

Perception of Anger in French as Foreign Language Experimental Protocol and Preliminary results

Catherine Mathon, Sophie de Abreu, Daniela Perekopska

EA333 “Atelier de Recherches sur la Parole”

Université Paris 7 – Denis Diderot

2 place Jussieu, 75251 Paris Cedex 05 France

{mathon;deabreu;perekopska}@linguist.jussieu.fr

Abstract

Learners of a foreign language must perceive the emotions of their interlocutor. They must also learn to reproduce the prosodic patterns associated with different emotions in the target language, otherwise the communication fails. Our project deals with 3 main questions: How are emotions perceived in a foreign language? Will a learner be able to reproduce such an emotion, and, if so, how? How will these (re)productions be recognized by native speakers? We concentrated on the study of the emotion “Anger”. This paper aims to show whether prosody provides enough information to allow students of French as a Foreign Language (FFL) to recognize this emotion. The perceptual test presented here is original because it is based on the use of a corpus of spontaneous French, containing ‘real’ emotions. We focus here on the first stage of our research: the results of the perception of anger by Czech and Portuguese speakers. We insist on the methodology as well as the experimental protocol in our work, as they represent the main framework of our project.

1. Introduction

The study we present intersects two domains of research: the prosodic characterization of emotions in speech and the area of learning and teaching a 2nd language. Communication is the link between these two domains. It has been one of the most important subjects of research in the field of foreign language teaching for decades [1]. Communication is very useful to describe real language skills between people, while remaining closely linked with emotions and attitudes. Little work has been done, however, on the perception and production of emotions in a foreign language, even though it is an essential skill for a 2nd language learner to acquire: a learner of a foreign language needs to perceive the emotions of her or his interlocutor in order to react in an appropriate way in a given situation. Furthermore, research on second language teaching insists on the necessity to use more concrete and authentic spoken material in classes [3]. The aim of this paper is to show the importance of prosody in detecting emotions in spontaneous speech. This work is part of a larger project which consists of comparing the perceptual ability of students of FFL in identifying emotions. We claim that it is necessary to include prosody in a communicative method of language teaching. We present the elaboration of a perceptual test based on spontaneous speech in French (radiophonic telephone hoaxes) as well as the results of the perception of the emotion of anger by Portuguese and Czech learners of French as a 2nd language.

The originality of the method used to prepare the perceptual test rests on the fact that it is based on a corpus of spontaneous French, instead of readings and academic speeches (see the numerous works in this domain [2]). Our corpus has determined the way we treated these perceptual tests, and this is why we describe it in detail below.

2. Experimental Protocol

2.1. Corpus

One of our main goals was to build a perceptual test based on ‘real’ emotions and not ‘fake/played’ ones. Everyone is aware of the difficulty of obtaining emotions such as anger in laboratory conditions and it was therefore important to work on spontaneous ‘everyday’ speech. An interesting source for our corpus was found in radiophonic hoaxes (online Fun Radio <http://www.funradio.fr>) : A radio presenter calls professionals (such as bakers or taxidermists) or institutions (such as high schools) and plays the role of a client who asks a question that does not correspond to the situation. For example, the animator calls a taxidermist and asks for a cab, thus creating a situation where the professional tries to explain the mistake while the presenter acts as if he does not understand the problem. Eventually, this misunderstanding leads the victim to express anger. We collected 1 hour 4 mins of speech (24 programmes). We have chosen not to analyze the production of the presenter for two main reasons: firstly, it is not possible to interpret his speech at the same level as the speech of the victim since he is aware of the hoax and is playing for his audience. Secondly, we assume that the type of anger the presenter expresses is different to that expressed by the victim: only the victim, unaware of the hoax, expresses “real” emotion, whereas the animator is simulating his emotion. Since our aim is to test the perception of authentic emotions by FFL learners, we concentrated our work on the speech of the victim. We transcribed 15 dialogues (the other dialogues were not expressive enough and the victim had doubts about the hoax), using Transcriber, which represents about 40 mins of speech. We first annotated each sentence of each turn using 3 types of labels: Anger (associated with a number from 1 to 5 to mark the degree of anger), Neutral Attitude, Other Emotion (associated with a number from 1 to 5 to mark the degree of this other emotion). These three labels constituted our main categories. The first annotation, based on our intuition and our competence as French native speakers, allowed us to harmonize the emotive charge of the sentences presented to the appreciation of native speakers in the pretest presented below.

2.2. Pretest

The first difficulty we faced was in the attribution of a degree to the emotion conveyed by the sentences of the corpus. 5 native speakers of French were asked to decide if the sentences they heard were said with anger, a neutral attitude, or with another emotion. If the sentence was pronounced with anger, the subjects had to evaluate the degree of this emotion by rating it from 1 to 5. The material used for the pre-test was 'raw', that is, not modified.

