

Intensive Gestures in French and their Multimodal Correlates

Gaëlle Ferré, Roxane Bertrand, Philippe Blache, Robert Espesser & Stéphane Rauzy.¹

¹ Aix-Marseille Université – CNRS (UMR 6057)

Laboratoire Parole et Langage

29, avenue Robert Schuman

13621 Aix en Provence – France

gaelleferre@yahoo.fr, {Roxane.Bertrand, Philippe.Blache, Robert.Espesser, Stephane.Rauzy}@lpl.univ-aix.fr

Abstract

This paper relates a pilot study on intensive gestures in French – e.g. gestures which accompany speech and participate in the highlighting of some discourse elements which the paper means to determine. The study is based on spontaneous French informal conversation and the intensive gestures correlates we looked at pertained to the morphological, prosodic and gestural dimensions.

Index Terms: Reinforcement, intensive gestures, prosody, morphology, hand gestures

1. Introduction

Works on Conversational Agents [1] and Talking Heads [2] have recently known a considerable development, so that present focus is set on the constant improvement of human-machine communication. It has been noticed that when certain types of gesture accompany speech, communication is more efficient. For instance [2] and [3] found evidence of a better perception of focus when intonational focus is accompanied by eyebrow raising. And this is indeed consistent with [4] and [5] on natural speech which established a link between F0 variations (as well as the expression of accented segments) and eyebrow raising. [6] concentrated on gaze and showed that when a speaker gazes at his own hand gesture, then his partner in interaction pays more attention to the gesture. The function of gaze in this case could be to signal the importance of the hand gesture it highlights [7] and this doesn't seem to be very different from the function of eyebrow raising when it accompanies intonational focus. Very few has been said however of other head movements or facial expressions which seem to us to play exactly the same role. A systematic study of such gestures in natural speech, analyzed from the perspective of the gestures themselves rather than from the perspective of verbal content or intonation, could be of use in the improvement of Conversational Agents' performances.

As stated in [8] (p. 105) the gestures or expressions under study "amplify the presence of an already existing information in order to reinforce its impact on the user" (our trans.). The question is: What kind of information? What is amplified by the presence of an intensive gesture?

This paper is based on a corpus of spontaneous French in which all the head movements, gaze directions and facial expressions which serve as reinforcing cues have been annotated and related to hand gestures and speech in its morphological, discursive and prosodic dimensions.

2. Corpus and methodology

The *Corpus of Interactional Data* ([9], <http://crdo.up.univ-aix.fr/corpus.php?langue=fr>) is an audio-video recording of 8 hours of spontaneous spoken French involving 8 pairs of speakers of the same gender (1 hour of recording per session). The corpus has been manually fully transcribed in enriched orthography. From this first transcription, we automatically annotated the 8 hours in phonemes and aligned phonemes with the speech signal. We then recovered larger units such as words, which have been used for the automatic annotation of morphological categories (with LPL-Suite [10]).

At the prosodic level, we annotated (using Praat [11]) prosodic units, pitch contours and the different types of accents. For this particular study, we focused on accents which will be described in more detail below.

These annotations have been done separately on the audio files and constitute the basis of our present project which consists in the annotation and processing of the corpus from a multimodal perspective. The annotation of the gestures made by the participants is being done manually using ANVIL [12] as shown in Figure 1.

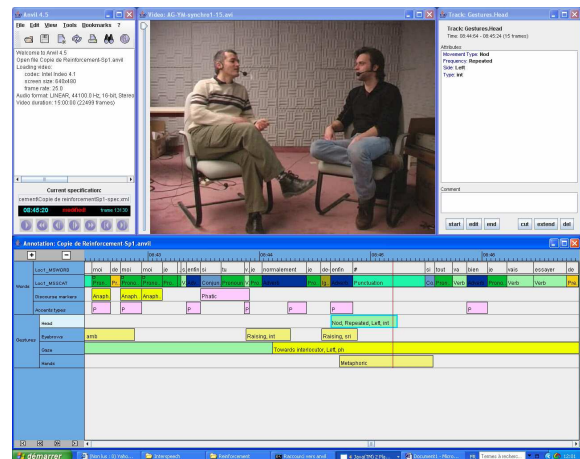


Figure 1: Anvil annotation for Speaker 1 (on the left).

