Etiquetage prosodique semi-automatique des corpus oraux

Estelle Campione et Jean Véronis

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

Résumé – Abstract

La transcription manuelle de la prosodie est une tâche extrêmement coûteuse en temps, qui requiert des annotateurs très spécialisés, et qui est sujette à de multiples erreurs et une grande part de subjectivité. Une automatisation complète n'est pas envisageable dans l'état actuel de la technologie, mais nous présentons dans cette communication des outils et une méthodologie qui permettent une réduction substantielle du temps d'intervention manuelle, et améliorent l'objectivité et la cohérence du résultat. De plus, les étapes manuelles nécessaires ne demandent pas une expertise phonétique poussée et peuvent être menées à bien par des étudiants et des « linguistes de corpus ».

The manual transcription of prosody is an extremely time-consuming activity, which requires highly specialised experts, and is prone to errors and subjectivity. Full automation is not achievable in the current state of the technology, but we present in this paper a technique that automates critical steps in the process, which results in a substantial annotation time reduction, and improves the objectivity and coherence of the annotation. In addition, the necessary human phases do not require a highly specific training in phonetics, and can be achieved by syntax students and corpus workers.

Mots-clefs – Keywords

corpus, prosodie, étiquetage / corpus, prosody, tagging

1 Introduction

La transcription de l'oral spontané[1] pose de multiples problèmes. De nombreux linguistes (par exemple Blanche-Benveniste & Jeanjean, 1987 ou Leech, McEnery & Wynne, 1997) ont fait

[1] Nous utilisons ce terme par commodité, pour l'opposer à la « parole de laboratoire », généralement lue ou associée à des tâches délimitées. Il y aurait cependant beaucoup à dire sur la notion de spontanéité à l'oral (voir par exemple Blanche-Benveniste & Bilger, 1999).

remarquer que le système de ponctuation de l'écrit était loin d'être satisfaisant pour transcrire l'oral, bien que ce système ait été utilisé pour transcrire de nombreux corpus (par exemple le *British National Corpus*). Leech, McEnery & Wynne (1997: 90) qualifient la transcription de l'oral avec les règles de l'écrit habituel de « pseudo-procédure, dont la seule excuse est le coût prohibitif qu'il y aurait à tenter quoi que ce soit d'autre ». Taylor (1996) étudie la transcription orthographique ponctuée du Lancaster/IBM Spoken English Corpus (SEC), et montre que 47,2% des frontières prosodiques ne correspondent pas à une ponctuation, tandis que 17,1% des ponctuations ne correspondent pas à une frontière prosodique. De plus, quand une frontière est réellement marquée, la correspondance entre son type et la ponctuation utilisée est loin d'être bi-univoque. Blanche-Benveniste & Jeanjean (1987) insistent également sur le fait que la ponctuation écrite impose au transcripteur des décisions de regroupement ou de coupure pas toujours fondées et préjudiciables à l'étude dans la mesure où elles plaquent sur l'oral le modèle canonique de la phrase écrite « suggérant [ainsi] une analyse avant de l'avoir faite » (Blanche-Benveniste & Jeanjean, 1987 : 142).

A cause de ce constat, diverses équipes, dont la nôtre, ont développé des conventions de transcription qui n'utilisent pas les ponctuations de l'écrit. Pour des raisons pratiques, ces conventions marquent souvent un sous-ensemble très limité de phénomènes prosodiques. Notre équipe a pris, par exemple, une position minimaliste, puisque seules les pauses sont marquées (Blanche-Benveniste & Jeanjean, 1987; Blanche-Benveniste, 1990). Toutefois, ce type de transcription n'est pas totalement satisfaisant non plus. D'une part, de nombreux phénomènes portant par exemple sur la segmentation de l'oral (particulièrement importante pour le traitement de la parole), ou sur la relation entre syntaxe et prosodie, ne peuvent pas être étudiés, ce qui est regrettable pour des corpus oraux. D'autre part, les transcriptions résultantes comportent des ambiguïtés relativement gênantes. Ainsi, en français, de nombreux marqueurs de discours ou éléments phatiques peuvent aussi appartenir à une autre catégorie et remplir une tout autre fonction (pronoms, coordinations, adverbes, etc.). Dans le fragment ci-dessous, par exemple, il est difficile de déterminer dans la transcription non ponctuée si *quoi* est un élément phatique, ou s'il est pronom, complément du verbe *savoir* :