A total of 81 sentences were chosen, of which 5 were training sentences. The subjects listened to the 81 sentences in random order. The beginning of each sentence was indicated by a specific sound and they had 2 seconds to note their answer. This pre-test revealed that choosing between 'anger' and 'another emotion' was difficult for the subjects. It appears that it was difficult for them to conceptualize 'emotion' (for instance, understanding what anger or emotion meant) especially in the context of a test, and they were disturbed. We therefore decided to eliminate the category of 'another emotion' from the perceptual test. The pre-test was mainly useful to select the sentences to be used for the final test. We kept sentences where judgments were clearly defined by a majority. For 'Anger', we kept only those sentences for which we had at least 80% of 'Anger' as an answer. From the pre-test, we selected 13 sentences judged as 'Anger' and 13 judged as 'No Anger' as well as 5 training sentences.

3. Perception test

3.1. Stimuli

In order to mask the segmental information, we had to isolate the prosodic information from other linguistic parameters. One way to achieve this is through re-synthesis. However, as we did not want to modify the spontaneous prosody of the sentences, it was not a satisfactory solution: re-synthesis would not have preserved the authentic character of the document. Another way to hide linguistic content is with low-pass filtering. The problem is that this method eliminates energy in high frequencies. Since energy in high frequencies is an important parameter of anger [4], [5], we assumed that it was a decisive perceptive parameter that could not be dismissed from our analysis.

We finally chose the 'white noise' method [6] to hide the linguistic content of sentences and keep only the prosodic information. To each chosen sentence extracted from the corpus after the pre-test, we added white noise, using the software Soundforge 7.0. The created white noise had the same length as the original sentence. We derived the intensity of the white noise from the intensity of the original speaker's voice. We finally mixed the sentences and the white noise to obtain our stimuli. It is interesting to point out that the constructed stimuli were perceived by the listeners as sounds of bad quality. This perception can be compared to what is obtained with low-pass filtering which gives the impression of a damaged soundtrack. We believe that adding white noise is an appropriate way of neutralising the effects of segmental information and to evaluate the part of prosodic perception in the perception of emotion. The listeners heard each sentence twice in order to verify the coherence of the

answers, added to a statistic verification of the significant value of the results. Then they were randomized and preceded by 5 training ones. The test, which contains 57 stimuli, is about 8 minutes long.

3.2. Task

The task the subjects had to accomplish was twofold: the listeners had to decide if the stimulus conveyed anger or not, and if so, to evaluate the degree of anger. They were advised beforehand that the quality of the sound was bad, in order to avoid the need for an adaptation.

3.3. Subjects

Three groups of listeners were invited to do the perceptual test: the first, composed of 10 native speakers of French (6 women and 4 men) represented our control group. The Portuguese speakers, 7 women and 3 men, our test group, were students in the Literature Department of the University of Lisbon, from the same class. Their level corresponds to a B2 level, according to the European portfolio of language [7]. The Czech speakers, 8 women and 2 men, were students of the University of West Bohemia of Pilsen. They also have a B2 level. It was important to get listeners of an intermediate level of French as we consider that beginners do not have enough knowledge in order to interpret the degrees of anger, while advanced students would have too much knowledge about prosody to show significative results.

3.4. Interface

We believe that the way a test is presented to the listeners can have an important influence on the results of a perceptual test. We therefore were very careful in controlling its effects on the test. Moreover, since we placed our work in a teaching perspective, we wanted to build an interface with the teaching of languages, and in particular emotions. The main perceptual test, based on the stimuli described above, was presented on a computer. The interface was written in HTML, using EasyPhp. Results were then extracted to a text format. One of the advantages of the interface was its practical aspect since results could be extracted automatically. But the main characteristic of this kind of interface is that the subjects are at the same time in a friendly and in a restrictive environment. As opposed to a paper format that would allow hesitations or modifications of choices with deletions, the electronic format allows for just one choice restricted by time. We were also interested in testing the robustness of the presentation of such a test on the Internet. So the interface was designed to allow us to control exactly what the speakers do while taking the test, the various 'clicks' they perform, and the time allocated for each step. We controlled, for instance, the impossibility of going back and redo a step (an identification number was attributed to each session of the test and we took into account only the sessions in which the speaker entered the personal information), the impossibility of giving an answer to the next step without having given one for the previous one, and the duration of the step. Finally, extra attention was paid to

the aspects of learning a foreign language while creating the interface. Another difficulty we faced was in deciding which language to use in the interface. Since the subjects were learners of French, they understood French, but it seemed to us more controlled to give them the instructions in their native language, to make sure there was no confusion about the task. The decision was even more difficult to come to with the questions used during the test since the test is about French. It was important to avoid a cognitive overload due to constant code switching. We finally proposed resorting to visual instructions during the test, using images that internet users are familiar with: the emoticons. We used three emoticons: one representing anger, another something other than anger, and the third to hear the sound. The three icons were introduced in the instructions for the task. We also used another technique to avoid cognitive overload: when a step was completed, it turned grey, indicating the impossibility of redoing it. In the figure below, for example, the emoticon 'not anger' is shaded grey, which indicates that it is no longer a possible choice. It also shows that the scale for rating the force of the emotion only appears once the option 'anger' is validated. This interface was designed with flexibility in mind so that it could be used to test other languages and other types of emotion, and to be applied in the area of foreign language teaching and learning.