The advantage of this annotation tool is that it allows the importation of Praat annotation tracks. Its XML output structure also makes it easy to import annotations made with other tools (as was the case of our morphological annotations) provided they are also in XML. It then becomes easier to visualize the different dimensions and to query the output file. The last reason is that automatic annotation needs to be corrected manually (as was the case for the morphological

dimension) and using ANVIL as an editor renders this task easier.

2.1. Gestures

2.1.1. Annotation of intensive gestures

For the annotation of intensive gestures, we selected 15 min in an interaction between 2 male speakers, this annotation process being particularly long. We proceeded in two steps.

First, we annotated eyebrow and head movements as well as gaze direction in terms of form without assigning any function to the movements.

The functions were annotated in a second step by two annotators. Among these functions were turn initial/final, backchannel, answer, phatic and intensive. We ended up with the following count and distribution for gestures perceived as intensive (Table 1). The first thing to be noticed is that a similar proportion of eyebrow raises and head shakes are perceived as intensive in the sense that they don't add meaning but rather highlight some parts of discourse without any participant feedback expectancy on the part of the speaker, in which case the gestures would have been coded as phatics.

Table 1. *Intensive gesture types and their distribution in the 15 min of recording.*

Gesture type		Nb	%
Eyebrows	raising	47	28.6%
	frowning	11	6.7%
Gaze		15	9.1%
Head	shake	46	28%
	nod	13	7.9%
	jerk	7	4.2%
	tilt	6	3.6%
	turn	1	0.6%
	waggle	1	0.6%
	other	17	10.3%
	Total		164

2.1.2. Annotation of hand gestures

We coded hand gestures using the McNeill typology [13], also described in [14]. The categories we used were beats, deictics, iconics, emblems and metaphors. For the time being, we did not go into further details such as gesture phases although we plan to do it in the future.

2.2. Morphology

At the morphological level, although the analyzer goes into many details in the annotation of the categories, we used a simplified version for the purpose of this study, with the following categories: Adjectives, Conjunctions, Determiners, Interjections, Nouns, Pronouns, Adverbs, Prepositions, Auxiliaries, Verbs, Ignored (morphemes which the analyzer could not decide on) and Punctuation (end of TCU¹; this category is used for the syntactic annotation).

¹ The CID is also annotated at the conversational level in TCUs (turn-constructive units) which are defined in the Conversational Analysis framework as "the smallest interactionally relevant complete linguistic unit" [15].

2.3. Prosody

Recent models of French accentuation ([16] among others) posit the existence of two types of accents in French: a primary accent (P) occurring on the final syllable of a full word (with a demarcative function) and a secondary (word initial) accent (S) occurring at the beginning of the word (with a rhythmic function). P and S accents are the complementary components of the metrical organization in French. Other types of accents, namely semantico-pragmatic accents, can also be found. Here we annotated as F (focalization) all the emphatic accents without distinguishing, in a first step, between contrastive and focal accent in the case of intensification for example.

2.4. Discourse markers

During the annotation for previous work, we had noted some discursive functions of words which we retained for this particular study. The categories we annotated are: connectors (words uttered by the speaker to link two TCUs, most of the time conjunctions or interjections), punctuators (words or phrases used at the end of the TCU), verbal phatics, anaphoras and cataphoras. Although we had no a priori hypotheses concerning these categories, we tested them with intensive gestures and the results are interesting all the same, so this is why we mention them here.

3. Results

3.1. Morphology

From earlier work on reinforcement in English [17] in which the methodology consisted in starting from the speech signal, we had noted that adverbs were often intensified by eyebrow raising or head nods, so we wanted to test this hypothesis first, from a gestural perspective this time.

