Transcribing intonational variation at different levels of analysis

Brechtje Post[1] *& Elisabeth Delais-Roussarie*[2]

[1]Research Centre for English and Applied Linguistics,
University of Cambridge, UK
[2]CNRS, UMR 7110 / LLF (Laboratoire de Linguistique formelle),
Université de Paris 7, France
bmbp2@cam.ac.uk, elisabeth.roussarie@wanadoo.fr

Abstract

In the transcription system for Intonational Variation (IVTS, derived from IViE), prosodic features are transcribed on (1) the rhythmic tier, (2) the local phonetic tier, (3) the global phonetic tier, and (4) the phonological tier. Each tier offers a range of labels which share a general architecture, but language-specific parameters determine which subset of labels a transcriber can choose from for the transcription of a particular language variety, and how the different tiers are associated with one another. In this paper, we will argue that the multi-linear architecture of IV-based systems offers transparency, flexibility and standardization, three key advantages in qualitative and quantitative studies of intonational variation across languages and language varieties.

1. Introduction

Research into prosodic phenomena such as accentuation and intonation often relies on discrete, symbolic representations of their phonetic realization. In such cases, the fact that these phenomena are continuous in nature, and that there is not always a generally accepted method of representing them symbolically, is a complicating factor. Moreover, prosodic features vary with factors such as speaking style, rate, and regional and socio-economic background. Hence, any transcription system that succeeds in encoding prosodic events discretely while taking variation into account will be of interest to a number of disciplines, both in linguistic analysis, as well as second language learning, automatic speech recognition, and other domains in which prosodic transcriptions could be applied fruitfully.

Although corpus studies have become increasingly prominent both in linguistics and in speech recognition, there is no standard for prosodic transcription that is sufficiently flexible to annotate prosodic variation. Well-known transcription systems like INTSINT [1] and ToBI [2], which share a number of basic features with IV-based systems, are not optimized for comparative analyses of prosodic variation:

- being based on acoustic analysis alone, INTSINT cannot take perceptual and linguistic knowledge into account
- intonational transcription in ToBI, by contrast, takes place at the phonological level, which makes it more appropriate for the transcription of varieties of languages which have been analyzed before, and for which the phonological inventory of contrastive intonational events is known
- INTSINT and ToBI focus on intonational phenomena at the expense of metrical phenomena, even though both systems were developed within the Autosegmental-Metrical framework, and
- ToBI - and to a lesser extent INTSINT as well - represent events at the local level of the phrase or the utterance and ignore the global discourse level.

In this paper, we will present the current prototype version of the Intonational Variation transcription system (IVTS), a machine-readable Autosegmental-Metrical transcription system for the transcription of intonational variation within and across language varieties. The system is directly based on IViE [3] (*Intonational Variation in English*; cf. [4] for German). IV-based systems have a number of advantages over systems like ToBI and INTSINT:

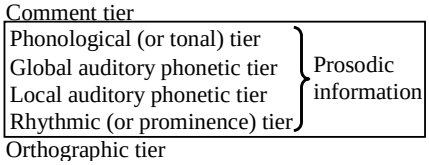
- IViE was specifically developed to analyze dialectal variation without requiring assumptions about the phonological identity of prosodic events, such as the metrical status of a prominent syllable or the phonological identity of a pitch movement
- the system can easily be adapted to include non-local phenomena such as downstep, changes in register, etc.
- the system's multi-linearity – transcribing phonetics, phonology and rhythm on different labeling tiers – offers transparency and flexibility (see Section 2 below)
- IViE labeling appears to be robust (for prominence: [5])

We will argue that IVTS can help bring the flexibility, transparency and standardization that are essential in comparative studies of prosodic variation. The discussion will be illustrated with examples from data which we are currently testing the system on: four French dialects [cf. 6] (recorded for the project *Phonologie du Français Contemporain*), and L2 English. We would like to stress, however, that the system is not intended as a transcription system for one specific language or language variety (cf. [3]), but our prototype system still needs extensive testing to establish to what extent it can accommodate data from other varieties and languages.

2. A multi-tiered system

Like IViE, the IV transcription system encodes orthographic and prosodic information on a number of transcription levels or *tiers*. IVTS has six tiers, four of which are used for the transcription of prosodic phenomena, as shown in (1).