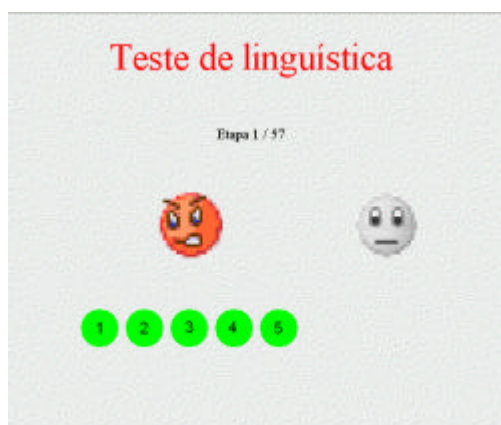
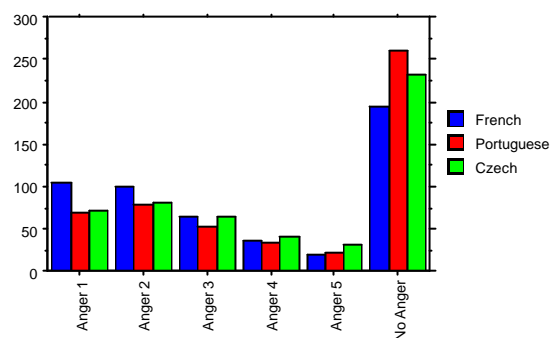


Figure 1: The interface of the perceptual test: examples of icons presented to the subjects.

4. Preliminary Results

In this section, we describe the preliminary results. We first compared the answers of the test for the three groups with the judgments of the pretest. The 13 sentences marked as 'angry' (based on the pretest) were recognized in 84% of the cases by French and Czech subjects; and at 77 % by the Portuguese. As mentioned earlier, we chose the stimuli of the test from the pretest majority judgments. For the pretest the subjects had to choose between 'Anger', 'Neutral Attitude', and 'Other Emotion'. If we look at the results more closely (stimulus by stimulus), we can see that the stimuli judged as 'Other Emotion' in the pretest became 'Anger' in the test. This could mean that the subjects detect an emotion in these stimuli and prefer to categorize it as 'Anger' rather than as 'No Anger'. Another hypothesis we can make is that the

emotional information carried by the prosody is strong, so the speaker has to weigh it with some more consensual semantic information. That is why a listener will perceive less anger when hearing lexical information. This aspect is being verified with a test based just on the transcription of the dialogue. As for the Portuguese group, the answers were closer to the judgments we obtained in the pretest. Portuguese students made a distinction between 'Anger' and 'No Anger'. The first graph shows a global distribution of the answers for all the sentences and without distinction between the language groups (Portuguese, Czech and French). We see that the main judgment for all the stimuli and the three language groups is 'No Anger'. We will return to this result in section 5. 'Strong Anger (4 and 5) is not really present (about 10%) while Anger 1 and 2 appear more often (about 30%). However, we have to weigh this observation: if the majority choice is "No anger", the sum of the other answers is higher. We will discuss a possible interpretation in the following section. If we compare the three groups they tend to the same results. But No Anger results are more important for Portuguese than for Czech and for French. French subjects spread their responses more widely into Anger 1, 2 and 3 categories.



Graph 1: Answers shared out by language group and by judgments.

The first difference we can observe between the three groups concerns the choice 'No Anger'. The Portuguese and Czech chose 'No Anger' more often than the French (50% for Portuguese and 44% for Czech vs 37% for French). But, French speakers detect weak anger (A1 and A2) more than the Portuguese and Czech. For A1 and A2 there is a 10% difference between French and Portuguese and of 9% between French and Czech.

From the analysis of these results, we can suggest that French perceive anger with more precision, while the Portuguese and the Czech groups classify weak anger more often as 'No Anger'. For strong anger (A4 and A5), we observe no significant difference between the French and Portuguese. The Czech group chose strong anger more often than the French and Portuguese (8% and 6% for A4 and A5 vs 7% and 4%). We can conclude that it is easier to distinguish strong anger than weak anger for the Portuguese and Czech learners of FFL. As noted previously, in order to verify the coherence of the answers of the listeners they had to judge

each stimulus twice. We tested this coherence with a Spearman Rank-Order Correlation test: for the French group, $.9 = 0.848$, for the Portuguese group $.9 = 0.833$, for the Czech group $.9 = 0.691$; in the three cases $p < 0.0001$. These results show a high coherence for each group between the first and the second answers.

