Semi-automatic tagging of intonation in French spoken corpora

Estelle Campione & Jean Véronis

Equipe DELIC
Université de Provence
29, Avenue Robert Schuman, 13100 Aix-en-Provence (France)
Estelle.Campione@up.univ-aix.fr, Jean.Veronis@up.univ-mrs.fr

Abstract

The transcription of spoken corpora using the punctuation of the written language is far from satisfactory. On the other hand, the manual transcription of prosody is an extremely time-consuming activity, which requires highly specialised experts, and is prone to errors and subjectivity. Automating the prosodic transcription of corpora can be interesting both in terms of effort reduction, and in terms of objectivity of the markup. Full automation is not achievable in the current state of the technology, but we present in this paper a technique that automates critical steps in the process, which results in a substantial annotation time reduction, and improves the objectivity and coherence of the annotation. In addition, the necessary human phases do not require a highly specific training in phonetics, and can be achieved by syntax students and corpus workers. The technique is applied to French, but most of the modules are language-independent, and have been tested on other languages.

Keywords: spoken corpora, annotation, prosody, intonation, French

1. Introduction

The transcription of spoken corpora is a difficult issue. It has been noted many times that transcriptions using written language punctuation are unsatisfactory and misleading (Blanche-Benveniste & Jeanjean, 1987; Leech, 1997), since the set of written punctuations is far from parallel to that of prosodic phenomena in speech. Leech (1997:90) calls the transcription of spoken language using ordinary orthography (and written punctuation) "a pseudo-procedure the only excuse for which is that it would be prohibitively expensive to attempt anything else".

Because of this inadequacy, some teams like ours have developed transcription conventions that do not make use of any of the written punctuations. For reasons of feasibility, these conventions usually mark only a very limited subset of prosody phenomena. In our case, for example, a minimalist stance has been taken and only pauses are marked (Blanche-Benveniste & Jeanjean, 1987; Blanche-Benveniste, 1990). However, this type of transcription is not entirely satisfactory either, because ambiguities appear in the resulting "text". In French, for example, most discourse markers can belong to another category and fulfil another function. Very often meaning and context are not enough to find the correct interpretation, which requires prosodic clues. In the example below, *quoi* can be a discourse marker (more or less similar to *you see* or *I mean*), but also a pronoun (*what*). The unpunctuated fragment, *je ne sais pas quoi* can therefore be interpreted either as *I don't know what* or as *I don't know, you see*:

> écrire un un petit euh **je sais pas quoi** un petit recueil qui qui explique comment les étapes qu'il faut suivre

Another common ambiguity consists in "floating" segments, usually complements, that can be attached to what comes either before or after in the utterance (see Bilger *et al.*, 1997):

> elle arrive moi je m'en vais **à une demi-heure près** on travaille pas ensemble

In a noticeable proportion of cases, the interpretation spontaneously adopted by corpus users is the wrong one. Sometimes they do not even notice the double reading.

The only satisfactory solution would be to faithfully transcribe the prosody of utterances[1], but this was attempted only for a handful of corpora (e.g. the London-Lund Corpus or the Lancaster/IBM Spoken English Corpus). Prosodically transcribed corpora are totally lacking for most languages (such as French). The reasons for this shortage lies in the difficulty of prosodic transcription stressed by many authors. It is highly time-consuming and requires from the annotators a type of phonetic-oriented competence that is not common among syntax scholars. In addition, the very subjective nature of prosodic labelling reduces the trustworthiness of the results or requires careful control by counter-experts thus increasing the cost yet more.

Automating the prosodic transcription of corpora can be interesting both in terms of effort reduction, and in terms of objectivity of the markup. Full automation is not achievable in the current state of the technology, but we present in this paper a technique that automates critical steps in the process, which results in a substantial annotation time reduction, and improves the objectivity and coherence of the annotation. In addition, the necessary manual steps do not require a highly specific training in phonetics, and can be achieved by syntax students and corpus workers. The technique is applied to French, but most of the modules are language-independent, and have been tested on other languages.

2. Overview

Many schools have developed prosodic theories and annotation schemes and there is no consensus on a transcription system. It has even been said that each new monograph introduces a different coding system (Hirst, 1979; Mertens, 1990:159). ToBI (Silverman *et al.,* 1992) has gained a wide popularity for American English, but it is not easy to adapt to other languages, even to other varieties of English (Nolan & Grabe, 1997; Leech, 1997). ToBI labelling also relies on linguistic judgements made by experts and is consequently difficult to carry out automatically, although research in that direction is underway (Wightman & Ostendorf 1992; Ostendorf & Ross, 1997).

