turn (head shakes used as backchannels are much less frequent than nods when the topic of conversation is not controversial). Second, it is quite relevant that the proportion of head nods in contexts of prosodic emphasis is not lower than the proportion of nods in the rest of the corpus considering the high frequency of nods used as backchannels. It is quite certain that if head nods were examined only when the participant is holding the speech turn (therefore excluding backchannels), then they would probably show a stronger link with prosodic emphasis. The difference with studies mentioned in section 2.3 may also be due to a difference in gesture coding as scholars do not necessarily distinguish nods from beats.

Results for hand gestures show that the proportion of beats is higher in the presence of prosodic emphasis than in the rest of the corpus, which is in agreement with McNeill who states that “a beat may highlight words whose occurrence is relevant for a larger narrative purpose” (1992:41). It is also in agreement with the aim of prosodic focalization and this explains why beats are more frequently found with this type of prosodic emphasis. We also found that iconics are produced in a lower proportion with prosodic emphasis than in the rest of the corpus. Iconics are described by McNeill as gestures which “present images of concrete entities and/or actions” (1992:39). Since prosodic emphasis (and especially focalization) is very frequently applied to degree adverbs, it is not surprising that iconics do not co-occur with prosodic emphasis. We expected iconics to rather co-occur with syntactic constructions which aim at highlighting an item of discourse (mostly NPs). However, we noticed that among the 18 iconics that reinforce syntactic constructions, 5 co-occur with animate agents or patients (when the gesture anticipates the predicate in the main clause, therefore co-occurring with the dislocated element), whereas 13 co-occur with inanimate agents or patients, with which they are in a relation of lexical affiliation. Since a large number of the NPs highlighted by the syntactic constructions are either proper nouns or pronouns, it is therefore not surprising that we didn’t find a higher proportion of iconics with fronted constructions than in the rest of the corpus. Instead, the proportion of metaphorics is higher. This is quite consistent with the role of syntactic fronting: these constructions are mostly used to (re)introduce some item in the discourse of the speaker and metaphorics precisely demarcate the different units in discourse organization. What syntactic fronting does in discourse is probably not so much linked with semantic presentation than with information structure.

The fact that we found a distinction in the type of gesture which accompany marked syntactic and prosodic constructions as opposed to unmarked ones gives us some insight into the conceptualization and verbalization of the message. If gestures were planned before grammatical and phonological encoding as postulated by de Ruiter’s Sketch Model presented at the beginning of section 2.3, then one would expect to find no difference in between marked and unmarked constructions. Therefore, some form of interaction is necessary between the speech formulator and the gesture generator to allow the expression of concomitant highlighting of some elements

of speech. The highest proportion of co-speech gestures that accompany prosodic highlighting being concomitant with focalization, it would sound difficult to posit that gesture essentially plays a role in lexical retrieval since focalization is a process that is at the exact opposite of word search. For this reason, we are not in favor of Krauss' Lexical Gesture Process Model although he proposes that gestures are planned during speech formulation and they play a role in lexical retrieval in the case of word search. The Growth Point Theory proposed by McNeill supposes that gesture and speech constitute a single integrated system. This model also seems to be ruled out in a theory of markedness in which a language has a choice between two types of marked constructions and this choice has an impact on the type of gesture selected when desired. At last, Kita & Özyürek's Interface Model seems to correspond more closely to the data explored in the present study since although in this model, gesture is planned at the conceptualization stage of the message, there is an interaction between the planner and the modules involved in speech production at later stages as well. Yet, in the conception of the models evoked in this section, the gestures taken into consideration were mostly representational gestures (iconics and some metaphors illustrating abstract concepts). In this study, the gestures produced together with speech were of a different nature: head and hand beats, head shakes and metaphors expressing discourse organization rather than illustrating abstract concepts. It would be possible that there exist different "gesture planners": a planner for representational gestures which would be consulted at the conceptualization stage of the message and other planners for "grammatical and prosodic" gestures. This could explain why the latter type of gesture can be superimposed on the former type.

Besides, among the different models presented above, none mentions a pragmatic module which is lacking in our opinion. There seems to be agreement on the fact that what speaker says and what speaker thinks are congruent, although we know for certain that the speaker's message does not necessarily match their thoughts and is constrained by other – pragmatic – factors. A pragmatic module is necessary as well to account for the fact that discourses are not made of unconnected utterances: any utterance in discourse is connected to a previous one and some gestures must be analyzed in connection to previous gestures too. At last, the addition of a pragmatic module to the existing models would also account for interactive gestures used in face-to-face interaction (backchannels and turn-taking or turn-allocation gestures).