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

Une autre ambiguïté courante apparaît sous forme de « segments flottants » (voir Bilger *et al.,* 1997), généralement des compléments, qui peuvent être rattachés soit à ce qui précède, soit à ce qui suit :

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

Bien sûr, dans un certain nombre de cas, l'habitude et l'expérience des utilisateurs peut leur permettre de lever l'ambiguïté en fonction du contexte (Blanche-Benveniste & Jeanjean, 1987 : 140). Deux problèmes demeurent cependant. D'une part, des utilisateurs moins expérimentés (étudiants-linguistes, apprenants du français, etc.) risquent de rencontrer de nombreuses difficultés et de faire de mauvaises interprétations[2]. D'autre part, la

[2] Nous avons constaté que même des linguistes habitués à la manipulation du corpus font des erreurs d'interprétation, souvent sans même apercevoir l'ambiguïté possible.

désambiguïsation (quand elle est possible) repose sur des indices faisant intervenir le sens, une connaissance du monde et des contraintes pragmatiques qui sont totalement hors de portée des systèmes qui exploiteraient les corpus automatiquement.

Il est donc tout à fait souhaitable que la prosodie puisse être notée dans les corpus oraux. Nous ne prétendons pas, bien sûr, qu'elle résolve toutes les ambiguïtés, pas plus que la ponctuation ne résout toutes celles de l'écrit, mais son rôle de « ponctuation de l'oral » ne peut plus guère être ignoré si l'on souhaite étudier sérieusement la langue spontanée, notamment en vue du traitement par des machines.

A l'heure actuelle, seuls quelques corpus de taille significative sont disponibles pour l'anglais (tels que London-Lund Corpus [Svartvik *et al.,* 1982] ou le SEC [Knowles, Wichmann & Alderson, 1996]). Pour le français, des corpus spontanés assez brefs ont été transcrits par quelques équipes, mais ils sont peu ou pas disponibles pour le reste de la communauté, et font l'objet de systèmes de transcription disparates. La raison de cette carence provient de la difficulté, maintes fois soulignée, de la transcription prosodique, qui demande un temps considérable, et fait appel à une compétence phonétique très spécialisée peu courante parmi les « linguistes de corpus ». De plus, la transcription prosodique est d'une nature éminemment subjective, qui réduit la fiabilité des données résultantes et impose des relectures par des annotateurs multiples accroissant encore le coût global de la tâche : d'après l'étude de Pickering, Williams & Knowles (1996) sur le SEC, les annotateurs sont en désaccord sur la présence de frontières prosodiques dans 27% des cas et leur taux d'accord sur les étiquettes de tons et d'accents est seulement de 55% meilleur que le hasard[3].

Automatiser l'étiquetage prosodique des corpus serait donc du plus grand intérêt, à la fois en termes de coûts et d'objectivité de l'annotation. Bien sûr, une automatisation complète n'est pas envisageable dans l'état actuel de la technologie, mais nous présentons dans cette communication des outils et une méthodologie qui permettent une réduction substantielle du temps d'intervention manuelle, et améliorent l'objectivité et la cohérence du résultat. De plus, les étapes manuelles nécessaires ne demandent pas une expertise phonétique poussée et peuvent être menées à bien par des étudiants et des « linguistes de corpus ».

