INTONATION AND SYNTAX: ANOTHER POINT OF VIEW

INTONATION IN THE (LINGUISTIC) SYSTEM

When we speak, when we read, even silently, a musical movement inevitably accompanies our words, constituting sentence intonation. For a long time, linguists have relegated the tune of the sentence to the study of emotions and social attitudes, with the exception of a very limited role given to intonation in the language system, namely the indication of the declarative or interrogative modality of the sentence.

The emergence of industrial applications of linguistic studies, such as text to speech synthesis and automatic speech recognition raised new and challenging questions in which the role of sentence intonation could not be ignored further. As a result, intonation, which was considered as resulting from a pure emotional and attitudinal mechanism, has now to be evaluated in the light of possible interactions with linguistic entities encoded by intonation.

In fact, this is not very different from the other phonological objects such as vowels and consonants, which obviously have to be produced through muscular activities interacting with the speaker general emotional state.

Most recent studies on intonation and emotions are heavily depending on empirical data. Typical studies merely comment various statistical studies on global or local aspects of the data, such as the average laryngeal frequency, the variations of tempo, and the like. Doing so these efforts maintain the old tradition, denying phonological character to intonational objects. But intonation is as linguistic as vowels and consonants. Its minimal units are sensitive in their realization to the emotional state of the speaker, yet they maintain a rigorous system of contrast with each other as other phonological units do.

HISTORICAL DEVELOPMENTS - ToBI OR NOT ToBI

Most if not all recent studies of sentence intonation adopts the theoretical views developed by Pierrehumbert-Beckman (ref), including the ToBI set of notation tools to describe intonation events. We will depart from this dominant approach to propose a different way to describe sentence intonation in languages such as French, Italian, Spanish, Romanian, English and European and Brazilian Portuguese.

Since the ToBI approach is so well spread, particularly in North American studies conducted not only on American English, but also on British English, Australian English, Dutch, Japanese, German, Greek, and even French (ref), we will devote a large part of

this first chapter on the theoretical and practical implications of the use of the ToBI notation for sentence intonation.

In short (the interested reader can find a complete reference and training material in ref), ToBI, which stand for "Tone and Break Index", proposes a set of symbols to transcribe prosodic events in the sentence in four tiers:

ToBI's HIGHS AND LOWS

- A tonal tier to represent pitch events;

- An orthographic or phonetic transcription tier ;

- A break indices tier, where perceived speech breaks (pauses) are judged on a scale from 1 to 4, relatively to the perceived coherence of units involved;

- A miscellaneous tier.

The tonal tier contains melodic events transcribed as H high and L pitch levels, and are usually attached to pitch accents and to (syntactic) boundaries (*boundary tones*). Phonetically, High and Low tones are interpreted as melodic targets for the actual acoustic fundamental frequency (laryngeal frequency) movements in the sentence. Phonetic variants include:

For stressed syllables:

H*   *peak accent*;
L*   *low accent*;
L* +H   *scooped accent*;
L + H* *rising peak accent*;
H +!H* *downstep*.

For sentence stress:

L-   *low, intermediate constituent boundary tone;*
H-   *high, intermediate constituent boundary tone;*
!H-   *high and downstepping.*

Boundary tones

L%   *low and sentence final;*
H%   *high, at the end of an intermediate constituent;*
%H   *high at the beginning of the sentence.*