More importantly, ToBI-like systems are too detailed for many linguistic purposes, especially for syntactic studies. Leech (1997:89) makes a distinction between *spoken language corpora* and *speech corpora,* the former consisting of usually large, naturally occurring samples of continuous language or discourse, the latter referring to "laboratory speech", usually words or sentences out of context. While a "narrow" transcription can be useful and even necessary for speech corpora, and phonetic studies, a high density of fine-grained prosodic symbols (typically one or more per word) is likely to be difficult to read and to blur the important facts in spoken corpora. As far as intonation is concerned, for example, most of the smaller melodic movements are the result of local phonotactic rules and contraints (such as the number of syllables in a word or the position of the lexical stress), and do not reflect communicative choices from the speaker (for French, see for example Hirst & Di Cristo, 1984; Di Cristo, 1999a and b). The following utterance, for example, can be either conclusive:

(la maison du voisin)$_{\mathbf{L}}$
[*the neighbor's house*]

or continuative:

(la maison du voisin)$_{\mathbf{H}}$

In the first case, the speaker uses a final falling tone (**L**) to show that he/she is temporarily finished, and that this is a place for the interlocutor to take over. In the second case, the speaker marks his/her intention to continue by means of a final rising tone (**H**), which implicitly invites the interlocutor not to interrupt.

In both cases, the inner movements of the sequence are controlled by the syntactic and lexical organisation of the utterance, and the speaker has no real choice concerning them (if no word is accented):

(la maison) $_{\mathrm{H\ L}}$(du voisin) $_{\mathbf{H}}$

If the utterance becomes longer, phonotactic rules impose a breakdown into smaller, less prominent prosodic groups:

---

[1] Of course we do no claim that it would solve all ambiguities in spoken corpora (no more than punctuation solves all ambiguities in the written language).

(la maison) $_{\text{H L}}$(du fils) $_{\text{H L}}$(du voisin) $_{\mathbf{H}}$
[*the neighbor's son's house*]

It seems to us that, apart from being the only feasible approach, a broad prosodic transcription, marking only the major prosodic events is sufficient and more readable for most uses of spoken corpora.

We will be as theory-neutral as possible, and simply consider that utterances are composed of consecutive *prosodic segments*, delimited by pauses and large pitch movements. These segments can be prosodically autonomous or depend on their neighbours. In the example above, we made the implicit assumption that the segment *la maison du voisin* was isolated from the rest of the discourse, for example by a long pause (marked --) :

(la maison du voisin)$_{\mathbf{H}}$ --

However, the same sequence of words with the same rising intonation could be prosodically coupled with the next segment, as in the following dislocation :

(la maison du voisin)$_{\mathbf{H}}$ (elle a brûlé) $_{\mathbf{L}}$ --
[*the neighbour's house, it burned down*]

Prosodic segments must therefore be grouped together into *prosodic units.* Prosodic units have an internal prosodic cohesion, and are independent from each other in the discourse flow.

Our goal is therefore to detect the prosodic segments, to tag them and group them into prosodic units. Five steps are involved in the process:

(1) pauses are automatically detected in the speech signal;

(2) the fundamental frequency or F_0 curve is stylised in order to eliminate its smaller, irrelevant details;

(3) the stylised curve is reduced to a sequence of discrete symbols encoding the pitch movements;

(4) the recording is orthographically transcribed and synchronised to pauses and major pitch movements;

(5) the sequence of melodic movements is filtered and translated to a final prosodic coding.

The orthographic transcription is entirely manual. The other steps are automatic, but require some hand correction, as summarised in Figure 1.