Nous visons une transcription prosodique « large », qui délimite seulement les unités prosodiques et marque les mouvements mélodiques majeurs. Un étiquetage plus fin (par exemple un symbole prosodique ou plus par mot) est sans aucun doute intéressant pour les études phonétiques, mais il serait peu lisible et peu utilisable pour les utilisations des corpus que nous envisageons (études syntaxiques ou pragmatiques). Notre technique se décompose en plusieurs étapes, que nous décrivons dans les sections suivantes : (1) stylisation de la courbe de fréquence fondamentale (F_0) ; (2) discrétisation des mouvements mélodiques ; (3) segmentation et transcription ; (4) codage prosodique. Nous appliquons notre technique au français, mais plusieurs des modules sont indépendants de la langue et ont été testés sur l'italien, l'espagnol, l'allemand et l'anglais.

---

[3] Calcul que nous avons effectué à partir des données brutes publiées par les auteurs, à l'aide du coefficient kappa (Cohen, 1960).

2 Stylisation de la courbe de F_0

La courbe de F_0 est la combinaison d'une composante macroprosodique qui dépend de la structure lexicale, syntaxique et pragmatique de l'énoncé, et d'une composante microprosodique qui dépend seulement de la séquence particulière de phonèmes (abaissement dû aux constrictives, etc.). Styliser la courbe de F_0 consiste à factoriser ces deux composantes, et à extraire la seule composante macroprosodique. L'algorithme utilisé dans cette étude, MOMEL (MOdélisation MELodique), dont on pourra trouver les détails techniques dans Campione, Hirst & Véronis (2000), réduit le contour intonatif à une série de points cibles, qui représentent les mouvements mélodiques pertinents (figure 1). Cet algorithme est indépendant de la langue, et ne nécessite aucune pré-segmentation du signal (comme, par exemple, D'Alessandro & Mertens, 1995, au niveau syllabique). Une fois interpolée par une courbe spline quadratique, la séquence de point cibles permet une régénération d'un contour de F_0 qui n'est pas perceptuellement distinguable de l'original, mis à part quelques erreurs de détection qui doivent être corrigées manuellement. Une évaluation quantitative sur un corpus important en cinq langues (Campione, 2001) a montré que l'algorithme produisait environ 5% d'erreurs, qui pour la plus grande partie consistaient en erreurs systématiques pouvant être regroupées en deux ou trois grands types, en particulier des points cibles manquants lors des transitions avec des pauses, ce qui suggère qu'une amélioration notable est possible dans les versions futures.

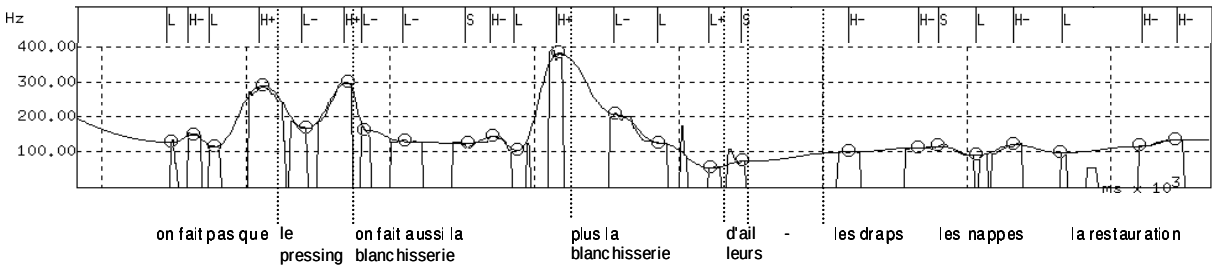

Figure 1. Stylisation et discretisation de la courbe de F_0

L'algorithme de stylisation est indépendant de la langue et ne demande ni pré-segmentation du signal (par exemple en syllabes), ni corpus d'entraînement. De plus, la correction des points cibles ne demande pas une expertise phonétique poussée. Le signal de parole peut être écouté fragment par fragment, et l'original peut être comparée à une version re-synthétisée à partir de la courbe stylisée. Si les deux diffèrent du point de vue perceptif, les points cibles peuvent être déplacés à l'aide d'une interface graphique jusqu'à ce que la re-synthèse soit jugée identique à l'original.

