INTONATION OF SPONTANEOUS SPEECH IN FRENCH

Philippe Martin
University of Toronto, Toronto, Canada

ABSTRACT

Phonosyntactic models of intonation are usually inferred from read speech data, and bear no or limited expected validity for spontaneous speech. To investigate this commonly accepted assumption, we analyzed a sample corpus of spontaneous speech recorded in various interviews broadcast on a French radio station. Experimental results were confronted to theoretical predictions, pertaining to the encoding of the prosodic structure in French and to constraints governing the relationship between the syntactic and the prosodic structures. The data involve pitch accents characteristics (rising and falling contour distribution), stress and syntactic clash conditions, eurhythmy and number of syllables in the prosodic word.

1. INTRODUCTION

Spontaneous speech is often characterized by numerous manifestations of disfluency (breakdowns in the acoustic continuum, lengthening of non-significant units, phones and syllables, breaks in Fo and intensity lines, abrupt tempo changes, empty and filled pauses, errors in syntactic markers and lexical access, etc.). Despite these possibly perturbation effects, they are a priori no reasons to believe that spontaneous speech is not generated by the same grammar that governs other styles of speech production such as reading or reciting. Although many researchers state that prosodic prediction may be of limited applicability with spontaneous speech [6], this papers presents some evidence of the contrary, and attempts to propose some explanations accounting for the variety of spontaneous speech intonation.

The prosodic grammar used here differs somewhat from the commonly used theory derived from the work of Pierrehumbert and others [7], and used worldwide [1]. In particular, pitch accents are assumed to function as markers of a multi level prosodic structure, and are described by features that pertain strictly to this function. Other observed prosodic facts (such as tone boundaries, declination, etc.) are considered as phonetic variations, and are not part of the phonological description.

2. THE MODEL TESTED

The prosodic grammar tested for spontaneous speech differs from dominant North American varieties essentially by assuming that the prosodic structure is a hierarchy of prosodic words indicated by pitch accents (a prosodic word is a stream of prosody which contains one and only one stress). Boundary tones are viewed as phonetic facts, although in French, this difference is of little consequence as stressed syllables are usually in final position.

The prosodic grammar specifies that pitch movements accounting for the prosodic structure, and the relationship between the syntactic and prosodic structures are governed by specific constraints. Among those constraints, syntactic clash (preventing configurations as (A [B] [C] D), where A, B, C and D are prosodic words grouped by syntax into (A B) and (C D) and by prosody into (BC), stress clash, maximum length of the prosodic word and eurhythmy.

An important aspect of the theory is neutralization, where pitch accents can have their function in the prosodic structure suspended, and therefore can have some or all of their phonetic features not realized. A classical example of this in French is the neutralization of the interrogative modality contour when another interrogative marker is present in the sentence, as for example in

\[ A \text{ quelle heure mangez-vous ?} \]

Neutralized declarative pitch contour

where the final pitch accent can be falling or rising without changing the sentence interrogative modality. Indeed, the rising feature of the contour correlated with the interrogative modality is made redundant by the presence of the interrogative morphological marker quelle.

At the sentence level, the prosodic structure in French is indicated by a contrast in melodic slope to the right, which gives in the example below the following predicted sequence of pitch contours (in this representation, the assumed prosodic structure is represented by square brackets, and organizes hierarchically the prosodic words A, B, ..., N):

\[
\begin{array}{c}
\text{Neutralized contours} \\
\text{1st level rising} \\
\text{2nd level contour} \\
\text{Final falling}
\end{array}
\]

Neutralized declarative pitch contour

In this example, A and B, terminated by a falling contour, form a larger prosodic group with C, bearing a rising contour, E and F form another group ended by a rising contour; M and N form the last group terminated by the falling declarative contour. At the next level, ABC and EF, ended by a rising contour, form the sentence with the final group MN, ended with a falling contour.

Other functions are realized by pitch contours, to indicate
- the declarative or interrogative modality of the sentence, and their variants (evidence, doubt, command and surprise)
- the "propos-thème" division of the sentence (distinct from theme-rheme), as in c'est le lapin que j'ai acheté, with a falling pitch contour on the final syllable of lapin, vs. a falling contour on the final syllable of the sentence
- emphasis on one word, usually by a rising contour on the first syllable of the emphasized word.

