# SPONTANEOUS DIALOGUE : SOME RESULTS ABOUT THE F0 PREDICTIONS OF A PRAGMATIC MODEL OF INFORMATION PROCESSING

M. Bessac and G. Caelen-Haumont Laboratoire CLIPS - IMAG
Université Joseph Fourier, B.P. 53
38041 Grenoble Cedex 9, FRANCE

Tel. +33/0 4 76 51 45 10, FAX: +33/0 4 76 44 66 75, E-mail: Mariette.Bessac@imag.fr, Genevieve.Caelen@imag.fr

# ABSTRACT

This paper presents the first results of a semanticpragmatic model which assigns a specific label to the relevant words of dialogue utterances and predicts their F0 value. The originality of this work lies in the kind of utterances the model has been designed for: dialogue utterances. The labels of the model represent the degrees of both the expected/unexpected and known/unknown aspects of the lexical information while the predicted value of F0 represents the corresponding weight of that information.

The aim of this work is 1) to observe the real values of F0 for each label and 2) to compare the prediction of the model to the real values. The real values correspond to the 3 relevant F0 indices (Maximum F0,  $\Delta$ F0 and mean F0). In this paper, only the levels 2 and 3 are discussed because they represent most of the population.

#### **INTRODUCTION**

The first model of information analysis was developed by Prince in [7] and [8]. This model was based on written utterances but was soon adapted to oral speech and prosody by Brown in [3] and Caelen-Haumont in [4] and [5]. In those works, the analysis concerned read aloud utterances or texts.

The model we present is a new adaptation of the upper cited models. This time, it concerns spontaneous oral dialogues. This kind of speech introduces new characteristics (intention, interaction, etc.) which in turn increase the difficulty of the analysis.

The first two sections of this paper present the observation corpus and the semantic-pragmatic model. The methodology and collection of F0 indices appear in section 3. The results are presented in section 4. Finally the conclusion and perspectives end up this paper.

### 1. CORPUS

Spontaneous dialogue has not been observed for a long time and only a few corpora exist. Among them, none could fulfil the constraints of prosodic and pragmatic domains. This is why a new corpus has been elaborated and recorded for this study.

#### **1.1 Elaboration of the corpus**

The corpus had to be adapted to the various planned analyses. They consisted in a prosodic analysis which required perfect acoustic conditions for the recording, a pragmatic analysis which required the interaction of 2 speakers and various degrees of situational conflicts, etc. Then it was decided to have 2 speakers and 2 tasks to be performed in collaboration. The speakers acted as a tourist and a tourism office employee. The first task consisted in changing some elements of the tourist's old plan to make it current. In the second task the speakers had to plan some visits and check out the route to follow. The difficulty of the second task was increased by the opposite aims of both speakers - the tourist wanted to do sports and the tourism office employee had to incite him/her to visit the cultural places- and by the number of pedestrian streets, road works, one-way streets, etc.

### 1.2 Recording and transcription

The couples of speakers were placed back to back in a semi-anechoic room. They had a little training with the experimenter before the recording of their conversation.

The 12 speakers (6 men and 6 women) produced 6 conversations, all about 15 minutes long.

The transcription of the 6 conversations required one year, especially because the voice overlappings were numerous and had to be noted carefully. A new kind of transcription was elaborated at that occasion. It meant to represent the production of both speakers in parallel, just like a music score. It was the best way to represent the intervention of each one according to the other.

#### 1.3 Analyses

After the transcription, each speaker's speech is split into speech turns and the succession of dialogue utterances is restored. Then the pragmatic and lexical analyses are achieved. The pragmatic analysis, which consists in assigning to each speech turn its dialogue act and enounciative modality (cf. [6]), will not be further presented here. The lexical analysis is not classical because many elements (more than 25%), which are due to the dialogue itself and to the interaction between the 2 speakers, cannot be properly analysed by a classical lexical analysis. Then it has been adapted to the kind of vocabulary speakers use in dialogue (cf. [2]). The main topic of this article, the semantic-pragmatic analysis, consists in selecting the lexical relevant words of the transcriptions and confront them to the model designed for this purpose. Then the semantic-pragmatic analysis has to occur after the lexical analysis. The model is presented in section 2.

