



Annotation prosodique et typologie

Elisabeth Delais-Roussarie, Brechtje Post, Cristel Portes

► To cite this version:

Elisabeth Delais-Roussarie, Brechtje Post, Cristel Portes. Annotation prosodique et typologie. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA), Laboratoire Parole et Langage, 2006, 25, pp.61-95. <hal-00308392>

HAL Id: hal-00308392

<https://hal.archives-ouvertes.fr/hal-00308392>

Submitted on 30 Jul 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANNOTATION PROSODIQUE ET TYPOLOGIE

Elisabeth Delais-Roussarie*, Brechtje Post**, Cristel Portes

* CNRS, UMR 7110, *Laboratoire de Linguistique Formelle (LLF), Université de Paris 7.*

** *Research Centre for English and Applied Linguistics, University of Cambridge, Cambridge, UK.*

1. Introduction

L'accroissement des travaux de linguistique consacrés à la prosodie et aux relations qu'elle entretient avec le discours et la grammaire, d'une part, et le développement des approches sur corpus, d'autre part, force à s'interroger sur la façon de représenter de façon symbolique et discrète les événements prosodiques continus. Cette question de l'encodage et de la représentation symbolique et discrète des phénomènes suprasegmentaux n'est pas neutre, surtout si on se la pose en tentant d'établir un parallèle avec la représentation de niveau segmental proposé par l'IPA (*International Phonetic Association* et *International Phonetic Alphabet*). Ce système de codage repose sur plusieurs hypothèses concernant le continuum de parole et son analyse, à savoir : a) certains aspects de la parole sont linguistiquement pertinents, tandis que d'autres ne le sont pas. Cela est intégré dans la possibilité de faire des transcriptions larges (ou phonémiques) et des transcriptions étroites (ou phonétiques) ; b) le continuum de parole peut être en partie représenté comme une séquence de segments.

Il est important de noter que le seul présupposé notoire que fasse l'IPA est que le continuum de parole peut être représenté sous forme de segments discrets. En revanche, le recours à un symbole distinct pour représenter un son ou un segment ne fait pas nécessairement de ce dernier un phonème. Considérons, par exemple, la transcription proposée dans le *Handbook of the International Phonetic Association* (1999 : 78-81) pour la lecture d'un texte du français :

(1) *La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort. Quand ils ont vu un voyageur qui avançait, enveloppé dans son manteau, ils sont tombés d'accord...*

[labiz e ləsələ'sə dɪspyte || ʃak ɛ̃syʁɑ̃'kiletə ləplyfɔ̃ʁ || kɑ̃til zɔ̃ v y
ɛ̃ vwajœʁ ki savɑ̃se || ɑ̃vlope dɑ̃ sɔ̃ mɑ̃to || i: sɔ̃ tɔ̃be dakɔ̃ʁ]

DELAIS-ROUSSARIE, Elisabeth ; POST, Brechtje ; PORTES, Cristel (2006), Annotation prosodique et typologie, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, vol. 25, p. 61-95.

Dans l'exemple (1), le pronom *ils* est transcrit par [i:] bien que l'opposition de longueur ne soit pas phonémique en français, contrairement à ce qui se passe dans d'autres langues comme le hausa (langue de l'Afrique de l'Ouest) où /i/ et /i:/ sont deux phonèmes distincts.

L'une des sources d'erreurs ou de discussion tient à la difficulté de la tâche de transcription. Le transcripteur peut en effet, dans sa façon d'encoder les sons entendus, faire des choix qui s'expliquent soit par sa connaissance du système phonologique de la langue à transcrire soit par sa méconnaissance de cette langue. Dans le premier cas, il peut être amené à choisir de transcrire les segments non pas tels qu'ils sont effectivement réalisés, mais comme des phonèmes. Dans le second, il peut être induit en erreur par sa propre langue (*cf.* idée du filtre phonologique formulée par Troubetzkoy, 1939).

Pour évaluer ou développer tout système de transcription des phénomènes prosodiques, il est donc important d'avoir conscience de ces limitations. En revanche, deux questions doivent être posées : a) le système repose-t-il sur des présupposés théoriques importants qui pourraient avoir des répercussions sur l'analyse qui peut être faite ultérieurement ? b) dans quelle mesure le système permet-il d'encoder les phénomènes prosodiques dans toute leur complexité, même lorsque la langue à transcrire n'est pas encore décrite ?

Pour tenter de répondre à ces questions et pour réfléchir à la façon de transcrire symboliquement la prosodie, nous nous proposons dans cet article de comparer quatre systèmes de transcription des phénomènes prosodiques : l'IPA et les symboles spécifiques pour l'encodage des phénomènes suprasegmentaux, INTSINT (Hirst *et al.*, 2000), ToBI (Beckman *et al.*, 2005) et IVTS, un système de transcription élaboré à partir de IViE (Grabe *et al.*, 2001 ; *IViE labelling guide* : <<http://www.phon.ox.ac.uk/IViE/guide.html>>. Nous nous fixons un double objectif :

- présenter chaque système de transcription en insistant sur les présupposés à partir desquels il est développé et sur la façon dont il fonctionne ;
- évaluer la façon dont ces différents systèmes peuvent encoder/ transcrire certains types de phénomènes prosodiques particuliers, notamment l'accentuation¹, la mélodie et le registre.

L'article sera organisé comme suit. Dans une première partie, nous présenterons les caractéristiques principales des quatre systèmes. Une attention particulière sera accordée à IVTS

1. Dans la section consacrée à l'accentuation, nous nous intéresserons principalement à l'accentuation métrique (*stress patterns*), mais nous serons amenées à dire que les syllabes proéminentes peuvent renvoyer à des événements distincts.

dans la mesure où ce système est moins connu. Dans une seconde partie, trois phénomènes prosodiques souvent rencontrés dans les langues du monde seront présentés (accentuation, mélodie et registre). Nous tenterons de voir comment ils peuvent être encodés dans les différents systèmes. Pour mener à bien cette réflexion, nous avons travaillé sur deux types de données :

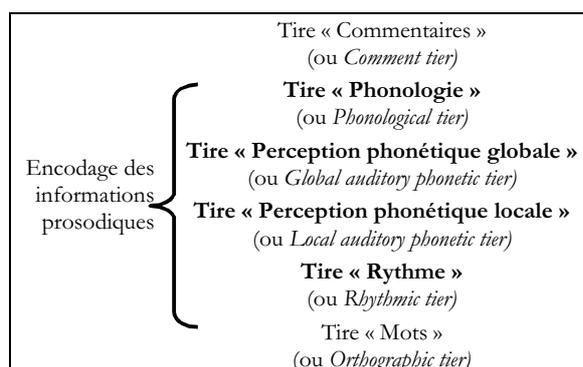
- des données de trois variétés du français (français de Marseille, français de Liège et français de Périgueux) qui ont été enregistrées dans le cadre du projet PFC (*cf.* Durand et Lyche, 2003) et qui comprennent des extraits de lecture de texte et des extraits de conversation informelle et formelle ;
- des données de deux variétés de l'occitan (occitan de Brive et occitan de Nontron) qui consistent en un extrait de texte lu par des conteurs.

2. Les systèmes de transcription / annotation prosodique

2.1. IVTS : caractéristiques fondamentales et modalités de fonctionnement²

Tout comme IViE ou TOBI, le système IVTS a recours à plusieurs tires d'annotation pour encoder les différentes informations prosodiques et linguistiques. En IVTS, une transcription est organisée sur six niveaux, parmi lesquels quatre sont utilisés pour transcrire les phénomènes prosodiques. La transcription prend donc la forme suivante :

Tableau 1
Six niveaux de transcription



2. Les caractéristiques d'IVTS présentées dans cette section sont reprises des présentations faites dans Post *et al.* (2006), Post et Delais-Roussarie (2006). Il est clair néanmoins que certains aspects du système peuvent être revus et améliorés, notamment si IVTS ne permet pas d'encoder de façon adéquate certains phénomènes (*cf.* section 3 de cette contribution).

Chacun des niveaux sert à encoder des informations particulières. La tire *Mots* est utilisée pour effectuer un alignement entre les mots produits et les portions de signal qui leur correspondent. Dans les autres tires, les étiquettes sont alignées avec des points déterminés sur le signal comme :

i) le milieu d'une syllabe perçue comme 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 en fonction de la façon dont s'opère l'ancrage des phénomènes mélodiques.

Sur la tire *Rythme*, l'étiquette **P** indique que la syllabe marquée est plus proéminente que les syllabes adjacentes. Cela peut se caractériser au niveau acoustique par un allongement de la durée, un mouvement mélodique, etc. (*cf.* Auer et Gilles, 2003). 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 *Perception phonétique locale* et *Phonologie*. Ceci étant, cette dernière opération n'est pas une nécessité.

La tire *Perception phonétique locale* est utilisée pour transcrire la forme des mouvements mélodiques réalisés sur les syllabes proéminentes ainsi que sur les syllabes adjacentes.

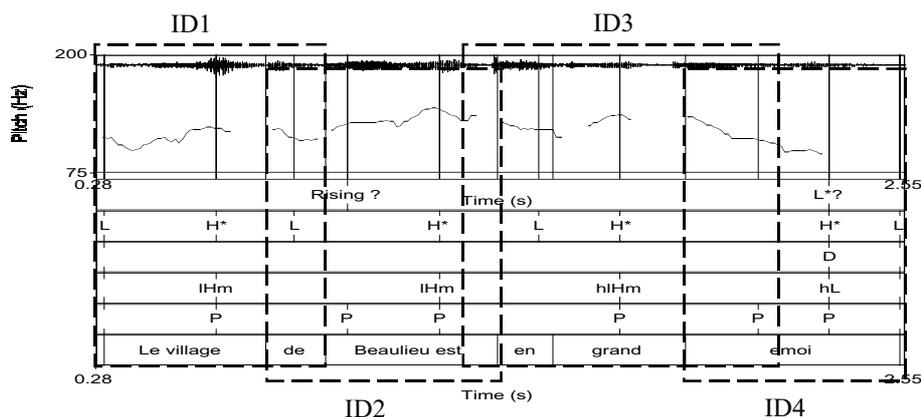


Figure 1
Découpage en domaines d'implémentation accentuel (ID)

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 *le village de Beaulieu est en grand émoi* se décompose en quatre ID, un pour chaque syllabe proéminente marquée par un mouvement mélodique.

Comme le montre la figure 1, 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 syllabe finale qui n'apparaissent dans aucun autre ID. Les niveaux retenus pour l'encodage mélodique sont le niveau 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, suivi d'une légère descente, dont le sommet est atteint sur la syllabe marquée **P** (niveau H). Notons que les étiquettes représentent des valeurs relatives, puisqu'elles dépendent des réalisations au sein d'un unique ID. Par exemple, les *H* (ou *h*) ont des fréquences fondamentales assez différentes d'une étiquette à l'autre, de même le *m* de ID1 est de hauteur comparable au *l* de ID2.

Les phénomènes mélodiques se réalisant sur des empan qui dépassent le cadre d'un ID sont encodés sur la tire « **Perception phonétique globale** ». Ainsi, les *resets* 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**. On peut y ajouter des symboles comme *IP* pour indiquer de quel domaine il s'agit.

Lorsque les événements notés sur les tires « **Perception phonétique locale** » et « **Perception phonétique globale** » sont considérés comme des événements qui font partie du système intonatif phonologique, 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 transcrire³. IVTS propose

3. Sur ce point, il est important de noter que la notion de contraste peut avoir des acceptions différentes selon qu'on privilégie une approche morphémique à base sémantique ou une approche basée sur la forme. L'exemple que nous proposons est principalement extrait de Post (2000), mais d'autres conceptions phonologiques pourraient être proposées et utilisées pour encoder les informations phonologiques.

simplement un ensemble de primitives tonales parmi lesquelles chaque transcripateur peut faire son choix (*cf.* Post *et al.*, 2006) pour une discussion plus détaillée).