3 Discrétisation des mouvements mélodiques

La deuxième étape consiste en une conversion des points cibles en un alphabet discret de symboles représentant les mouvements mélodiques. Nous avons développé un modèle mathématique qui permet une réduction de la courbe initiale à un alphabet de sept symboles sans réduction significative de l'information (figure 2). Ce modèle est basé sur l'observation que la distribution des points cibles est d'allure approximativement normale (Campione & Véronis, 1998). On pourra trouver une description détaillée dans Véronis & Campione (1998).

		Type de mouvement		
		montant	*nul*	*descendant*
Amplitude	*Forte*	**L+**		**H+**
	Moyenne	**L**	**S**	**H**
	Faible	**L-**		**H -**

Figure 2. Alphabet des mouvements mélodiques[4]

Ces symboles n'ont aucune visée phonologique, et consistent seulement dans une représentation extrêmement compacte de la courbe de F_0. Nous avons montré, à l'aide d'une évaluation sur un corpus étendu (4h20 de parole, 50 locuteurs, 5 langues), que le codage obtenu permettait une re-génération de 99% des points cibles à moins de 2 demi-tons des points originaux (Véronis & Campione, 1998). Les courbes de F_0 résultantes sont virtuellement impossibles à distinguer des originaux, ou en tous cas, lorsqu'une légère différence de hauteur est perceptible, elle n'influe jamais sur l'interprétation linguistique (syntaxique ou pragmatique) de l'énoncé.

Le modèle a des propriétés intéressantes. En particulier, des mouvements de même amplitude (en demi-tons) n'ont pas nécessairement le même codage, selon la place à laquelle ils se trouvent dans le registre du locuteur, ce qui est cohérent avec le fait que les variations vers les extrêmes du registre demandent plus d'effort articulatoire que les variations dans la zone médiane. Le modèle prédit aussi de façon correcte l'effet de déclinaison observé dans la parole, sans pour autant nécessiter un paramètre de déclinaison spécifique.

4 Segmentation et transcription

Le signal est tout d'abord automatiquement segmenté aux pauses (silencieuses). Malgré sa simplicité apparente, cette tâche est loin d'être facile à réaliser par des moyens automatiques : des pauses longues peuvent être interrompues par des bruits ambiants (ce qui est fréquent hors conditions de laboratoire), tandis que les pauses courtes sont difficiles à distinguer des plosives (par exemple, nous avons repéré des instances de pauses respiratoires pouvant être aussi brèves que 60 ms, ce qui est plus court que la plupart des plosives). Nous avons utilisé dans cette étude un détecteur de pauses basé sur la détection de F_0, qui s'avère raisonnablement robuste par rapport au bruit ambiant. En utilisant un seuil de 350 ms, très peu de fausses détections se produisent. Cependant, quelques pauses très brèves doivent être ajoutées manuellement. Cette correction n'est pas très difficile, grâce à l'utilisation d'un éditeur graphique qui permet la visualisation du signal et une réécoute segment par segment.

Les segments obtenus sont ensuite redivisés automatiquement lorsqu'ils contiennent des mouvements mélodiques majeurs (codés **H+** ou **L+**), en position autre que finale. Dans l'exemple de la figure 1, par exemple, le segment « *on fait pas que le pressing* » est coupé en deux parties (« *on fait pas que* » **H+** « *le pressing* »)[5].

---

[4] Signification des symboles : H(igher), L(ower), S(ame).

[5] Cet exemple, qui contient quatre points de segmentation autre que les pauses, a été choisi parce qu'il concentre plusieurs phénomènes intéressants en un temps très court, mais il ne représente pas la norme.

