Vers un système multilinéaire de transcription des variations intonatives

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 et elisabeth.roussarie@wanadoo.fr

ABSTRACT

In the paper, we will present a transcription system for Intonational Variation (IVTS), derived from IViE. The prosodic features are transcribed on i) the rhythmic tier ; ii) the local phonetic tier ; iii) the global phonetic tier ; and iv) 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. In this paper, we will argue that the multi-linear architecture of IViE-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

La difficulté des recherches en prosodie s'explique par le fait que i) les phénomènes comme l'accentuation ou l'intonation sont par nature continus, et donc difficiles à représenter de façon discrète ; ii) les caractéristiques prosodiques varient en fonction d'éléments comme le débit ou l'origine socio-géographique des locuteurs. Aussi tout système de transcription qui permet d'encoder les événements prosodiques de façon discrète tout en prenant en compte la variation est intéressant dans de nombreux domaines : la linguistique, l'apprentissage des langues secondes, le traitement automatique de la parole, etc.

Bien que les travaux sur corpus se soient considérablement développés ces dernières annnées tant en linguistique qu'en traitement automatique de la parole, il n'existe pas à ce jour de système de transcription prosodique standardisé qui soit suffisamment flexible pour traiter la variation et encoder les phénomènes prosodiques dans des dialectes non encore décrits. Même si des systèmes de transcription connus et fréquemment utilisés comme INTSINT [1] ou ToBI [2] partagent de nombreuses caractéristiques avec des systèmes dérivés de IViE [3] (*Intonational Variation in English*), ils ne sont pas conçus pour travailler dans une perspective comparative et variationniste, et cela pour plusieurs raisons : i) INTSINT, en partant d'une analyse exclusivement acoustique, ne tient pas compte de données perceptives et linguistiques ; ii) ToBI, de son côté, présuppose que l'inventaire phonologique des phénomènes intonatifs soit établi puisque le transcripteur encode directement les informations au niveau phonologique ; iii) INTSINT et ToBI privilégient nettement les phénomènes intonatifs au dépens des phénomènes métriques, bien que les deux systèmes soient développés dans le cadre métrique-autosegmental ; iv) ToBI, et dans une moindre mesure INTSINT, représente le niveau local de la phrase ou de l'énoncé plutôt que le niveau global du discours.

Dans ce papier, nous présentons une première version d'un système de transcription prosodique appelé IVTS (*Intonational Varitaion Transcription System*). Ce dernier est développé dans le cadre métrique autosegmental (cf. [4] et [5]) et repose sur l'idée que pour analyser adéquatement les phénomènes intonatifs, il est préférable de clairement distinguer le niveau phonologique du niveau phonétique. IVTS doit permettre d'encoder les variations intonatives afin de comparer des langues, mais aussi différents dialectes d'une même langue. Il repose sur une adaptation directe d'IViE (cf. [3], mais aussi [6] pour une application à l'allemand) qui se justifie par le fait que les systèmes comme IViE offrent de nombreux avantages en comparaison de systèmes comme ToBI ou INTSINT : i) étant conçus pour analyser les variations dialectales, ils ne présupposent pas de connaissances sur le statut phonologique de tel ou tel événement prosodique (statut métrique des syllabes proéminentes, identité phonologique d'un mouvement mélodique, etc.) ; ii) ils peuvent facilement être adaptés pour permettre l'encodage de phénomènes non locaux comme les changements de registre ou les downsteps ; iii) du fait de leur structure multilinéaire (l'encodage d'information phonétique, métrique et phonologique se faisant sur plusieurs tires ou niveaux), ils offrent plus de transparence et de flexibilité (cf. section 2 ci-après) ; iv) ils se montrent très robustes comme en témoignent les travaux menés dans le cadre d'IViE sur les proéminences (cf. [7]).

