L'ANALISI
LINGUISTICA E LETTERARIA

2

ANNO XVI   2008

Proceedings of the IADA Workshop
Word Meaning in Argumentative Dialogue

Homage to Sorin Stati

Milan 2008, 15-17 May
VOLUME 2

edited by G. Gobber, S. Cantarini, S. Cigada, M.C. Gatti & S. Gilardoni

EXPRESSION OF ANGER
IN SPONTANEOUS FRENCH DIALOGUES

CATHERINE MATHON

1. *Introduction*

In the field of emotions and especially in the numerous studies about the vocal expression of emotions, the corpus is a central problem. A great number of experiments are based on simulated emotional expressions with actors (professional or not) playing roles. This allows for a tighter control of the quality of the recordings as well as a selection of the emotion to be acted, and finally a control of the lexical content in the expressions.

Some studies are conducted on elicited speech under laboratory. It consists of putting subjects in situations that may cause an emotional reaction. In most condition subjects have to perform a specific task, which can limit the lexical content of the utterances produced. The scenarios vary according to studies. They may be based on a video game (Johnstone & Scherer 1999) or a computer task.

Recently more and more researchers have insisted on the necessity of using natural emotional speech. Some are conducted using multimedia data (Chung 2000; Mathon 2007). Some corpus is also obtained from call-centers (Vidrascu & Devillers 2005). This methodology allows scientists to work on "authentic" emotions. However, spontaneous emotional speech still causes many problems because of a lack of overall control parameters. Indeed, to obtain these data, there are ethical as well as practical difficulties such as:
  – the signal is quite often of poor quality, because of bad recording conditions;
  – the situation of speech cannot always cover a wide range of emotions; indeed it is often difficult to find a discourse type favouring emotion expression.
  – lexical content of the statements is not controlled.

This last point leads the researchers using natural speech data to take into account not only vocal features (i.e. fundamental frequency parameters, speech and articulation rate, high frequency energy, intensity...) but also segmental information and speaker behaviour.

Thus, I assume that in natural speech data, emotion is expressed through both segmental and supra-segmental parameters. A multimodal analysis is conducted in order to verify this.

My second hypothesis in this study is that male and female speakers do not use the same means to express anger.

The first part of this paper describes the corpus on which the study is based, and how the emotive charge of the corpus has been validated; it measures also the real influence of both segmental and supra-segmental information on emotion detection. Then, I present the results of a multimodal analysis, connecting pragmatic contexts, segmental features (lexical and morphosyntactic) and supra-segmental cues (F0 registers). This analysis reveals a difference between female and male speakers in the management of a conflictual dialogue; there is a strong relationship between the management of a speaker's emotion on the one hand, and linguistic (segmental and supra-segmental) strategies s/he uses to express anger on the other hand.

2. *Corpus*

The study is based on a corpus of natural dialogues recorded from a radio program. The radio presenter calls professionals and provokes a situation of miscommunication by playing the role of a client and asking something which doesn't fit the situation. Eventually, this miscommunication leads the victim of the hoax to express anger.

The corpus consists of twelve dialogues transcribed using Transcriber 4.0.

Table 1: *Characteristics of the corpus*

Dialogues	12
Speakers	13M, 14F
Speaker turns	765

The entire corpus was labelled with three types of labels:
- N for Neutral State (concerning principally the speaker turns at the beginning of the exchange),
- A for Anger (associated with a scale from 1 to 5, depending on the intensity of emotion),
- OE for Other Emotion. This last label indicates all the speaker's affective states, which are not anger.

The labelling task, made by the experimenter herself and another linguist, consisted in listening to all the dialogs, and reading the transcriptions of the speaker turns simultaneously and deciding which was the expressed emotional state, according to the

conventions above. In case of disagreement, the speaker turns were re-evaluated until the annotators came to an agreement. The productions of the radio presenter were not taken into account since we assumed that his productions could contain a part of acted out speech.