(1) Six levels of annotation in IVTS



Unlike IViE, IVTS has an auditory phonetic transcription tier which can be used to record global intonational events such as compression of the speaking range, upstep, downstep, etc.

2.1. An example of an IV-based transcription: French

Figure 1 gives an example of an application of IVTS to French. The speech sample is an extract of the corpus *Phonologie du Français Contemporain*, and the utterance *Le village de Beaulieu est en grand émoi* 'The village of Beaulieu is in uproar' was read by a middle-aged male speaker of Belgian French from Liège. In the figure, the six annotation tiers listed in (1) above are time-aligned with the fundamental frequency trace and the pressure wave. This example was generated by Praat, but other speech processing software can be used for this purpose.

As can be seen in Figure 1, the individual words spoken are aligned with the corresponding interval in the speech signal on the orthographic tier. The labels of the other tiers are aligned with particular points rather than intervals, and in this example, those points are (1) the centre of the prominent syllables, (2) the boundaries of major intonational domains (marked on all prosodic tiers), and (3) on the comment tier, in the vicinity of the event the comment relates to. Different alignments may be relevant for other languages, depending on language-specific association conventions that govern the mapping between the tonal and segmental structures.

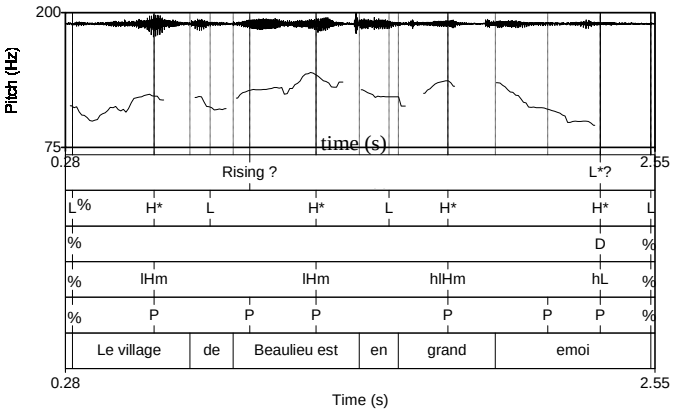


Figure 1: *An IVTS transcription of a Belgian French utterance*

The prominent syllables are labeled 'P' on the rhythmic tier in Figure 1, indicating that the syllable is more salient than the syllables that surround it, which may be signaled by for instance a change in pitch, greater intensity and/or longer duration (see [6] for a discussion of our assumptions about prominence and accentedness in the French version of IVTS). Thus, 'P' for prominence implies actual phonetic salience, and not necessarily an abstract structural property of a word or word group, such as lexical stress. Also, when 'P' labels mark a syllable that is associated with a pitch movement, they will coincide with labels on the phonetic and phonological tiers, but as the 'P' labels on *Beaulieu* and *émoi* in Figure 1 show, 'P' labels do not have to be associated with pitch movements. Instead, the auditory impression of the pitch trace in these examples suggests that pitch is interpolated from a preceding target to a target that follows the syllable in question (e.g. from low *de* to high *–lieu* in the case of *Beaulieu*). Nevertheless, the syllables sound relatively loud and long, and they are therefore marked by a 'P' label.

The auditory impression of the pitch movements that are realized on the prominent syllables is transcribed on the local and global phonetic tiers. By abstracting away from the acoustic detail of the speech signal at an auditory phonetic level, we provide what is effectively equivalent to a narrow IPA-type transcription of pitch events (cf. [2]).

On the local phonetic tier, the transcription focuses on the shape and alignment of pitch movements relative to the prominent syllables marked on the rhythmic tier, rather than global aspects such as pitch range. The transcriptions are syllable-based, where the auditory impression of pitch in the prominent syllable is labeled relative to that of the preceding and following syllables in its Implementation Domain (ID), if there are any. How an ID is defined in a specific language depends on the principles that govern the mapping between tonal events and segmental structure. In French, the ID contains (1) the prominent syllable, (2) any preceding syllables up to the preceding prominent syllable or IP boundary, and (3) the syllable following the prominent syllable at issue, if there is one. For example, the utterance of Figure 1 has four Implementation Domains, one for each prominent syllable that is marked by a pitch movement, as is shown in Figure 2.