5. Discussion

One of the main results presented above is that only a few sentences were recognized as strong anger. This is not a very surprising result, even though it had not been shown experimentally before. One explanation of this result lies in the specificity of the situation of communication selected (radiophonic hoaxes) compared to the daily life situations in which anger can occur, where the cases of real strong anger are rare. Considering a continuum from low to high anger, the occurrences more often take place in the lower part. If we recall, the presenter calls up professional people at work, a context in which the latter cannot always express anger freely. The cases of strong anger presented here are due to the express ability of the presenter to irritate people, but these cases are isolated.

These observations bring us to conclude that it would be more important to analyze what happens around boundaries between two categories (Anger 1-Anger 2, A2-A3 etc). The prosodic parameters seem to give very subtle information, and studying what happens in extreme cases will not teach us very much that is new, as the results reveal. That explains somehow the results of the French listeners who spread their responses more widely into 'low anger' categories (A1, A2). The high number of Portuguese and Czech answers in the 'No anger' category, as opposed to the French answers, is the indication of a much rougher and approximate perception of prosody.

This explains the majority of answers in the 'low anger' end of the scale. The high number of Portuguese and Czech answers in the 'No anger' category, as opposed to the French answers, is the indication of a much rougher and approximate perception of prosody. Czech listeners gave more answers in 'Anger' 4 and 5. It could hint at a difference in the expression of anger in French and in Czech cultures. This will be verified at a later stage when we look at the prosodic parameters of anger in these languages and work on the re-production of prosodic patterns of anger by Czech and Portuguese learners. It also appears that the Portuguese subjects are not able to give a subtle characterization of anger when it is not strong in the foreign language. In this case, it is more the categorisation in degrees than the recognition of anger that raises difficulties. The reasons underpinning our choice of methodology to isolate the prosodic information were stated earlier, but we are aware of the inherent disadvantages, and the questions such a choice may bring forth. Indeed, one of the difficulties in using white noise to hide segmental information is in finding the appropriate level of intensity to hide the linguistic information without masking the prosody. One way to control the amount of segmental information perceived by the speakers, in spite of our efforts to match the intensity of the white noise with the original sentence, would be to evaluate the part played by lexical perception in the perception of anger. To this end, it

would therefore be useful to make an subsequent perception test based on the script of the selected sentences. The results could then be compared with the original perception tests and our results be qualified.

6. Conclusions

In this paper, we described an experimental protocol for an interactive perceptual test of the prosody of emotions. We showed that prosodic information is enough to allow subjects to recognize anger. Portuguese and Czech FFL learners were able to distinguish 'Anger' from 'Not Anger' and to give an appropriate evaluation. Work is in progress to finalize the perceptual test in a written context, that is to say without sound, in order to check the influence of the segmental information. The same experiment conducted with different methods to hide the segmental information of the stimuli, such as low-pass filter, could be interesting and complementary.

Furthermore, we are proceeding to finalise the prosodic analysis of the original stimuli. When completed, we want to correlate the results with the judgments of the listeners. This study is part of a global project whose aims are:

- To extend the perceptual test to other language groups and to other emotions.
- To work on the reproduction of the sentences with anger prosodic patterns by the foreign language speakers, correlated with an analysis of the prosody of their native language.

7. References

- [1] Beale, J., Is communicative language teaching a thing of the past ?, *Babel*, Vol. 37, No. 1, Winter 2002, pp. 12-6.
- [2] Galazzi E. ; Guimbretière E., *Intonation et attitudes: une question de perception*, Studi di Linguistica, Storia della lingua Filologia francesi, Edizioni dell'Orso, Milan, 1994.
- [3] Holec, H. Des documents authentiques, pour quoi faire ? *Mélanges CRAPEL* - N°20 – 1990
- [4] Bänziger, T. and Scherer, K. R., Relations entre caractéristiques vocales perçues et émotions attribuées. *Actes des Journées Prosodie 2001*, Grenoble, France, 119-124, 2003.
- [5] Scherer, K. R, How emotion is expressed in speech and singing, *Proceedings of ICPHS 95*, Stockholm, vol.3, p. 90-96, 1995
- [6] Miller and Licklider The intelligibility of interrupted speech, (*J. Acoust. Soc. Am.* 22: 167-173, 1950)
- [7] http://www.enpc.fr/fr/international/eleves_etrangers/porfolio.pdf