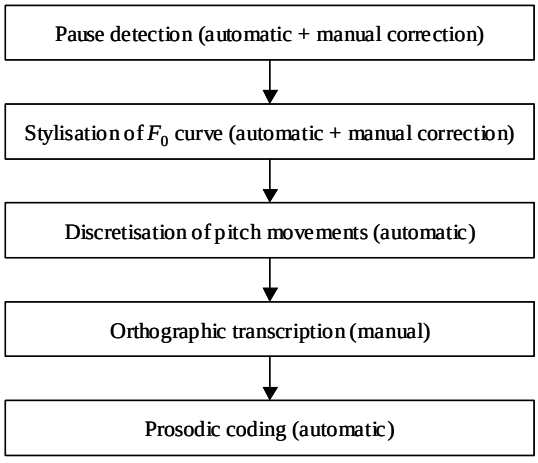

Figure 1. Overview of transcription and annotation process

3. Pause detection

The first step of the annotation process consists in an automatic detection of (silent) pauses. As simple as it may seem, this task is far from straightforward to perform by automatic means. Long pauses can be interrupted by ambient noise (frequent outside laboratory conditions). On the other hand, very short

pauses are extremely difficult to distinguish from plosives. For example, some very brief breathing pauses can be as short as 60 ms, which is shorter than many occurrences of plosives.

We use a pause detector based on F_0 detection, which behaves reasonably well in terms of robustness to ambient noise. Using a threshold of 350 ms, very few false detections occur. However, some very short pauses are not detected, and must be added by hand. The task is not particularly difficult, with the help of a graphic editor that enables the visualisation of the signal and playback segment by segment. Typically the manual correction takes ca. one hour for a 15 minutes recording, and does not require a highly specialised phonetic expertise. At the end of this phase, silent pauses are categorised into three groups:

(1) very short pauses (< 350 ms, coded ^)

(2) short pauses (≥ 350 ms and < 1.5 s, coded -)

(3) long pauses (≥ 1.5 s, coded --).

Even if no subsequent prosodic treatment is planned, this technique leads to a much greater reliability in pause transcription than direct transcription from a tape recorder, and is therefore advisable in any spoken corpus work. Pause transcription is a very difficult exercise when done entirely manually, and we have noticed that most linguists, even highly competent ones, tend to miss many pauses, especially when they are coupled with other phenomena (such as hesitation or syllable lengthening). For example, in the fragment that we will use as a running example in the next sections:

> on fait pas que le pressing on fait aussi la blanchisserie **plus la blanchisserie d'ailleurs** - les draps les nappes la restauration

the pause (marked with a dash) was missed by the (skilled) linguists who transcribed and verified the corpus.

The part in bold is three-way ambiguous. The form *d'ailleurs* is either a locative adverb (*from elsewhere*) or a discourse marker (similar to *actually*) which can be attached to what comes before or after. The three interpretations are therefore:

> 1. *We don't do only dry cleaning, we also do the laundry* ***plus laundry from other places****: sheets, tablecloths, catering...*
> 2. *We don't do only dry cleaning, we also do the laundry.* ***More the laundry actually****: sheets, tablecloths, catering...*
> 3. *We don't do only dry cleaning, we also do the laundry.* ***More the laundry. Actually*** *sheets, tablecloths, catering...*

The pause is an important clue for the disambiguation, but is not sufficient in itself. The intonation contour must be taken into account.

4. Stylisation of the F_0 curve

F_0 curves can be seen as the combination of a macroprosodic component governed by syntactic and pragmatic rules, which reflects intonational intention of the speaker, and a microprosodic component which is entirely dependent on the effects of the particular phonemes in the utterance (lowering of F_0 for voiced obstruents, etc.). Stylization consists in extracting the macroprosodic component from the F_0, while the microprosodic component is factored out. Various stylization methods have been proposed since the sixties (Cohen & t'Hart, 1965; t'Hart, Collier, & Cohen, 1990; D'Alessandro & Mertens, 1995; Fujisaki & Hirose, 1982; Taylor, 1994; etc.), and rely on more or less complex models.

The method used in this work (MOMEL, standing for *MOdélisation de MELodie*) was proposed by Hirst & Espesser (1993) (see also Hirst, Di Cristo & Espesser, 2000). It has some appealing features compared to other methods:

(1) it is language-independent;

(2) it does not require any pre-segmentation of the signal (e.g. in syllables);

(3) it does not require any training on the data;

(4) it performs automatically with a very good success rate;

(5) the stylised curve is perceptually undistinguishable from the original.