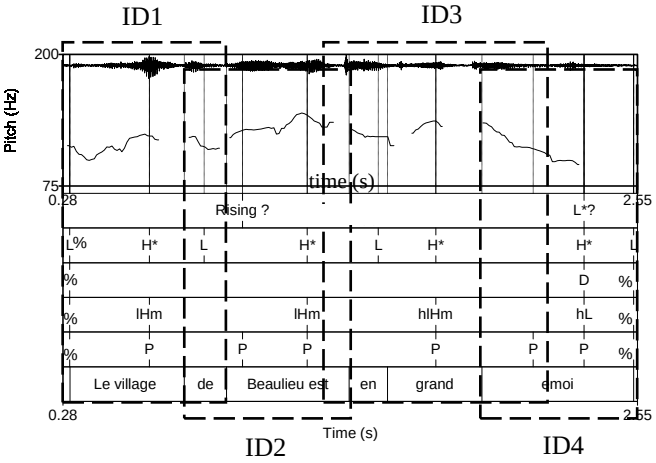


Figure 2: *Four IDs in a Belgian French utterance*

The example in Figure 2 also shows that IDs overlap by one syllable unless the prominent syllable coincides with the right edge of an Intonation Phrase, as is the case for ID4 here. Pitch levels are high, mid or low, and a capital marks the pitch level of the prominent syllable. The combination 'lHm', for instance, describes a rising slope from low to a high target in the prominent syllable, followed by a mid value that is in between the two (but note that the labels are relative and do not correspond to absolute values in the speaking range in any way; 'M' or 'm' is only used for a third value in a label that already contains 'H' or 'h' and 'L' or 'l').

Pitch changes that occur across different IDs can be transcribed on the global phonetic tier. For instance, register resets might occur when a speaker introduces a new topic in a conversation, which can be labeled 'R' on the global phonetic tier. Another example is the lowering of subsequent accentual peaks which could be indicated by a 'D' for downstep.

When the events transcribed on the local and global phonetic tiers are judged to be phonological in nature, they will also be labeled on the phonological tier. How a pitch event is categorized at this level depends on the inventory of contrastive contours of the language variety that is being

transcribed, and on the particular phonological analysis that is adopted, if one exists for the variety in question. For instance, a variety may have contrasting L and H phrase tones, or it may have downstepped accents that are phonologically distinct from non-downstepped ones. IVTS merely proposes a set of tonal primitives that the transcriber can choose from, as shown in (2) (see [6] for further details).

(2) Set of labels on the phonological tier

Tones		*Modifiers*	*Boundaries*	
H*	H	^ : upstep	%H	H%
L*	L	! : downstep	%L	L%
+		>: spreading	%	%

In the analysis of Figure 1, all pitch movements that are associated with prominent syllables are transcribed as H* tones, and the absence of the '!' modifier for the final H* accent indicates that downstep is considered to be phonetic rather than phonological in nature in this dialect (but note that this analysis is queried on the comments tier). In other words, we tentatively adopt the view here that pitch movements which occur at the beginning and at the end of the accentual domain (Phonological Phrase [7] or Accentual Phrase [8]) are phonologically equivalent in this variety of French, as has been proposed for standard French [9][7]. An alternative would be to treat the two as distinct tonal events (cf. [8]), or as boundary markers (cf. [10]), in which case different labels will be selected from the inventory given in (2).

The comments tier is used for remarks. In the example of Figure 1, both comments query the prosodic analysis that is proposed for this utterance, reflecting the fact that the analysis of the dialect of Liège is still in its early stages. The tier can also be used to record information like halving and doubling errors, overlap between speakers, etc.

2.2. Why four prosodic tiers?

Using four prosodic tiers to annotate different aspects of the prosodic realization of a speech sample has a number of advantages. First, by explicitly unfolding prosodic variation at four different levels, taxonomic differences between language varieties or languages can be described [11]. That is, languages differ in terms of the distribution of prominent syllables in the utterance, the phonetic realization of intonation contours, and in the phonological inventory of contours, and these differences can all be directly recorded on the corresponding labeling tiers. In addition, if there are differences in the phonotactic restrictions on the way in which tones or contours can be combined, this information can be derived from the phonological tier.

Second, it brings some advantages in terms of the consistency and transparency of transcriptions [11]. Analyzing a language's intonation system involves making choices about the nature of observed auditory phonetic differences (are they contrastive?), and about the economy, exhaustivity and transparency of the phonological account proposed. If the variety has not been analyzed before, the labels of the different tiers can be used as a tool to draw up hypotheses about phonetic variation and phonological contrasts. Since labelers have a record of the information which has led to a specific analysis, they can go back to this information at a later stage, when they have gathered comparative evidence from samples of different speakers and different contexts. This method proved successful in the IViE project [11] (see also [12]).