Table 2: *Proportion of speaker turns by label*
(N = Neutral State; A1 = Anger degree 1; A2 = Anger degree 2; A3 = Anger degree 3;
A4 = Anger degree 4; A5 = Anger degree 5; OE = Other Emotion)

Labels	Ratio (%)
N	52
A1	12
A2	15
A3	9
A4	3
A5	1
OE	8

Table 2 shows the proportion of the speaker turns by labels. 40% of the victim's turns were labelled as anger, across the five degrees of the anger scale. Mild anger seems to be more present than strong anger, probably because of the socio-professional context of the dialogs.

3. *Perception test*

The labelling of the corpus permitted to select which speaker turns were to be tested perceptively. A pre-test was conducted with 5 French listeners of 81 out of the 765 initial speaker turns, in order to extract the 26 turns which made up the final stimuli for the perception test. The pre-test was useful to show up the presence of anger in the corpus, as well to verify the most relevant segmental or supra-segmental cues for the detection of emotion. The final stimuli were tested in three different conditions by a total of 49 French listeners:
– in the first condition, 26 listeners could access both segmental and supra-segmental information;
– in the second condition, 13 readers could access only the linguistic content of the speaker turns;
– in the third condition, 10 listeners could access only the prosodic information. The segmental content of the speaker turns had been hidden using white noise.

The subjects task was to indicate if the acoustic signal they were listening to or the transcription they were reading (depending on the test condition) conveyed anger or not. If so, they were to evaluate the degree of anger on a scale from 1 to 5.

Figure 1: *Responses depending on degrees of anger for the three perception conditions*
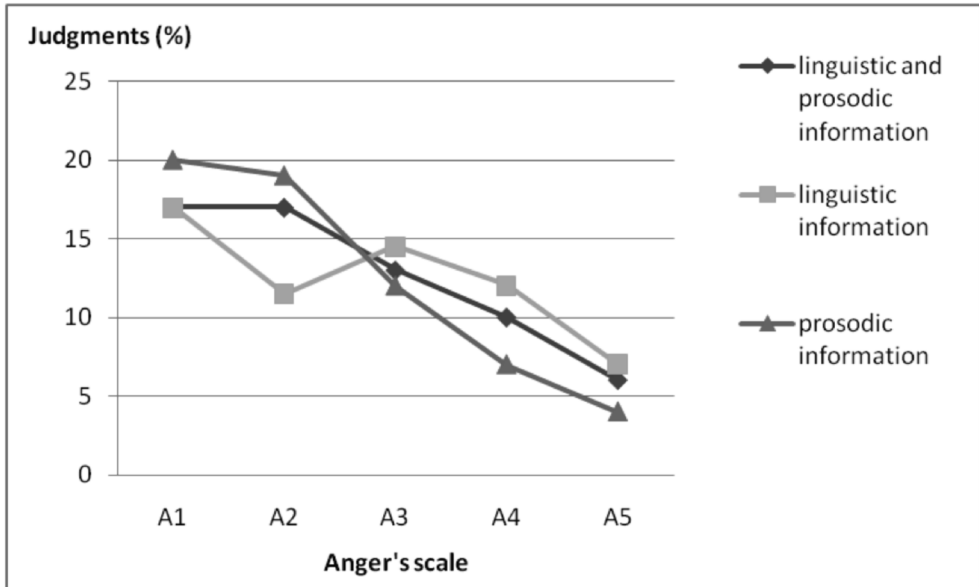

Figure 1 shows the answers for each degree of anger of the three conditions studied. These results confirm the first perceptive annotations: mild anger is more often detected, while strong anger seems to be less frequent. Prosodic information underlines mild anger, since in the third condition, listeners chose more often degrees 1 and 2 of anger. On the contrary, linguistic information focuses on strong anger (degrees 4 and 5).

