Multi-lingual modelling of intonation patterns

Daniel Hirst*, Albert Di Cristo*, Martine Le Besnerais**, Zohra Najim*, Pascale Nicolas* & Pascal Roméas*.
*Institut de Phonétique d'Aix, URA CNRS 261 Parole et Langage,
Université de Provence, Aix en Provence, France
**Universidad Autónoma de Barcelona, Spain

ABSTRACT

An account is given of an ongoing research programme developping a general method of investigation applicable to the intonation systems of different languages. At present, four languages in particular are being studied. : French, English, Spanish and Arabic. Four levels of prosodic representation are distinguished : underlying phonological, surface phonological, phonetic and acoustic. Each level is required to be interpretable in terms of the immediately superior and inferior levels. Automatic and semi-automatic procedures are being developped for this task.

INTRODUCTION

In this paper we present preliminary results of an ongoing research programme developping a general method of investigation applicable to the intonation systems of different languages. At present, four languages in particular are being studied. Three of these are Indo-European languages : French (Romeas 1991; Di Cristo in press; Nicolas 1992 in progress), English (Hirst in press), and Spanish (Alcoba et al. 1993; Murillo & Alcoba in press; Le Besnerais in progress), while one is a non Indo-European language : Arabic (Benkirane in press, Najim & Hirst in press; Najim in progress).

These four languages present interesting differences in their prosodic systems. From a phonological point of view it has been claimed that the intonation systems of English and French differ in at least two parameters.

P1. Whereas English accent groups are traditionally held to be organised into left-headed structures, with initial prominence (Pike 1945, Jassem 1952, Abercrombie 1964), accent groups in French are regularly structured the opposite way, into right-headed structures with final prominence (Wenk & Wioland1982; Hirst & Di Cristo 1984).

P2. A second parameter concerns the tonal sequence usually found on phrase internal Tonal Units (=accent groups) in the two languages. It has been argued (Hirst & Di Cristo 1984, in press) that while this tonal sequence is most commonly [H L] in unmarked declarative utterances in English, in similar French utterances it is usually [L H].

There is some evidence (Hirst & Di Cristo in press) that the first of these two parameters, P1, might in fact characterise a difference between Germanic and Romance languages in general. From this point of view a comparison of results obtained on English and French with those obtained on Spanish provides an interesting control of this prediction while the inclusion of a non Indo-European language should make it possible to see how far such a parameter can be applied to languages from other linguistic phyla.

Levels of representation and description.

At one extreme we may distinguish an abstract level of cognitive representation (phonological) and at the other the level of observation of physical data (acoustics, physiology etc.). It has been argued (Hirst 1992) that between these we should distinguish at least two intermediate levels, the level often referred to as "phonetic transcription", which is in fact a level of surface phonology, and the essentially hybrid level of phonetic representation, where *phonetics* is taken to constitute the interface between the cognitive (phonological) and the physical (acoustic) levels.

Each level of description is required to satisfy the *Interpretability Constraint* : each level *i* must be able to be interpreted on both levels *i+1* and *i-1* when such levels exist.

One of the most general questions we hope to be able to address in this research is to what extent the prosodic variability which is to be observed between different languages can be attributed to language specific parameters on each of the different levels.

Phonological representation of intonation patterns

While the exact nature of such a representation is of course unknown we assume,as in a number of recent models, that an intonation pattern can be derived from a phonological structure to which language specific templates associate appropriate tonal segments (for discussion cf Hirst & Di Cristo in press).

Phonetic modelling of F0 curves.

A number of different techniques have been developped in recent years for automatically generating fundamental frequency patterns of synthetic speech from symbolic input. Less research has been devoted to the inverse problem : the automatic coding of fundamental frequency patterns by symbolic output. There have been even fewer attempts to "model" such patterns where the output of the automatic coding is directly usable as input for the automatic synthesis system. Such an automatic modelling system (described in more detail in Hirst & Espesser in press) has recently been developped in our Institute. The output of the programme MOMEL is a sequence of target points <Hz, ms>, which we claim constitute an appropriate phonetic representation of the F0 curve. These target points can be used to generate a quadratic spline function giving a very close approximation to the smooth continuous F0 curve observed on fully sonorant segments of speech (Hirst 1980). The residual micromelodic profile can be stored separately and then added to the quadratic spline to obtain very high quality speech synthesis (Di Cristo & Hirst 1986).

