Intonation's many functions

Philippe Martin
French Studies
University of Toronto

Abstract

For the non-linguist, intonation is often thought of as a feature of human voice which conveys emotions and attitudes. Yet Intonation has another – although considered marginal by linguists – role in speech: to indicate the declarative or interrogative sentence modality and its many variants (whose linguistic status is often debatable). But intonation is also heavily conditioned by linguistic rules, specific to each language, respected by the speaker even in the most severe physical conditions of speech production.

Intonation in the (linguistic) system

 When we speak, when we read, even silently, a musical movement inevitably accompanies our words, constituting the sentence intonation. For a long time, linguists have relegated this music of the sentence to the study of emotions and social attitudes, with the exception of a very limited part given in the language system, with 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 other phonological objects such as vowels and consonants, which obviously have to be produced through muscular activities interacting with the speaker general emotional state. Their realizations follow linguistic rules but reflect as well the emotions and attitudes of the speaker.

Still, most recent studies on intonation and emotions are heavily depending on empirical data, and not on any phonological insight. Typical studies merely comment various statistical study on global or local aspects of intonation, such as the average laryngeal frequency, the variations of tempo, and the like. By 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 - such as pitch contours located on stressed syllables - 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.

To speak is to imitate

Indeed, we learn to speak by imitating surrounding speech behavior, and from this activity global and local grammar rules making us capable to produce speech sequences we never heard before. Many of our social activities, if not all (sexual activity is the most notable exception as models to imitate are normally not freely offered) are acquired by a similar process.

Walking is a good example: the human child has innate possibilities to learn walking. Yet, the actual implementation of walking is clearly done by imitation. Whereas all humans move by walk by putting one foot in front of the other, and displacing the body weight from one foot to the other, the actual details of these complex movements are specific to each social group, as perspicacious children on Greek beaches have been observing for a long time: English walkers do not walk like Americans or French do.

This is comparable to any realization of gesture in human activity, and speech production as a succession of complex gestures is not an exception. This is why phonology was clearly established as a separate science from phonetics. The technique to avoid confusion between both in the description is very simple, and based on the concept of function. To go back to our example of walking, the phonological description would retain only what satisfies to the given function (i.e. to move the body by walking), whereas the phonetic description would give details of the gestures and muscular activities involved. Unfortunately, contemporary phonology, particularly in the domain of intonation, has lost this perspective in its scientific production, and mixes both levels, making difficult and opaque the actual phenomena involved.

To speak is to submit

To speak is a submissive activity, physiological and social:

-    Physiological as the speech source results from a controlled recuperation of the expiration activity in the human respiratory cycle, the resulting energy going to a complex filtering and resonance process in the vocal tract, or in the production of mini-explosions (for stop consonants).

-    Social as the primary speaker energy (emotional) directly linked to the body physiological state is filtered by social conventions, local in the speech act, or group at large in the overall speaking process.

In this sense, speaking as any other human socialized activity results from the filtering (shaping) of primary impulses which have to undergo social filtering

but which nevertheless constitute the only source of energy. What is remarkable is that, like in the example of walking, some large liberty is given to each individual as long as the social constraints are respected. Going over these constraints leads to various sanctions, from pair mockery to severe legal penalties.

To speak is a demanding physical exercise

Speech is a constrained activity: to produce speech sounds, we use organs designed for others function. The lungs produce the airflow necessary to make noise, but are primarily there to provide enough oxygen to the blood vessels. The larynx vocal folds function essentially to prevent food to enter accidentally the respiratory tube, and sometimes to provide a better stiffness in our body structure when we lift heavy objects. The tongue, the teeth and the lips provide all the mechanisms essential to food ingestion. When we speak, all these organs are used in another mode, and interact with their primary mode if we conduct the primary activity when we speak.

We all know it is impolite (and inefficient) to speak while eating, but how about breathing? Indeed we have to maintain the respiratory cycle when we speak. Since the two processes are incompatible, we switch continuously between speaking and breathing, trying to accommodate both activities concurrently. For breathing we modify our respiratory cycle: in normal mode the cycle gives an equal time to inspiration of air and expiration, in speech mode, the inspiration time is minimal, and the expiration maximal. Since speech is essentially produced during expiration, this change in cycle will allow us to speak longer, between two intakes of oxygen.

On the other hand, our speaking activity has to be modified as well, and will be constrained by the quantity of air we accumulated in our lungs during the inspiration process. To reduce the perturbing effect of inspiration on speech, we learn during language acquisition where to place the inspiration – and the speech interruption – on the time scale, and how much air to take in, as govern by our prediction of what we need to say what we plan to say next. During our speech turn, we place these interruptions – the so called respiratory pauses – at strategic places in our discourse, where they can be confused – or hidden – by silences normally used to mark the end of a large syntactic group, or of a sentence.