More details on this prosodic grammar can be found in [2], [3], [4] and [5].

3. EXPERIMENTAL PROCEDURE

Five speakers were recorded on the French radio station France Inter during approximately 1-minute each. All speakers were interviewed about a subject they knew well, and the speech stream recorded was not interrupted by the interviewer. The recorded material was then processed by the signal analysis program WinPitch [10]. This Windows program is optimized for accurate pitch tracking and for phonetic measurements with a user-friendly interface. In fact the data were obtained and analyzed in real time on the Internet and gave excellent pitch curves despite the signal compression inherent to the transmission. With this software program, syllabic segmentation was particularly easy to achieve.

The following measurements were taken for each of the 5 speakers:
- Number of sentences
- Number of prosodic words in the sentence
- Presence of a Propos-Thème division
- Presence of emphasis, implication (evidence) contours
- Presence of prosodic parenthesis.

For each prosodic word:
- Duration
- Number of syllables
- Pitch contour (rising, falling, neutralized).

4. EXPERIMENTAL RESULTS

As differences in style of discourse were anticipated, experimental results were kept separate for each of the 5 speakers, CA, G1, G2, G3 and PV, and are summarized in table 1.

This table contains, for each speaker:
Phrase: number of sequences ended by a falling declarative modality contour
Syllable/PW: average number of syllables in one prosodic word
PW Duration: average duration (in ms) of one prosodic word
Number of PW: total number of prosodic words
Propos-Thème: number of propos-thème divisions
Level 1, 2, 3: number of prosodic groups belonging to different levels of the prosodic structure that display the expected melodic slope contrast
Neutralization: number of prosodic words with a flat pitch accent
Implication: number of pitch accents with an evidence bell pitch curve
Parenthesis: number of parenthetical sequences
Syntactic clash: number of syntactic clashes.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>CA</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrases</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Syllable/PW</td>
<td>3.7</td>
<td>3.5</td>
<td>3.3</td>
<td>3.6</td>
<td>3.5</td>
</tr>
<tr>
<td>PW Duration</td>
<td>661</td>
<td>690</td>
<td>635</td>
<td>651</td>
<td>624</td>
</tr>
<tr>
<td>Number of PW</td>
<td>67</td>
<td>70</td>
<td>68</td>
<td>34</td>
<td>66</td>
</tr>
<tr>
<td>Propos-thème</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Level 1</td>
<td>21</td>
<td>18</td>
<td>15</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>Level 2</td>
<td>26</td>
<td>13</td>
<td>8</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Level 3</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Neutralization</td>
<td>11</td>
<td>32</td>
<td>41</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Implication</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Emphasis</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Parenthesis</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Syntactic clash</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Prosodic data for each speaker

The first remarkable fact this table shows is the relatively long size of the sentences produced by all speakers (with the possible exception of PV), for a roughly equal total duration. The average duration, and the average number of prosodic words (or stressed syllables) contained in a single sentence is much longer than in written or read speech. Writers such as F. Celine among others have already noticed this long time ago, when they attempt to evoke spontaneous discourse in their writings.

The distribution of prosodic words is very similar for all speakers in terms of number of syllables. The average is 3.55 syllables per word, with an average duration of 653 ms. This distribution appears in Fig. 1.

![Distribution of Prosodic Words](#)

The overall distribution of average syllable length in function of the number of syllables of the prosodic word is given in Fig. 2.
The propos-thème division, expected to be a salient feature of spontaneous speech, is almost never used, except by speaker PV. Implication, manifested by a bell curve contour on either the sentence final or the first level group final syllables is used by almost all speakers, but parenthesis has been observed only 3 times. Finally, stress clashes are only found with speaker G3, in sequences where he uses only one level prosodic structure.