Third, IV-based transcription systems facilitate comparative analyses between speakers, speaking styles, dialects or languages, because they standardize transcriptions. Each prosodic tier offers a range of primitives that the transcribers can combine to transcribe a particular variety or language. Depending on a number of basic parameters along which languages vary (e.g. iambic vs. trochaic feet which will affect the language's Implementation Domain), a subset of labels can be constructed which will need to be motivated explicitly. At the phonological level, choices will potentially be informed by conflicting theoretical assumptions, but the general architecture of the system and the provision of label sets to choose from should facilitate comparisons between transcriptions that are made by different transcribers, and encourage further discussion of the assumptions that underpin the choices that have been made.

Fourth, distinguishing a local and a global phonetic level of transcription brings the advantage that discourse phenomena can be transcribed, and that languages which do not rely primarily on localized anchoring points for pitch events, such as metrically strong syllables, can be better accommodated.

As with any type of transcription, the transcriber's objectives will have to determine to what extent all tiers are actually exhaustively labeled in a particular study, since transcribing speech data is very time-consuming. In particular, when a variety has been analysed before, transcription at a subset of levels will often be sufficient (as in ToBI, which does not normally include tiers for auditory phonetic labeling).

2.3. Taxonomy of prosodic differences

2.3.1. Cross-varietal differences

Our pilot study of cross-varietal differences in three dialects of French has confirmed that IV-based transcription systems can be applied successfully to reveal a range of taxonomic differences in intonation languages, not only in English, but also in languages that are prosodically quite different, like French [6].

Even though the sample was limited, the study in [6] showed that at the rhythmic level, the varieties of French spoken in the Alsace – and probably also those in Belgium – differ from many other varieties of French in that the penultimate syllable before a major intonational boundary can be rhythmically prominent in addition to the final syllable (see e.g. the 'P' label on *é-* in *émoi* in Figure 1). In standard French, final prominences are only very rarely immediately preceded by another prominent syllable, and if they are, this syllable will normally be marked by a change in pitch.

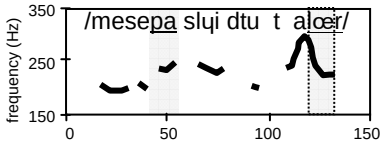


Figure 3: *A fall from a penult in standard French*

An example is given in Figure 3, where the utterance *Mais c'est pas celui de toute à l'heure* 'But it isn't the one we had earlier' is realized with a falling movement to the middle of the speaking range from a peak on the penultimate syllable (from [7], where it is analyzed as a fall from an unaccented onto an accented syllable). Regardless of how one chooses to

analyze this contour, the pre-final syllable is marked by a change in pitch, unlike the Alsatian and Belgian pre-final prominence.

Southern varieties, which can also have pre-final prominence, differ from Alsatian and standard French since (1) final prominences occur exclusively in words that end in a schwa such as *village* (not normally pronounced in standard, Belgian and Alsatian French), and (2) their occurrence is not conditioned by the presence of a major intonational boundary, as appears to be the case in Alsatian. For instance, when we compared productions of the word *village* in phrase-final position in Marseille and Liège, we found that both had a peak on the /a/ of *village*, thus marking the same syllable as prominent. For the speaker from Liège this syllable coincided with the word boundary, but for the speaker from Marseille, it did not. IVTS will clearly capture this difference on the local auditory phonetic tier. In Marseille, if the phrase ends in a rise, pitch will stay high or continue to rise on the post-accentual syllable, which would be transcribed as lHh or lMh respectively. By contrast, in Liège – as in any other variety without final schwa – the rise would end on the accentual syllable, transcribed as lH (or mLH if both the low and the high target occurred in the final syllable).

At the phonological level, French dialects also show considerable variation. Belgian French, for instance, shows evidence of a contrast between rising and falling accents within Intonation Phrases. In standard French, falls only appear to occur at major boundaries ([7], but see [13]).