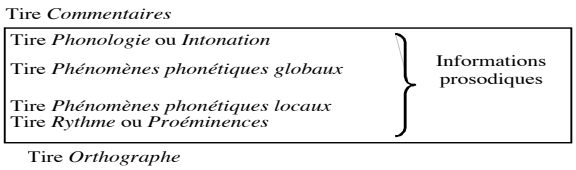
Nous allons montrer qu'un système comme IVTS permet d'avancer vers plus de flexibilité, de transparence et de standardisation, ce qui est essentiel pour mener des études prosodiques dans une perspective comparative et variationiste. Cette idée sera illustrée par des exemples précis extraits de données variées sur lesquelles nous travaillons actuellement : quatre dialectes du français (cf. [8]) et des données d'acquisition (anglais langue seconde). Pour finir, il est important de noter que ce système a vocation à être

adapté et utilisé pour étudier plusieurs langues ou variétés d'une même langue, même si, pour le moment, il n'a été testé que sur un petit échantillon de données du français.

2. UN SYSTÈME MULTILINÉAIRE

Le système de transcription IVTS encode différentes informations prosodiques et orthographiques sur six niveaux (ou *tires*) distinct(e)s, parmi lesquel(le)s quatre sont spécifiquement consacré(e)s à l'annotation prosodique.

(1) les six niveaux de transcription retenus dans IVTS



2.1 Un exemple d'annotation en français

La figure 1 représente une utilisation d'IVTS pour transcrire des données du français extraites du corpus PFC. Il s'agit de l'énoncé *le village de Beaulieu est en grand émoi* lu par un homme originaire de la région de Liège (Belgique). Les six niveaux d'annotation mentionnés sous (1) sont alignés temporellement avec le signal et la courbe de fréquence fondamentale. Cette transcription a été effectuée sous Praat, mais elle pourrait l'être avec tout autre logiciel d'analyse et de visualisation du signal de parole.

Comme on le voit sur la figure, les différents mots produits sont alignés avec les portions (ou intervalles) de signal qui leur correspondent sur la tire *orthographe*. Dans les autres tires, les étiquettes sont alignées avec des points déterminés sur le signal comme : i) le milieu d'une syllabe proéminente ; ii) les frontières de domaines intonatifs, etc. Sur la tire *Commentaires*, les points retenus pour aligner les commentaires au signal correspondent aux zones sur lesquelles portent les commentaires. Les différents alignements retenus le sont en fonction de leur pertinence dans la langue traitée et de la façon dont s'opère l'ancrage des phénomènes mélodiques.

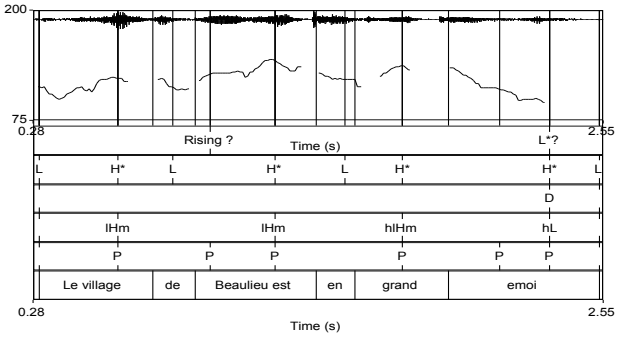


Figure 1 : Exemple d'annotation avec IVTS

Sur la tire *Rythme*, l'étiquette **P** indique que la syllabe marquée est plus proéminente que les syllabes adjacentes (cf. fig. 1). Cela peut se caractériser au niveau acoustique par un allongement de la durée, un mouvement mélodique, etc. (cf. [6]). Notons que **P** signale une saillance perceptive, mais pas nécessairement une propriété structurelle abstraite du mot ou du groupe de mots comme un accent lexical. De plus, si les étiquettes **P** affectent des syllabes sur lesquelles des mouvements mélodiques particuliers se réalisent, elles vont être alignées avec des étiquettes sur les tires *Phénomènes phonétiques locaux* et *Phonologie*. Ceci étant, cela n'est pas une nécessité comme on peut le voir dans la fig. 1 avec les étiquettes **P** associées aux syllabes initiales des mots *Beaulieu* et *émoi*.