Tableau 2
Ensemble d'étiquettes pour le niveau phonologique

Tons		Modificateurs	Frontières	
H*	H	^ : <i>upstep</i>	%H	H%
L*	L	! : <i>downstep</i>	%L	L%
+		> : <i>propagation</i>	%	%

Pour finir, la tire « Commentaires » est utilisée pour noter les remarques diverses.

La figure 2 ci-dessous représente une utilisation d'IVTS pour transcrire des données du français. 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 que nous venons de décrire brièvement sont alignés temporellement avec le signal et la courbe de fréquence fondamentale. Cette transcription a été effectuée sous PRAAT (Boersma et Weenink, 2005), mais elle pourrait l'être avec tout logiciel d'analyse et de visualisation de la parole.

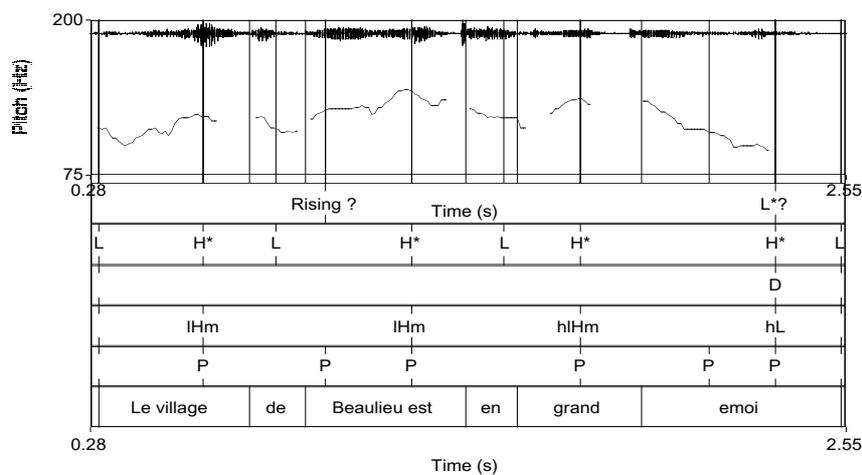


Figure 2
Exemple de transcription IVTS en français (Liège)

Il est important de noter que le fait d'encoder les informations prosodiques sur plusieurs niveaux offre de nombreux avantages, lesquels ont été mentionnés dans les publications consacrées à IViE (Grabe et Post, 2004). Tout d'abord, en utilisant explicitement quatre niveaux pour annoter les phénomènes prosodiques, le système permet de décrire les variations dialectales de nature taxonomique (Ladd, 1996). En effet, la variation (régionale) de la prosodie en français peut relever de différences qui affectent la distribution des syllabes proéminentes, la réalisation phonétique des contours intonatifs (implémentation phonétique), ou l'inventaire phonologique des contours présents dans une variété. Toutes ces différences peuvent être directement encodées dans la tire d'annotation dont elles relèvent. En outre, s'il existe des restrictions dans les possibilités d'agencement des contours, elles peuvent être dérivées à partir de la tire « **Phonologie** ». Deuxièmement, le recours à plusieurs niveaux d'annotation prosodique permet d'obtenir plus d'uniformité et de transparence dans la réalisation des transcriptions. Comme les transcrip-teurs peuvent être amenés à analyser des variétés qui n'ont jamais été décrites auparavant, il est intéressant d'avoir accès aux différents types d'informations (perceptives, phonologiques ou rythmiques) qui ont conduit à faire telle ou telle analyse. Le processus dans son ensemble n'en est que plus transparent. De même, lorsqu'un inventaire phonologique des contours n'est pas encore établi, les transcrip-teurs peuvent d'abord travailler sur la tire « **Perception phonétique locale** » où ils encodent les différents mouvements mélodiques qu'ils perçoivent, et ensuite tenter de proposer une analyse phonologique. Ceci étant, ils auront toujours la possibilité de remettre en cause les hypothèses qu'ils auront formulées pour l'analyse phonologique.

2.2. IPA et codage prosodique⁴

Initialement, le système de transcription de l'IPA a été développé pour encoder les informations segmentales, comme le montrent d'ailleurs les présupposés sur lesquels il se fonde. Ceci étant, un ensemble de symboles a été proposé pour encoder certains phénomènes prosodiques, qu'ils soient de nature métrique ou, au contraire, tonale. Pour encoder les phénomènes métriques, l'IPA propose deux symboles distincts, « ' » et « ˊ » : le premier pour représenter les syllabes recevant un accent primaire et l'autre celles recevant un accent secondaire. Ces symboles se placent juste avant la syllabe proéminente. Une autre série de symboles est utilisée pour représenter les découpages

4. La section 2.2. expose les modalités d'encodage des phénomènes prosodiques dans le cadre de l'IPA. Ceci étant, elle pourrait également s'appliquer à l'extension prosodique de l'IPA lisible sur des ordinateurs, à savoir SAMPROSA. Dans ce modèle, les symboles retenus et les événements annotés sont comparables à ce qui est fait avec l'extension prosodique de l'IPA. Pour avoir des informations sur SAMPROSA, le lecteur peut consulter le site suivant : <http://www.phon.ucl.ac.uk/home/sampa/samprosa.htm>.

prosodiques. Deux niveaux de structuration sont retenus : le pied (ou groupe mineur) représenté par le symbole « | » , et le groupe intonatif majeur représenté par le symbole « || » . Pour les phénomènes de nature tonale, l'IPA a deux séries de symboles, l'un pour les tons statiques⁵, l'autre pour les tons modulés⁶. Ces symboles sont conçus pour transcrire les tons lexicaux dans des langues comme le chinois (*cf. Handbook of the International Phonetic Association, 1999*) ; en revanche, ils ne permettent pas un encodage des phénomènes intonatifs phrastiques. Dans ce cas, quelques symboles à large portée peuvent être utilisés : le symbole du *downstep* ↓, celui de l'*upstep* ↑, celui d'un mouvement descendant ↘ et celui d'un mouvement montant ↗.

Passons maintenant aux types de problèmes que pose ce système de transcription si nous voulons tenter d'encoder les phénomènes prosodiques phrastiques d'une langue, qu'elle soit ou non décrite. Le premier problème concerne l'encodage des phénomènes accentuels métriques. Pour annoter les syllabes accentuées, IPA propose de faire une distinction entre accent primaire et accent secondaire, c'est-à-dire de travailler au niveau phonologique. En effet, cette distinction suppose que le fonctionnement métrique de la langue soit connu, et donc qu'il soit possible de dire si une prééminence est un accent primaire ou secondaire. Dès lors, la possibilité, qui existe au niveau segmental, d'opérer une distinction entre transcription large et transcription étroite semble impossible ici.

La seconde difficulté résulte dans l'utilisation des symboles de segmentation en groupes prosodiques. L'IPA propose d'établir une distinction entre deux niveaux de structuration prosodique : le pied (ou groupe mineur) et le groupe intonatif. Ceci étant, rien n'est dit sur les critères qui interviennent dans la définition de ces groupements. Dans bien des travaux, le pied correspond à un constituant métrique de base qui se construit donc à partir d'une distinction entre syllabes proéminentes et syllabes non proéminentes : il contient une séquence de syllabes dont l'une est accentuée (*cf. entre autres, Hayes, 1995*). Mais une étude des différents exemples dans le

5. Une synthèse des symboles représentant les tons statiques est proposée dans le tableau ci-dessous :

Extra Haut	Haut	Moyen (mid)	Bas	Infra bas
ě	é	ē	è	è̇
ǀ	ǁ	ǂ	ǃ	Ǆ

6. Les différents symboles représentant les tons modulés sont donnés dans le tableau ci-dessous :

Contour montant	Contour descendant	Contour montant de Mid à Haut	Contour Montant de bas vers Mid	Contour Montant-descendant
ě	è	ē	è̇	ě̇
ǀ	ǁ	ǂ	ǃ	Ǆ

Handbook of the International Phonetic Association (1999) montre que les choses sont loin d'être aussi uniformes. Dans le cas du catalan, le choix entre les deux niveaux semble être subordonné à l'usage ou non d'un contour terminal : si la séquence s'achève par une continuation, la frontière de groupes mineurs est utilisée, sinon c'est celle des groupes majeurs. Pour la transcription du français, les frontières de groupes majeurs sont utilisées aussi bien après un contour de continuation qu'après un contour final. Le choix entre l'un ou l'autre de ces symboles s'avère donc souvent délicat et dépendant du transcripateur. Notons d'ailleurs que ce problème de segmentation est constant dans la majorité des systèmes de transcription, sauf peut-être avec ToBI et les *break indices* qui constituent un inventaire plus solide et plus robuste.

Le dernier point délicat est l'encodage des phénomènes intonatifs dont la réalisation se fait sur une séquence ou un empan. Bien que des symboles soient proposés pour encoder les *downsteps*, les *upsteps* ou les mouvements globaux, nous ne sommes pas en mesure de savoir comment ils doivent être assignés. Est-ce à toute syllabe portant un mouvement intonatif, et donc, de façon relative ? Ou est-ce sur l'ensemble du domaine sur lequel se réalise le phénomène de *downstep*, de montée intonative, etc. ? Dans ce dernier cas, il reste alors à déterminer comment encoder les domaines sur lesquels portent ces symboles !

Tentons maintenant de faire avec l'IPA une transcription, que nous commenterons, de l'énoncé proposé dans la figure 2 :

(2) *Le village de Beaulieu est en grand émoi* (locuteur de Liège)
 [lə.vi.'laʒ | də.bo:.'ljø | e.tã.'grã: | te:.'mwa ˩]

En (2), les syllabes [bo:] et [te:] ne sont pas marquées comme portant un accent (sans doute secondaire), contrairement à ce que propose la transcription en IVTS (fig. 2). Nous avons plutôt choisi de noter ces syllabes comme allongées. En réalité, cela résulte du fait que nous ne sommes pas en mesure de décider si ces proéminences relèvent du niveau phonologique ou non, et donc si elles doivent être encodées comme des accents secondaires. En revanche, la perception de la proéminence résultant d'un allongement, il est aisé d'indiquer dans la transcription segmentale le fait que les voyelles sont allongées.

Comme on le voit, l'extension prosodique de l'IPA n'offre pas une souplesse comparable à ce qui se fait au niveau segmental. Il paraît difficile d'encoder avec ses symboles la prosodie d'une langue non encore décrite phonologiquement. Les supposés théoriques inhérents au système sont donc beaucoup plus lourds : d'une part, le niveau phonologique est le seul à être pris en compte ; d'autre

part, la prosodie est surtout vue comme un élément de contraste lexical, comme l'indiquent les symboles tonaux.

2.3. INTSINT

L'école aixoise propose un modèle théorique et des outils d'annotation des systèmes intonatifs dont la description la plus exhaustive est présentée dans Hirst, Di Cristo et Espesser (2000). L'originalité de cette approche consiste à proposer un outil d'annotation automatique à deux niveaux : un niveau de représentation « phonétique » généré par l'algorithme MOMEL (MOdeling MELody) et un niveau de représentation « phonologique de surface » automatiquement transcrit au moyen de l'alphabet INTSINT (International Transcription System for INTonation).

2.3.1. L'algorithme MOMEL et la représentation « phonétique » de la fréquence fondamentale

Développé dès 1993 (Hirst et Espesser, 1993), cet algorithme transforme la courbe discontinue issue de la détection brute de la fréquence fondamentale en une courbe continue conçue comme sa résultante intonativement pertinente. Le rôle de l'algorithme est de séparer la composante « macroprosodique » de la composante « microprosodique », écartée comme linguistiquement non pertinente. En sortie, MOMEL génère un ensemble de points définis par un couple localisation temporelle/F₀ ; ces points sont ensuite reliés par une fonction spline quadratique dont ils représentent les zones d'inflexion (les sommets et les vallées).