Starting from the conceptualization of a message, the model we propose in *Figure 2* contains a lexicon module, which associates a mental word lexicon with representational gestures or features of representational gestures as well as emblems (conventional gestures). In the pragmatic module, the speaker decides on the placement of the message in the overall interaction and its possible link with previous utterances. At this stage, they may select interactive gestures as described above, but also pointing gestures with an anaphoric function. The pragmatic module also serves in the possible selection of a gesture or gesture features made together with a previous utterance (gesture repetition) linked to the actual concept. The grammar module allows the selection of discourse organization gestures such as the metaphors presented in this study, but also enables speakers to align representational gestures with the verbal grammatical encoding of the message as is presented in Kita & Özyürek (2003). At last, the phonology module enables the

encoding of the message into sounds, therefore selecting the necessary articulatory gestures, as well as the encoding of a prosodic pattern with the possible selection of prosodic gestures like beats. At this stage, if no other gesture was selected in the previous modules, a simple hand beat is possible as a prosodic gesture. If a hand gesture was selected in the previous modules, then a hand beat may be superimposed on this gesture, the speaker's hands staying in the configuration of the representational gesture. If the gesture features (like particular movement types) do not allow the superimposition of a beat, another articulator like the head may be selected to perform the beat, which does not mean that a head movement may not be selected in the absence of hand gesture.

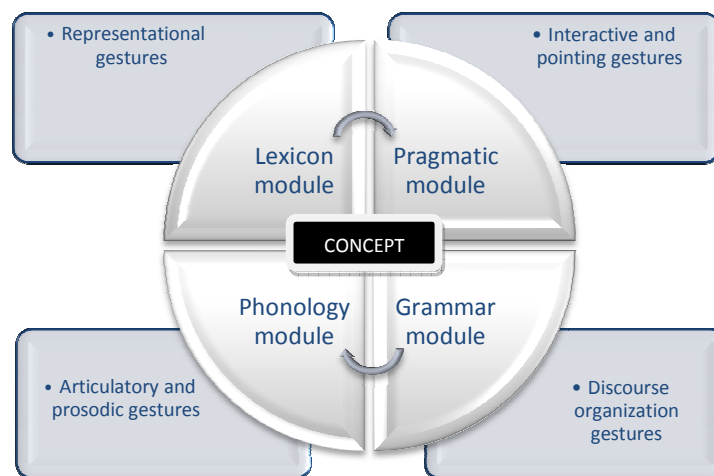


Figure 2: The modular model.

The modular model is presented as cyclic rather than as a strict hierarchy to account for the fact that speakers may for instance begin the phonological encoding of a concept right after the word has been selected in the lexicon module, and then realize that the grammatical encoding is wrong which generates a new grammatical start from the speaker.

6. Conclusion

Drawing on a corpus of conversational French, this paper has shown the interaction between three possible ways of highlighting discourse elements: this can be done in syntax with the use of fronted constructions such as dislocation, topicalization, (pseudo-)cleft and presentative constructions in the verbal mode. It can also be achieved with prosodic emphasis in the vocal mode and gesture reinforcing in the visual mode. We have shown that on the whole, the three types of highlighting are complementary and are very rarely used in conjunction, and agree with Lacheret (2003:159) who states that “what is not encoded in syntax is encoded in prosody”. When several types of highlighting co-occur, then speakers show a strong preference for the double-marking of prosodic emphasis and gesture reinforcement, whereas syntactic highlighting

is generally not associated with any other type of marking apart from reinforcement by metaphoric hand gestures. With the exception of hand beats which are regularly associated with prosodic focalization, however, the different types of prosodic emphasis and syntactic constructions did not show any preference for any accompanying gesture. This may be due to the fact that the subdivision into categories reduces the number of occurrences per category and many of them could not be tested statistically. It is then an encouragement to increase the amount of annotated corpus, in the hope that a greater number of occurrences would enable us to find patterns which we could not possibly find in this study.