4. *Multimodal analysis*

The next step was to conduct a systematic analysis of all the speakers' turns in order to list the relevant linguistic features for anger detection. We focalized on three types of analysis:
1.   an analysis of the pragmatic context of the speakers' turns. Different speech acts were identified;
2.   an analysis of the lexical and morphosyntactic items relevant for anger detection;
3.   an analysis of the F0 registers.

4.1 Pragmatic analysis

The pragmatic analysis was conducted using both the transcription and the audio data with the help of Transcriber 4.0, that followed the course of the dialogues. Male and female productions were analyzed separately in order to verify if there was a gender difference in the

management of the dialogues. Figures 2 and 3 show the pragmatic contexts found in the corpus and their division depending on the speakers' gender.

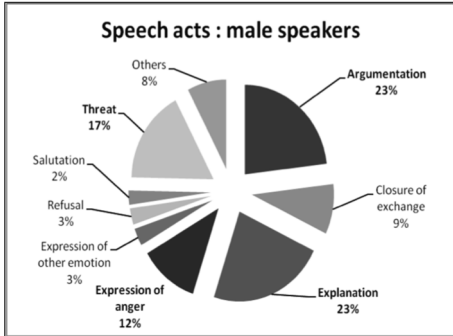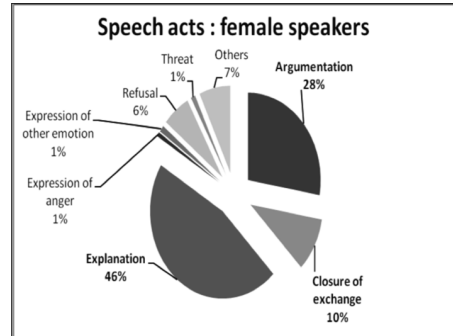Figure 2: *Proportion of speech acts (%) in male speakers' productions*

Figure 3: *Proportion of speech acts (%) in female speakers' productions*

Looking at speech acts in both male and female speaker productions, we can see that explanations and argumentations take an important place in their discourse. Male speakers also express threat and anger, while female speakers do not.

Figures 4 and 5 show the 3 speech acts: "explanation", "expression of anger" and "threat" in male and female speaker productions depending on degrees of anger.

Figure 4: *Proportion (%) of 3 speech acts (explanation, expression of anger, threat) in male speakers' productions*
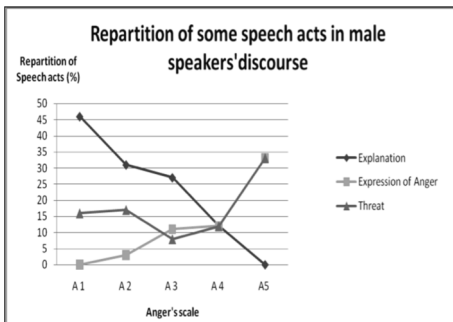
Figure 5: *Proportion (%) of 3 speech acts (explanation, expression of anger, threat) in female speakers' productions*
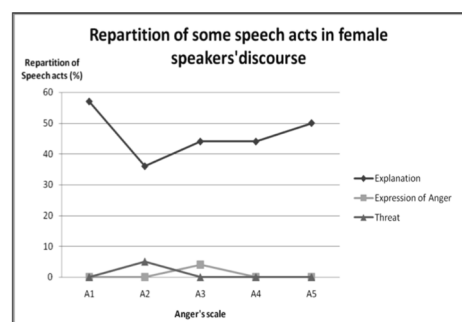
We can observe that the proportion of "explanations" decreases in male speaker discourse depending on the increase in anger. On the contrary, the proportion of "explanations" stays at a high level in female speaker discourse on all degrees of anger.

In male speaker discourse, the proportion of "expression of anger" and "threat" increases depending on the increase in anger. On the contrary, the proportion of "expression of anger" and "threat" stays at a low level in female speaker production.