5. INTERPRETATION OF DATA

5.1. Modality
Declarative modality falling contours reach the lowest frequency values of the sequences, and are easily identified perceptually. Contrary to what was to be expected, they do not bear any implication variant (i.e. evidence in the declarative case), which are located instead on first level rising contours for all speakers. This could be explained by the large number of prosodic words in a single sentence: placed on the final contour, an evidence marker could only appear once at the end of a long sequence, whereas its presence on multiple first level contours appears much more efficient to signal evidence.

5.2. "Propos-Thème"
Propos-thème division of the sentence is less frequent than expected, as this specific use of intonation contours in French is often described as a characteristic of spontaneous, non-formal speech. Only speaker PV uses it 4 times, in relatively short sentences, with a theme containing only one prosodic word. This could imply that propos-thème division is only effective for short or medium length sentences, and is not felt appropriate by speakers for long sentences.

5.3. Emphasis
Again, emphasis is commonly expected to be a feature of spontaneous speech. Nevertheless, our random choice of speech streams reveal only one speaker (CA) using as much as 7 emphatic stresses, all located on the first syllable of multisyllabic prosodic words.

5.4. Prosodic Word
The number of syllables of the prosodic words is distributed as shown in table 1, with an average of 3.55 syllables per word, the majority of which have 2, 3 or 4 syllables. The distribution of prosodic word duration in function of their number of syllables appears in table 2. This table reflects results already described earlier [8], with an exponentially decreasing duration with the number of syllables, which makes French a syllable timed language only for sequences larger or equal to 2 unstressed and one stressed syllables. The larger duration observed for one syllable prosodic words is linked to the stress clash condition, always present since the proceeding vowel is necessarily stressed in French.

5.5. Pitch Contours
In French, stress is normally located on the last stressed syllable of the prosodic word, and is usually manifested by a longer syllable duration, and either by a pitch variation, or, in the case of neutralization, by a flat pitch. According to the theoretical predictions tested, the rising or falling variations are used to indicate the grouping of the prosodic word or group to larger units, ending with a pitch accent whose slope is the opposite of the pitch accent slope ending the prosodic word or group. All speakers did use this mechanism as predicted, but at different degrees, pertaining to the number of levels encoded in the prosodic structure. G3 for instance, used the simplest prosodic coding, which consists of a 2 imbedded levels of enumeration, irrespectively of the syntax. His second (out of 2) sentence was composed with 32 prosodic words, either ended by a rising contour, or composed with 2 or more prosodic words with a neutralized flat contour or a falling contour. By contrast, speaker CA uses a more sophisticated prosodic structure, implying sometimes the use of 3 levels of grouping, with realization of appropriate melodic contrasts at each level. As expected, rising contours pertaining to different levels of prosodic groups do contrast by the amplitude of melodic variation (reflected by the perceived stress level).

5.6. Prosodic Structure
Table 1 gives the degree of complexity of prosodic structures possibly reflected by the complexity of the syntactic structures of the sentences. Indeed, speaker CA is very familiar with interviews, and uses prosody to more effectively structure his discourse than speaker G3, who is at the other end of the scale of complexity. In this latter case, prosodic words, or groups of 2 words, are mostly enumerated, and those groupings are not congruent with syntax. By contrast, speaker CA carves his prosodic structure in 3 levels, in order to correspond closely to syntax, and help the bootstrapping process of syntactic decoding.

5.7. Syntactic clash
Again one could expect the occurrence of more syntactic clashes, which are only observed for speaker G3. The other speakers did not produce any clashes, and produced prosodic structures always congruent with syntax.

6. CONCLUSION
Although arbitrary by nature, this random set of samples of spontaneous speech exhibits interesting characteristics in terms of usage of a language specific prosodic grammar. As a bootstrapping device for syntactic decoding, the prosodic
structure appears clearly as a general purpose frame in which the speaker will attempt to insert and fit the preplanned and possibly constantly revised syntactic structure as the speech generation process takes place.

Experimental data illustrate the use of different strategies by the speakers (possibly under the influence of read speech style), to structure prosodically their sentences. In almost all cases however, the prosodic encoding mechanism, specific to French, is applied to hierarchically structure the speech stream.

REFERENCES