The technique consists in reducing the intonation contour to a series of target points, which represent the relevant pitch movements (Figure 2). Once interpolated by a quadratic spline curve (unvoiced segments are interpolated so that the resulting curve presents no discontinuities), the series of target points produces an F_0 contour perceptually undistinguishable from the original, apart from a few detection errors that must be corrected by hand. A quantitative assessment showed that the algorithm produces about 5% of errors. A large part of these errors (approximately 3%) were moreover systematically of two or three different types, in particular missing targets in transitions from voiced to voiceless segments of speech, which suggests that an improved algorithm could probably eliminate the majority of them (Campione, forthcoming).

Again, the correction is easy to perform and does not require a specialised training. The signal can be played segment by segment, and the original can be compared to the version re-synthesised using the stylised curve. If the two differ perceptually, the target points can be moved via a graphic interface until the re-synthesis is judged similar to the original. The correction phase takes around one hour for a 15 minute recording, as for the pause detection. For the moment, the two are done separately due to the use of different tools, but we plan to integrate them in the future, thus reducing the total correction time substantially.

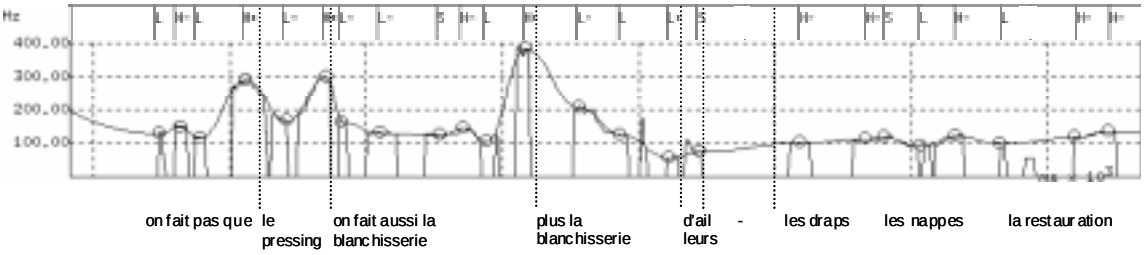

Figure 2. Stylisation and discretisation of the F_0 curve

5. Discretisation of pitch movements

The next step consists in converting the target points into a sequence of discrete symbols encoding the pitch movements. The requirements we set for this operation are as follows:

(1) the set of symbols should be as small as possible;

(2) it should be possible to generate from the sequence of symbols an F_0 curve perceptually undistinguishable from the original;

(3) it should be language independent.

After experimentation with several coding systems, including INTSINT (Hirst, 1991; Hirst & Di Cristo, 1998), we have developed a mathematical model that enables a reduction of the initial curve to an alphabet of 7 symbols without substantial loss of information. The model is based on the observation that the distribution of target points is approximately normal (Campione & Véronis, 1998). Details of the model are outside the scope of this paper, but the reader can find a description in Véronis & Campione (1998).

The alphabet of symbols is as follows:

L+ large falling movement;

L medium falling movement;

L- small falling movement;

S very small or null movement;

H- small rising movement;

H medium rising movement;

H+ large rising movement.

These symbols have no phonological value, and consist only in an extremely compact representation of the F_0 curve. We showed, by an evaluation on a large multilingual database (4 hours 20 minutes of speech, 50 speakers, 5 languages), that the encoding enables regeneration of ca. 99% of points at less than 2 semi-tones than the original (Véronis & Campione, 1998). The F_0 curves re-generated from the encoding using the mathematical model are therefore virtually undistinguishable perceptively from the original.

The model has interesting properties. In particular, movements of the same amplitude in (semi-tones) do not necessarily have the same coding, depending on the place at which they occur in the speaker's range, thus reflecting the fact that pitch variation towards the extremes requires more articulatory effort than pitch variation in the speaker's medium area. As a consequence, the model also predicts the downdrift effect which is actually observed in speech (Figure 3) without requiring a specific downdrift parameter.

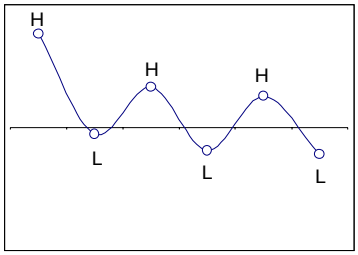

Figure 3. Downdrift effect

6. Orthographic transcription

The speech signal, already segmented at pauses (see section 3), is further segmented at large pitch movements (coded **H+** or **L+**). In the example above, four new breakpoints are inserted (three **H+** and one **L+**)[2], thus delimiting six segments:

_____ (**H+**) _____ (**H+**) _____ (**H+**) _____ (**L+**) _____ (**pause**) _____ (**pause**)

The corpus is then orthographically transcribed, using a graphic interface that enables playback segment by segment. During this phase, two additional prosodic phenomena are manually encoded, but *only when they occur at the end of a segment*:

(1) accents (coded *)

(2) final syllable lengthening (coded :).