We applied a proportion test to find out whether adverbs are associated with intensive gestures oftener than any other morphological category. The test was highly significant ($\chi^2=54.50$, $p<<0.001$ for speaker 1, $\chi^2=33.23$, $p<<0.001$ for speaker 2; see Table 2 below for the number of occurrences and the distribution of morphological categories of words associated with an intensive gesture²). If we look in more detail at the type of adverb associated with an intensive gesture, what comes out is that there is a much higher proportion (33.5%) of intensive gestures accompanying lexical adverbs such as "très" (very), "super" (super in adverb position), "normalement" (normally) than the negation particle "pas" (*not*) counted as an adverb (11.4%). Gestures to be met with the negation particle are rather head shakes, whereas eyebrow raising most of the time accompanies lexical adverbs. Yet, since it was shown in Table 1 that the proportion of eyebrow raises and head shakes are almost similar, there is not a one-to-one correspondence between negation and shake or between lexical adverbs and eyebrow raising.

Grouping the occurrences for the two speakers, we also tested the intensive gestures associated with verbs and interjections with a proportion test. The results were indeed significant but

² In order to avoid generalizing from speaker specific tendencies, we made sure to check, for each proportion test, that the result showed the same tendency for the two speakers.

the proportion of verbs and interjections is lower this time with an intensive gesture, e.g. verbs and interjections are less often associated with an intensive gesture than any other morphological category (verbs: $\text{Chi}2=7.98$, $p=0.004$; interjections: $\text{Chi}2=6.99$, $p=0.008$). We did not test the other categories for which we did not have enough occurrences. They would have to be tested on a larger corpus.

Table 2. *Morphological categories of words associated with an intensive gesture.*

Category	Nb	%
adverb	71	44.90%
adjective	9	5.60%
conjunction	4	2.50%
determiner	4	2.50%
noun	16	10.10%
preposition	1	0.63%
pronoun	7	4.40%
punctuation	1	0.63%
interjection	23	14.50%
verb	22	13.90%

3.2. Discourse markers

As explained in section 2.4, we annotated discourse markers without any preconceived idea of the results. It appeared in our query that a high number of connectors were in fact associated with an intensive gesture and we applied a proportion test as well. The result was that connectors are oftener associated with an intensive gesture than any other discourse marker ($\text{Chi}2=6.49$, $p=0.01$ for both speakers). When one looks at the type of intensive gesture associated with connectors however, one can find no one-to-one mapping since all types of eyebrow or head movements are associated with this marker. The association of intensive gestures and connectors is however interesting regarding the syntactic structure of the utterances: without having tested the syntactic units which are not yet ready, and since connectors are uttered at the beginning of the TCU, we can predict that intensive gestures are met mostly on the first NP of the utterance, in subject position. This hypothesis will hopefully be verified in future work.

3.3. Hand gestures

Due to the small number of occurrences of deictics and emblems, we did not test these hand gestures statistically (see Table 3). We however applied a proportion test to metaphorics, iconics and beats. The prediction here was that since beats highlight some steps in the discourse structure [13], they would start off favourites — although the proportion of metaphorics highlighted by an intensive gesture is higher (there is a much higher proportion of metaphorics whatever the context in our video clip).

Table 3. *Number of occurrences and percentage of hand gestures reinforced by an intensive gesture.*

Hand gestures	Nb	%
beats	11	11.40%
deictics	3	3.10%
emblems	3	3.10%
iconics	12	12.50%
metaphorics	67	69.80%

The proportion test showed that in fact there are many more metaphorics associated with an intensive gesture than any other kind of hand gesture ($\text{Chi}2=7.11$, $p=0.007$). Tests for beats and iconics were not significant (Beats: $\text{Chi}2=0.54$, $p=0.45$; Iconics: $\text{Chi}2=0.079$; $p=0.77$).

As for the correspondence between intensive gestures and hand movements, one finds a one-to-one mapping between intensive gaze and hand movements although the latter may be reinforced by other types of intensive gestures as well.