So in practice, the melodic events get represented on the tone tier by a sequence of High and Low tones, and their variants. In the following example (taken from ref):
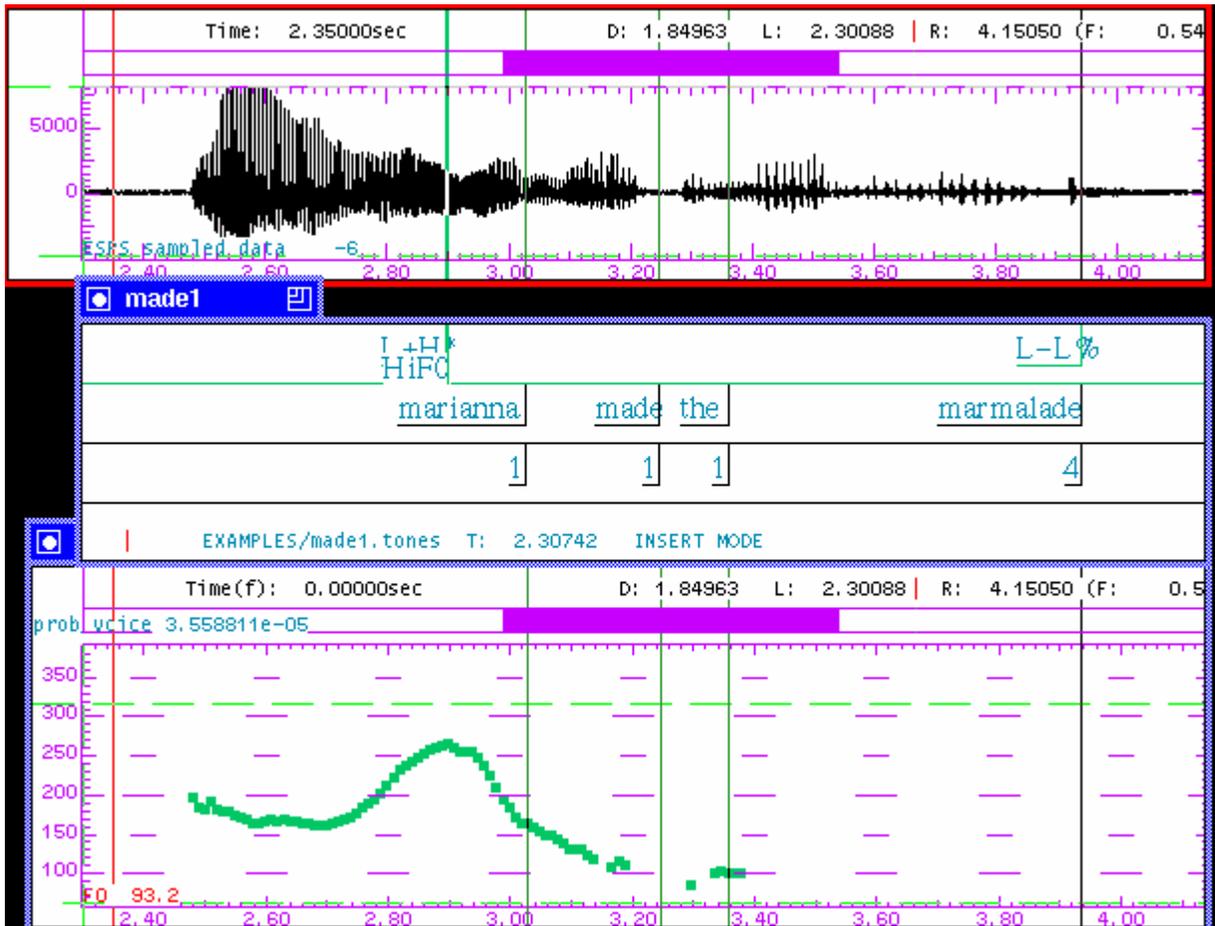

**Figure 1**: An example of a ToBI transcription

The black curve on top corresponds to the speech signal; the green curve at the bottom represents the fundamental frequency (an acoustical measure of the laryngeal frequency from the speech signal) and gives a visual aspect of the melody of the sentence; the middle image shows the four tiers used in ToBI, L +H*   L-L% for the tone sequence, "Marianna made the marmalade" the text, "1   1   1   4" the perceived coherence at the group boundaries. The last tier contains comments. (from ref).

WHAT'S WRONG (IF ANYTHING) WITH TOBI?

Although extremely popular, to the point that prosodic studies not using it are often neglected and discarded from the research mainstream by the research community, the ToBI description on which phonological models of intonation are based, suffers from similar drawbacks as API transcription. First, it mixes phonetic and phonology in its use:

as H and L tones are assigned on stressed syllables and syntactic boundaries, the assignment of course relies on the phonological premises that these points in time are a priori important. Equally not phonologically neutral appears the alignment of sequences such as *H*, L*, L*+H, L+H** and *H+ !H** on narrow and broad focus stressed syllable. Obviously, the use of ToBI symbols according to their definition implies the acceptation of the underlying phonological theory which is part of their definition.