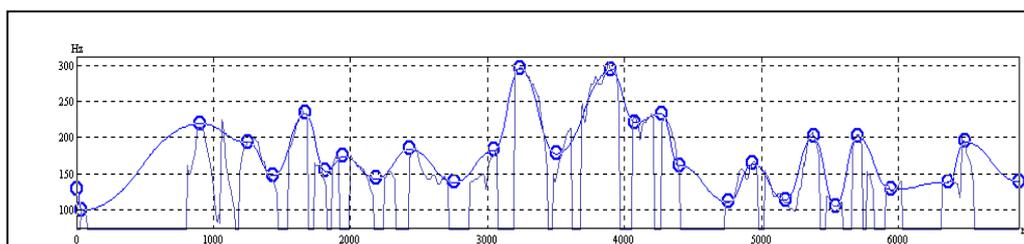


Figure 3
*Points-cibles et courbe continue générée par l'algorithme MOMEL
à partir de la détection de la fréquence fondamentale*

2.3.2. L'alphabet INTSINT et la représentation « phonologique de surface » de l'intonation

L'alphabet INTSINT consiste en un jeu de huit symboles tonals abstraits. Trois de ces symboles codent des tons « absolus » qui délimitent l'étendue globale du registre du locuteur sur l'empan d'une « unité intonative » (correspondant à l'unité maximale de phrasé prosodique pour ce modèle) : il s'agit des symboles T pour *Top*, M pour *Mid* et B pour *Bottom*. Les cinq autres symboles codent les tons « relatifs » dont la valeur dépend de celle du ton précédent. Les tons relatifs se distribuent en deux sous-catégories : les tons non itératifs (H pour *higher*, S pour *same* et L pour *lower*) et les tons itératifs (U pour *upstepped* et D pour *downstepped*). Outre les symboles orthographiques, un second jeu de symboles iconiques est utilisé préférentiellement au sein des transcriptions orthographiques de textes, où ces symboles ne peuvent pas être confondus avec les lettres de l'alphabet. La correspondance entre symboles iconiques et symboles orthographiques est donnée dans le tableau 2 ci-dessous.

Il est possible d'obtenir le codage INTSINT automatique de données de parole au moyen d'un algorithme tenant compte de deux paramètres supplémentaires : la clé et le registre. L'un et l'autre dépendent à la fois du locuteur et de l'énoncé. Les tons absolus T et B sont définis comme les limites du *registre* tonal du locuteur qui se répartit de façon symétrique autour de la *clé* qui définit la valeur du ton M.

Tableau 2
Symboles orthographiques et symboles iconiques pour INTSINT (adapté de Hirst et al., 2000)

		Positifs	Neutres	Négatifs
Absolus		T ↑	M ⇒	B ↓
Relatifs	Non itératifs	H ↑	S →	L ↓
	Itératifs	U <		D >

2.3.3. Avantages et inconvénients du système de transcription aixois

2.3.3.1. Avantages

Le principal avantage que présente ce système de transcription de l'intonation est de permettre l'annotation automatique des données. Dès lors que l'on dispose au préalable du découpage des données en unités intonatives (UIs), il est possible d'obtenir automatiquement une version fiable de la courbe MOMEL et du codage INTSINT des énoncés que l'on veut décrire. Pour éviter le codage manuel des unités intonatives, les utilisateurs du système aixois recourent généralement au

découpage automatique des données en intervalles inter-pauses, réputés comparables aux UIs, en détectant les silences dont la durée excède 200 ms.

Un deuxième avantage provient de l'inter-convertibilité entre une séquence de points-cibles issue de l'algorithme MOMEL et leur codage INTSINT. Il est donc possible de tester l'effet acoustique de n'importe quelle séquence de tons codés avec INTSINT en synthétisant un énoncé auquel est appliqué la courbe MOMEL correspondant au codage INTSINT testé. L'inter-convertibilité fait de ce système un outil exploratoire des variations intonatives très facile d'accès.

L'avantage le plus important pour ce qui concerne la problématique du présent article est que l'annotation peut être réalisée indépendamment de la connaissance préalable que l'on a de la phonologie intonative de la langue étudiée. Ceci permet des comparaisons inter-langues qui sont illustrées dans Hirst et Di Cristo (1998).

2.3.3.2. Inconvénients

Lorsqu'on utilise ce modèle, il importe d'être conscient que même la courbe MOMEL est le produit d'une modélisation qui émet des hypothèses sur la nature micro- ou macro- de l'information prosodique. Ces hypothèses ne font pas nécessairement l'unanimité dans la communauté des chercheurs.

Bien que l'unité intonative ait fait l'objet de définitions précises (voir par exemple Di Cristo et Hirst, 1996), les critères qui président à son identification dans les corpus à annoter sont peu explicites. Ce problème se pose néanmoins à l'ensemble des systèmes de transcription. Nous aurons l'occasion de revenir sur ce point (voir notre conclusion).

Selon nous toutefois, le principal inconvénient du modèle réside dans sa conception de l'interface phonétique/phonologie. Les deux points suivants en illustrent la difficulté :

- la nature « phonologique de surface » du codage INTSINT pose un certain nombre de problèmes théoriques. Selon Hirst (2005), la nature phonologique des tons codés par l'alphabet INTSINT est due essentiellement à leur caractère discret. Ces tons sont supposés « distinctifs », mais il n'est pas précisé de quelle manière ils le sont. Par exemple, il n'est pas certain que la substitution d'un ton B à un ton M dans une séquence de tons formant un contour intonatif, ou que l'insertion d'un ton U entre un ton L et un ton H ait des conséquences sur l'interprétation qui sera donnée du contour. Or il nous semble que ce point est crucial dans un système d'annotation de l'intonation.
- Enfin, la volonté proclamée d'éviter l'annotation auditive manuelle au profit de l'annotation automatique, bien qu'elle puisse être versée en partie au crédit du modèle, (voir section a) ci-dessus) a son revers : n'est-ce pas justement la transcription auditive qui donne accès à la nature phonologique ou non d'une forme lorsque le système phonologique de la variété étudiée n'est pas encore établi ?

2.4. ToBi

ToBi (Silverman *et al.*, 1992 ; Beckman *et al.*, 2005) constitue la première tentative réussie pour développer un système d'encodage prosodique approprié pour la transcription de grands corpus de parole. Ce système a été largement adopté pour transcrire une bonne dizaine de langues aux systèmes prosodiques assez divers, comme l'anglais, le coréen, le grec, le mandarin, et le suédois (*cf.* Jun, 2005a). Ceci témoigne du fait qu'un seul modèle prosodique peut caractériser avec succès des langues de typologie prosodique très différente (*cf.* Jun, 2005a : 1) ; cela renforce donc également la validité des objectifs que nous nous fixons avec IVTS. Comme le souligne d'ailleurs Jun, une telle approche facilite les comparaisons inter- et intra-linguistiques sur le plan de la typologie prosodique ainsi que l'étude des relations entre les différentes parties de la grammaire (la sémantique, la syntaxe, *etc.*).

Tout comme IVTS, ToBi se fonde directement sur les principes de la théorie Métrique-Autosegmentale (Bruce, 1977 ; Pierrehumbert, 1980 ; Gussenhoven, 1984, 2004 ; Ladd, 1996), c'est-à-dire qu'il y a une structuration intonative sur plusieurs niveaux d'analyse, notamment au niveau phonologique, où se manifestent les phénomènes linguistiquement distinctifs, et au niveau phonétique qui concerne la réalisation de ces phénomènes, ce niveau dépendant de la structure segmentale, du locuteur, du contexte, de l'attitude et de l'émotion, *etc.* Par ailleurs, l'intonation est analysée en termes de niveaux distinctifs (des points-cibles représentés comme des tons H et L), plutôt qu'en termes de configurations dynamiques. Les tons s'associent avec les syllabes accentuées et les frontières de groupe. Les valeurs F0 effectives sont dérivées au cours de l'implémentation phonétique des représentations phonologiques.

Le système ToBi de base propose une annotation sur quatre tires : une tire « Mots », une tire *Break Indices* pour la transcription de quatre degrés de frontières prosodiques, une tire « Tons » pour l'encodage tonal, et une tire « Commentaires » (Silverman *et al.*, 1992 ; Beckman *et al.*, 2005). Les tires sont temporellement alignées avec un enregistrement audio et une représentation graphique de la fréquence fondamentale. ToBi n'a pas encore été adapté pour le français, mais la figure 4, ci-dessous, donne une approximation d'une transcription probable de l'énoncé analysé auparavant avec IVTS et IPA.

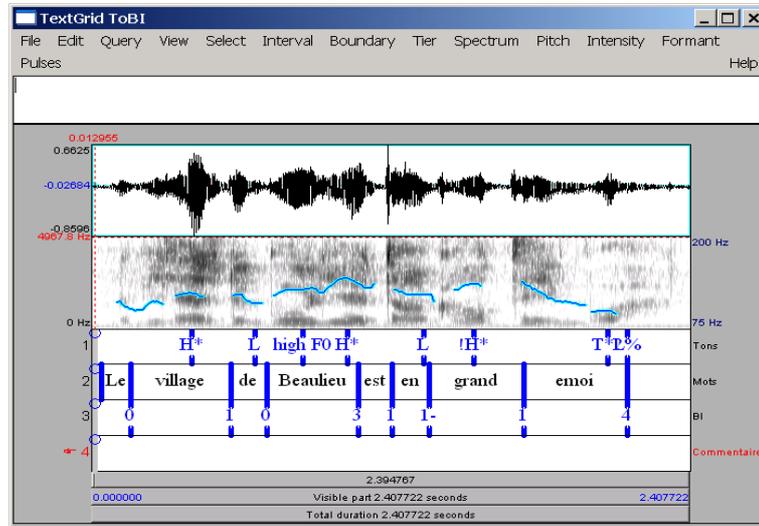


Figure 4
Exemple de transcription ToBI provisoire en français (Liège)

La tire « Commentaires » se trouve tout en bas de la figure. Les annotations sur la deuxième tire *Break Indices* montrent que les quatre niveaux de frontière sont réalisés dans cet exemple. L'étiquette 1 indique que le transcritteur n'est pas sûr qu'il s'agisse d'une frontière 1 ou 2. Viennent ensuite la tire « Mots », et, tout en haut, la tire « Tons ». Il est à noter que, tout comme pour la tire « **Phonologie** » en IVTS, l'annotation sur la tire « Tons » dépend de l'analyse des contours contrastifs adoptée pour le dialecte à transcrire. Dans le cas présent, l'analyse phonologique n'a pas encore été réalisée. Aussi, l'annotation tonale dans la figure sert-elle simplement à illustrer le genre de symboles qui pourrait figurer sur cette tire. Ainsi, nous avons marqué les syllabes proéminentes hautes par H^* , analyse qui fait l'hypothèse que le français de Liège a des syllabes accentuées. Provisoirement, les valeurs F0 minimum ont été annotées L , mais dans ToBI, ces tons devraient probablement être identifiés ou bien comme des tons de phrase (intermédiaire, accentuel ou phonologique ; cf. par exemple Jun et Fougeron, 2002), ou bien comme des accents bitonaux ($L+H^*$; cf. D'Imperio *et al.*, 2006, pour des accents de ce type en position finale de syntagme intonatif). Quoi qu'il en soit, ToBI insiste sur le fait qu'on ne peut pas produire de transcription tonale sans que le transcritteur connaisse le système tonal de la langue en

question (*cf.* Beckman *et al.*, 2005 : 12) ; ainsi, la tire « Tons » ne devrait pas avoir été remplie pour cet énoncé.

Une comparaison avec la transcription IVTS donnée auparavant dans la figure 1 met en évidence l'importance pour IVTS de l'héritage de ToBI, mais elle révèle aussi deux différences importantes :

- IVTS a plus de tires, notamment deux tires phonétiques ;
- la tire *Break Indices* en ToBI et la tire « Rythme » en IVTS, bien qu'elles se ressemblent puisqu'elles servent à indiquer des frontières prosodiques, ne couvrent pas les mêmes phénomènes prosodiques, et elles trouvent leur motivation dans des objectifs différents.