Another important information results from the transcription itself, and consists in filled pauses (hesitations) which are transcribed as special lexical items (*euh*).

These cues are necessary for the correct interpretation of pitch movements in the last step (see section 7). In the example above, the segment by segment transcription yields:

on fait pas que* (**H+**)

le pressing (**H+**)

on fait aussi la blanchisserie (**H+**)

plus la blanchisserie (**L+**)

d'ailleurs (**pause**)

les draps les nappes la restauration (**pause**)

Orthographic transcription using this strategy is less time-consuming than the usual technique with a tape recorder, and more reliable. The pre-segmentation in small units, that can be replayed at will, facilitates the transcriber's task, and helps avoiding errors (missing hesitations, repeats, etc.).

[2] This is not the norm. For the sake of brevity, the example was chosen because it contained several interesting phenomena in a short time span.

Transcribing and correcting a 15 minutes recording takes about two hours, as opposed to three hours with no assistance.

7. Prosodic coding

We distinguish two types of prosodic events, depending on their scope. A first type of prosodic event is of a punctual nature: it consists of a change of pitch or a plateau at the end of the segment, regardless of the sequence of pitch movements in that segment. In the final segment of our example:

> les draps les nappes la restauration

the rising movement on the last syllable of *restauration* (**H-**) is enough to indicate continuation (in the context of the following pause). The sequence of preceding movements in the segment consists of a rising-falling alternation, governed by syntactic and lexical constraints.

Other prosodic events bear on an entire segment. In this case, the normal rising-falling alternation is replaced by a continuous rising, falling or flat sequence on the entire segment. In the example above, the sequence

> plus la blanchisserie

is composed only of falling movements (**L- L L+**). It has a communicative role since it indicates, in this particular case, that the speaker has an afterthought and corrects what she has just said. It is confirmed by the next segment

> d'ailleurs

whose flat intonation (coded **S**) is typical of a discourse marker that "tags on" the preceding segment. In French, the combination of a downstepped intonation followed by a plateau is a common strategy for self-correction or retrospective message modification. The available prosodic information (including presence of accent) therefore disambiguates among the three interpretations listed in Section 3, in favour of the following:

> *We don't do only dry cleaning, we also do the laundry.* ***More the laundry actually****: sheets, tablecloths, catering...*

The first step of the algorithm marks the prosodic events at the end of each segment using three codes:

- ↗ rising (**H-**, **H**, **H+**)
- ↘ falling (**L-**, **L**, **L+**)
- → flat (**S**)

Example :

> les draps les nappes la restauration ↗

Braces indicate that a prosodic event bears on an entire segment (for segments of more than two syllables):

> {plus la blanchisserie}↘

The output of this first step on our example yields:

> on fait pas que* ↗ le pressing ↗ on fait aussi la blanchisserie ↗ {plus la blanchisserie}↘ d'ailleurs → – –
> les draps les nappes la restauration ↗ – –

The second step of the algorithm consists in grouping together the prosodic segments into prosodic units, or, to put it another way, to detect the prosodic boundaries between units. This step requires that all the available clues are taken into account:

(1) pitch movements;

(2) silent pauses;

(3) accents;

(4) final syllable lengthening;

(5) filled pauses (hesitations).

The interaction of these clues is complex. For example, in French, a falling movement is perceived as a conclusive boundary when it is followed by a short or long pause, but not if it is preceded by a syllable lengthening, or a filled pause. On the other hand, rising movements followed by a short or long pause mark continuative boundaries, even if they are preceded by a syllable lengthening, or a filled pause.