3.4. Prosody

Although it has often been said in the literature that eyebrow movements are associated with stress, we found that stress may be highlighted by any type of head and eyebrow movements. Moreover, when one looks at the distribution of stress types (P, S, and F) when accompanied by an intensive gesture, one finds that no particular type of intensive gesture is in a one-to-one correspondence with any type of stress. Yet as shown in Table 4, intensive gestures seem to be highly associated with primary stress. But in fact, primary stress wasn't significant ($\text{Chi}2=2.27$, $p=0.13$) and this was kind of a surprise. We cannot say that the proportion of primary stresses associated with an intensive gesture is higher than the proportion of primary stresses in the whole extract. Secondary stress wasn't significant either ($\text{Chi}2=0.05$, $p=0.8$). We then applied a proportion test to focal accents for both speakers. The test wasn't significant ($\text{Chi}2=2.02$, $p=0.15$). Focal accents are not associated with intensive gestures oftener than other types of accents.

Table 4. *Accent types on words associated with an intensive gesture.*

Stress type	Nb	%
P	73	64.60%
S	15	13.20%
F	25	22.10%

4. Discussion

As a summary, the principal results showed a link between gestural reinforcement and connectors, metaphorics and adverbs. No link could be established between any accent type and reinforcement.

More precisely connectors are more liable to be associated with an intensive gesture which can be interpreted as a contribution "to discursive planning" as was stated in [5]. Indeed, connectors are produced at the very beginning of the TCU. This is also consistent with the type of hand gesture they accompany since metaphorics are representational gestures which are produced to express concepts [14]. This first result is important since to our knowledge no particular study has yet shown that this type of hand gesture has more chance to be highlighted in discourse than other gestural phrases. Intensive gestures also regularly accompany words such as temporal conjunctions which show the links between the different ideas expressed.

Then, as we expected intensive gestures are found in the proximity of adverbs. Apart from the negation particle which is most of the time associated with a head shake¹, the adverbs

¹ In the cases where the negation particle is accompanied by eyebrow raising, we will have to check whether there also is a focal accent on the particle, which would show that the

intensified are degree adverbs such as "super", etc. It appears that when the speaker introduces a degree adverb, he doesn't want it to pass unnoticed and intensifies it with a gesture. As stated in [5], it is a way to say "Listen, what I'm saying is important". It would be interesting to see if an intensified degree adverb is perceived with a higher degree than when it is only expressed in words.

For accentuation and more specially focalization, we cannot confirm previous results establishing a link between prosody and gestures. The difference between the quoted studies and our analysis is that we started from the gesture itself, to which we attributed a particular function, whereas previous studies rather started from intonational variations to which any type of gesture may be associated whatever its function. We are then induced to think that when there is an intensive gesture, it may be sufficient to highlight parts of discourse since it plays the same role as intonational devices (such as stress). The "Listen, what I'm saying is important" effect can be obtained either with intonational or gestural devices or both.

5. Conclusion

In this pilot study, we analyzed 15 minutes of a video corpus from a multimodal perspective, concentrating on intensive gestures in spontaneous French. The dimensions we looked at in relation to these gestures were the morphological categories, the stress types, the discursive markers and the hand gestures. All the linguistic dimensions were either annotated directly in ANVIL or annotated with other tools and then imported in ANVIL for edition and query.

The study showed that intensive gestures are more liable to accompany degree adverbs and negation particles, metaphoric gestures and connectors. Considering this, we concluded that the gestures we looked at — which were head and eyebrow movements as well as gaze direction — rather played a discursive role of intensification, especially since none of these gestures were associated with any specific stress type. The study also shows that intensive gestures are not redundant in their expression of emphasis: the segments they highlight do not fall under intonational focalization, for instance, with which they are in complementary distribution.

Our further steps will be to first check these results on a larger video file involving more speakers. We also want to add the full syntactic annotation to our data file. Since connectors were much preferred in the environment of intensive gestures, we can predict that utterance initial constituents will also respond to the proportion test. It would also be interesting to determine the weight of the intensive gesture, especially when for instance, it reinforces a degree adverb that bears a focal accent. It would lastly be interesting to cross-examine the distribution of focal accents (for instance their association with discourse markers or particular morphological types).