Le corpus est alors orthographiquement transcrit, à l'aide d'un éditeur permettant une réécoute segment par segment. Pendant cette phase, deux phénomènes prosodiques supplémentaires sont encodés manuellement, mais seulement s'ils apparaissent à la fin d'un segment[6] : les accents (codés *) et les allongements sur la syllabe finale (codés :). Une autre information importante résulte de la transcription elle-même, et est constituée par la présence de pauses remplies (hésitations) qui sont notées par des items lexicaux spéciaux (*euh*). Ces différents indices sont nécessaires à l'interprétation des pauses et des mouvements mélodiques dans l'étape suivante. La transcription de l'exemple de la figure 1 donne :

on fait pas que (**H+**)*
*le pressing (**H+**)*
*on fait aussi la blanchisserie (**H+**)*
*plus la blanchisserie (**L+**)*
*d'ailleurs (**pause**)*
*les draps les nappes la restauration (**pause**)*

5 Codage prosodique

Comme il a été mentionné en introduction, nous ne cherchons pas à produire un étiquetage « étroit », qui noterait toutes les informations prosodiques des énoncés, mais seulement un étiquetage « large », consistant en une segmentation et un marquage des mouvements mélodiques majeurs. Les mouvements mélodiques ayant déjà été repérés et codés, le problème principal est un problème de segmentation : il s'agit de décider quels indices prosodiques sont *démarcatifs* et lesquels sont *cohésifs*.

Nous considérons en effet (comme de nombreux auteurs) que les énoncés sont constitués d'*unités prosodiques* successives, c'est-à-dire de segments dont la prosodie ne dépend pas du segment précédent ou suivant. Ces unités sont aussi des unités de communication, dans la mesure où elles déterminent des points d'articulation marqués par des signaux utilisables dans la gestion des tours de parole entre locuteurs, dont les plus évidents sont un *signal de continuation* (le locuteur a encore quelque chose à dire et ne souhaite probablement pas être interrompu) et un *signal de conclusion* (le locuteur a terminé provisoirement et marque un point où le locuteur peut intervenir).

Ainsi, dans le corpus dont l'exemple de la figure 1 est extrait, la locutrice enchaîne une série d'unités continuatives clairement marquées par une intonation finale montante (notée ↗ : (« *ben je travaille dans un pressing* », etc.), jusqu'à unité conclusive marquée par une intonation finale fortement descendante (notée ↘ : « *comme partout* ») – où l'interlocutrice ne prend d'ailleurs pas le tour de parole – puis réenchaîne une série d'unités continuatives (figure 3). Chacune de ces unités est autonome du point de vue prosodique, en ce sens que la prosodie des segments voisins peut être changée sans que l'unité considérée soit affectée du point de vue syntaxique ou communicatif. Il est facile de vérifier par resynthèse que la modification de l'intonation finale d'une séquence quelconque de la figure 3 n'affecte pas l'interprétation du reste de l'énoncé.

---

[6] pour des raisons de gain de temps et de fiabilité. Rien n'empêcherait d'en marquer la totalité. Les accents et allongements internes aux segments ne sont pas, toutefois, pertinents pour la segmentation en unités.

A l'intérieur des unités prosodiques, les mouvements mélodiques importants délimitent des segments qui ne sont pas autonomes. Au contraire, ces mouvements mélodiques ont une valeur cohésive qui lie les différents segments entre eux du point de vue syntaxique, macro-syntaxique ou accentuel. Par exemple, dans l'unité prosodique correspondant à la figure 1, le mouvement mélodique ascendant fort (**H+**) sur *pressing* indique une construction macrosyntaxique binaire classique, de type contrastif (cf. Blanche-Benveniste, 1990 : 124-125), ce qui est renforcé par l'accent sur *que* :

on fait pas que le pressing ↗ on fait aussi la blanchisserie*

De même, dans la suite de l'énoncé (« plus la blanchisserie d'ailleurs »), un patron mélodique caractéristique, constitué d'un segment entièrement descendant (*downstep,* noté {...}↘) suivi d'une intonation plate indique une construction macro-syntaxique de type noyau+affixe[7] (Blanche-Benveniste, 1990 : 126sqq)[8] :