ToBI n'admet pas de tires phonétiques perceptives, parce que la fréquence fondamentale – qui fait obligatoirement partie de toute transcription en ToBI – est vue comme la seule représentation phonétique des événements tonaux qui est pertinente pour l'analyse (*cf.* la discussion dans Beckman *et al.*, 2005 : 36-41). Du point de vue pratique, ajouter des tires phonétiques alourdit considérablement la tâche du transcripateur. Du point de vue théorique, ToBI rejette le type de représentations proposé par IVTS en argumentant qu'il n'est pas possible d'encoder le signal de façon discrète sans faire des hypothèses théoriques dépendantes de la langue à propos du phénomène étudié. Dans la mesure où leur tire « Tons » a le statut d'une transcription phonologique large, cela signifie que ce modèle met à jour l'inutilité, voire l'impossibilité, d'une transcription phonétique étroite intermédiaire entre le niveau acoustique du signal (la courbe de F0) et la phonologie de surface. Au contraire, nous pensons que, en IVTS, les tires phonétiques offrent l'avantage de donner accès de façon systématique et transparente à l'encodage de la variation régionale et de langues peu connues. En outre, la tire globale permet de transcrire des phénomènes prosodiques dont le domaine est plus large que ceux qui sont traditionnellement étudiés dans les recherches linguistiques sur l'intonation. Toutefois, les tires phonétiques peuvent être omises dans une analyse utilisant IVTS, par exemple dans le cas d'une étude dans laquelle les systèmes prosodiques sont bien connus et dans laquelle la variation phonétique est sans intérêt, ou quand il n'y a pas de comparaisons inter- ou intra-linguistiques.

La seconde différence concerne l'annotation des syllabes proéminentes et des frontières prosodiques. Normalement, les proéminences ne sont pas annotées en ToBI, sauf si elles sont réalisées avec un accent tonal transcrit sur la tire « Tons » (voir cependant ToBI en mandarin, Peng *et al.*, 2005). Pourtant, en français, comme en anglais, il peut y avoir des syllabes proéminentes sans accent mélodique, mais marquées par exemple par un allongement de la durée syllabique. Plusieurs études ont montré que, parfois, il s'agissait d'une variable dialectale (Grabe et Post, 2004 ; Post et Delais-Roussarie, 2006, aussi section 3.1. plus bas). Ainsi, la séparation des niveaux phonétique et

phonologique permet l'encodage explicite d'événements rythmiques qui ne sont pas nécessairement tonals, ce que ToBI autorise pour le groupement en domaines.

En outre, IVTS ne spécifie pas tous les niveaux de rupture que spécifie ToBI, pour des raisons pratiques aussi bien que théoriques. Si rien dans la langue à transcrire n'incite à poser quatre niveaux distinctifs, et si le but du transcripteur est de fournir une transcription des événements tonals plutôt que de représenter la structure prosodique et les groupements de mots, marquer la frontière de chaque mot n'apporte pas d'information pertinente. En IVTS, rien n'empêche l'usage de symboles de frontière choisis en fonction du niveau de rupture perçue (p. ex. %1 %2 sur la tire « Rythme ») si cela facilite l'étude des phénomènes prosodiques dans le corpus à transcrire. D'ailleurs, la tire *Break Index* de ToBI mélange réalisation et motivations structurelles, car le transcripteur est supposé annoter le poids perceptif des frontières qu'il entend, mais en même temps, les valeurs 0 – 4 sont supposées correspondre aux relations morphe-syntaxiques et tonales qui sont en jeu dans l'énoncé (Beckman *et al.*, 2005 : 21-23).

Notons que le fait qu'il y ait plus de tirs en IVTS ne signifie rien en soi. Par exemple, pour le système ToBI grec (Arvaniti et Baltazani, 2005), les auteurs ajoutent une tire supplémentaire pour la segmentation en mot prosodique (transcrit en symboles segmentaux IPA), afin d'annoter les processus phonologiques de parole continue, mais aussi pour indiquer où se trouvent les mots prosodiques dans la parole, car ces découpages sont intimement liés à l'accentuation. Un autre exemple est le système ToBI développé pour le coréen, dans lequel Jun (2005b) propose de distinguer entre forme tonale sous-jacente et forme tonale de surface ('phonologique' et 'phonétique' dans la terminologie de Jun). Pour le mandarin (Peng *et al.*, 2005), les auteurs proposent un codage des proéminences syllabe par syllabe qui se fait indépendamment de la notation des *break indices*. Comme en IVTS, le choix en ce qui concerne le nombre de tirs et leur contenu est laissé à la libre appréciation du transcripteur en ToBI, en fonction de la langue à transcrire.

3. Transcription et phénomènes prosodiques particuliers

3.1. Les phénomènes accentuels

Cette section a pour but de voir comment sont encodés les phénomènes accentuels dans les systèmes de transcription qui ont été présentés dans la section précédente, IPA, ToBI, INTSINT et IVTS. Nous entendons par phénomène accentuel tout phénomène se réalisant par une proéminence ou une saillance syllabique, qu'elle résulte d'une variation mélodique, d'un allongement ou d'une modification de timbre. Dans la parole, ces saillances peuvent renvoyer à deux types de faits distincts :

- à des accents pragmatiques, c'est-à-dire des accents réalisés sur un syntagme ou un lexème, afin de lui conférer un statut ou une signification particulière (les accents pragmatiques peuvent être réalisés pour indiquer un contraste (3), pour insister, pour exprimer un sentiment sur un point (4), *etc.*).

- (3) a. A : Qu'est-ce que tu as offert aux enfants ?
 B : J'ai donné un LIVRE à VALÉRIE et un DISQUE à PIERRE
- b. A : Il va à BOlogne, et non un VOlogne.
- (4) a. C'est éPOUvantable !
 b. c'est déGUEUlasse !

- à des accents métriques qui n'ont aucune signification, mais qui participent à la construction des schèmes métriques⁷.

- (5) (ces enFANTS) (sont arriVEs) (en reTARD).

Bien qu'il soit nécessaire de distinguer ces deux types de proéminences, dans la mesure où leur réalisation et leur distribution ne sont pas régies par les mêmes principes, les systèmes de transcription ne doivent pas poser d'*a priori* sur la nature des saillances perçues. De plus, la manifestation phonétique de ces proéminences ne varie pas nécessairement en fonction de leur catégorie, puisqu'elle est également influencée par le contexte, la position linéaire dans l'énoncé, notamment par rapport au focus, *etc.*

Parmi les quatre systèmes, INTSINT et ToBI (en tous cas dans sa version 'classique'⁸) ne permettent pas d'encoder clairement les phénomènes accentuels, avant que leur phonologie ne soit connue. INTSINT n'est pas en mesure d'encoder ce type d'information principalement pour deux raisons : i) d'une part, le système vise à transcrire les aspects mélodiques, et cela sans tenir compte de phénomènes prosodiques liés à des variations de timbre, d'intensité ou de durée (bien que certaines tentatives allant dans ce sens existent, *cf.* Hirst et Auran, 2005) ; ii) d'autre part, la détermination des points-cibles nécessaires à une représentation discrète de la courbe de fréquence fondamentale est faite uniquement sur une base acoustique, et non d'après des informations perceptives ou linguistiques. Or, le caractère proéminent d'une syllabe relève davantage de la

7. Les accents métriques correspondent à ce qui est appelé *stress*, mais ils peuvent être réalisés de différentes façons selon les langues : variation de timbre, variation de durée, mouvement mélodique, *etc.* En outre, dans le cas du français, les accents métriques primaires ont parfois été analysés comme des frontières, dans la mesure où ils remplissent une fonction démarcative et sont réalisés sur les syllabes finales des lexèmes.

8. Certains travaux menés dans la mouvance de ToBI ont recours à une tire particulière pour l'accentuation métrique (*stress*), comme, par exemple, l'adaptation de ToBI à la transcription prosodique du chinois mandarin (*cf.* Peng *et al.*, 2005).

perception que de l'acoustique. Il peut résulter d'une somme d'indices (allongement, modification de timbre, modification de l'intensité, etc.).

ToBI ne permet pas non plus d'encoder aisément ces phénomènes, dans la mesure où ce système a surtout été développé pour encoder les informations intonatives et les niveaux de constituance (notamment avec les indices de rupture ou *break indices*). Parmi les accents, seuls ceux qui ont un marquage mélodique seront donc notés. En outre, l'encodage en ToBI nécessite une connaissance préalable du système métrique et mélodique de la langue transcrite. Dès lors, le système ne peut pas être utilisé pour transcrire des données dans une langue non décrite, afin de proposer ensuite une analyse du système métrique.

Dans l'extension prosodique de l'IPA, l'encodage des informations accentuelles se fait, comme dans ToBI, au niveau phonologique. Les symboles reposent sur une distinction entre accent primaire et accent secondaire, distinction qui ne peut être établie que lorsque le système métrique de la langue est connu. De plus, le système de transcription ne permet pas d'encoder des syllabes proéminentes qui résultent d'un marquage sémantico-pragmatique comme en (3) et en (4), d'autant que ces dernières ne coïncident pas nécessairement avec les syllabes portant un accent métrique. De même, le système ne permet pas de noter des proéminences qu'il est plus difficile à catégoriser. Considérons, par exemple, les données sous (6).

(6) *Le premier ministre a en effet décidé de faire étape dans cette commune.*

- a. Dans nos données, un locuteur marseillais a réalisé cet énoncé de la façon suivante⁹:

Le PRemier miNISTRe a EN effet déciDE de faire éTAPe dans cette coMMUNE au COURS de sa tourNEE dans la réGION en fin d'aNNEe.

- b. Le locuteur belge, originaire de Liège, a réalisé cet énoncé comme suit :

Le preMIER miNISTRe a en effET déciDE de faire éTAPe dans cette COMMUNE au COURS de sa tourNEE dans la réGION en FIN d'aNNEe.

Le locuteur marseillais a réalisé des accents, généralement décrits comme primaires car frappant la syllabe finale des lexèmes dans *ministre, décidé, étape, commune, cours, tournée, région* et *année*. En revanche, il a réalisé un accent sur *en* qu'il est difficile de catégoriser, mais cela pourrait résulter d'une volonté de respecter le principe de bipolarité dans les syntagmes *le premier ministre* et *a en effet décidé* (cf. Di Cristo, 1998). De son côté, le locuteur de Liège n'a pas fait les mêmes choix. Tous les accents sont en position finale de lexème. Néanmoins, une étude plus fine est nécessaire pour savoir si certains d'entre eux relèvent de l'application du principe de bipolarité (cf. en particulier, les accents sur *premier, cours* et *fin*). Une transcription qui impose dès le départ de distinguer entre

9. Les syllabes transcrites comme proéminentes sont en lettres majuscules.

accents primaire et secondaire ne laisse pas la possibilité de réfléchir en tenant compte de l'ensemble des énoncés réalisés. En outre, le locuteur belge tend à réaliser un accent uniquement marqué par un allongement de la durée sur les syllabes pénultièmes des syntagmes intonatifs majeurs (cf. les accents réalisés sur les syllabes initiales de *Beaulieu* et de *émoi* dans la figure 2, ainsi que celui réalisé sur la syllabe initiale de *commune* en (6)).

Ces accents non standard ne peuvent pas aisément être encodés dans des systèmes qui présupposent une connaissance du système métrique (ToBI ou IPA), ou qui favorisent le marquage intonatif (INTSINT), dans la mesure où certains de ces accents, dont ceux réalisés sur les pénultièmes des syntagmes intonatifs majeurs, reposent sur un allongement.

