

Multimodal Analysis of Anger in Natural Speech Data

Catherine Mathon

EA333 ARP, Université Paris Diderot, UFRL case 7003, 2 place Jussieu, 75251 Paris Cedex
mathon@linguist.jussieu.fr

ABSTRACT

This paper reports a study on anger detection and expression in French, conducted on natural speech data. Perceptual tests showed that both linguistic and prosodic cues could convey information about the affective state of the speaker. Pragmatic, segmental and supra-segmental analyses of the corpus were conducted in order to reveal the pertinent cues for the detection of emotion and the classification of anger in degrees.

Keywords: Emotion, Natural speech data, Prosody.

1. INTRODUCTION

In researches carried out with acted speech data [1] [2], sentences are generally chosen in order to be semantically neutral. While natural speech studies have to take into account not only vocal features but also segmental information and speakers' behaviours.

The first part of this paper describes how the emotive charge of the corpus has been validated, and measures the real influence of both segmental and supra-segmental information on emotion detection. Then, we present the results of a three-level analysis, connecting pragmatic contexts, segmental features (lexical and morphosyntactic) and supra-segmental cues (F0 registers). The analysis reveals a difference between female and male speakers in the management of a conflictual dialog. We show that there is a strong relationship between the management of a speaker's emotion on the one hand, and linguistic (segmental and supra-segmental) strategies he uses to express anger.

2. CORPUS

The study reported in this paper makes use of a corpus of natural dialogs recorded in a radio program. The radio presenter calls on professionals and leads 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 fifteen transcribed dialogs using Transcriber 4.0.

Table 1: Characteristics of the corpus

| | |
|---------------|----------|
| Dialogs | 15 |
| Speakers | 13M, 14F |
| Speaker turns | 765 |

The entire corpus was labeled using three types of labels:

- N for Neutral State (concerning principally the speakers turns of 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 labeling task consisted in listening all the dialogs, and reading simultaneously the transcriptions of the speaker turns and deciding which the expressed emotional state (among the ones proposed by the labelling system).

The labeling task was conducted by the experimenter and another linguist. 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 speech.

Table 2: Repartition of labels (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 repartition of the speakers' turns according to the labels. 40% of the victim's turns were labeled 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. PERCEPTIVE TESTS

The labeling of the corpus permitted to select the speakers' turns to be tested perceptively. Then we conducted a pretest among 5 French listeners on 81 out of the 765 initial speaker's turns, in order to extract 26 turns which composed the final stimuli for the perceptive test. The pretest was useful in the validation of the presence of anger in the corpus, as well as in the verification of 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 to both segmental and supra-segmental information.
- In the second condition, 13 readers could access only to the linguistic content of the speaker turns.
- In the third condition, 10 listeners could access only to the prosodic information. The segmental content of the speaker turns had been hidden using white noise.

Subjects' task was to indicate if the stimulus they listened to or they read (depending on the test condition), conveyed anger or not, and if so, to evaluate the degree of anger (on a scale from 1 to 5).

Figure 1: Repartition of perceptive answers depending on degrees of anger, for the three perceptive conditions.

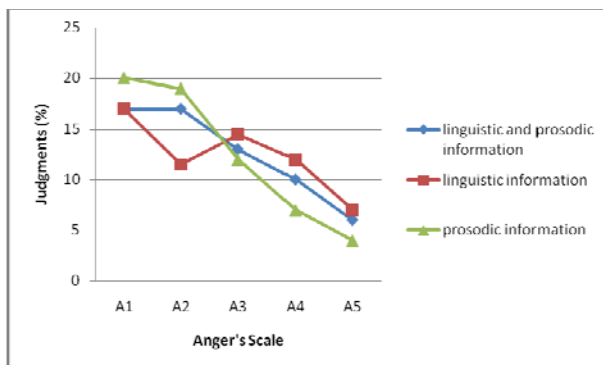


Figure 1 shows the repartition of answers for each degree of anger, in the three conditions. These results confirm the first perceptive annotations: mild anger is more often detected, while strong anger seems to be rarer. Prosodic information stresses mild anger, since in the third condition, listeners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

Considering these results, we may suppose that lexical and generally segmental information is more relevant for strong anger detection, while prosodic features are more reliable for mild anger.