In order to maximise readability, prosodic units are separated by paragraph marks. Each prosodic unit is preceded by its start time in seconds. In addition, redundancies are removed. Since accents always appear with a rising intonation, the arrow following the accent mark (*) is removed. In the same way, flat or falling movements before an internal pause are not marked (unless they bear on an entire segment), because they can be interpreted only as hesitation. An example of output on a larger sample is given Figure 4.

```
L1
      0.0   voilà
L2
      2.9   ben je travaille dans un pressing ↗
            –
      4.1   on fait pas que* le pressing ↗ on fait aussi la blanchisserie ↗ {plus la blanchisserie}↘
            d'ailleurs – les draps les nappes la restauration ↗
            ––
      17.2  on fait beaucoup de colonies beaucoup de: – de choses comme ça on travaille pour la
            police pour la gendarmerie euh – on travaille pour beaucoup de monde ↗
            –
      24.1  on a beaucoup de marchés ↗ donc c'est pas évident ↗
            –
      26.6  {parce qu'il y a des jours où il y a:} ↘ – pas de boulot ↗ il y a des jours où il y a du
            boulot ↗
            –
      29.4  comme partout ↘
            –
      31.5  donc on est deux ↗
            –
      34.2  moi et ma collègue Hayat ↗
            ––
      36.0  on s'entend bien ↗ on a une bonne ambiance dans l'entreprise donc je pense que c'est
            quand même assez: – assez bien ↗
            –
      41.6  {quand il y a une bonne entente}↘ parce que le boulot faut faut reconnaître on n'y va pas
            par plaisir ↗
            ––
      45.1  on y va par obli*gation – euh donc euh – moi je touche à aux deux ↗ à la blanchisserie
            et au pressing ↗
            ––
      60.1  parce que ma collègue n'a pas la la qualification au niveau du pressing donc c'est pour
            ça qu'elle y touche pas pour le moment ↗
            ––
```

Figure 4. Example of prosodic transcription

8. Conclusion and future work

The strategy and algorithms outlined in this paper enable semi-automatic transcription of prosodic information in spoken corpora. The transcription aimed at is a “broad” prosodic transcription, in which only the major prosodic events are annotated. We claim that a broad transcription is more suitable for corpus-based syntactic and pragmatic studies, since most of the smaller melodic movements are the result of local phonotactic rules and lexical contraints. A “narrow” annotation (in addition to being impossible to carry out on a large scale) would be difficult to read and unnecessary for many purposes.

At the moment, several separate tools are used in the various phases of the transcription. An obvious direction for further development would consist in integrating these tools into a single “prosodic

annotation workstation". This would reduce the transcription time substantially, since the two phases of manual correction (pause detection and F_0 stylisation) could be merged, and accomplished during the orthographic transcription itself. We estimate that, using an integrated environment, the transcription and correction time of a 15 minute recording could be reduced to about three hours. This figure is similar to the time currently required for the orthographic transcription and correction alone using a simple tape recorder. Therefore, it seems possible to add very useful prosodic information to spoken corpora at little or no extra cost. In addition, our strategy provides a segment by segment alignment of the transcription with the audio signal, which can be useful for many purposes (e.g. listening to the fragment corresponding a concordance line).

Other directions of research are concerned with the fine-tuning of the various tools. For example, the discretisation of the pitch movements uses two parameters, the mean frequency and variance of target points for a speaker. Currently, these values are computed for the entire recording, but one of the features of spontaneous speech is the presence of switches in speaking style, with bursts of greater (or smaller) variation in F_0. A relatively simple pre-processing could enable us to segment the recording into sections of coherent F_0 mean and variance. Other tools could also be included in the processing chain. For example, we have started experimenting the use of a filled pause detector, which could assist in the transcription of this important parameter. Although the detector used was developed for Japanese (Goto, Itou & Hayamizu, 1999), and would require tuning and adaptation for French, preliminary results are encouraging.

Acknowledgements

In this study we made use of the *Transcriber* software developed by Claude Barras (DGA) and the *Signaix* speech tools developed by Robert Espesser (CNRS). We are particularly indebted towards Robert Espesser for his help and assistance throughout this project. We thank Masataka Goto for making his filled pause detector available to us, although extensive experimentation has not been possible yet. We are also grateful to Claire Blanche-Benveniste, Albert Di Cristo, José Deulofeu, Daniel Hirst and Frédéric Sabio for their advice and comments. Remaining errors are of course ours.

References

Bilger M, Blasco M, Cappeau P, Pallaud B, Sabio F, Savelli M-J 1997 Transcription de l'oral et interprétation ; illustration de quelques difficultés. *Recherches sur le français parlé*, 14:57-86.

Blanche-Benveniste C (ed) 1990 Le français parlé : études grammaticales. Paris, CNRS éditions.