### 2. THE SEMANTIC-PRAGMATIC MODEL

#### 2.1 Elaboration

As mentioned above, the semantic-pragmatic model (CPD) is based on Prince's taxonomy of information (cf. [7] and [8]) which had already been adapted to oral speech by Brown (cf. [3]) and then by Caelen-Haumont (cf. [4] and [5]). This time, the adaptation concerns dialogue utterances: the three kinds of information proposed by Prince (new, inferred, given) are split into 11 categories. The most important changes concern the inference domain where a distinction is made between a direct reasoning (when a continuity exists from a concept

to the next one) and an indirect reasoning (when one or several steps are necessary to go from a concept to the next one). This analysis gives a specific weight to each relevant unit, according to their degree of knowledge and expectation: 1 to 4, plus an extra one (5) for unpredictable information.

The concerned items are location adverbs, nouns, compound nouns, compound verbs, etc. The structure of the model and the weights are presented in Figure 1.

The assigned weight is supposed to predict a range of F0 level: for instance a very unexpected word will get a higher F0 value than a repeated word.

The classification and values are decided by the expert.

#### 2.2 Structure

The first level of analysis corresponds to Prince's classification. The definitions of the 11 final categories are given just after Figure 1. Such a display could be modified with the future work on other conversations.



Figure 1: structure of the CPD model

### 2.2.1 New information

The new information, which corresponds to newly mentioned elements, is divided into 2 sub-categories: (a) the unexpected and (b) the expected information.

- (a) The unexpected information is very rare. It occurs when the speaker mentions something totally unexpected and unrelated to what has been said before.
- (b) The expected information concerns newly mentioned elements which can be expected from the situation but cannot be inferred from what has been said before.

#### 2.2.2 Inferable information

The inferable information has been divided into 4 categories. The first 2 ones (c) and (d) correspond to newly mentioned information while the next 2 ones (e) and (f) represent already mentioned elements.

All elements are linked to what has been said before and/or to the situation.

In comparison to another element which is linked in space, time, context or situation,

- (c) the element is new
- (e) the element has already been mentioned

In comparison to another element which is not directly linked, in space, time, context or situation,

- (d) the element is new
- (f) the element has already been mentioned.

### 2.2.3 Available information

The available information does not always concern already mentioned elements. It may correspond to (g) elements of the plan, (h) elements known in advance or (i) a repetition by a synonym.

- (g) The element is new but is visible on the tourist's plan.
- (h) The element is new but both speakers know it before the recording.
- (i) The element is new but it is the synonym of an already mentioned element. Then it is like a repetition.

The available information also concerns already mentioned elements either (j) quite long ago or (k) just before.

- (j) The speakers have been talking about the rink, then the swimming pool and the restaurant, etc. and start talking again about the rink.
- (k) The element has just been mentioned and is repeated.

### 3. METHODOLOGY

#### 3.1 Labelling

The labelling is first achieved theoretically on the text (cf. Figure 2) and transposed to the speech signal (cf. Figure 3).



DdC	Nb.	alors $(=$ then $)$	pphp		
		vous (= you)	pper		
		êtes (= are)	verb		
		à y la (= at the)	prel		
		borne (= spot)	subc	new	CA 2
		à côté d'la (= near the)	prel		
		gare (= station)	subc		CC 2
		j'(= I)	pper		
		imagine (= suppose)	verb		
		là (= there)	pphf		

Figure 2: example of theoretical labeling

The pragmatic, lexical and semantic-pragmatic analyses appear on the same file.

Codes for Figure 3 D/FDdC = beginning/end of a confirmation question D/FCA = beginning/end of a new element, known in advance D/FCC = beginning/end of a known element, inferable by contiguity

The curve of the bottom window corresponds to the F0 values.



Figure 3: example of speech signal labelling

The labelling is performed with the help of the WaveEdit application (cf. [1]). It is a very long phase of the work but it is necessary to collect the F0 indices.

### 3.2 F0 indices

The collection of the F0 indices<sup>1</sup> corresponds to the measures of the values of F0 between the couples of labels. For each dialogue act and each labelled word, the

Minimum F0, Maximum F0,  $|\Delta$ F0| and mean F0 are measured and the values are reduced to a 4-level scale. Then 4 values are available for each word and each act, but only the results of the words are presented in this paper.

The indices and 4-level reduction have been observed and validated in early works (cf. [4]). The collection of all the F0 values has been tested through several methods such as the measure according to the regression line but the rough measures have given the best results.

<sup>1.</sup> Because the direction of the slope does not matter, the relevant information is better represented by the absolute value of  $\Delta$ F0. However in this paper  $|\Delta$ F0| is written as  $\Delta$ F0.

### 4. **RESULTS**

The model is being observed for 1 dialogue (about 15 minutes with 2 speakers). The values of the indices have been calculated in 1/8 tones for each speaker. The reduction to 4 levels has been automatically computed from the values of the labelled lexical words.