In the first degrees of anger, we observe that speakers, male and female, try to explain the situation of the call, in order to fit with the demand of the client (the radio-presenter). Male speakers also try to threaten and intimidate the fake client, while female speakers prefer to close the dialogues. Even in strong anger, female speakers continue explanations and argumentations while male speakers express strong anger.

The behaviour differences between male and female speakers concerning the management of the situation of communication are also based on different uses of lexical and suprasegmental cues.

4.2 Lexical and Morphosyntactic Analysis

The second level of the analysis consisted in a systematic review of all the lexical and morphosyntactic cues for anger detection (Kerbrat-Orecchioni 2000). All these features were assigned to different classes and occurrences were counted. The proportion of occurrences of each class of features in the different degrees of anger was calculated and converted to percentages.

Five types of morphosyntactic features were identified: interjections, disfluencies, negative forms, verbal forms in the imperative, and other modality orders. Male and female speaker productions were analyzed separately. The negative form is the most occurring feature for both male and female speakers and is present in all degrees of anger. Interjections are also present in an important proportion (from 23% to 39% of the speaker turns) but preferably in the first degrees of anger. Moreover, the proportion of imperative verbal forms increases with the degrees of anger in male speakers productions (from 17% to 47% of the male speaker turns). This last result may be connected with threatening behaviour by male speakers.

Lexical features can be organized in six classes:
– words expressing anger directly;
– words expressing anger indirectly;
– colloquial words, since I assumed that a speaker who feels anger tends to use a more informal language;
– insults, swear words, which imply that the speaker expresses a strongly negative emotion;
– words that imply a closure of exchange;
– and finally all adverbs or coordinating conjunctions which imply an opposition.

In male productions, the most relevant feature is colloquial words. Their proportion in male speaker turns increases depending on degrees of anger, from 2% (A1) to 67% (A5). A regular increase in the proportion of opposition conjunctions is also observed, from 6% (A2) to 33% (A5). The only reliable lexical feature for female productions is the opposition conjunction. A regular increase in its proportion for the female speaker turns is observed, from 9% (A1) to 50% (A5). Female speakers do not use lexical markers at all like colloquial words

or insults. They do not even use words that express anger, directly or indirectly. These results are connected with female behaviour observed in the dialogues. Indeed they prefer to explain and argue which explains why the proportion of conjunction of opposition is relevant in female speech productions. But anger detection in female productions implies that they use other ways than lexical means to express anger. Perhaps a reliable feature for female anger detection will be supra-segmental and this will be treated in the next section.

4.3 Prosodic Analysis

In this study I focused on F0 parameters. Intensity and energy features were not taken into account because of the nature of the corpus. F0 measures were extracted automatically with the help of the WinPitchPro software. This software takes the transcriptions and signal segmentations first made with Transcriber which assigned a layer to each speaker. WinPitch-Pro recognises all the layers created with Transcriber and treats them separately. F0 was extracted from all the speaker turns (at a time sampling rate of 20 ms).

The minimum, maximum, mean and range of F0 for each turn were statistically computed. The voice amplitude of each speaker i.e. the delta difference between the maximum and the minimum of fundamental frequency, was divided in four equal registers: Low (L), Medium-Low (ML), Medium-High (MH), and High (H). The F0 values of these registers vary from one speaker to another one. F0 means were calculated for each turn using F0 automatic extractions. Then each value was classified in the corresponding register. The proportion of speaker turns (%) was calculated for each register and each degree of anger. This method can be used to compare voices which are significantly different, such as male and female voices.

Just as with the pragmatic and lexical strategies, the results of the classification of speaker turns, depending on degrees of anger and F0 registers, were examined separately in male and female productions. Figures 6 and 7 show the proportion (%) of each voice register depending on each degree of anger, for male (Figure 6) and female (Figure 7) speakers.