La tire *Phénomènes phonétiques locaux* est utilisée pour transcrire la forme des mouvements mélodiques réalisés sur des syllabes proéminentes ainsi que sur les syllabes adjacentes. L'accent est mis ici sur la configuration mélodique et sur les modalités d'alignement de ces mouvements intonatifs. Les phénomènes mélodiques plus globaux tels le registre ou le downstep ne sont pas encodés à ce niveau. La transcription des mouvements mélodiques se fait sur des bases perceptives et auditives, et non à partir d'une analyse acoustique de la fréquence fondamentale. Elle s'effectue à partir de l'écoute attentive d'une portion de signal correspondant au **domaine d'implémentation accentuel** (noté ID). L'extension de ce domaine varie selon les langues puisqu'elle dépend de la façon dont s'effectuent les associations tonales. En français, tout ID comprend i) la syllabe proéminente notée P ; ii) toutes les syllabes qui la précèdent jusqu'à la syllabe proéminente précédente ou jusqu'à la frontière d'un domaine intonatif majeur ; et iii) la syllabe qui la suit immédiatement. D'après cette définition, l'énoncé présenté dans la figure 1 se décompose en quatre ID, un pour chaque syllabe proéminente marquée par un mouvement mélodique.

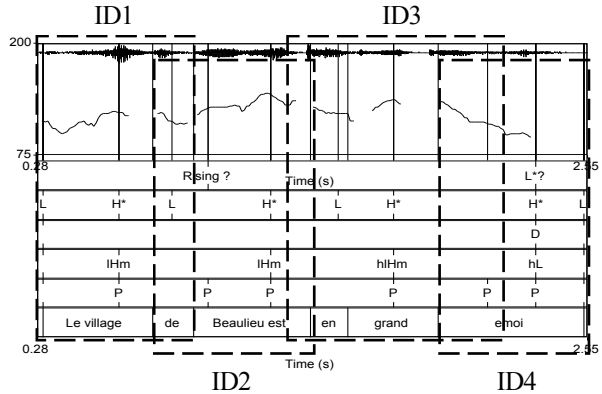


Figure 2: Exemple de segmentation en IDs

Comme le montre la figure 2, deux ID consécutifs partagent généralement une syllabe. Ainsi la syllabe *de* dans *le village de Beaulieu* appartient aux deux premiers IDs. Les IDs situés en début et en fin de domaines intonatifs ont respectivement une syllabe initiale et une syllabes finale qui n'apparaît dans aucun autre ID. Les niveaux retenus pour l'encodage mélodique sont haut (H ou h), moyen (M ou m) et bas (L ou l) et sont tous relatifs. En outre, ces niveaux sont notés en lettres majuscules dès lors que leur cible est

alignée sur le noyau d'une syllabe proéminente. La séquence *lHm*, par exemple, indique un mouvement mélodique montant du niveau bas vers le niveau haut, ce dernier étant atteint sur la syllabe marquée P. Ce mouvement est ensuite suivi d'une légère descente. Notons que les étiquettes représentent des valeurs relatives, puisqu'elles dépendent des réalisations au sein d'un unique ID.

Les phénomènes mélodiques se réalisant sur des empans qui dépassent le cadre d'un ID sont encodés sur la tire *Phénomènes phonétiques globaux*. Ainsi, les resettings qui sont parfois réalisés lorsque le locuteur introduit un nouveau thème de discours seront notés R sur cette tire. De même, un phénomène de downstep se réalisant au sein d'un domaine intonatif sera noté D.

Lorsque les événements notés sur les tires *Phénomènes phonétiques locaux* et *Phénomènes phonétiques globaux* sont considérés comme phonologiques, ils sont aussi encodés sur la tire *Phonologie*. La façon d'encoder les phénomènes mélodiques à ce niveau d'annotation dépend de l'inventaire phonologique des mouvements mélodiques contrastifs retenu pour la langue (ou variété) à transcrire. IVTS propose simplement un ensemble de primitives tonales parmi lesquels chaque transcripteur peut faire son choix (cf. [8] pour une discussion plus détaillée).