Purely phonetic representation by ToBI is also debatable. Consider the following examples of melodic variations:
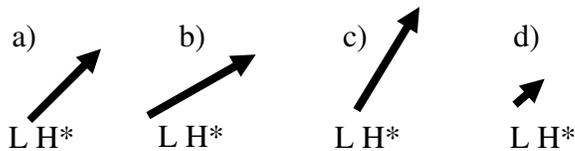
All the melodic variations a), b), c) and d) are transcribed by the sequence LH*, whereas the amplitude and the duration of melodic variations are different in each case.
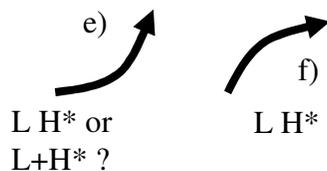
By the same token concave and convex melodic contours such as e) and f) could receive the same representation LH*. Obviously there is an underlying justification to consider these contours as identical, but this justification is closely linked to the phonological model in hand, and again is never explicitly given at start.

This is a classical problem in phonology: should the description precede the explanation, or the explanation the description? In practice, scientific activity proceeds by navigating back and forth between description and explanation, but eventually, and this is particularly true in theoretical physics, presents its final results in an hypo-deductive way, as data are actually revealed through hypothesis, and therefore support them. Applied to ToBI, this means that the principle of transcription should result from the principles of description. In a sense, this is the case, as for instance H* and L* have to be assigned to stressed syllables (but stress can be a phonological as well as a phonetic concept, and as phonetic can be claimed as universal). By contrast, the assignment of boundary tones is clearly linguistic and not phonetic, as obviously based on some morpho-syntactic definition of units ended by boundaries.

Examples taken from a variant applied to Dutch and called ToDI (ref) show that opinionated and undercover phonological vision can produce some "data twisting" to force data to fit into the theoretical model. For instance, Dutch sentences presented

always start with %L, whereas the melodic curve is actually falling, flat or rising:

a)　　　　　　　　　b)

%L　　　　　　　　　%L

Indeed, the melodic contour starts from a low level in the first case a), and reaches a low level in b), which justifies the notation %L,

%L　　H*L　　(H*L)　　　　　　H*L　　　　　　　L%
Om de　juiste　kruiden en　　　　planten te　vinden

%L　　H*L　　　　　　　　　　　　　　　　　　L%
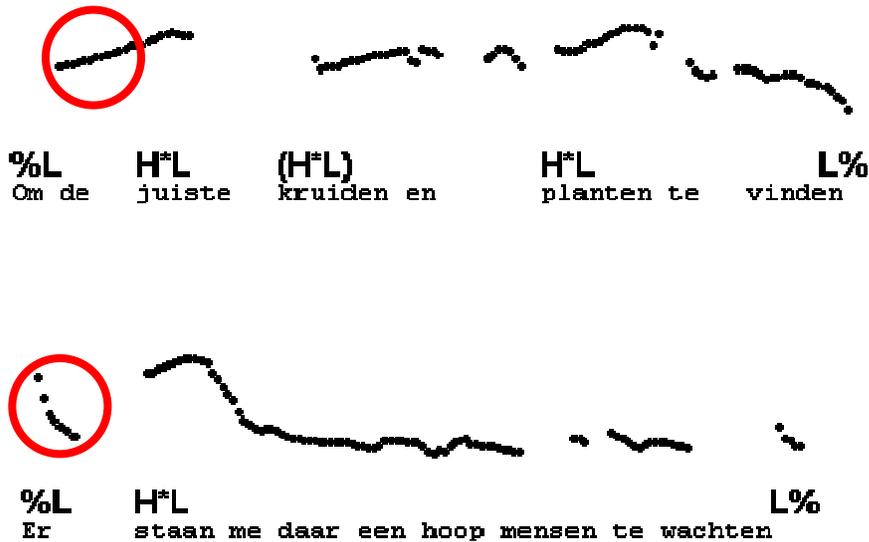Er　　staan me daar een hoop mensen te wachten

Figure 2: Two distinct fundamental frequency contours (circled in red) get identical ToBI transcription.