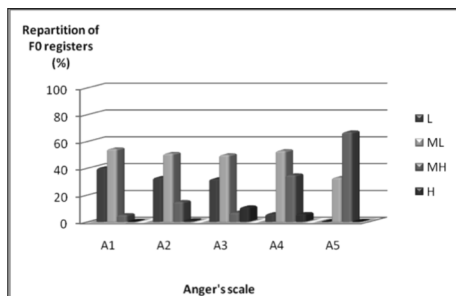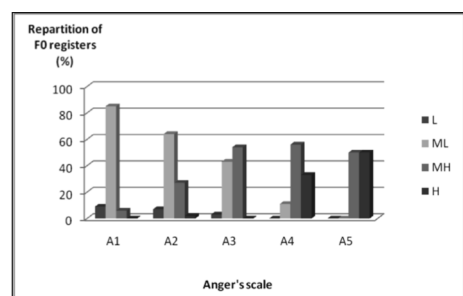Figure 6: *Proportion of F0 registers for male speakers*

Figure 7: *Proportion of F0 registers for female speakers*

For the first degrees of anger, the register used by both gender is the Medium-Low register. However male speakers also use the Low register, while female speakers prefer to go up to the Medium-High register. For both male and female speakers, there is a global increase of the F0 register depending on the degrees of anger, but this movement comes earlier in female speech productions (A3) than in male speaker turns (A4). For strong anger turns (A4 and A5), female speakers use MH and H registers, while male speakers stay in ML and MH registers.

The multimodal analysis showed that female speakers do not use any lexicon which implies impoliteness, or a direct expression of anger. Instead of using lexical information to communicate anger, female speakers use voice registers more easily. On the contrary, male speakers express anger by threatening and insulting the fake client. The lexical information, in this case is meaningful enough to communicate the affective state. It implies that male speakers do not need to use a higher voice register to express anger. Furthermore, they tend to intimidate the interlocutor. It has been demonstrated that intimidation and threatening need a low register (Demers 2003).

5. *Strategies of emotion expression*

Through the multimodal analysis described in the precedent section, it has been shown that there is some sort of a trade-off between segmental (i.e. lexical and morphosyntactic) and supra-segmental (prosodic) levels.

   In conclusion, I will show examples of melodic curves, illustrating this trade-off between the two levels. Just as with the multimodal analysis, male and female productions are examined separately.

   Figure 8 is an example of the melodic curve of a French woman's expression of anger (labelled as Anger 4).

Figure 8: *Melodic curve (WinPitch Pro) of a French woman's expression of anger (labelled as Anger 4). The horizontal axis is the time axis in seconds and the vertical axis is the F0 in Hertz (Hz)*
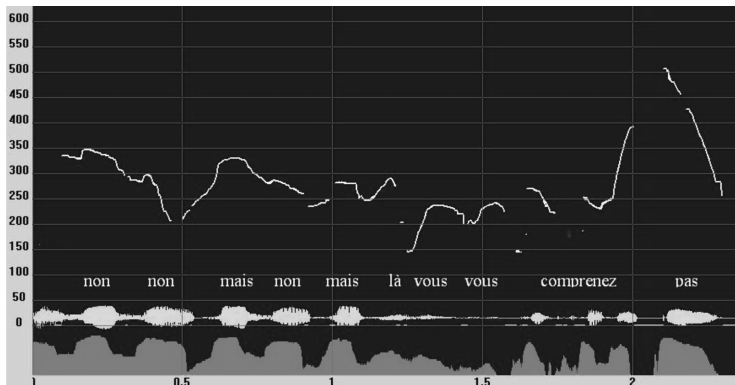
The content of the statement "non non mais non mais là vous vous comprenez pas", which can be translated as "no no but no but there you you don't understand", is not meaningful as far as the expression of anger is concerned. Indeed, considering the lexical content, there is no expression of anger, no swear words. There are just some conjunctions of opposition and negative forms. In compensation, the melodic pattern presents strong pitch variations with steep rises and falls. The F0 range is very wide going from 150 Hz to 500 Hz. Precisely on the last syllable the voice of the speaker decreases from high to medium-low register. Even in the lowest points of the melodic curve, the speaker voice does not fall under medium-low register.