Dans IVTS, les phénomènes accentuels, et plus précisément les saillances perceptives, sont encodés dans la tire « **Rythme** ». L'étiquette P est en effet assignée à toute syllabe proéminente, et cela quelle que soit la nature de la proéminence (métrique ou non), et quelle que soit sa manifestation prosodique (dans la figure 2, les proéminences sur les syllabes initiales de *Beaulieu* et de *émoi* sont notées P, bien qu'aucun mouvement mélodique ne soit associé). La seule chose qui nécessiterait d'être approfondie est de voir comment catégoriser phonologiquement ces différentes proéminences, certaines pouvant être purement phonétiques. Il serait sans doute judicieux d'ajouter une tire « Phonologie Métrique » dans laquelle serait encodé en accents primaire et secondaire (ou non primaire) tout accent participant à la construction des schèmes métriques, quelle que soit la façon dont ils sont réalisés. Les accents pragmatiques seraient alors encodés par défaut dans la tire « **Phonologie** ». Des réflexions dans ce sens doivent donc être menées afin de renforcer la flexibilité et la richesse du système.

3.2. Les phénomènes mélodiques

Comme nous l'avons expliqué dans la section 2, chaque système de transcription encode différemment le contour mélodique en fonction des points de vue théoriques sur lesquels il se base¹⁰. INTSINT propose un encodage de tout mouvement mélodique à deux niveaux, indépendamment de sa fonction. L'algorithme MOMEL identifie les points-cibles sur la fréquence fondamentale ; ces derniers sont ensuite transformés en tons discrets INTSINT (par un second algorithme ou par un codage manuel). Bien que la représentation INTSINT soit aussi une représentation autosegmentale, le statut des tons est crucialement différent des représentations tonales de ToBI et d'IVTS. ToBI et IVTS se rapprochent sur ce point, car pour ces modèles, les

10. Comme l'IPA ne permet pas d'encoder des événements intonatifs, mais se limite aux accents primaires et secondaires, et aux tons lexicaux, nous ne l'avons pas inclus dans cette discussion.

mouvements mélodiques associés aux syllabes proéminentes sont distingués des mouvements de frontière, contrairement à ce qui se fait pour INTSINT. Dans TOBI et IVTS, les points-cibles du contour de F0 sont interprétés comme des cibles phonétiques dérivées des tons de la représentation phonologique. Dans cette dernière, chaque ton fait partie d'un accent mélodique ou d'un événement de frontière. Les configurations de tons qui peuvent être réalisées au cours du déroulement de l'énoncé sont ainsi contraintes à chaque moment par la structure prosodique sous-jacente, représentée par des accents et des groupements de mots.

Le statut peu clair des tons dans INTSINT a pour conséquence qu'on ne sait pas très bien dans quelles circonstances une différence de ton devient phonologiquement distinctive (voir section 2.3.2. ci-dessus). En conséquence, un contour mélodique comme celui qu'illustre la figure 5 (voir *infra*) pourrait être analysé de deux façons distinctes. La figure montre deux réalisations du contour montant-descendant réalisé à la fin d'un syntagme intonatif (dernier mouvement mélodique avant le symbole %L dans la tire *phonological*). Dans la partie haute de la figure, la totalité du mouvement montant-descendant est réalisée sur la syllabe finale alors que dans la partie basse de la figure, le mouvement monte sur la syllabe pénultième et descend sur la syllabe finale qui a pour noyau un schwa. Les deux mouvements ont des réalisations acoustiques différentes : trois cibles peuvent être identifiées dans la partie haute de la figure (*i.e.* L, H et L) alors qu'on en identifie 4 dans la partie basse (L, T, L et S).

Comme le montre notre transcription dans les tires phonétique et phonologique de la figure 5, nous avons analysé ces deux mouvements comme appartenant à la même catégorie phonologique mais ayant deux réalisations phonétiques différentes qui reflètent une différence d'alignement par rapport à la frontière du syntagme (IHL *vs* mHI dans la tire *local phonetic*) et une différence de hauteur du pic mélodique par rapport à l'accent précédent (*Downstep and Upstep*). C'est en définitive cette différence *phonétique* qu'encode INTSINT : la différence d'alignement est reflétée par l'opposition entre L (en haut) et LS (en bas) ; la différence de hauteur par l'opposition entre H (en haut) de T (en bas).

Bien qu'INTSINT soit différent de TOBI et d'IVTS à cet égard, sur d'autres points il ressemble à TOBI. Les deux systèmes encouragent en effet le transcripneur à étayer prioritairement sa décision sur des informations acoustiques plutôt que perceptuelles. INTSINT prévoit bien une représentation phonétique, mais dans ce modèle, les points-cibles sont dérivés automatiquement à partir du contour de F0 et ne représentent donc pas nécessairement une information auditivement pertinente. Or cette information est particulièrement importante lorsqu'on analyse des langues qui n'ont pas encore été décrites (Nolan et Grabe, 1997). En effet, deux tracés de fréquence fondamentale qui sont différents visuellement peuvent être associés à la même impression auditive, soit parce que l'auditeur neutralise

auditivement les différences microprosodiques associées (*auditory sensation*, Crystal, 1969), soit parce que la différence auditive, bien que perçue, n'affecte pas l'interprétation du tracé mélodique (*auditory interpretation*, Crystal, 1969).

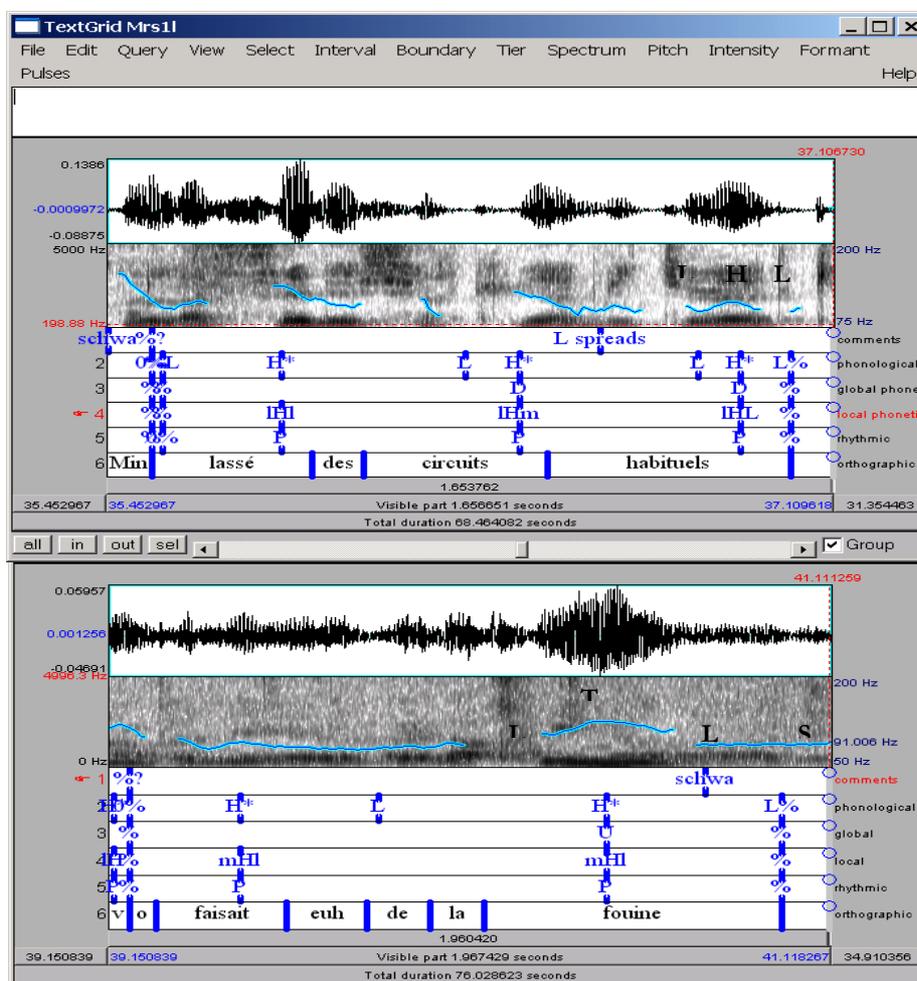


Figure 5
Deux réalisations phonétiques différentes d'un mouvement montant-descendant en français de Marseille

Dans le modèle INTSINT, l'algorithme MOMEL n'élimine que les variations micromélodiques (voir cependant la section 2.3.1.). Quant à ToBI, ses transcrip-teurs ne font pas la différence entre des variations mélodiques perceptivement saillantes et des événements tonals phonologiquement pertinents puisqu'ils ne disposent pas d'un niveau d'encodage phonétique (autre que la courbe de F0 elle-même).

L'absence d'un niveau phonétique auditivement pertinent n'implique pas seulement qu'INTSINT et ToBI puissent rencontrer des difficultés quand l'acoustique et le perceptif se contredisent. Elle implique en outre, pour ToBI, que l'étiquetage relève obligatoirement du niveau phonologique et ne puisse être réalisé que lorsque le système tonal d'une langue est déjà connu (*cf.* Grabe et Post, 2004, mais aussi Beckman *et al.*, 2005). Nous pensons que pour analyser le type de données utilisées dans cet article, l'usage d'annotations discrètes mais susceptibles de représenter tous les événements mélodiques perceptivement saillants peut fournir un outil d'analyse puissant. Ainsi, dans la figure 5 ci-dessus, l'annotation IVTS rend les décisions du transcrip-teur explicites à chaque étape : une comparaison entre la tire phonétique locale et la tire phonologique montre que deux réalisations phonétiques différentes sont interprétées comme une même représentation phonologique. Dans ToBI, ceci ne serait pas évident puisque le transcrip-teur n'a pas à rendre compte de la manière dont il perçoit le contour. Les différences d'alignement ne sont pas clairement exprimées non plus.

Un avantage supplémentaire avec la tire « **Perception phonétique locale** » est qu'elle permet d'établir une comparaison qualitative rapide pour les réalisations phonétiques des accents mélodiques entre des jeux de données différents (Nolan et Grabe, 1997). Le fait qu'on rencontre beaucoup de micro-variations acoustiques dans la réalisation des accents mélodiques, même quand ils sont alignés de façon similaire avec la structure segmentale, peut masquer des similarités entre mouvements. Pour illustrer ce point, la figure 6 ci-dessous montre trois exemples de mouvements à la finale de syntagme intonatif dans le français de Périgueux. Dans tous les cas, on peut observer une montée sur la syllabe pénultième proéminente, mais il s'avère que trois choix sont possibles pour la réalisation du contour sur le schwa suivant : a) la F0 continue de monter (panel supérieur), b) la F0 continue au même niveau de hauteur (panel moyen), c) la F0 redescend (panel inférieur). Comme le montre la transcription, nous avons provisoirement étiqueté les trois mouvements comme trois catégories tonales différentes : (L)H*H%, (L)H*L% et (L)H*0% (c'est-à-dire pas de spécification tonale différente pour la frontière ; voir Grabe, 1998). On peut ensuite chercher dans les données transcrites les étiquettes phonétiques correspondant aux accents pénultièmes de fin de syntagme intonatif pour confirmer ou infirmer cette analyse provisoire. En d'autres termes, on peut aussi utiliser la transcription phonétique pour formuler des hypothèses sur les relations entre les catégories intonatives et leurs réalisations phonétiques. De telles hypothèses doivent bien sûr

être testées quantitativement, comme cela fut fait pour l'hypothèse des trois spécifications tonales des tons de frontières en français parisien dans Post (2000). Cette hypothèse était issue de données de corpus et fut vérifiée au moyen de tests de perception.