To the best of our knowledge, such a study had not been conducted before from a gestural perspective and we hope that the information it gives on what intensive gestures intensify will be of some use for the improvement of human-machine communication.

eyebrow movement intensifies the focal accent rather than the particle as a verbal marker. Cross-associations of cues will be the object of future work, it could not be done now for time and space reasons.

6. References

- [1] Cassell, J., Pelachaud, C., Badler, N., et al., "Animated Conversational: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents". SIGGRAPH '94 Proceedings:413-420, 1994.
- [2] Krahmer, E., Ruttkay, Z., Swerts, M., and Wesslink, W., "Pitch, eyebrows and the perception of focus", Speech Prosody 2002 Proceedings:443-446, Laboratoire Parole et Langage, Aix-en-Provence, 2002.
- [3] Krahmer, E., Ruttkay, Z., Swerts, M., and Wesslink, W., "Perceptual evaluation of audiovisual cues for prominence", Proceedings of 7th International Conference on Spoken Language Processing (ICSLP 2002):1933-1936, Vol. 3, s.l., Denver, Colo., USA, 2002.
- [4] Guaïtella, I., "Mélodie du geste, mimique vocale", *Semiotica*, Vol. 3/4, no. 103:253-276, 1995.
- [5] Guaïtella, I., Santi, S., and Cavé, C., "Links between Eyebrow Movements and Voice Variations : an Experimental Investigation", *Gesture*, Forthcoming.
- [6] Gullberg, M., and Holmqvist, K., "Keeping an eye on Gestures: Visual Perception of Gestures in Face-to-Face Communication", *Pragmatics and Cognition*, 7, 1999, p. 35-63.
- [7] Barrier, G., Caelen, J., and Meillon, B., "La visibilité des gestes : paramètres directionnels, intentionnalité du signe et attribution de pertinence", *Workshop Francophone sur les Agents Conversationnels Animés (13-14 juin)*:113-123, 2005.
- [8] Nicolas, E. C., Pelachaud, C., and Pelé, D. "Vers un modèle de régulation de l'attention", *Workshop Francophone sur les Agents Conversationnels Animés (13-14 juin)*:101-110, 2005.
- [9] Bertrand, R., Blache, P., Espesser, R., & al., "Le CID: Corpus of Interactional Data – protocoles, conventions, annotations", *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix en Provence (TIPA)*, 25:25-55, forthcoming (2007).
- [10] Van Rullen, T., "Vers une analyse syntaxique à granularité variable", PhD Thesis, University of Aix-Marseille I, December 2005.
- [11] Boersma, P., and Weenink, D, "Praat : doing phonetics by computer (release 4.3.14)", <http://www.praat.org/>, 2005.
- [12] Kipp, M., "Anvil 4.0. Annotation of Video and Spoken Language". <http://www.dfki.de/~kipp/anvil>, 2003-2006.
- [13] McNeill, D., *Hand and Mind. What Gestures Reveal about Thought*, Chicago University Press, Chicago, 1992.
- [14] Cassell, J., "A Framework for Gesture Generation and Interpretation", in R. Cipolla & A. Pentland (eds.), *Computer Vision in Human-Machine Interaction*, Cambridge University Press, New York, 191-215, 1998.
- [15] Selting, M., "The construction of 'units' in conversational talk", *Language in Society*:477-517, Vol. 29, 2000.
- [16] Di Cristo, A., "Le cadre accentuel du français : essai de modélisation", *Langues*:184-205, vol. 2, no. 3, 1999.
- [17] Ferré G., "Degrés d'intensité exprimés à l'oral. Du discours à la gestualité en passant par la prosodie", in M. Noailly & F. Lefeuvre (eds), *Actes du 17ème Colloque du CerliCO : Intensité, Comparaison, Degré*, Presses Universitaires de Rennes, 13-26, 2004.