Figure 9: *Melodic curve (WinPitch Pro) of a French man's realisation of anger (labelled as Anger 4). The horizontal axis represents the duration in seconds and the vertical axis the level of fundamental frequency in Hertz (Hz)*
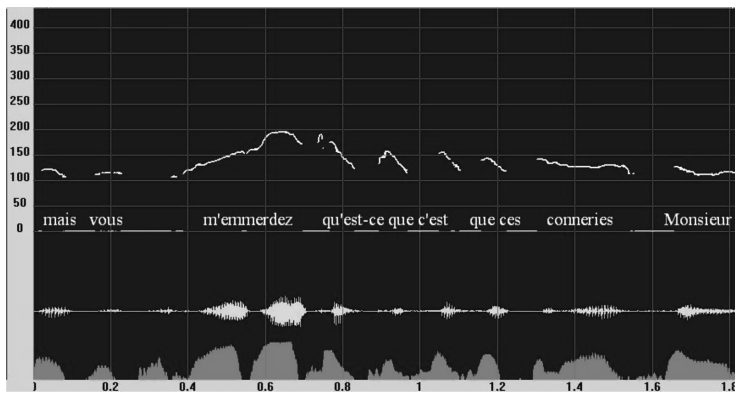

Figure 9 is an example of the melodic curve of a French male's expression of anger (labelled as Anger 4). The content of the statement "non mais vous m'emmerdez qu'est-ce que c'est que ces conneries Monsieur" which can be translated as "you're pissing me off what are these stupid things Sir", is meaningful considering of the expression of anger. Indeed the speaker strongly expressed anger by using swear words. In contrast, the prosodic information seems to be not relevant. The voice of the speaker stays in the low and medium-low registers. The contour is quite flat except on "m'emmerdez" ("pissing me off").

6. *Conclusion*

This study brings out the linguistic and paralinguistic speaker strategies to express anger, depending on the behaviour they adopt in a situation of conflict. It reveals a sort of trade-off between lexical and prosodic features.

Moreover, I found there is a gender influence in the management of conflict which influences the segmental and supra-segmental parameters the speakers use to express anger. This research shows a need to explore all communication features of emotion and underlines

the trading relationships between linguistic and paralinguistic (segmental and supra-segmental) features that are used to convey emotion and attitudes.

Modern technology should increase the use of multimodal (visual, acoustic, semantic, pragmatic) analysis in this domain.

References

Chung, Soo-Jin (2000). L'expression et la perception de l'émotion extraite de la parole spontanée: évidences du coréen et de l'anglais. Thèse de l'Université Paris 3 (Sorbonne Nouvelle).

Demers, Monique (2003). La voix du plus fort. Étude acoustique sur le registre vocal en tant qu'indicateur sociolectal et dialectal en français spontané. In: Demers, M. (ed.). Registre et voix sociale. Québec: Nota Bene, 196.

Johnstone, Tom & Klaus R. Scherer (1999). The effects of emotions on voice quality. In: Proceedings of 14th ICPhS, San-Francisco, 1999, 1-7 August, 2029-2032.

Kerbrat-Orecchioni, Catherine (2000). Quelle place pour les émotions dans la linguistique du XXème siècle? Remarques et aperçus. In: Plantin, C., M. Doury & V. Traverso (eds.). Les émotions dans les interactions. Lyon: Presses Universitaires de Lyon, 33-74.

Mathon, Catherine (2007). Multimodal Analysis of Anger in Natural Speech Data. In: Proceedings of 16th ICPhS, Saarbrucken, 2007, 6-10 August, 2117-2120.

Vidrascu, Laurence & Laurence Devillers (2005). Detection of real-life emotions in dialogs recorded in a call center. In: Actes de Interspeech 2005, Lisboa, 2005, 4-8 September, 1841-1844.