Speaking is thus a very demanding, if not disturbing, physical activity, as it prevents us to feed and even to breathe normally, and can be only ideally performed when we are at a total rest. Therefore any physical activity, any physical implication of our body will influence the way we produce speech, as will any emotional fluctuation.

Intonation is a serious thing, it has a structure

Popular knowledge of intonation gives many roles to the music of the sentence: convey emotions, attitudes, and so on, marginally the ability to indicate the declarative or interrogative modality of the sentence, but rarely gives any importance to the indication of the structure, syntactic, semantic or other.

Therefore, the musical activity of the speaking subject is felt as mainly extra linguistic, and as such free to follow dominant non-linguistic mood and attitudes of the speaker. It can be shown that on the contrary, while speaking, a system of intonative contrasts (specifically in terms of syllable duration and melodic variation) has to be maintained to carry out the linguistic activity and ensure the success of the information encoding and decoding.

This is true even in adverse speech production conditions, such as created by intense physical activity, extreme emotional distress, etc., when these conditions prevent us to meet the numerous physical (muscular) constrains involved in speech production. We all know some of these conditions by experience, when we run out of breathe after a long run or swim, of while crying following a severe emotional shock. Of course all speech production units such as the phoneme realizations will be affected in those conditions, but the most likely to be disturbed will be the elements whose production involves time and rhythm.

The emotional state is of course reflected on the acoustic properties of speech along the whole sentence. Anger with intense activity will involve a higher sub-glottal pressure with higher fundamental frequency and pitch variations; sadness and depression with a low physical activity will produce a lower intensity, lower frequency and reduced pitch movements, and so on. Laugh is probably the only mode where speech production is impossible, as it takes over completely the respiratory control necessary to regulate phonation airflow.

As all other phonological units, such as vowels and consonants, intonation contours located on (or around) stressed syllables take part in a double process:

1. A process of phonological contrasts, in which each unit has simply to be sufficiently different from any other unit that can appear in its place.

2. A process of conformity or imitation, by which each unit has to emulate, imitate, as close as possible to a model given by the dominant socio-geographic model of the speaker.

Phoneticians have been generally dealing with the second process, trying for instance to describe empirically the detailed acoustical features of some regional variety of intonation. Phonologists, at least those with some structural background, aim to explain the minimal set of features that should be found in speech contours in order to fulfill its assumed function(s).

Dominant North American intonation researchers, however, tend to mix both attitudes, as they generally start as phoneticians to get a sufficient empirical description, and then as phonologists to find rules that would generate all the well-formed sequences of contours in a sentence. Of course, acting this way, and using an a priori notation system such as ToBI to represent empirical data, their approach is somewhat biased from the beginning, as they will only take into account in their phonology the data retained in their representation system in the first place. But let's not tell them as such contradiction makes them very unhappy and can lead at times to violent reactions.

For our part, we will start the opposite way, in giving roles – functions – to sentence intonation which should appear as reasonable hypotheses, but as such which cannot be proved (in mathematical logic they would be called axioms). These hypotheses pertain to the following linguistic content:

1. The prosodic structure of the sentence, a hierarchical structure that might resemble to the syntactic structure, but which differs from it in many aspects. The units organized by both structures are not identical – the prosodic structure handles prosodic words (or stress unit) with one and only one (non emphatic) stress), and the hierarchical organization of corresponding stressed and syntactic units is not necessarily congruent.

Despite this, the prosodic structure is associated to the syntactic structure with the following constraints:

1. A stress clash condition, which prevents two stressed syllables to be adjacent (unless they are separated by enough "space", such as a pause or a complex articulatory sequence (ex.: a [skp] sequence of consonants);

2. A syntactic clash condition, which forbids two minimal prosodic units (i.e. stress groups) to be reunited in the same higher prosodic group if they are dominated in the syntactic tree by two distinct nodes;

3. A maximal number of syllables conditions, which forces stressable syllables to be effectively stressed if the number of consecutive non-stressed exceeds a number of 7 or 8 (depending on the actual speech rate produced by the speaker). This conditions applies to languages such as French, where stressable syllables are not necessarily stressed);

4. An eurhythmic rule, which either a) tends to balance the number of syllables in prosodic groups at the same level in the prosodic structure (at the expense of its congruence with the syntactic structure), or b) modifies the rhythm of same level prosodic groups by accelerating groups with a larger number of syllables and slowing down those with a fewer number of syllables. The congruence with the syntactic structure is then not affected in this case.

The following example taken from French will perhaps help to clarify these functions and constraints.