Blanche-Benveniste C, Jeanjean C 1987 *Le français parlé : transcription et édition.* Paris, Didier Erudition.

Campione E, Véronis J 2001 Une évaluation de l'algorithme de stylisation mélodique MOMEL. T*ravaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence*, 19 [in press].

Campione E, Véronis J 1998 A statistical study of pitch target points in five languages. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sidney, pp 1391-1394.

Cohen A, t'Hart J, 1965 Perceptual Analysis of Intonation Pattern. In *Proceedings of 5ème Congrès International d'Acoustique*, Liège, 1-4.

D'Alessandro C, Mertens P 1995 Automatic Pitch Contour Stylisation Using a Model of Tonal Perception. *Computer, Speech and Language*, 9:257-288.

Di Cristo A 1999a Le cadre accentuel du français : essai de modélisation : première partie. *Langues*, 2(3):184-205.

Di Cristo A 1999b Le cadre accentuel du français : essai de modélisation : seconde partie. *Langues*, 2(4): 258-269.

Fujisaki H, Hirose K 1982 Modelling the dynamic characteristics of voice fundamental frequency with application to analysis and synthesis of intonation. In *Proceedings of 13th International Congress of Linguists*, Tokyo, pp 57-70.

Goto M, Itou K, Hayamizu S 1999 A Real-time Filled Pause Detection System for Spontaneous Speech Recognition. In *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech '99)*, Budapest, pp.227-230.

Hirst, DJ 1979 The transcription of English intonation. *Studia Phonetica*, 17:29-39.

Hirst, DJ 1991. Intonation models. Towards a third generation. In *Proceedings of ICPhS*, I, pp 305-310.

Hirst DJ, Di Cristo A 1984 French Intonation : a parametric approach. *Die Neueren Sprachen*, 83(5):554-569.

Hirst DJ, Di Cristo A 1998 A survey of intonation systems. In Hirst DJ, Di Cristo, A (eds). *Intonation Systems: A Survey of Twenty Languages.* Cambridge, Cambridge University Press, pp 1-44.

Hirst DJ, Di Cristo A Espesser R 2000 Levels of representation and levels of analysis for the description of intonation systems. In Horne M (ed), *Prosody: Theory and Experiment*. Dordrecht, Kluwer Academic Publishers.

Hirst DJ, Espesser R 1993 Automatic Modelling of Fundamental Frequency Using a Quadratic Spline Function. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 15:75-85.

Leech G, McEnery A, Wynne M 1997 Further Levels of Annotation. In Garside R, Leech G, McEnery A (eds), *Corpus Annotation : Linguistic Information from Computer Text Corpora*, London, Longman, pp 85-101.

Mertens P 1990 Intonation. In Blanche-Benveniste C (ed), *Le français parlé: études grammaticales.* Paris, CNRS Edition, pp 159-176.

Nolan F, Grabe E 1997. Can 'ToBI' transcribe intonational variations in British English? In *Proceedings of ESCA Workshop Intonation: Theory, Models and Applications*, Athens, pp 259-262.

Ostendorf M, Ross K 1997 A multi-level model for recognition of intonation labels. In Sagisaka, Campbell and Higuchi (eds), *Computing Prosody*. Springer, Berlin, pp 291-308.

Silverman K, Beckman M, Pitrelli J, Ostendorf M, Wightman C, Price P, Pierrehumbert J, Hirschberg J 1992 ToBI: a standard for labelling English prosody. In *Proceedings of ICSLP'92*, Banff, pp 867-870.

Svartvik J (ed) *The London Corpus of Spoken English: Description and Research.* Lund, Lund University Press.

t'Hart J, Collier R, Cohen A 1990 *A Perceptual Study of Intonation, an experimental-phonetic approach to speech melody*, Cambridge, Cambridge University Press.

Taylor LJ, Knowles G 1988 *Manual of information to accompany the Spoken English Corpus.* Technical Report, UCREL, University of Lancaster.

Taylor P 1994 The Rise/Fall/Connection Model of Intonation. *Speech Communication*, 15(1,2):169-186.

Véronis J, Campione E 1998 Towards a reversible symbolic coding of intonation. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sidney, pp 2899-2902.

Wightman CW, Ostendorf M 1992 Automatic Recognition of Intonational Features. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, pp 221-224.