An example of two different fundamental frequency (melodic) curves at the beginning of a sentence (circled in red), the first rising and the second falling, treated as identical by ToDI as %L (From ref)

Studies in intonation have been the domain of phoneticians for a long time. Traditionally indeed many phonologists denied any linguistic role other than semantic to intonation (Mathesius, 1929, see Rossi, 1999 for discussion), whereas phoneticians, encouraged by the advent of sophisticated acoustical analysis tools, felt that intonation could be a legitimate and worthwhile target for their investigations. Consequently, strong empiricism characterized most phonetic studies of intonation, whose typical approach consisted of collecting large amount of acoustical or perceptual data according to somewhat vaguely defined criteria. Statistical tools helped then to reduce these data to more manageable sets, while auditory simulated process would convey some perceptual validity. Finally an appropriate formal model ensures a further reduction of the data complexity. The model is considered satisfactory if capable of regeneration of all the data obtained at the start of the process.

It is not too hard to prove that this kind of research activity is by essence circular: such models appear ultimately as a representation of the experimental data, and as such do not convey any other theoretical truth than the data themselves (Badiou, 1969). As the data are gathered according to some empirical principle, the model that gives account of these data is empirical as well. Literature on intonation research shows abundant examples of these two faces of the empirical process of scientific activity. Dominant theories for instance would build explanatory models on transcriptions made according some independently defined principles, such as ToBI (Beckman and Ayers, 1997). The adequacy of these models refers to the data that were used to build them, and those data and their collection are therefore central in the process of phonological description.

Figure 3: Flow of discovery process in a typical empirical approach

The hypothetical-deductive approach proceeds differently: the description is extracted from data according to the hypothesis made at start. The flow of discovery process is thus
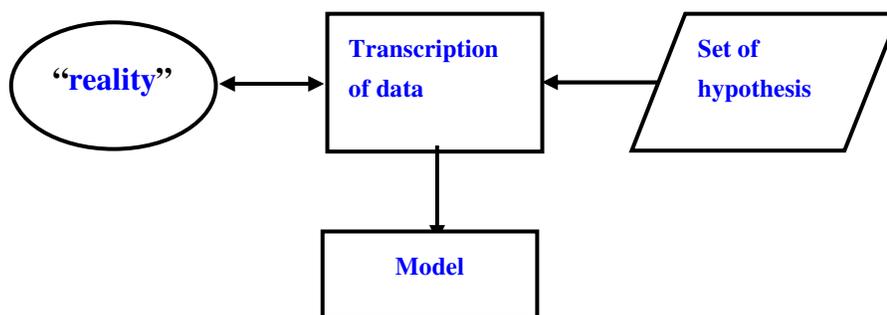

Figure 4: Flow of process in a hypothetical-deductive approach

With this approach, pertinent data are "extracted' from the experimental material according to a set of rules and principles derives from the starting hypothesis, which acts as a "grid" to retain and filter data pertinent to the starting principles.

As stated above, an intonation model is an instantiation of a hypothetical-deductive process. To approach the method progressively, let's start with the simple and well-known correlation existing between intonation and the sentence modality.

The utterance establishes a specific relationship between the speakers and the other participants of the speech act. This relationship, the modality, can be a priori classified according to various grids and classes, from the simplest involving declaration and interrogation, to more complex ones involving subtle degrees of social relationship, of speech act context, etc. In most linguistic systems, various markers, syntactic, morphologic, as well as the tone of voice indicate sentence modality. Recent studies (Cresti, Martin, Moneglia, 1999) discuss a more general approach to modality, imbedded in a general theory of speech acts.

Let's consider the simplest case, where sentence modality can be either declarative or interrogative. So the classes of relationship between the speaker and the audience are reduced to either delivering information or requesting information. The traditionally imperative modality is therefore considered here as a variant of declaration (the lack of proper imperative form in the verb system in French gives other arguments to consider imperative as a variant of declaration).

Consider the following example *tu viens* in some neutral context:

Declarative   ***tu viens.***   Interrogative: ***tu viens ?***

The absence of other markers (syntactic, morphological, or present in the context or in the situation) forces intonation to function as the only marker of the declarative or interrogative modality. We can then expect to discover some significant differences between the two intonation contours correlated with these modalities, differences manifested in the prosodic data. To find out which are the features involved, we have to turn to experimental data (of course our example is trivial, and the result has been known for a long time). If we are not willing or can distinguish between declarative and interrogative contours, we can then ask for the help of modern technology, i.e. to the acoustical analysis of the sentence, which should reveal quickly where the differences lay. Using the program WinPitch described above, we obtain typical displays of fundamental frequency in function of time, as given in fig. 7.
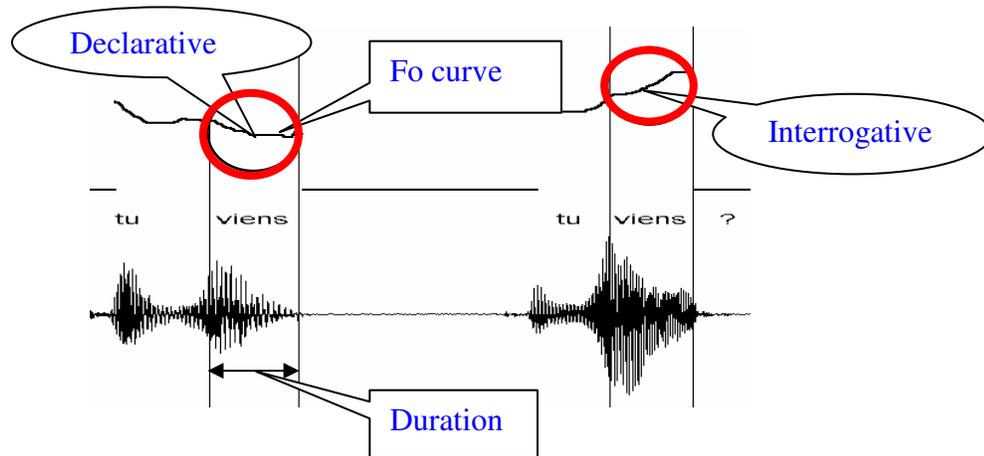
Figure 5: Fundamental frequency curve for interrogative modality contour

At this point, we introduce an important constraint in our way to look at the data. We will introduce a "filter" that would extract from acoustical fundamental frequency, intensity and duration only segments corresponding to (effectively) stressed syllables. Those parameters have been shown for a long time to be correlated with stress in most languages, and stressed syllables are the central feature of a unit introduced later, the prosodic word. In fact, stress is central as always present even when the sentence is reduced to a minimal form with a sequence of syllables containing one stress, or just one syllable, necessarily stressed.

Our simple example has two syllables, and the last one (as French phonology predicts) is stressed. Opposing the acoustic manifestations of the declarative and interrogative modalities, the most prominent acoustic feature appears to be fundamental frequency, falling in the declarative case and rising for interrogative. If Rising is chosen as marked feature of the modality contour, we have then the following simple system:

Declarative	Interrogative
- Rising	+ Rising

As all other phonological markers, the modality contour can be neutralized in it's function if another marker of modality is present in the sentence. This is the case if the so-called imperative morphological form is used for the verb, as in *viens* and when the inversion subject-verb or the *est-ce que* locution is used to indicate the interrogative modality of the sentence **viens-tu** ? **est-ce que tu viens ?**

It's function being suspended as redundant, the melodic contour does not have to manifest the feature + Rising, as shown on the following acoustical curve
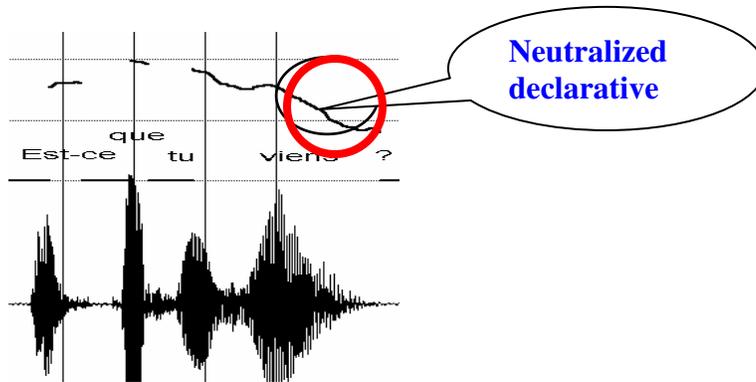

Figure 6: Fundamental frequency curve for neutralized interrogative modality contour

References

Badiou, A. (1969). *Le concept de modèle*, Maspéro, Paris.

Beckman, M. and Ayers, G. (1997) *Guidelines for ToBI Labeling*,
    Web site http://www.ling.ohio-state.edu/~tobi/

Beckman, M. and Pierrehumbert, J. (1986) "Intonational Structure in English and
    Japanese", *PhY*, 3, 255-310.

WinPitch (1997) http://www.winpitch.com