Le dernier point que nous voudrions aborder à propos de la transcription des mouvements mélodiques dans les systèmes de transcription que nous comparons ici concerne le fait qu'aucun d'eux ne distingue un niveau sous-jacent et un niveau de surface en phonologie (cette distinction existe dans INTSINT mais ce qui correspond au niveau « phonologique de surface » dans ce modèle relève plutôt du niveau phonétique selon nous : voir la section 2.3.3., b) et l'analyse de la figure 5 ci-dessus ; pour ToBI voir aussi Jun, 2005a). Aussi, au niveau sous-jacent, les éléments tonals sont transcrits sous la forme qu'ils sont supposés avoir dans l'inventaire des formes contrastives d'une langue. Un exemple fictif d'un tel inventaire pourrait être un ensemble d'accents mélodiques H*L et L*H qui s'associent avec les syllabes proéminentes (ce qui pourrait être prédictible dans une certaine mesure en fonction de la structure métrique sous-jacente) et de deux tons de frontière L% et H%, qui pourraient s'associer avec les frontières de syntagmes intonatifs dans une langue donnée. La 'grammaire intonative' (du locuteur) choisirait des éléments dans l'inventaire tonal et les associerait avec la structure segmentale. Ils seraient alors soumis à des règles phonologiques qui pourraient altérer ou supprimer les spécifications tonales, soit optionnellement ou obligatoirement (Gussenhoven, 1984, 2004, pour l'anglais). Par exemple, la « suppression de ton » de Gussenhoven (1984) peut supprimer un ton L sous-jacent dans un accent mélodique H*L en néerlandais. Si cette règle de la suppression de tons est appliquée au premier de deux accents H*L consécutifs dans un syntagme, le patron résultant sera le « chapeau plat » de la tradition IPO (t Hart *et al.*, 1990).

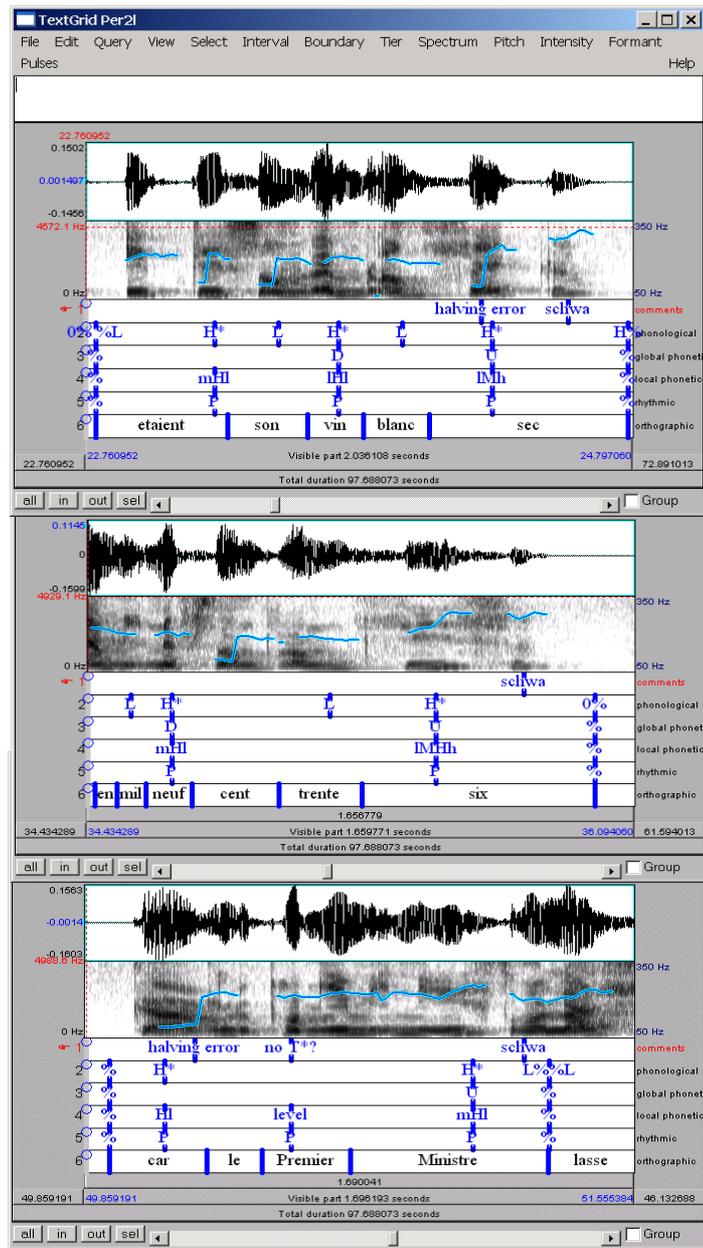


Figure 6
F0 montante, en plateau et descendante après un accent montant en français de Périguenx

La tire de transcription dédiée à la phonologie de surface devrait donner les éléments tels qu'ils sont réalisés une fois les opérations phonologiques, comme la suppression de tons, appliquées. Le niveau phonologique de surface est un niveau phonémique et c'est probablement le niveau que la plupart des tires « Tons » dans ToBI et dans les systèmes qui se fondent sur IViE transcrivent. Nous aimerions insister sur le fait que ce niveau n'est pas identique au niveau phonétique auditivement pertinent dans IVTS, qui encode auditivement les détails saillants du contour mélodique. Bien que ce niveau consiste en une abstraction à partir du signal acoustique dans le sens où il enregistre les événements auditifs qui sont potentiellement des faits phonologiques plutôt qu'acoustiques, il n'encode pas uniquement les événements qui sont effectivement phonologiquement distincts les uns des autres (soit au niveau phonologique de surface, soit au niveau sous-jacent). Notez que bien que ni ToBI ni IVTS ne proposent de transcrire un niveau sous-jacent et un niveau phonologique de surface, l'un et l'autre des systèmes sont parfaitement capables d'intégrer différents niveaux phonologiques quand ils le jugent utile. Comme cela a été mentionné plus haut, la version coréenne de ToBI propose de distinguer deux niveaux pour la transcription des tons (Jun, 2005b).

3.3. Les phénomènes liés au registre

Dans les travaux sur l'intonation, la notion de registre concerne le placement des cibles tonales sur l'échelle de fréquence. Deux dimensions sont parfois distinguées dans la littérature, le niveau du registre (*register level*) et l'étendue du registre ou gamme tonale (*register span*). Le niveau du registre renvoie au niveau global de la hauteur pour une séquence ; sa mesure varie d'un auteur à l'autre. Par exemple, Patterson (2000) propose une mesure basée sur les cibles L uniquement alors que Hirst (2005) utilise cibles hautes et basses pour établir ce qu'il appelle la clé du registre du locuteur. L'étendue du registre correspond à la distance entre cibles hautes et cibles basses.

3.3.1. Les phénomènes observés

On observe une distorsion importante entre le grand nombre de phénomènes distincts liés à la notion de registre et recensés dans la littérature et la place relativement modeste qu'ils occupent dans les systèmes de transcription. Nous donnerons une brève définition de chacun de ces phénomènes avant d'examiner les problèmes que pose leur transcription dans les différents systèmes.

La déclinaison est définie comme un effet global et graduel d'abaissement des cibles hautes sur le domaine de la phrase entière (ou de l'énoncé). Ce phénomène est lié à des contraintes physiologiques, notamment à la diminution de la pression de l'air entre deux phases respiratoires.

La notion de *downstep* fait référence à un phénomène phonologique qui conditionne l'abaissement d'une cible H relativement à la cible H qui la précède. Ce phénomène peut être récursif et affecter

plusieurs cibles hautes consécutives. Deux phénomènes distincts peuvent venir interrompre l'abaissement des cibles hautes : i) l'*upstep*, qui est défini comme le retour des cibles H à une valeur plus haute après une séquence de *downsteps* mais à l'intérieur du même domaine phonologique (Truckenbrodt, 2002) ou sans qu'intervienne un changement de topique discursif ; ii) le *reset*, qui désigne le même retour des cibles hautes à une valeur élevée mais à l'occasion d'un changement de domaine (Truckenbrodt, 2002) ou d'un changement de topique discursif. La figure 7 illustre la différence entre ces trois notions sur un exemple emprunté au français.

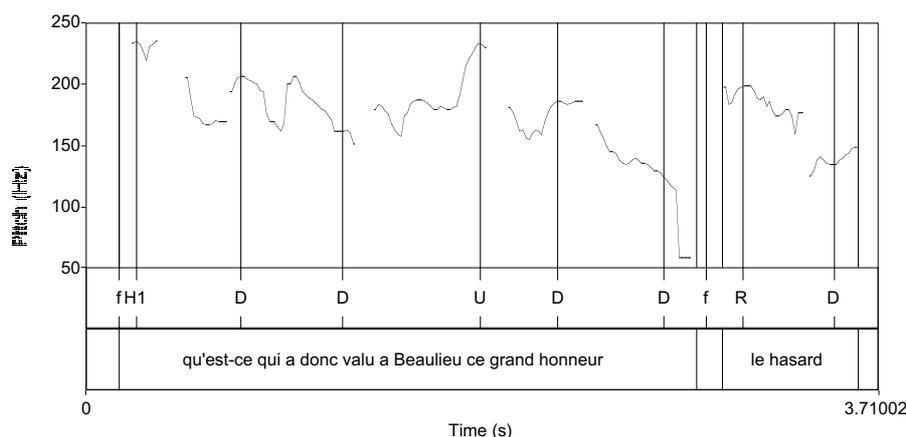


Figure 7

Upstep (U), downstep (D) et reset (R) dans un enchaînement de deux énoncés en français.
 « H1 » note le ton H initial, « f » signale les frontières du domaine pertinent pour la distinction de l'upstep et du reset (du moins selon l'analyse de Truckenbrodt, 2002).
 Les points annotés correspondent au croisement de la courbe de F0 avec les lignes verticales en pointillés.

L'abaissement final (*final lowering*) est un phénomène observé en fin d'énoncé. La notion est définie par rapport à celle de *downstep* dans les travaux de Liberman et Pierrehumbert (1984) sur l'anglais. Elle renvoie à un abaissement plus important que celui qui serait attendu en suivant la pente des *downsteps* successifs qui précèdent. Un effet identique a été mis en évidence en espagnol mexicain (Prieto *et al.*, 1996).

Enfin, la compression du registre correspond à une diminution importante, voire à une disparition, des variations de la fréquence fondamentale sur certaines séquences de la phrase. Ce phénomène a été observé dans plusieurs langues (japonais, anglais, français, *etc.*). En français, il concerne tout particulièrement la séquence post-focale de certains énoncés (Di Cristo, 1998). La compression du

registre n'implique pas pour autant la disparition du phrasé de la séquence concernée qui se trouve alors marqué par des variations de durée de la syllabe (Di Cristo et Jankowski, 1999).

3.3.2. Transcription des phénomènes liés au registre

Les différents phénomènes que nous venons de définir ne reçoivent pas un traitement équivalent dans les transcriptions de la prosodie.

Par exemple, seuls les modèles superpositionnels de l'intonation (Fujisaki et Nagashima, 1969) modélisent explicitement la déclinaison comme une dimension indépendante. Ce phénomène n'est pas annoté dans les autres approches en raison de son origine physiologique qui le rend non pertinent linguistiquement, justification que nous retenons comme recevable.