4. THREE-LEVELS ANALYSIS

4.1. Pragmatic analysis

The next step in our approach was to conduct a systematical 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.

The pragmatic analysis was conducted by the experimenter from both the transcription and the audio data with the help of Transcriber 4.0, which allowed to follow the course of the dialogs. Male and female productions were analyzed separately, in order to verify if there was a gender difference in the management of the dialogs.

Table 3 shows the pragmatic contexts found in the corpus and their repartition depending on the degrees of anger.

Table 3: Repartition of speech acts (%) depending on anger degrees and speaker's gender (M = Male speakers; F = Female speakers; A1 = Anger degree 1; A2 = Anger degree 2; A3 = Anger degree 3; A4 = Anger degree 4; A5 = Anger degree 5)

| | A1 | | A2 | | A3 | | A4 | | A5 | |
|-----------------------------|----|----|----|----|----|----|----|----|----|----|
| Speech Acts | M | F | M | F | M | F | M | F | M | F |
| Argumentation | 9 | 9 | 15 | 29 | 19 | 18 | 41 | 33 | 33 | 50 |
| Closure of Exchange | 9 | 12 | 14 | 16 | 11 | 11 | 12 | 11 | 0 | 0 |
| Explanation | 46 | 57 | 31 | 36 | 27 | 44 | 12 | 44 | 0 | 50 |
| Expression of anger | 0 | 0 | 3 | 0 | 11 | 4 | 12 | 0 | 33 | 0 |
| Expression of other emotion | 2 | 0 | 6 | 2 | 8 | 4 | 0 | 0 | 0 | 0 |
| Refusal | 4 | 9 | 0 | 7 | 5 | 4 | 6 | 11 | 0 | 0 |
| Salutation | 8 | 0 | 2 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| Threat | 16 | 0 | 17 | 5 | 8 | 0 | 12 | 0 | 33 | 0 |
| Others | 9 | 12 | 15 | 4 | 9 | 19 | 6 | 0 | 0 | 0 |

In the first degrees of anger, we observed 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 dialogs. Even in strong anger, female speakers continue explanations and argumentations, while male speakers express their strong anger. The differences of behaviors between male and female speakers concerning the management of the situation of communication are also based on different uses of lexical and supra-segmental cues.

4.2. Segmental analysis

The segmental analysis consisted of a systematical review of all the lexical and morphosyntactic cues for anger detection. All these features were assigned to different classes, and their occurrences were counted. The proportion of occurrences of each type of features in the different degrees of anger was calculated and converted in percentage.

4.2.1. Morphosyntactic cues

We identified five types of morphosyntactic features: interjections [3], disfluencies [4], negative forms [5], verbal forms in the imperative, and other order modalities. Male and female speakers' productions were analyzed separately.

Negative form is the most occurring feature for both male and female speakers, present in the 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 grows 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 the male speaker's threatening behavior.

4.2.2. Lexical Cues

Lexical features can be organized in six classes:

- words expressing anger directly,
- words expressing anger indirectly,
- colloquial words, since we assumed that a speaker who feels anger tends to use a more informal language
- insults, swearwords, 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 familiar words. Their proportion in male speaker turns grows depending on degrees of anger, from 2% (A1) to 67% (A5). A regular growth of the proportion of opposition's conjunctions is also observed, from 6% (A2) to 33% (A5).

The only reliable lexical feature for female productions is the conjunction of opposition. A regular growth of its proportion in the female speaker turns is observed, from 9% (A1) to 50% (A5). Female speakers do not use at all lexical markers like familiar words or insults. They do not even use words that express anger, directly or indirectly. These results are connected with the female behaviors observed in the dialogs. 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 the lexical cues to express anger. We suppose that the reliable feature for female's anger detection will be supra-segmental.