The work presented here also gives an insight into the current theories in prosody and gesture studies. French has been described by some researchers in prosody as differing from other languages like English or German in that it uses “a single strategy to signal focus, namely phrasing” (Féry, 2001:153). In the same paper, the author later states that “French makes use of prosodic phrasing to a much greater extent than other languages, because it is the only device to realize focus domains and other discourse structural entities that it has at its disposal” (op. cit., pp 154-155). We can agree on the fact that French offers more possibilities of syntactic fronting than other languages like English — and not only with the marked syntactic constructions presented in this paper, but also of adverbial, modal and other clause types which are often produced in a sequence in sentence-initial position, especially at the introduction of a new topic or sub-topic⁹. Yet, the present study showed that the (re)-phrasing of discourse entities as Intonational Phrases is not the only device used by the French to signal focus since focus may also be expressed at Accentual Phrase level with prosodic emphasis and this is not different from what occurs in languages like English. Actually, the speakers in our corpus of spontaneous French produced more occurrences of focal accents than of syntactic fronting (386 vs 314 occurrences respectively). In view of the gesture types that may accompany either syntactic fronting (metaphoric hand gestures) or prosodic emphasis (preferentially beats), it sounds more reasonable to say that whereas prosodic emphasis marks narrow focus (as opposed to broad focus expressed by final nuclear stress), syntactic fronting does not mark focus but rather plays a pragmatic role in discourse structure. For fronted constructions to mark focus, prosodic highlighting and/or gesture reinforcing must be present as well, although this is not the preferred way of highlighting discourse elements in spoken French. The fact that syntactic fronting and prosodic emphasis have often been associated in previous research on French is largely due to the experimental data eliciting contexts in which both were required. This does not mean that speakers favor this type of combination in ordinary speech.

In terms of the theories currently proposed in gesture studies, we have seen in the discussion section of this paper that since marking is possible either at the grammatical level or at the

⁹ Sentences of the type “moi, ma sœur, son ami, je ne l’ai pas encore rencontré” (lit. *me, my sister, her boyfriend, I haven’t met him yet = I haven’t met my sister’s boyfriend yet*) or, “nous, tu sais, au cinéma, hier, on a mangé une glace” (lit. *us, you know, at the movies, yesterday, we had an ice-cream = we had an ice-cream at the movies yesterday, you know*) are not at all unusual in spoken French, although their use is restricted to particular contexts which would be too long to describe here.

prosodic level in French, and since both types of marking entail the use of different gesture types, it means that the gestures accompanying the types of utterance considered here cannot be decided on by the speaker prior to the grammatical encoding. Kita & Ozyürek's Interface Model (2003) seems to correspond more closely to the data presented in this paper since it proposes that linguistic factors influence gesture production. They did not enquire into prosodic factors but the results obtained here show that prosodic factors also influence the choice of gesture production. Our opinion is that whereas representational gestures can be selected at a very early stage in the message production, other gesture types — discourse organization, beats, interactive gestures (Bavelas, 2000) — are selected at later stages and this explains why a hand beat can be superimposed on a representational gesture. The modular model we propose is based on previous models and associates different gesture types with different modules of speech production in a cyclic way instead of a strict hierarchy in order to account for production errors. The lexicon module in the model associates both words and representational gestures with a concept. The pragmatic module enables speakers to place the utterance in the overall interaction, but also to tie their concept to previous utterances in the turn or across turns. The grammar module provides the grammatical packaging of the message and is the site for the production of discourse organization gestures as well as the alignment for representational gestures and grammar constraints in a language. The phonology module enables the aural encoding of the message, both in terms of phonemes at the segmental level and prosody at the suprasegmental level. Prosodic gestures like beats may be selected at this stage, either in the form of a superimposition of the beat on another gesture, or in the selection of an independent prosodic gesture.

Interestingly, the study has drawn up a wells of questions, that could be answered in further research. For instance, during the annotation process, we noticed that the speaker's gaze tended to be oriented towards the co-participant while he/she was producing prosodic emphasis. We could not check this systematically, but it would be interesting to know more of gaze direction during the production of syntactic fronting or prosodic emphasis. It would also be quite interesting to enquire into some features of the hand gestures that reinforce syntactic and prosodic emphasis: are gestures produced with particular amplitude, velocity or hand shape for instance? The current corpus allows this type of analysis, but once again, more data may be needed in syntax and prosody for any pattern to emerge. Also to be enquired into is the fact that for instance, pronominal dislocation (of the type *me, I...*) – being extremely frequent in spoken French – does not emphasize the dislocated item to the same degree as other types of syntactic fronting. We noticed that for pronominal dislocation to acquire a real discourse contrast value in spoken French, it had to be accompanied with prosodic emphasis on the dislocated pronoun, which it sometimes does. With more occurrences, it would be possible to better understand the role of these syntactic constructions in speech, which is necessarily different from what we find in written French. And indeed, few studies have been conducted on spoken French in this respect: in a qualitative study, Stark (1999) presented an analysis of syntactic fronting in spoken dialogues, from a pragmatic viewpoint but to our knowledge, no such study has been carried out in a