{plus la blanchisserie} ↘ *d'ailleurs* →

De nombreux auteurs ont montré que la combinaison des différents indices prosodiques (pauses, allongement syllabique, *euh*) changeait leur valeur perceptive et communicative. L'interaction de ces indices est complexe. En particulier, la présence d'une marque d'hésitation (allongement syllabique ou *euh*) change radicalement la valeur des pauses (silencieuses) et des mouvements mélodiques. Ainsi, les pauses sont un indice démarcatif majeur des unités mélodiques, mais elles n'ont cette valeur que si elles ne sont pas précédées d'une marque d'hésitation, et si elles ont une valeur relativement brève. Candéa (2000) montre ainsi qu'à partir d'un certain seuil (qu'elle situe à environ 1,8 ou 2 s), une rupture linguistique se produit même en cas d'hésitation (obligeant par exemple à une reprise du segment en cours). De même, un mouvement descendant est perçu comme frontière conclusive, sauf s'il est précédé d'une marque d'hésitation. Par contre, les mouvements ascendants suivis par une pause sont toujours perçus comme continuatifs, même s'ils sont précédés d'une marque hésitation.

L'algorithme de codage final prend en compte ces différentes contraintes, et produit un étiquetage du type de celui de la figure 3. Pour des raisons de lisibilité, certaines redondances sont supprimées : la combinaison * ↗ est simplement notée * (l'accent se réalisant toujours avec un mouvement ascendant), tandis que l'intonation plate ou légèrement descendante portant sur les marques d'hésitations n'est pas transcrite (c'est en effet le contour par défaut de ces marques). Les pauses courtes (≥ 350 ms et < 1.5 s) sont codées – et les pauses longues (≥ 1.5 s) sont codées –□–. Les (rares) pauses ultra-brèves (< 350ms) sont codées ^ (elles ne sont jamais considérées comme démarcatives). Les unités prosodiques sont marquées par des

---

[7] encore appelée *rhème+postrhème* (Morel & Danon-Boileau, 1998), *body+tag* (Biber *et al.,* 1999), etc.

[8] Il est intéressant de noter le rôle désambiguïsateur de la prosodie dans ce cas. En l'absence d'indication prosodique, le fragment « *plus la blanchisserie d'ailleurs les draps les nappes la restauration* » peut s'interpréter de trois façons, *d'ailleurs* pouvant être considéré soit comme un complément locatif (= *la blanchisserie d'un autre endroit*), ou, comme un marqueur de discours, mais dans ce cas, celui-ci peut s'attacher soit à ce qui précède, soit à ce qui suit.

changements de paragraphes, et un premier champ indique le point de départ de chaque unité en secondes par rapport au début de l'enregistrement.

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

Figure 3. Exemple d'étiquetage prosodique semi-automatique

6 Conclusion et perspectives

La technique présentée dans cette communication a été testée sur dix enregistrements, représentant environ une heure de français oral spontané (cinq hommes et cinq femmes). La notation résultante apporte des informations extrêmement précieuses pour l'exploitation des corpus, et notamment l'étude des phénomènes syntaxiques et macro-syntaxiques. Ce premier corpus étiqueté nous permet de commencer un réglage plus fin des paramètres, en particulier des seuils de pauses (dont il faut examiner la relation avec le débit du locuteur, etc.). On pourra également étudier comment les paramètres doivent être changés sur l'axe temporel pour un même corpus : on sait en effet que l'oral spontané se caractérise par des « basculements » (*switches*) entre des passages de débit et de variation mélodique différents. Les performances des modules existants sont tout à fait honnêtes, comme nous l'avons mentionné, mais une direction de développement futur consiste bien évidemment à chercher une amélioration, en particulier sur le module de stylisation et le détecteur de pauses, de façon à minimiser le travail de correction manuelle. D'autres modules peuvent également être

ajoutés : nous avons commencé à tester un module de détection automatique des marques d'hésitation, qui, bien que développé pour le japonais (Goto, Itou & Hayamizu, 1999), semble fournir des résultats prometteurs pour le français.