2.3.2. Cross-linguistic differences

Our findings of cross-varietal differences at various levels of analysis in French mirror earlier findings for British English [11]. For instance, at the phonetic level, one and the same phonological category can be realized in different ways when there is very little sonorant material for the movement to be realized on. Thus, a fall on a word like *shift* will be truncated in Leeds, and compressed in Cambridge English. The difference is reflected in a difference in labels on the IViE phonetic tier: lH-m for the former, and mH-l for the latter.

These IViE labels are different from the labels we used to transcribe our French data [6]. Since the domains that are relevant in prosodic structure are different in French and English, the labels of the phonetic tier are of necessity different, too. Similarly, since the phonology of French intonation is different from English, the set of labels representing the inventories of intonational categories differ. Nevertheless, the same principles underlie the transcription systems used for the two languages, which facilitates cross-linguistic comparisons.

Cross-linguistic comparisons are also relevant in studies that investigate the role of transfer in L2 intonation. For instance, Spanish intermediate level learners of British English appear to produce prenuclear falls that are appropriate to the context, but their phonetic implementation is sometimes totally inappropriate, with alignments that are more characteristic of rises than falls in native English (data from a study by Dolores Ramírez-Verdugo, Universidad Autonomá de Madrid, and the first author).

3. Conclusion and discussion

The transcription system for Intonational Variation allows for the discrete encoding of prosodic events through a multi-level analysis, both at a local and global level. Unpacking the prosodic realization of utterances in this way offers the transparency, flexibility and standardization necessary in comparative studies of prosodic variation. We have given some examples of how IV-based transcription systems can be applied successfully to chart cross-varietal taxonomic differences in French as well as English, and to identify intonational characteristics of L2 learners of English.

The prototype system described here requires extensive further testing on a wider variety of languages to establish whether its current architecture is adequate and what label sets need to be included. We are also planning to explore how software such as MOMEL, a program which assigns labels to turning points on the basis of the fundamental frequency trace, may help speed up the transcription process.

4. References

[1] Campione, E.; Hirst, D.; Véronis, J., 2000. Automatic stylisation and symbolic coding of F0: implementations of the INTSINT model. In *Intonation. Research and Applications*, A. Botinis (ed.). Dordrecht : Kluwer.

[2] Beckman, M. E.; Hirschberg, J.; Shattuck-Hufnagel, S., 2005. The original ToBI system and the evolution of the ToBI framework. In *Prosodic Typology* : *The Phonology of Intonation and Phrasing*, S.-A. Jun (ed.). Oxford: Oxford University Press.

[3] Grabe, E.; Post, B.; Nolan, F., 2001. Modelling intonational variation in English: The IViE system. In *Prosody 2000*, S. Puppel; G. Demenko (eds.). Poznan: Adam Mickiewicz University, 51-58.

[4] Auer, P.; Gilles, P., 2003. *Prosodic Variation in German between Areality and Pragmatics: an Overview*. Unpublished manuscript.

[5] Kochanski, G; Grabe E.; Coleman, J; Rosner, B., accepted. *Loudness predicts prominence; fundamental frequency lends little. JASA.*

[6] Post, B; Delais-Roussarie, E.; Simon, A-C., In press. Développer un système de transcription des phénomènes prosodiques. In *Bulletin PFC 4*, G. Caelen-Haumont ; A-C. Simon (eds.).

[7] Post, B., 2000. *Tonal and phrasal structures in French intonation* (PhD thesis). The Hague: Holland Academic graphics.

[8] Jun, S.-A.; Fougeron, C., 2002. Realizations of accentual phrase in French intonation. *Probus* 14, 147-172.

[9] Di Cristo, A.; Hirst, D., 1996. Vers une typologie des unités intonatives du français. *XXIème J.E.P., Avignon 1996*, 219-222.

[10] Welby, P., 2002. The realization of early and late rises in French intonation: A production study. *Speech Prosody 2002*, 695-698.

[11] Grabe, E.; Post, B., 2002. Intonational Variation in English. In *Proceedings of the Speech Prosody 2002 Conference*, B. Bel; I. Marlin (eds). Aix-en-Provence: Laboratoire Parole et Langage, 343-346.

[12] Grabe, E.; Kochanski, G.; Coleman, J., accepted. The intonation of native accent varieties in the British Isles: Potential for miscommunication? In K. Dziubalska-Kolaczyk and J. Przedlacka (eds.), *English pronunciation models: a changing scene*. Peter Lang.

[13] Martin, P., 1975. Analyse phonologique de la phrase française. *Linguistics* 146, 35-76.