Models and Reality: From Intonation Theory to Intonation Curves

Philippe Martin
Experimental Phonetics Laboratory
University of Toronto
21 Classic Ave. Toronto, Ont. Canada M5S 2X6

ABSTRACT
There are a few theories of intonation to account for various phonological facts in different languages such as English, French, Swedish, etc. Depending on their respective framework (glossematics, generative or linear phonology, etc.), these theories provide some understanding of observable phenomena of acoustic or auditory nature, such as the evolution of fundamental frequency in the sentence, the perception of syllable pitch and intensity, and so forth. Nevertheless, the apparent large number of discrepancies that appear between the theoretically predicted facts and the physical reality often baffle researchers dealing with the acoustical nature of the speech signal. In their eyes, any theoretical insight on prosody may seem useless. The object of this paper is to discuss some of the reasons for these discrepancies through examples, and to provide a better understanding of what an intonation theory delivers.

A THEORY: WHY?

As for other linguistic facts, prosodic descriptions come in three different flavors: empirical, formal and axiomatic.

1) Empirical: The data collected through experimental phonetic analysis are processed and organized to find interesting correlation. Ex.: Experimental data show a positive correlation between the amplitude of the fundamental frequency variation and the number of syllables of read sentences.

Graphs are commonly used to represent the results of this kind of analysis;

2) Formal: A common sense principle governs the presentation and the interpretation of the experimental data. Ex.: For possibly unknown reasons, speakers uttering a long sentence have a tendency to differentiate their word accents with larger variations of pitch. Therefore larger variations of fundamental frequency are to be expected in long sentences, as (clearly) shown in the data.

Models are often used to represent this analysis, together with appropriate mathematical formula. In a sense, formalism appears to be a variant of empirical analysis.

3) Axiomatic: A (common sense) principle controls the description process and is used consistently to organize and interpret the data. Ex.: A prosodic structure does exist in the sentence and prosodic contours indicate its characteristics. A direct consequence of their existence is found in the amplitude of variation of the fundamental frequency in sentences with a large number of syllables.

Somehow, only the axiomatic approach delivers a theory, as scientific knowledge produced by empiricism and formalism is in reality governed by the selection of experimental data. In other words, the starting principle (the axiom) determines a set of discovery principles that allow the classification and the interpretation of the data. Here a set of statements (theorems) summarizes the results of the analysis.

A THEORY: WHICH?

Let us consider an axiomatic theory of intonation applied to French. The axiom postulates that a prosodic structure organizes prosodic words in the sentence, as the syntactic structure organizes syntactic units (words). If it exists, this prosodic organization is indicated by some prosodic markers belonging to the prosodic words. To track those markers down, we will consider a sentence containing only the smallest unit, i.e., one syllable. This syllable corresponds to the smallest prosodic word, which must contain (at least) one prosodic marker: the word accent. Prosodic markers are therefore equated to word accents, revealed by pitch contours. The number of prosodic words in a sentence is thus equal to the number of word stresses (not emphatic or any other kind of stress), which must be in contrast to each other to properly encode the postulated prosodic structure. By describing these contrasts with an appropriate set of binary features (for example), the pitch contours lose their abstractness by a choice of characteristics such as the slope of fundamental frequency (rising vs. falling), duration, intensity, etc.

In order to force the prosodic markers to reveal their inner mechanism, a technique commonly used is to consider sentences with no other hierarchical organization than prosodic. Common examples are structured enumerations like telephone numbers, multiplication tables, etc. The prosodic grammar found at that stage can then be used to analyze more complex cases where the sentence is organized by other mechanisms (e.g., syntax) (See more details in Martin, 1987).