Pour la première fois de sa vie le général a décoré son meilleur chameau

(For the first time of his life, the general has decorated his best camel)

All the rules given above lead to a final result, among all possible sequences of melodic contours. Once the stress assigned to specific syllables, the principle of contrast of slope (which applies only to French) defines on each stressed syllable a pitch contour with a slope inverse to the slope of its immediately dominant contour on its right. In a group (A B), the pitch contour on A will be rising if B is falling, and falling if B is rising. Applied iteratively on the above example, assuming that the prosodic structure is congruent o the syntactic structure, we have:

(Pour la première fois) (de sa vie) (le général) (a décoré son meilleur chameau)
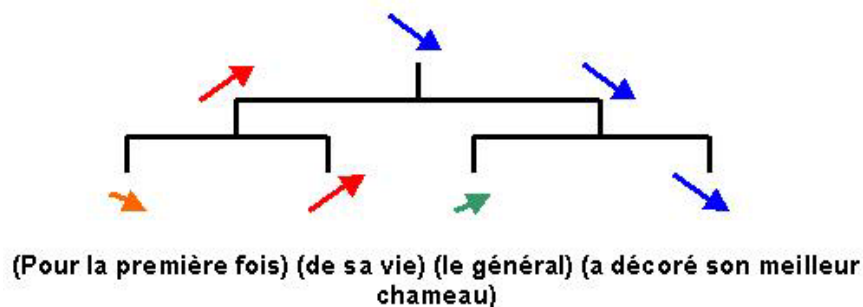
First the dominant contour of modality is assigned to the root of the prosodic structure (red arrow). Secondly, at the left first level node a rising contour (in red) opposed to the dominant blue modality contour is assigned. Finally we assign at the last level, a falling contour (in brown) dominated by the red rising contour and a rising contour (in green) dominated by its right neighbor.

The resulting sequence of pitch contours assigned to effectively stressed syllables of the example is thus fall, rise, rise and fall. A further constrain, linked to the stress levels, differentiate by the level of pitch amplitude (i.e. the excursion range of the fundamental frequency on the voiced part of the syllable) contours of same slope which could appear at the same place (i.e. on the same stressed syllable) encoding thus another prosodic structure. The first falling contour for instance must be different from a falling declarative modality, and the second (red) contour has to contrast with the third (in green) as located at another level in the prosodic structure.

This sequence of melodic contours is specific to French. An Italian equivalent of the example, showing a very similar syntactic structure, would give totally different results (Martin, 1999).

The many features of prosodic contours

Although limited in numbers, more than one prosodic feature can be used by the speaker in a specific linguistic system to encode the system of contrasts existing in the prosodic structure, first to ensure the specificity of stressed syllables opposed to un-stressed, second to indicate the hierarchical organization of the prosodic "chunks" in the sentence.

In order of frequency of use, we have:

- Syllable duration, specifically used to contrast un-stressed vowels with stressed ones, and frequently to signal the final stressed syllables among the non-final ones;

- Fundamental frequency movement, rise of fall;

- Syllable intensity;

- Fundamental frequency level, contrasting with the average level in the sentence along the declination line;

- Specific pitch movements along the overall contour in the stressed syllable, such as a contour final limited fall for rising contours, and limited rise for falling contours;

- Vowel quality, modifying slightly (on considerably in the case of languages such as English or European Portuguese) the quality of the vowel;

- etc.

While speaking, some of these features may be too expensive to use due to a severe competition with physical or emotional constrains. A depression state of the speaker is a classical example, as reflected by a slower speech rate and reduced pitch change in the course of the sentence. All stressed prosodic contours will show a very limited frequency excursion or be completely flat, and duration will be the most important feature used to contrast prosodic contours in the sentence.

Another example is whispered speech. No vocal fold vibration takes place and voicing of stressed syllables cannot be performed. Stressed syllables contrast then by some remaining features, such as duration and intensity.

Opposite examples pertain to speech production in noisy environments. The speaker then can exaggerate the coding of stress, using all possible features allowed in the linguistic system, large frequency excursions and duration contrasts, possible use of idiosyncratic pitch movements, etc. An extreme example is given by singers of pop music, who have to invent original pitch variations on their stressed (and un-stressed) syllables in order to stand out vocally from the usually very loud background and develop a proprietary voice style easily recognizable.

Intonation many hats

In summary, as for any other phonological unit such as vowels and consonants, the speaker has to comply with linguistic rules as well as with imitation rules. Linguistic rules are followed in order to convey the information encoded in the discourse by helping the listener in the syntactic parsing process by a appropriate prosodic structure, and imitation rules are followed to mark the belonging of the speaker to a specified social and geographic group.

References

- Fonagy I. (1983), *La Vive Voix*, Paris, Payot.
- Léon P. (1993), *Précis de phonostylistique, parole et expressivité*, Paris, Nathan.
- Martin, Ph. (1999) Prosodie des langues romanes : Analyse phonétique et phonologie », Recherches sur le français parlé, Publications de l'Université de Provence, 1999 No 15, pp. 233-253.
-
- Martin, Ph. (2000) "L'intonation en parole spontanée", "Revue Française de Linguistique Appliquée", Vol. IV-2, pp. 57-76, Paris.