(2) Ensemble d'étiquettes pour le niveau phonologique

Tons		*Modifieurs*	*Frontières*	
H*	H	^ : upstep	%H	H%
L*	L	! : downstep	%L	L%
+		>: propagation	%	%

Pour finir, la tire *commentaires* est utilisée pour noter les remarques diverses.

2.2 Pourquoi quatre tires pour l'encodage prosodique ?

Recourir à quatre tires distinctes pour annoter différents aspects de la réalisation prosodique offre de nombreux avantages. Tout d'abord, en utilisant explicitement quatre niveaux pour annoter les phénomènes prosodiques, le système permet de décrire les variations dialectales de nature taxonimique. Comme les différences entre les langues peuvent affecter la distribution des syllabes proéminentes, la réalisation phonétique des contours intonatifs ou l'inventaire phonologique des primitives tonales, elles peuvent être encodées directement sur la tire dont elles relèvent.

Deuxièmement, le recours à plusieurs niveaux d'annotation prosodique permet d'obtenir plus d'uniformité et de transparence dans la réalisation des transcriptions. Analyser le système intonatif d'une langue impose de faire des choix sur i) le caractère contrastif ou non d'un mouvement mélodique ; ii) l'économie et la pertinence du système dans son ensemble, etc. Si la langue étudiée n'a pas encore été décrite, le recours aux différentes tires permet d'élaborer des hypothèses sur lesquelles il sera toujours possible de revenir par la suite.

Troisièmement, de par leur uniformité, les systèmes de transcription prosodiques dérivés de IViE permettent de mener facilement des comparaisons entre des langues, des dialectes (ou variétés) d'une même langue, voire entre des styles. Un ensemble d'étiquettes et de primitives est en effet proposé pour chaque tire, le transcripteur pouvant alors les combiner pour décrire telle ou telle variété, voire même en ajouter d'autres. Il est clair que les choix retenus au niveau phonologique résulteront souvent d'hypothèses théoriques particulières. Néanmoins, l'architecture globale du système et l'accès aux autres tires doivent faciliter les comparaisons.

Quatrièmement, la distinction entre les phénomènes phonétiques locaux et les phénomènes phonétiques globaux permet i) d'encoder les phénomènes discursifs ; ii) de traiter les langues dont le système intonatif ne repose pas une association entre événement tonal et proéminence syllabique.

2.3 Taxonimie des différences prosodiques

2.3.1 Différences cross-dialectales

L'étude pilote que nous avons menée sur trois variétés de français a confirmé le fait qu'un système dérivé de IViE est capable de révéler un grand nombre de différences taxonimiques entre dialectes, et cela dans des langues aussi différentes que le français ou l'anglais (cf. [8] et [9]).

Bien que limitée, notre étude [8] a montré qu'au niveau rythmique, les variétés de français parlées en Alsace, et probablement aussi en Belgique, se distinguent de nombreuses autres variétés de français par le fait que les syllabes pénultièmes en position finale de domaine intonatif majeur sont souvent proéminentes, et cela même si les syllabes finales le sont aussi (cf. l'étiquette P associée à la syllabe [e] de *émoi* dans la fig. 1). En français standard, les syllabes finales métriquement distinguées ne sont que très rarement précédées de syllabes proéminentes, et si elles le sont, cela se caractérise généralement par un mouvement mélodique (contra l'exemple de la fiG. 1). Les parlers méridionaux, où les syllabes pénultièmes peuvent aussi être proéminentes, diffèrent à la fois du français standard et des variétés d'Alsace dans la mesure où i) les proéminences sur les syllabes pénultièmes n'apparaissent que lorsque les mots se terminent par un schwa comme dans *village* (non prononcé dans les variétés non méridionales) ; et ii) l'apparition de telles proéminences n'est pas conditionné par la présence d'un frontière de domaine prosodique majeur, comme cela semble être le cas en Alsace. Ainsi, si nous comparons les productions du mot *village* en position finale d'énoncé dans les variétés de Marseille et de Liège, nous voyons que, dans les deux cas, la syllabe proéminente a pour noyau le [a]. Mais pour le locuteur de Liège, cette syllabe est également la dernière syllabe du mot et de la phrase, tandis que pour celui de Marseille, ce n'est pas le cas. Ces différences sont

clairement encodées dans IVTS au niveau de la tire *phénomènes phonétiques locaux.*