Au contraire, *downstep* et *upstep* sont présents dans les quatre systèmes de transcription que nous comparons ici. Que ce soit dans l'IPA, dans INTSINT, ou dans la plupart des travaux utilisant ToBI, ces phénomènes sont traités comme des événements locaux et relatifs, puisqu'un ton sera abaissé (pour *downstepped*) ou relevé (pour *upstepped*) par rapport au ton qui le précède immédiatement. On trouve toutefois chez Truckenbrodt (2002) ainsi que dans les travaux du groupe aixois ProDiGE (Di Cristo *et al.*, 2004) une extension de la notion de *downstep* à un niveau plus global puisque le *downstep* peut concerner la relation entre domaines adjacents et non plus seulement entre tons. La figure 8, empruntée à Truckenbrodt, illustre ce point :

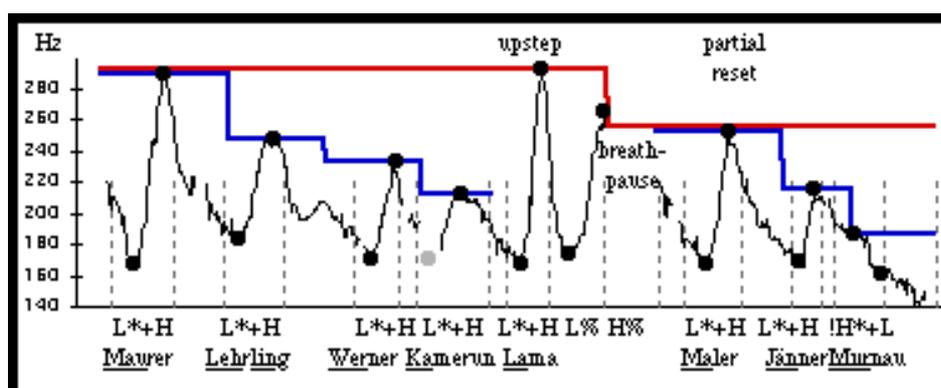


Figure 8

Application de la notion de downstep aux syntagmes intonatifs (intonational phrases=IP) dans le modèle de Truckenbrodt (2002).

Le reset partiel (partial reset) marque que la seconde IP est abaissée (downstepped) par rapport à la première (escalier supérieur).

Le downstep entre tons H adjacents est « enchâssé » (escaliers inférieurs) dans le précédent puisqu'il s'applique à l'intérieur de l'IP.

Dans la version actuelle d'IVTS, *downstep* et *upstep* sont encodés dans la tire phonétique globale, ce qui autorise théoriquement leur application à des domaines comme à des tons. Il reste toutefois à développer l'annotation de cette distinction. Dans un premier temps, celle-ci pourrait user d'un jeu de diacritiques associés aux symboles déjà disponibles : U (*upstep*), D (*downstep*) et R (*reset*). Par exemple, la comparaison des cibles codées H1 et R dans la figure 7 montre l'abaissement de la seconde par rapport à la première (comme la cible *partial reset* dans la figure 8 est abaissée par rapport à la première cible haute). Le *downstep* entre domaines (IP) qui est alors mis en évidence pourrait être annoté par un indice « d » ajouté à R : « R_d ».

À l'inverse du *downstep* et de l'*upstep* qui sont transcrits dans la plupart des modèles, le niveau et l'étendue du registre reçoivent un traitement beaucoup plus marginal. Ceci s'explique sans doute par le fait que les variations de ces deux dimensions sont en relation avec des aspects discursifs ou pragmatiques plutôt que syntaxiques. Notons cependant que :

- la version pour l'anglais de ToBI prévoit un diacritique « HiF0 » pour marquer la valeur maximum de F0 de chaque *intermediate phrase (ip)*, représentant le niveau de registre atteint par ce domaine ;
- dans le cadre du système de transcription INTSINT, le registre du locuteur joue un rôle plus important et se trouve représenté par les valeurs de F0 associées aux tons absolus sur l'empan d'une unité intonative (équivalent de l'*intonational phrase* dans ce modèle). Les ton T (*top*) et B (*bottom*) définissent respectivement les limites haute et basse de la gamme tonale ; le ton M (*mid*), intermédiaire entre les deux précédents, donne la « clé » du registre du locuteur, c'est-à-dire l'équivalent de son niveau moyen ;
- c'est dans les travaux du groupe aixois ProDiGE (Di Cristo *et al.*, 2004), que l'on trouve le système de transcription le plus développé pour rendre compte du niveau et de l'étendue du registre. ProDiGE propose une grille d'annotation plurilinéaire de la prosodie et du discours qui développe une annotation manuelle du niveau et de l'étendue du registre sur des bases perceptives. La grille comporte deux lignes réservées à l'annotation du registre. La première concerne le niveau du registre (ReLe pour *Register Level*) et permet d'affecter à chaque domaine (comportant un ou plusieurs syntagmes intonatifs) une des valeurs suivantes : « N » pour *normal*, « Rai » pour *raised* (élevé) et *Low* pour *lowered* (abaissé). La seconde ligne permet d'annoter les variations de la gamme tonale (ReSp pour *Register Span*) pour laquelle trois valeurs sont proposées : « N » pour *normal*, « Exp » pour *expanded* (étendue ou dilatée) et « Com » pour *compressed* (compressée). Ce système présente l'avantage de faire apparaître des variations pertinentes sur des empan discursifs et de les mettre en relation avec une analyse discursive proposée en regard. La figure 9 illustre l'application de ce système d'annotation à une courte séquence de parole spontanée.

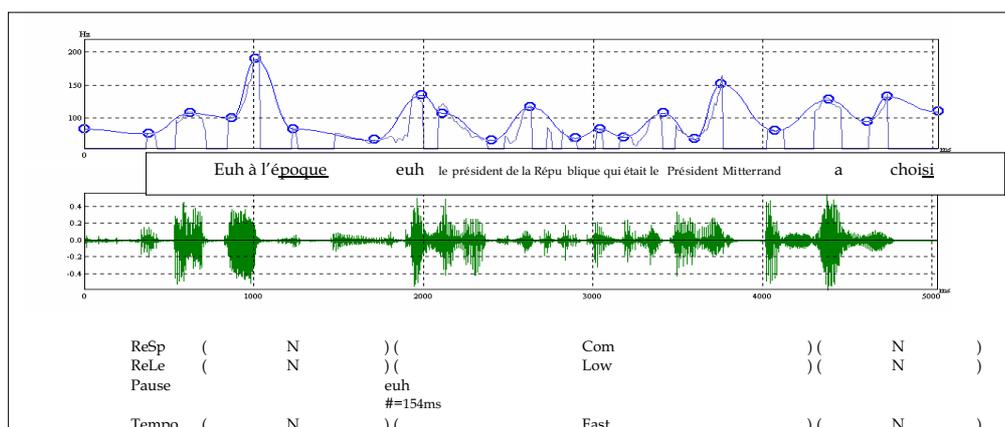


Figure 9

Exemple de séquence (ici un syntagme nominal sujet complexe) présentant un niveau de registre abaissé (ReLe=low) et une gamme tonale compressée (ReSp=com) entre deux séquences produites dans le registre moyen du locuteur.

Exemple emprunté aux travaux du groupe ProDiGE.

Le système IVTS propose quant à lui un symbole « S » s'inscrivant dans la tire « **Perception phonétique globale** » et désignant des variations de l'étendue du registre sans en préciser le sens (expansion ou compression). Il nous semble toutefois que la distinction proposée par ProDiGE entre le niveau et l'étendue du registre mérite d'être prise en compte. En effet, il n'est pas certain que ces deux dimensions co-varient de manière parallèle, en particulier dans la mise en forme prosodique de constituants parenthétiques ou incidents, ou de constituants relevant du fond, ou encore dans le discours rapporté. Le choix d'une mesure pour le niveau et pour l'étendue du registre ainsi que l'existence d'une annotation spécifique restent cependant à élaborer dans le cadre que nous proposons. En particulier, il faudrait prévoir une tire spécifique où des domaines plutôt que des points locaux pourraient être annotés comme présentant un registre abaissé ou relevé, une gamme tonale étendue ou compressée. La figure 10 illustre ce que pourrait être une telle annotation.

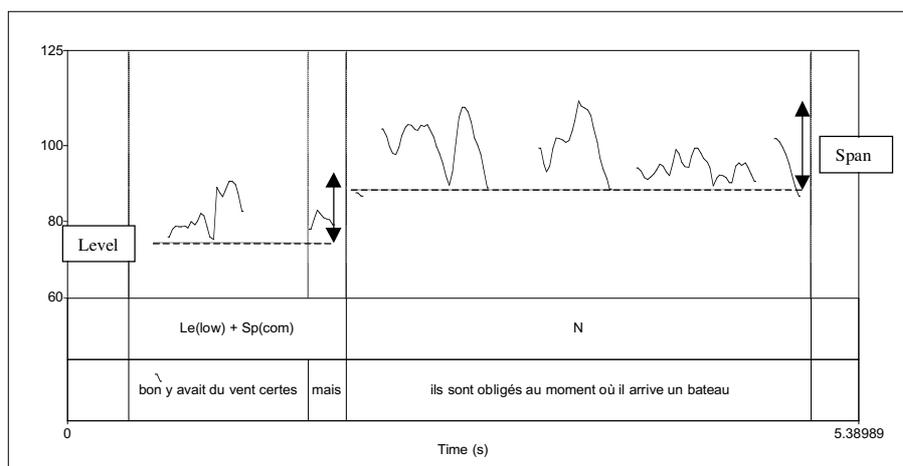


Figure 10

Proposition d'annotation des variations du niveau et de l'étendue du registre dans une nouvelle tire « registre » dans le système IVTS. Il faut noter que le connecteur discursif « mais » est produit avec un registre abaissé et compressé comme l'énoncé qui le précède.

4. Conclusion

Divers systèmes d'annotation prosodiques ont été développés ces quinze dernières années afin de répondre à une demande croissante en faveur d'outils consacrés à l'étude et au traitement de données de parole. Dans cette contribution, nous étudions dans une perspective typologique quatre systèmes d'annotation, qui sont tous capables de fournir des informations relatives aux faits prosodiques de façon quantifiable et discrète. En nous intéressant principalement à la transcription des phénomènes d'accentuation, de variation mélodique et de registre, nous avons identifié les points forts et les faiblesses de chacun des systèmes étudiés.

Le système de l'IPA (*Handbook of the International Phonetic Association*, 1999) est de loin le plus utilisé en typologie, et le plus connu. Il est parfaitement approprié pour transcrire, au niveau phonologique, les phénomènes métriques et tonals de nature lexicale, mais, de même que les faits tonals de frontière sont impossibles à encoder, il n'existe pas de symboles pour traiter de l'intonation.

De son côté, le système INTSINT (Hirst *et al.*, 2000) a deux avantages notables par rapport aux autres systèmes: i) une partie de l'étiquetage peut être faite automatiquement; et ii) il est possible de tester quels effets ont sur la perception des modifications de la représentation intonative, puisque les différentes représentations peuvent être réinterprétées et converties en signal acoustique.

Néanmoins, comme les représentations proposées reposent sur les seules informations acoustiques, l'étiquetage peut ne pas toujours refléter l'interprétation linguistique que ferait un locuteur natif en entendant le signal. L'un des à-côtés de cette base acoustique est que le système peut être utilisé pour n'importe quelle langue, peu importe que le transcrip-teur en connaisse le système prosodique, ou même, parle cette langue. Un autre problème avec ce système est que le statut des tons dans les représentations n'est pas spécifié, et, en conséquence, aucune distinction claire n'est proposée dans les transcriptions entre les configurations tonales phonologiquement distinctes (ou contrastives) et les tons qui ne participent pas à l'établissement de contrastes.

ToBI (Silverman *et al.* ; 1992, Beckman *et al.*, 2005) est sans aucun doute le système de transcription le plus utilisé pour l'intonation. Il a été développé par un groupe de chercheurs venant d'horizons et de disciplines variés. La tire *Break Indices* permet de fournir une transcription détaillée et fine des degrés de frontières prosodiques, en comparaison de ce qui est proposé dans les autres systèmes. Mais un désavantage de ce système est qu'il ne peut être utilisé que pour les langues ou les variétés pour lesquelles le fonctionnement du système prosodique est déjà connu et établi, puisque les phénomènes sont directement transcrits ou encodés au niveau phonémique. Cela peut constituer un sérieux inconvénient pour des études typologiques, dès lors que le linguiste se fixe comme objectifs i) de construire de toutes pièces un inventaire tonal, ii) d'expliquer comment des événements tonals forment les contours pertinents et comment ils sont alignés avec des frontières prosodiques ou des syllabes proéminentes, et pour finir iii) de formuler des hypothèses sur le fonctionnement d'une grammaire tonale.