multimodal perspective in French. Yet, the functioning and weight of reinforcing gestures which are not linked in any way to any other emphatic construction still have to be enquired into.

Acknowledgements

This research is supported by the French National Research Agency (Project number: ANR BLAN08-2_349062) and is based on a corpus and transcriptions made by various team members beside the author of the current paper, whom we would like to thank here. The OTIM project is referenced on the following webpage: <http://aune.lpl.univ-aix.fr/~otim/>. Many thanks as well to the anonymous reviewers for their useful comments on previous versions of the paper.

References

- Al Moubayed, S., Beskow, J., Granström, B., 2010a. Auditory visual prominence. From intelligibility to behavior. *Journal of Multimodal User Interfaces* 3, 299-309.
- Al Moubayed, S., Beskow, J., Granström, B., House, D., 2010b. Audio-Visual Prosody: Perception, Detection, and Synthesis of Prominence. In: Esposito, A.M. (Ed.), *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues*. Springer Verlag, Heidelberg, 55-71.
- Beavin Bavelas, J., 2000. Nonverbal Aspects of Fluency. In: Riegenbach, H. (Ed.), *Perspectives on Fluency*. University of Michigan Press, Ann Arbor, 91-101.
- Bertrand, R., et al., 2008. Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues* 49, 105-133.
- Blache, P., et al., 2009. Creating and Exploiting Multimodal Annotated Corpora: The ToMA Project. In: M. Kipp, et al. (Eds.), *Multimodal Corpora. From Models of Natural Interaction to Systems and Applications*. Springer-Verlag, Berlin, Heidelberg, 38-53.
- Boersma, P., and Weenink, D., 2005. Praat: doing phonetics by computer (release 4.3.14), <http://www.praat.org/>.
- Bolinger, D., 1986. Intonation and Gesture. *American Speech* 58(2), 156-174.
- Büring, D., 2007. Intonation, Semantics and Information Structure. In: G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces*. OUP, Oxford, 445-474.
- Calhoun, S., 2009. What makes a word contrastive? Prosodic, Semantic and Pragmatic Perspectives. In: D. Barth-Weingarten, et al. (Eds.), *Where prosody meets pragmatics: Research at the interface (Studies in Pragmatics 8)*. Emerald, Bingley, 53-78.
- Combettes, B., 1999. Thématization et topicalisation: leur rôle respectif dans l'évolution du français. In: C. Guimier (Ed.), *La thématization dans les langues*. Peter Lang, Berlin, 231-245.
- Creissels, D., 2004. Chapitre 17 : Topicalisation et focalisation. In: D. Creissels (Ed.), *Cours de syntaxe générale*. PUF, Paris, 1-16.
- Dahan, D., Bernard, J.-M., 1994. Production et perception des phénomènes prosodiques dans la parole : Comment identifie-t-on l'insistance malgré la variabilité interlocuteurs ? *Journal de Physique IV, Colloque C5, Supplément au Journal de Physique III, Vol. 4*, 501-504.
- de Ruijter, J.P., 2000. The production of gesture and speech. In: D. McNeill (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, 284-311.