Au niveau phonologique, les dialectes du français sont également très différents les uns des autres. La variété parlée en Belgique montre clairement un contraste entre des mouvements mélodiques montant et descendant à l'intérieur d'un domaine intonatif (pitch accent H* vs. L*). En français standard, les mouvements descendant de ce type apparaissent plutôt en position finale de domaine intonatif (cf. [10 et [11]).

2.3.2 Différences entre langues

Les résultats de nos études sur les variations dialectales du français montrent que les différences relèvent de plusieurs niveaux d'analyse. Cela était également le cas pour l'anglais britannique (cf. [9]). Ainsi, une même catégorie phonologique peut être réalisée différemment au niveau phonétique. Les recherches sur l'anglais britannique ont par exemple montré qu'un mouvement mélodique descendant sur un mot comme *shift* est tronqué dans les variétés de Leeds, et compressé dans celles de Cambridge. Dans IViE, ces différences se reflétaient dans des différences d'étiquetage sur la tire *phénomènes phonétiques locaux* : lH-m était retenu dans le premier cas, et mH-l dans le second.

Les étiquettes retenues pour IViE ne sont pas les mêmes que celles utilisées pour le français (cf. [8]). Cela s'explique par le fait que les domaines pertinents pour rendre compte de la prosodie de chacune des langues sont différents. De la même façon, l'inventaire des formes phonologiques retenus pour encoder la phonologie de chacune des langues diffère. Malgré tout, les systèmes de transcription utilisés pour traiter ces deux langues reposent sur les mêmes principes, ce qui favorise donc les études comparatives.

Les comparaisons entre langues sont également intéressantes à mener dès lors que l'on s'intéresse à la question des transferts dans l'acquisition/ apprentissage des langues secondes. Considérons le cas d'apprenants espagnols ayant atteint un niveau intermédiaire en anglais britannique. Il est fréquent qu'ils produisent des mouvements mélodiques descendant en position prénucléaire dans un contexte approprié, mais avec un alignement incorrect qui relève d'une erreur d'implémentation phonétique : l'alignement est en effet celui attendu pour les mouvements montants en anglais britannique standard (données extraites d'une étude menée par Dolores Ramírez-Verdugo, UAM, Madrid et B. Post).

3. Conclusion et discussion

Le système de transcription IVTS permet d'encoder de façon discrète les événements prosodiques sur plusieurs niveaux, qu'ils soient locaux ou globaux. Cette façon de procéder pour analyser les phénomènes prosodiques offre de nombreux avantages comme la transparence, la flexibilité et la normalisation : autant d'éléments essentiels pour mener à bien des études comparatives. Dans ce papier, nous avons fourni quelques exemples qui permettent de voir comment un système comme IVTS a permis i) de comparer des variétés du français et de l'anglais ; ii) d'identifier des difficultés d'acquisition de l'anglais langue seconde chez des hispanisants.

Nous sommes conscients que le prototype présenté ici nécessite d'être testé sur des données plus variées, et cela afin de vérifier si l'architecture globale du système est satisfaisante ou nécessite d'être modifiée. C'est ce que nous allons faire prochainement en travaillant sur d'autres variétés du français et sur des variétés de l'occitan. Par ailleurs, nous envisageons d'étudier l'apport que pourrait constituer un précodage automatique des données avec des logiciels comme Momel (cf. [1]).

Bibliographie

[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] Pierrehumbert, J., 1980. The phonology and phonetics of English Intonation. Ph.D. thesis M.I.T.

[5] Ladd, D. R., 1996. *Intonational Phonology* Cambridge: Cambridge University Press.

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

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

[8] 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.).

[9] 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.

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

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