A first interesting result is that the regression line is not worth taking into account in spontaneous speech: the calculation does not occur on the same number of words from an utterance to the other.

At the moment, the values of 257 words have been extracted from the corpus, and the analysis is restricted to the categories which have the greatest population. They correspond to levels 2 and 3. These categories are (c) with a level 3, (e) with a level 2 and (g) with a level 3.

Calculated on the labels and their relevant indices (Maximum F0,  $\Delta$ F0 and mean F0), a function which measures the information gain brought by each index in relation with the model categorisation, indicates that the information gain is poor for mean F0, quite important for Maximum F0 and rather strong for  $\Delta$ F0. This information is interesting because it agrees with previous works on reading (with an intelligibility constraint), where not only  $\Delta$ F0 supports the greatest information but is the most common index used by readers in that kind of reading. On the whole, in the frame of reading and spontaneous speech, it appears that  $\Delta$ F0, calculated at the lexical level, is meaningful and efficient.

In those conditions, as a comment of the first results, only the behaviours of Maximum F0,  $\Delta$ F0 indices are observed.

Comparing the weight assigned by the CPD model with the measured values, it appears that:

- concerning the (c) words

- 66% of the values of the model are connected with the  $\Delta$ F0 values,
- 70% of the values of the model coincide with the Maximum F0 values,
- concerning the (g) words
- 83% of the values of the model coincide both with the Maximum F0 and the  $\Delta$ F0 values.

These categories correspond to level 3, where on the whole, the level of coincidence between model and indices reaches 70% for  $\Delta$ F0 values, and 73% for Maximum F0 values. Up to now however, level 2 remains not completely explained by the model, as the level of coincidence does not reach 50%, which is locally the case for the (e) words. Further investigation will be carried first on the method of calculation. For instance, about the F0 range, which level is relevant, that of the labelled words or that of the whole speaker's range?

## **CONCLUSION AND PERSPECTIVES**

On the whole, it seems that for reading as for spontaneous speech, the mean F0 index does not carry relevant information at the lexical level. The Maximum F0 and the  $\Delta$ F0 indices are more significant. Moreover both are strongly related as  $\Delta$ F0 is calculated from the minimum F0 and Maximum F0 values.

As a conclusion, for the categories under examination in this paper, it seems that when the information can be reached by reasoning (first reasoning of the concept or assisted by the map), the indices values are more connected than when the information has already been stored. Conversely, it appears that, if these results cannot be improved by the method, it is because the stored information includes different kinds of storage in the work memory, which have not yet been taken into account by the CPD model. Therefore the model will be modified in that perspective. Besides the model may also be improved by giving more importance to the dialogic aspect of the conversation.

### REFERENCES

[1] M. Akbar, "WaveEdit, an interactive speech processing environment for Microsoft Windows platform", Proc. EUROSPEECH'97, Rhodes, 1997 (to be published).

[2] M. Bessac, et G. Caelen-Haumont, "Analyses pragmatique, prosodique et lexicale d'un corpus de dialogue oral homme-homme", JADT 1995, Actes des IIIèmes journées internationales d'analyse de données textuelles, Vol. 1, Rome, 1995, Rome : Eurograf 2000, pp. 363-370, 1995.

[3] G. Brown, "Prosodic structure and the given/new distinction", in "Prosody: Models and Measurements", Edited by A. Cutler and R. D. Ladd, Berlin : Springer Verlag, pp. 67-77, 1983.

[4] G. Caelen-Haumont, "Stratégies des locuteurs en réponse à des consignes de lecture d'un texte : analyse des interactions entre modèles syntaxiques, sémantiques et pragmatique et paramètres prosodiques", Thèse de doctorat d'état : Lettres : Université de Provence Aix Marseille I, 2 Vol., 1991.

[5] G. Caelen-Haumont, "Prosodie et Sens : une étude expérimentale", Editions du CNRS, Coll. Sciences Humaines, Paris (to be published).

[6] N. Colineau, *"Vers une compréhension des actes de discours"*, Grenoble : Institut National Polytechnique de Grenoble, 1994, Rapport de D.E.A. Sciences Cognitives.

[7] E. F. Prince, "Towards a taxonomy of given-new information" In : Radical Pragmatics, Edited by P. Cole. New-York : Academic Press, pp. 223-225, 1981.

[8] E. F. Prince, "Fancy syntax and "shared knowledge", *Journal of pragmatics*, Vol. 9, pp. 65-81, 1985.