IVTS résout ce problème en ayant recours à des niveaux de transcription phonétiques aussi bien que phonologiques. IVTS comme ToBI – et dans une certaine mesure INTSINT – souffrent cependant d'un manque de transparence en ce qui concerne l'utilisation correcte et l'interprétation des étiquettes (*cf.* Pirelli *et al.*, 2000), mais cela est sans conteste un trait inhérent à tout système de codage des variations tonales phonologiquement pertinentes. Aussi, même si nous pensons que IVTS peut devenir un standard pour transcrire la variation prosodique – au moins pour les langues à intonation – ce dernier n'est pas comparable au standard offert par l'IPA pour la transcription des segments. Les symboles segmentaux de l'IPA peuvent être utilisés pour proposer des transcriptions phonétiques étroites pour des langues qui n'ont pas encore été complètement décrites, même si, dans de tels cas, les transcrip-teurs peuvent parfois être amenés à proposer des transcriptions alternatives. Pour mener à bien une transcription prosodique en IVTS, le transcrip-teur doit toujours être en mesure non seulement d'identifier les syllabes proéminentes et les frontières des domaines prosodiques sur lesquels s'associent les mouvements mélodiques, mais également de définir les domaines d'implémentation qui sont nécessaires pour encoder les

mouvements mélodiques d'une langue (*cf.* Nolan et Grabe, 1997). Aussi, il est important de souligner que les annotations phonétiques en IVTS constituent des outils d'annotation, et que les représentations discrètes qui sont alors proposées ne reflètent pas directement la façon dont le locuteur perçoit effectivement les mouvements mélodiques.

Le recours à des niveaux de transcription supplémentaires (par l'ajout de tires d'annotation) est un des désavantages majeurs du système, puisque cela alourdit la tâche du transcripateur. Bien entendu, le type d'informations encodées et le degré de détails retenus dans une transcription sont, pour une large part, dictés par les objectifs visés. Aussi n'est-il pas toujours nécessaire d'encoder les informations à chaque niveau. Après avoir ajouté la tire « **Perception phonétique locale** » dans IViE, nous avons créé dans IVTS la tire « **Perception phonétique globale** » afin de rendre le système plus complet pour répondre aux objectifs que nous nous fixons. La liste des ajouts n'est cependant pas close : dans une perspective conversationnaliste, il peut être judicieux d'encoder d'autres types de phénomènes prosodiques comme les variations d'intensité ou les modifications de timbre, en ajoutant des tires spécifiques.

Comme les trois systèmes autosegmentaux d'annotation prosodique qui ont été discutés ne s'excluent pas mutuellement et ne sont pas incompatibles les uns par rapport aux autres, il est possible, selon nous, que chaque utilisateur emprunte à chacun des systèmes les traits qui lui paraissent les plus pertinents en fonction de ses objectifs et de ses besoins. Dans une étude dont le but est d'évaluer les effets des degrés de frontières prosodiques sur la réalisation des mouvements intonatifs, il peut être appréciable d'introduire un niveau d'annotation comparable à la tire *Break Indices* utilisée par TOBI. De même, pour accélérer les processus d'annotation, nous espérons à terme pouvoir utiliser MOMEL (ou le Prosogramme de Mertens, 2004) pour générer automatiquement une première transcription phonétique acoustique qui pourra être évaluée sur des bases perceptives et reprise manuellement pour développer la transcription en IVTS. De même, en utilisant un algorithme de conversion symbolique comme celui utilisé par INTSINT, il serait possible de générer de façon semi-automatique une première annotation phonologique, à partir du moment où, pour la variété analysée, la grammaire tonale et les modalités d'interface phonétique/ phonologie seraient connues. Ces différentes évolutions demandent toutefois des recherches plus approfondies.

5. Bibliographie

ARVANITI, A. & M. BALTAZANI. (2005). Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora, in S.-A. Jun (ed.), *Prosodic Typology : The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, p. 84-117.

- AUER, P & P. GILLES. (2003). *Prosodic Variation in German between Areality and Pragmatics: an Overview*, Unpublished manuscript.
- BECKMAN, M. E. ; J. HIRSCHBERG & S. SHATTUCK-HUFNAGEL. (2005). The original ToBI system and the evolution of the ToBI framework, in S.-A. Jun (ed.), *Prosodic Typology : The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, p. 9-54.
- BOERSMA, P. & D. WEENINK. (2005). *Praat: doing phonetics by computer* (version 4.3.23). Retrieved September 24, 2005 from <http://www.praat.org/>.
- BRUCE, G. (1977). *Swedish Word Accents in Sentence Perspective*, Lund: Gleerup.
- CAMPIONE, E., D. HIRST & J. VERONIS, (2000). Automatic stylisation and symbolic coding of F0: implementations of the INTSINT model, in A. Botinis (ed.), *Intonation. Research and Applications*, Dordrecht: Kluwer.
- CRYSTAL, D. (1969). *Prosodic systems and intonation in English*, Cambridge: Cambridge University Press.
- DI CRISTO, A. (1998). Intonation in French, in Hirst, D. & Di Cristo, A. (eds), *Intonation Systems: a survey of twenty languages*, Cambridge : Cambridge University Press.
- DI CRISTO, A. & HIRST, D. (1996). Vers une typologie des unités intonatives du français, *16es Journées d'Étude sur la parole*, Société française d'acoustique (ed.), Avignon, p. 219-222.
- DI CRISTO, A. & JANKOWSKI, L. (1999). Prosodic organisation and phrasing after focus in French, *Proceedings of XIVth ICPbS*, San Francisco: USA. 1565-1568.
- DI CRISTO, A. ; AURAN, C. ; BERTRAND, R. ; CHANET, C. ; PORTES, C. (2004). Outils prosodiques et analyse du discours, *Cahiers de l'Institut de Linguistique de Louvain*, vol. 28. p. 27-84.
- D'IMPERIO, M. ; BERTRAND, R. ; DI CRISTO, A. ; PORTES, C. (2006). The phonology and phonetics of prenuclear and nuclear accents in French, *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics* (août 28-30 : Athens, Greece), p. 121-124.
- DURAND, J. & C. LYCHE. (2003). Le projet « Phonologie du Français Contemporain » (PFC) et sa méthodologie, in E. Delais-Roussarie & J. Durand (eds), *Corpus et variation en phonologie du français : méthodes et analyses*, Toulouse : Presses Universitaires du Mirail.
- FUJISAKI, H. & S. NAGASHIMA. (1969). A model for synthesis of pitch contours of connected speech, *Annual Report Engineering Research Institute*, University of Tokyo, 28, p. 53-60.
- GRABE, E. (1998). *Comparative intonational phonology: English and German*. Doctoral dissertation, Max Planck Institut für Psycholinguistik, Nijmegen, and University of Nijmegen.
- GRABE, E.; B. POST & F. NOLAN. (2001). Modelling intonational variation in English: The IViE system, in S. Puppel, G. Demenko (eds), *Prosody 2000*, Poznan: Adam Mickiewicz University, p. 51-58.
- GRABE, E. & B. POST. (2004). Intonational variation in the British Isles, in G. Sampson & D. McCarthy (eds), *Corpus Linguistics: Readings in a Widening Discipline*, London and New York: Continuum International, p. 474-481.

- GUSSENHOVEN, C. (1984). *On the grammar and semantics of sentence accents*, Dordrecht: Foris.
- GUSSENHOVEN, C. (2004). *The Phonology of Tone and Intonation*, Cambridge: Cambridge University Press.
- Handbook of the International Phonetic Association*. (1999), Cambridge: Cambridge University Press.
- HAYES, B. (1995). *Metrical Stress Theory: Principles and Case Studies*, Chicago: University of Chicago Press.
- HIRST, D.J. (2005). Forms and function in the representation of Speech Prosody, *Speech Communication*, 46, p. 334-347.
- HIRST, D. & C. AURAN. (2005). Analysis by synthesis of speech prosody: the ProZed environment, *Proceedings of European Conference on Speech Communication and Technology*, Eurospeech, Lisbonne, Portugal, sept.
- HIRST, D.J. & ESPESER, R., (1993). Automatic modelling of fundamental frequency curves using a quadratic spline function, *Travaux de l'Institut de Phonétique d'Aix*, 15, p. 71-85.
- HIRST, D.J. & DI CRISTO A. (eds) (1998). *Intonation Systems: a survey of twenty languages*, Cambridge: Cambridge University Press.
- HIRST, D.J.; DI CRISTO, A. & ESPESER, R. (2000). Levels of representation and levels of analysis for intonation, in M. Horne (ed), *Prosody: Theory and Experiment*, Dordrecht: Kluwer Academic Publishers, p. 51-87.
- JUN, S.-A. (ed.), (2005a). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- JUN, S.-A. (2005b). Korean Intonational Phonology and Prosodic Transcription, in S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, p. 201-229.
- JUN, S.-A. & FOUGERON, C. (2002). Realizations of accentual phrase in French intonation, *Probus*, 14, p. 147-172.
- LADD, D. R. (1996). *Intonational Phonology*, Cambridge: Cambridge University Press.
- LIBERMAN M. & PIERREHUMBERT J. (1984). Intonational invariance under changes in pitch range and length, in Aronoff M. & Oehrle (eds), *Language sound structure*, Cambridge: Mass., MIT Press, p. 157-233.
- MERTENS, P. (2004). Le prosogramme : une transcription semi-automatique de la prosodie, *Cahiers de l'Institut de Linguistique de Louvain*, 30, 1-3, p. 7-25.
- NOLAN, F. & GRABE, E., (1997). Can TOBI transcribe intonational variation in British English? in *Proceedings of the ESCA Tutorial and Research workshop on Intonation: Theory, Models and Applications*. Athens, Sept.
- PATTERSON, D. (2000). *A linguistic approach of pitch range modeling*, PhD Thesis, University of Edinburgh.

- PENG, S., CHAN, M., TSENG, C., HUANG, T., LEE, O. J., & BECKMAN, M. (2005). Towards a Pan-Mandarin System for Prosodic Transcription, in S.-A. Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, p. 230-270.
- PIERREHUMBERT, J. (1980). *The phonology and phonetics of English Intonation*, Ph.D. thesis M.I.T.
- PIRELLI, V. ; SORIA, C. ; HEID, U. (2000). *MATE deliverable D4.1*: Validation of coding schemes and coding workbench. where retrieved from? <http://mate.nis.sdu.dk/about/deliverables.html>.
- POST, B. (2000). *Tonal and phrasal structures in French intonation*, The Hague: Holland, Academic graphics.
- POST, B. ; DELAIS-ROUSSARIE, E. ; SIMON, A-C. (2006). Développer un système de transcription des phénomènes prosodiques, in G. Caelen-Haumont ; A-C. Simon (eds), *Bulletin PFC 4*.
- POST, B. ; DELAIS-ROUSSARIE, E. (2006). Transcribing intonational variation at different levels of analysis, in R. Hoffmann et H. Mixdorff (eds), *Speech Prosody*, Dresden.
- PRIETO, P. ; SHIH, C. ; NIBERT, H. (1996). Pitch downtrends in Spanish, *Journal of Phonetics*, 24, p. 445-473.
- SILVERMAN, K. ; BECKMAN, M. ; PITRELLI, J. ; OSTENDORF, M. ; WIGHTMAN, C. ; PRICE, P. ; PIERREHUMBERT, J. ; HIRSCHBERG, J. (1992). TOBI: A standard for labeling English prosody, *Proceedings International Conference of Spoken Language Processing*, 2, p. 867-870.
- ’T HART, J. ; COLLIER, R. ; COHEN, A. (1990). *A perceptual study of intonation*, Cambridge: Cambridge University Press.
- TROUBETZKOY, N. (1939). Grundzüge der Phonologie, *Travaux du Cercle Linguistique de Prague*, VII, Reprint Goettingen: Vandenhoeck & Ruprecht.
- TRUCKENBRODT, H. (2002). Upstep and embedded register levels. *Phonology*, 19, Cambridge University Press, p. 77-120.