Dans l'état actuel des choses, l'étiquetage prosodique demande environ quatre heures pour 15 minutes d'enregistrement. Toutefois, la manipulation est alourdie par le fait que nous utilisons plusieurs éléments logiciels séparés, obligeant à trois passes complètes sur le corpus (correction de la stylisation, correction des pauses, transcription). Nous estimons que ce temps pourrait être réduit à trois heures si les différents modules étaient intégrés dans un environnement audio-graphique unifié permettant la correction et la transcription en une seule passe. Ce temps est équivalent à celui que mettent à l'heure actuelle les membres de notre équipe pour transcrire les corpus sans indication prosodique avec un simple magnétophone. En effet, le temps de correction est compensé par le confort et la fiabilité qu'amènent la segmentation en petites unités et la transcription à l'aide de l'environnement audio-graphique. Il semble donc que la semi-automatisation que nous proposons puisse, dans un futur relativement proche, être utilisée de façon opérationnelle pour la construction de grands corpus oraux annotés pour la prosodie.

Remerciements

Nous avons utilisé dans cette étude le logiciel *Transcriber* développé par Claude Barras (DGA) et les outils *Signaix* développés par Robert Espesser (CNRS). Nous sommes particulièrement reconnaissants à Robert Espesser pour son aide technique tout au long de ce projet. Nous remercions également Masataka Goto d'avoir mis à notre disposition son détecteur de pauses remplies (marques d'hésitation). Nous remercions enfin Claire Blanche-Benveniste, Albert Di Cristo, José Deulofeu, Daniel Hirst et Frédéric Sabio pour leurs conseils et commentaires.

Références

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* Harlow: Pearson Education Limited.

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

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

Blanche-Benveniste C., Bilger, M. (1999). « Français parlé – oral spontané ». Quelques réflexions. *Revue Française de Linguistique Appliquée, IV*(2), 21-30.

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

Campione, E., & Véronis, J. (2001). Une évaluation de l'algorithme de stylisation mélodique MOMEL. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence, 19*, sous presse.

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

Campione, E., Hirst, D., Véronis, J. (2000). Stylisation and symbolic coding of F0 : comparison of five models. In A. Botinis (Ed.), *Intonation: Models and Theories* (pp. 185-208). Dordrecht: Kluwer Academic Publishers.

Candéa, M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané. Etude sur un récit en classe de français.* Thèse de doctorat. Université Paris III.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20, 37-46.

D'Alessandro, C., Mertens, P. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language*, 9, 257-288.

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

Knowles G., Wichmann A., Alderson P. (Eds.) (1996). *Working with speech.* Harlow : Addison Wesley Longman Limited.

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

Morel M.-A., Danon-Boileau L. (1998). *Grammaire de l'intonation. L'exemple du français.* Gap : Ophrys.

Pickering B., Williams B., Knowles G. (1996). Analysis of transcriber differences in the SEC. In Knowles G., Wichmann A., Alderson P. (Eds.) *Working with speech.* Harlow : Addison Wesley Longman Limited.

Svartvik J. (Ed.) (1982). *The London Corpus of Spoken English: Description and Research.* Lund, Lund University Press.

Taylor L. (1996). The correlation between punctuation and tone group boundaries. In Knowles G., Wichmann A., Alderson P. (Eds.) *Working with speech.* Harlow : Addison Wesley Longman Limited.

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