- Dohen, M., Loevenbruck, H., 2004. Pre-focal Rephrasing, Focal Enhancement and Post-focal Deaccentuation in French. In *Proceedings of ICSLP 2004*, Jeju Island, Korea, 1313-1316.
- Dohen, M., Loevenbruck, H., 2005. Audiovisual Production and Perception of Contrastive Focus in French: a Multispeaker Study. In *Proceedings of Interspeech/Eurospeech*, Lisbon, Portugal, 2413-2416.
- Dohen, M., Loevenbruck, H., Cathiard, M.-A., Schwartz, J.-L., 2004. Visual perception of contrastive focus in reiterant French speech. *Speech Communication* 44, 155-172.
- Dooley, R.A., Levinson, S.H., 2000. *Analyzing Discourse: A Manual of Basic Concepts*. Sil International, Dallas.
- Ferré, G., et al., 2007. Intensive Gestures in French and their Multimodal Correlates. In *Proceedings of Interspeech 2007*, Antwerp, Belgium, [On CD-rom].
- Ferré, G., 2010. Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. In *Proceedings of LREC: Workshop on Multimodal Corpora*, Valetta, Malta, 86-91.
- Ferré, G., 2011. Functions of Three Open-Palm Hand Gestures. *Multimodal Communication* 1(1), 5-20.
- Féry, C., 2001. Focus and Phrasing in French. In: Féry, C., Sternefeld, W. (Eds.), *Audiatur Vox Sapientiae. A Festschrift for Arnim von Stechow*. Akademie-Verlag, Berlin, 153-181.
- Féry, C., 2003. Gradient prosodic correlates of phrasing in French. In: M. Trudel & M. Selig (Eds.), *Nouveaux départs en phonologie*. Niemeyer, Tübingen, 161-182.
- Gregory, M.L., Michaelis, L.A., 2001. Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics* 33, 1665-1706.
- 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 & Language* 48, 16-32.
- Kita, S., et al., 2007. Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes* 22(8), 1212-1236.
- Krauss, R.M., et al., 2000. Lexical gestures and lexical access: a process model. In: D. McNeill (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, 261-283.
- Krahmer, E., et al., 2002. Pitch, Eyebrows and the Perception of Focus. In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 443-446.
- Halliday, M.A.K., 1967. Notes on Transitivity and Theme in English, Part II. *Journal of Linguistics* 3:199-244.
- Herment-Dujardin, S., Hirst, D., 2002. Emphasis in English: a Perceptual Study Based on Modified Synthetic Speech. In: Bel, B., Marlien, I. (Eds.), *Speech Prosody 2002*. Laboratoire Parole et Langage, Aix-en-Provence, 379-382.
- Katz, J., Selkirk, E., 2011, Contrastive Focus vs. Discourse-New: Evidence from Phonetic Prominence in English, *Language* 87(4), 771-816.
- Kipp, M., 2001. Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech)*, Aalborg, Denmark, 1367-1370.

- Kohler, K.J., 2006. What is Emphasis and How is it Coded? In *Proceedings of Speech Prosody*, Dresden, Germany, 748-751.
- Krahmer, E., Swerts, M., 2001. On the alleged existence of contrastive accents. *Speech Communication* 34, 391-405.
- Krahmer, E., Swerts, M., 2007. The Effects of Visual Beats on Prosodic Prominence: Acoustic Analyses, Auditory Perception and Visual Perception. *Journal of Memory & Language* 57(3), 396-414.
- Krauss, R.M., et al., 2000. Lexical gestures and lexical access: a process model. In: D. McNeill (Ed.), *Language and Gesture*. CUP, Cambridge, 261-283.
- 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*. Oxford University Press, Oxford, 93-116.
- Lacheret-Dujour, A., 2003. Focalisation et circonstance. Que nous dit la prosodie du français parlé ? *Bulletin de la Société de Linguistique de Paris* XIII, 137-160.
- Lambrecht, K., 1994. *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. CUP, Cambridge.
- McNeill, D., 1992. *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press, Chicago and London.
- McNeill, D., 2005. *Gesture and Thought*. The University of Chicago Press, Chicago, London.
- Prévost, S., 2003. Détachement et topicalisation : des niveaux d'analyse différents. *Cahiers de Praxématique* 40, 97-126.
- Selkirk, E., 1978. On Prosodic Structure and its Relation to Syntactic Structure. In: Fretheim, T. (Ed.), *Nordic Prosody II*. Tapir, Trondheim, 111-140.
- Stark, E., 1999. Antéposition et marquage du thème (topic) dans les dialogues spontanés. In: Guimier, C. (Ed.), *La thématization dans les langues*. Peter Lang, Berlin, 337-358.
- Swerts, M., Krahmer, E., 2008. Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics* 36, 219-238.
- Wilmes, K.A., 2009. *Hands in Focus: Focus Marking by Speech Accompanying Gestures*. University of Osnabrück, Osnabrück. BA thesis.