Surface phonological modelling.

The target points obtained from the programme MOMEL are coded symbolically using the INTSINT transcription system (Hirst & Di Cristo in press). According to this system (used in half of the chapters in Hirst & Di Cristo (eds) in press), tonal segments are assumed to be of two types : absolute tones[T(op) M(id) B(ottom)], whose phonetic interpretation is assumed to be independent of the immediately preceding tone; and relative tones : [H(igher) L(ower) U(pstepped) D(ownstepped) S(ame)] whose phonetic interpretation is assume to be dependent on that of the immediately preceding tone.

A number of different options for automatically coding target points are being investigated. Among these are the use of a threshold for the distinction between absolute and relative tones; the use of syntagmatic constraints to ensure that only D and U are used for cases of iterative lowering or raising while T and H or B and L always correspond to peaks or valleys respectively. A further possibility is that of establishing an optimal coding of the target points, minimising the difference of variance between the model and the observed data.

The investigation of the association between the tonal segments and the segmental transcription raises a number of interesting issues and is at present in a very preliminary stage. Various models making use of different temporal parameters (word boundaries, syllable boundaries, vowel onset etc) are being explored.

Phonetic interpretation of surface phonology.

The phonetic interpretation of the symbolic coding of the F0 target points can be obtained by statistical analysis of the original dataset. Absolute tones are thus modelled as the mean value of the corresponding target points; relative tones are modelled by linear regression on the preceding target point, irrespective of its code. (Hirst, Nicolas & Espesser 1991). It remains to be seen what is the optimal domain for such interpretation.

Universal and language specific characteristics of the symbolic coding and the phonetic interpretation of such modelling will be investigated for the four languages.

Preliminary results

- Durational characteristics of Spanish

In order to test for evidence whether Spanish accent groups are organised into left-headed or right headed structures, measurements of syllable duration were made on three corpora : a list of declarative sentences, a portion of a continuous text and a portion of a continuous spontaneous monologue. Among the factors tested were position in the word, stressed/unstressed nature of syllable, number of syllables in the left-headed/right-headed foot, number of syllables in the left-headed/right-headed restricted foot (where a "restricted foot" is assumed not to cross word boundaries, cf the "narrow rhythm unit" of Jassem 1952, Jassem et al. 1984). Figure 1 shows the summed unexplained variance for four models : left-headed/right headed (LH/RH) unrestricted/restricted (/-R) accent groups. In all cases the variance of syllables not included in the relevant foot structures were added to the unexplained variance of the modelled structure so that the summed unexplained variances are comparable across models . The right-headed foot structure gave the best fit for each corpus, followed by the restricted right-headed foot-structure (Alcoba et al. 1992). The model based on the number of syllables in the word (not illustrated here) gave a worse fit for all three corpora.

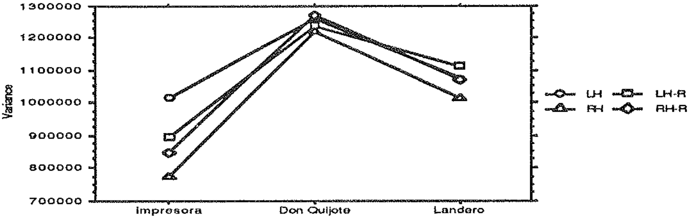

Figure 1 : Unexplained variance for four models of accent groups : Left/Right-headed [LH/RH], unrestricted/restricted [/-R] for three corpora in Spanish.

- Small corpus pilot study

A small corpus of 20 sentences was recorded for each of the four languages. The F0 curves of the sentences were modelled using MOMEL with manual correction when necessary. The target-points were then coded as INTSINT symbols. For temporal alignment of the tonal segments the onset of the stressed vowels was labelled as well as the word boundaries. It remains to be seen how far such a sparse labelling of the corpus, together with information concerning the number of syllables, is sufficient to account for the temporal localisation of the target points in the different languages. In particular the way in which these parameters can be related to the Left/Right-headed structures referred to above is of particular interest.