WHY IT WORKS

Such a hypothetico-deductive approach leads to the definition and the description of phonological prosodic contours and their relationships in the prosodic structure. Some of the immediate consequences are:

1) Fo variations: If the number of syllables in the sentence increases, so does the number of prosodic words. In the case where the prosodic structure acquires more levels, so does the number of contrasts between pitch accents. Contrasts using Fo will then have a tendency to use more of this material to encode the contrasts, and manifest larger Fo changes on the stressed syllables;

2) Penultimate prosodic contour: The prosodic grammar of French uses contrasts in Fo slope to mark a dependence to the right, which ensures that the penultimate prosodic word should always bear a rising contour (in the case of declarative sentences with a falling final contour). Being obligatory, this feature is thus neutralized, as shown in most experimental data;

3) First level contours: These contours often end a noun phrase and are rising only if preceded by at least another prosodic word. Indeed, only in this case a contrast falling/rising is at work to mark this part of the structure. If no prosodic contour precedes a noun phrase pitch accent, the rising Fo feature need not be present;

4) Prosodic features: Many acoustical, perceptual, or articulatory parameters are plausible candidates for manifesting the prosodic contrasts between contours predicted by grammar (A pure axiomatic attitude such as glossematics would even manipulate only abstract prosodic features). Although experimental data suggests that Fo slope, Fo swing and syllable duration are the most commonly used in read speech, other styles reveal the use of different sets of features (whispered speech constitutes an interesting case since it substitutes syllable pause duration for Fo contrasts).

WHY IT DOES NOT WORK

Experimental observations made on read sentences or on spontaneous speech are not always easy to relate or even to contradict theoretical predictions. Indeed, many processes other than just intonation grammar encoding the prosodic structure are involved in the production of prosodic parameters in a sentence. Some of the discrepancies commonly discussed are:

1) Prosody of competence vs. theory of performance: Syntactic descriptions normally use the competence of the speaker, whereas, curiously, most prosodic studies analyze only its performance. Except in the case of very specific studies, no grammatical description of a language is build from spontaneous speech data. Nevertheless, most researchers in prosody seem to rely only on those data. It is therefore no surprise that an intonation theory may appear inadequate when confronted with spontaneous speech data. Most syntactic descriptions would appear inadequate as well if tested in the same conditions;

2) Establishing the prosodic structure: Another source of problems stems from the fact that, except in rare cases, we do not have access to the prosodic structure other than the marking mechanism performed by the prosodic contours themselves. We can only postulate a certain homomorphy with the syntactic hierarchy, tempered by specific prosodic constraints (Such as the avoidance of stress clashes and the planarity of the prosodic structure);

3) Stressability, stress and destressing: A third reason for discrepancies pertains to the elusive nature of the stressed characteristics of a syllable in French. Whereas the stressability of a syllable can be clearly assessed, its effectively stressed or destressed nature is much more difficult to establish. Therefore, expected theoretical predictions on some contour features, say its rising Fo variation, rely on the effectively stressed nature of the corresponding syllable. This difficulty is somewhat alleviated by the (potentially dangerous) use of external insight provided by auditory of articulatory phonetics to help in establishing the stressed character of a syllable (see for example Di Cristo, 1980, on this subject).

4) Other prosodic features: Numerous other prosodic features are not taken into account by the theory: declination line, first syllable pitch movements, emphatic stress, etc. (See for instance Mertens, 1989, on this.) Again, this results from the choice of prosodic features that do participate to the marking of the structure.

APPLICATIONS

These considerations suggest that speech synthesis and language teaching are two typical applications that would benefit from a theoretical axiomatic approach. Text-to-speech synthesis obviously proceeds from the competence level, whereas speech recognition, dealing with spontaneous speech, can only make little or no use of a grammar of intonation.

CONCLUSION

By definition, an axiomatic approach sticks to its principles in its discovery process. This hopefully constitutes a guarantee that the results of the experimental analysis do reflect the starting hypothesis. All descriptions of prosodic contours, contrasts in Fo slope, syllable duration, etc., produced by the discovery process correlate with the existence of a prosodic structure. Obviously many other prosodic facts that are revealed experimentally and that do not participate to the assigned function are absent from the phonological description. In assessing the importance of those facts when considering specific applications leads to either change the theorems or to postulate other axioms.

REFERENCES

A. Di Cristo (1980), "Variabilité acoustique et intégration perceptive des cibles prosodiques", *Actes des XIèmes Journées d'Études sur la Parole* (Strasbourg), pp. 91-110.

Ph. Martin (1987), "Prosodic and Rhythmic Structures in French", *Linguistics,* Vol. 25, pp. 925-949.

P. Mertens (1989), *L'intonation du français, de la description linguistique à la reconnaissance automatique*, Doctorale dissertatie, Leuven Univ. , Dept. Linguistieck.