4.3. Supra-segmental analysis

We focused on F0 parameters for this study. Intensity and energy features were rejected, because of the nature of the corpus.

4.3.1. Methodology

F0 measures were extracted automatically with the help of the software WinPitchPro. This software takes into consideration the transcriptions and signal segmentations first made with Transcriber. WinPitchPro recognises all speakers created with Transcriber and treats them separately in specific layers. F0 was extracted from all the speakers' turns (at a sampling rate of 20 ms). Some statistics were performed, calculating the minimum, maximum, mean and range of F0 for each speaker's turn.

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 speaker to another one.

F0 means were calculated for each speaker turn, using F0 automatic extractions, and 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 allowed to compare voices which are significantly different, like male and female voices.

4.3.2. Results

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.

Tables 4 and 5 show the proportion (%) and the repartition of each voice register depending on each degree of anger, for male (Table 4) and female (Table 5) speakers.

Table 4: Repartition of F0 registers for male speakers

| | | F0 Registers | | | |
|------------------|----|--------------|------|-----|----|
| | | L | ML | MH | H |
| Degrees of Anger | A1 | 40,3 | 54,4 | 5,3 | 0 |
| | A2 | 33 | 51 | 15 | 1 |
| | A3 | 32 | 50 | 7 | 11 |
| | A4 | 6 | 53 | 35 | 6 |
| | A5 | 0 | 33 | 67 | 0 |

Table 5: Repartition of F0 registers for female speakers

| | | F0 Registers | | | |
|------------------|----|--------------|----|----|----|
| | | L | ML | MH | H |
| Degrees of Anger | A1 | 9 | 85 | 6 | 0 |
| | A2 | 7 | 64 | 27 | 2 |
| | A3 | 3 | 43 | 54 | 0 |
| | A4 | 0 | 11 | 56 | 33 |
| | A5 | 0 | 0 | 50 | 50 |

For the first degrees of anger, the register used by both gender of speakers is the Medium-Low register, but 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 growth of F0 register depending on degrees of anger, but this movement comes earlier in female speech productions (A3) than in male speaker turns (A4). For strong anger speakers' turns (A4 and A5), female speakers use MH and H registers, while male speakers stay in ML and MH registers.

5. DISCUSSION AND CONCLUSION

This study brings out the linguistic and paralinguistic strategies of speakers to express anger, depending on the behaviours they adopt, in a situation of conflict. Connected with the results obtained in the precedent analysis, we accounted for the strategies used by the speakers to express anger. We found that female speakers do not use any lexicon, which implies impoliteness, or a direct expression of anger. Instead of using segmental information to communicate anger, female speakers use voice registers more easily. On the other hand, 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, and intimidation, and threatening need a low register [6].

6. REFERENCES

- [1] Scherer, K. R., 1995. How emotion is expressed in speech and singing. Proc. 13th ICPhS Stockholm, 90-96.
- [2] Bänzinger, T., Scherer, K. R., 2001. Relations entre caractéristiques vocales perçues et émotions attribuées. Actes des Journées Prosodie 2001, Grenoble, France, 119-124.
- [3] Vidrascu, L., Devillers, L., 2005. Detection of real-life emotions in call centers. Proc. 9th Interspeech, Lisboa, September 4-8, 2005.
- [4] O'Connell, D. C.; Kowal S., 2005. Where do interjections come from? A psycholinguistic analysis of Shaw's Pygmalion. Journal of psycholinguistic research 35, 497-514.
- [5] Arrivé, M., 1994. Langage et psychanalyse, linguistique et inconscient: Freud, Saussure, Pichon, Lacan. Paris. Presses Universitaires de France.
- [6] Demers, M. 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. (Eds), *Registre et voix sociale*. Nota Bene, 196.