By way of illustration, the following figure shows the result of the procedure applied to one of the sentences from the Arabic corpus. :

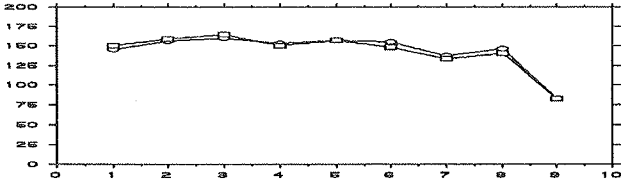

Figure 2 : Observed (circles) and predicted values (squares) for the Arabic sentence 'Kataba lwaladu alkabiiru risaalatan' (The big boy wrote a letter). The corresponding (automatically derived) INTSINT coding was [M U T L H D L H B].

The automatically derived quadratic spline targets (circles) were coded as the INTSINT sequence [M U T L H D L H B]. The statistical values of each tonal segment were then calculated on the complete corpus as described above and the predicted values were then derived (squares).

CONCLUSIONS

We have outlined a general methodology which is at present being developped for the comparative analysis of the intonation systems of different languages.

ACKNOWLEGEMENTS

This research was partly financed by the Franco-Spanish *Programme D'Action Intégrées* as part of a three-year project between the Universities of Provence and Barcelona entitled *Prosodie comparée du français et de l'espagnol contemporains.*

REFERENCES

Alcoba, S.; Di Cristo, A.; Hirst, D.J.; Le Besnerais, M.; Murillo, J. & Roméas P. (1993) *Rapport Final sur l'Action Intégrée Prosodie comparée du français de de l'espagnol contemporains.* (Unpublished report, Université de Provence).

Di Cristo (in press) "Intonation in French." in Hirst & Di Cristo (eds) in press.

Di Cristo, A. & Hirst, D.J. (1986) "Modelling French micromelody : analysis and synthesis." *Phonetica* 43, 11-30

Hirst, D.J. & Di Cristo A. (eds.) (in press) *Intonation Systems a Survey of Twenty Languages.* (Cambridge University Press; Cambridge)

Hirst, D.J. & Di Cristo A. (in press) "A survey of intonation systems" in Hirst & Di Cristo (eds) (in press).

Hirst D.J. & Espesser R. (in press) "Automatic modelling of fundamental frequency." *Travaux de l'Institut de Phonétique d'Aix*, 15.

Hirst, D.J. (1980) "Un modèle de production de l'intonation." *Travaux de l'Institut de Phonétique d'Aix* 7, 297-311.

Hirst, D.J. (1992) "Prediction of prosody : an overview." in G.Bailly & C.Benoît (eds) *Talking Machines : Theories, Models and Applications.* (Elsevier Science Publishers)

Hirst, D.J. (in press) "Intonation in British English" in Hirst & Di Cristo (eds) in press.

Hirst, D.J.; Nicolas, P. & Espesser, R. (1991) "Coding the F0 of a continuous text in French : an Experimental Approach." *Proc. ICPhS 12* (Aix), 5, 234-237.

Jassem, W. (1952) *Intonation of colloquial English.* (Panstowe Wydawnictwo Naukowe; Warswawa).

Jassem, W.; Hill, D. & Witten, I.H. (1984) "Isochrony in English speech : its statistical validity and linguistic significance." in Gibbon & Richter (eds) Intonation : Accent and Rhythm. (de Gruyter, Berlin)

Le Besnerais, M. (in progress) *Parámetros rítmicos para el estudio contrastado del francés y del español contemporáneos.* (Doctoral thesis, University of Barcelona).

Najim, Z. & Hirst, D.J. (in press) "Codage prosodique d'un corpus d'arabe littéral lu par un locuteur marocain." *Travaux de l'Institut de Phonétique d'Aix*, 15.

Najim, Z. (in progress) L'intonation de l'arabe littéral parlé au Maroc : analyse historique et expériementale. (Doctoral thesis, Université de Provence).

Nicolas, P. (1992) "Emergence de la structure intonative du texte lu en français." Actes du Séminaire Prosodie (La Baume lès Aix, October 1992), 103-112.

Nicolas, P. (in progress), *Apport de la prosodie à la parole de synthèse : cas du texte lu en français.* (Doctoral thesis, Université de Provence)

Romeas, P. (1991) *L'organisation prosodique des énoncés en situation de dialogue homme-machine simulé : théorie et données.* (Doctoral thesis, Université de Provence).

Wenk, B.J.. & Wioland, F. (1982) 'Is French really syllable-timed?' *Journal of Phonetics* 10 (2), 193-216.