

Chunking syntaxique, structure prosodique et réalisation spectrale des voyelles

Cédric Gendrot, Lab. de Phonétique et Phonologie, Université Paris3, CNRS, UMR7018
Kim Gerdes, Lab. de Phonétique et Phonologie, Université Paris3, CNRS, UMR7018

Le but de cette étude est de relier la réalisation spectrale des voyelles et la structure syntaxique en dressant un parallèle avec la structure prosodique. Ce travail s'inscrit dans un projet de plus grande envergure visant à reconstituer une hiérarchie prosodique d'après des analyses spectrales. Le corpus ESTER (Galliano et al., 2005), constitué d'enregistrements radio- et télédiffusés d'émissions journalistiques, a été retenu pour nos analyses. Il s'agit de parole publique : l'articulation, sans être soutenue, y reste bonne, afin que la parole puisse être partagée par une large audience. La parole ne peut pas être qualifiée de spontanée, il s'agit plutôt de parole préparée : on observe peu d'hésitations, peu de fragments de mots et les structures syntaxiques restent souvent proches du langage écrit. Le corpus utilisé correspond à environ 60 heures de parole utile extraites pour la majeure partie d'émissions de France Inter, enregistrées en 1998 et transcrites orthographiquement (600k mots et environ 800k voyelles orales).

Il a été montré sur de nombreuses langues que la durée des voyelles influe sur leur réalisation spectrale, et plus précisément leurs valeurs de formants. L'espace vocalique formé par les formants F1 (1er formant) et F2 (2ème formant) augmente progressivement avec la durée des segments analysés (cf. Figure1). Les variations de F1 et de F2 peuvent être interprétées – si on ne tient pas compte de l'effet des lèvres – en termes d'aperture/fermeture (corrélée à F1) et d'antériorité/postériorité (corrélée à F2). (pour le français, Gendrot et Adda-Decker, 2005, à paraître)

Les variations de durée des voyelles peuvent être imputées à un très grand nombre de facteurs tels que le style, le débit, le contexte phonémique, l'accentuation pragmatique, etc. La durée des voyelles est également dépendante de leur position prosodique (par exemple la position dans la syllabe, position dans le mot, position dans le groupe accentuel, position dans le groupe intonatif). Les quatre positions mentionnées ici sont fréquemment utilisées dans le cadre de la phonologie prosodique où elles forment une hiérarchie¹⁸. La littérature concernant la réalisation phonétique (principalement en terme de f0) des constituants prosodiques est importante et malgré tout parfois très discordante ; il apparaît tout de même que les frontières de ces différentes catégories prosodiques (de syllabe, de mot, de groupe, ...) sont sources d'allongements vocaliques et sont fréquemment corrélés à la présence de pauses pour les constituants prosodiques de haut niveau (groupes intonatifs, énoncé, paragraphe).

Au sein de notre projet, deux étapes ont déjà effectuées parmi l'analyse des catégories prosodiques : l'analyse des frontières de mots et l'analyse des frontières de constituants en présence d'une pause.

- Des analyses formantiques ont été effectuées en fonction de la position de la voyelle dans le mot. Les voyelles en syllabe finale de mots (potentiellement plus longues) ont été comparées aux voyelles en syllabe initiale et intermédiaire de mots (généralement non allongées, sauf en cas d'insistance), et ont été détectées comme ayant des positions plus périphériques dans l'espace acoustique. Ces positions lexicales sont fréquemment accompagnées de valeurs de f0 plus élevées. (Gendrot et Adda-Decker, 2006).

¹⁸ Différents cadres théoriques sont proposés mais afin de ne pas entrer dans ce débat, nous nous contentons d'utiliser les catégories pour lesquelles il existe un consensus large.

- Des analyses formantiques ont également été effectuées en fonction de la présence/absence de pauses dans l'entourage vocalique (figure 2) et ont révélé des positions vocaliques plus périphériques pour les voyelles en présence d'une pause (pause + voyelle ; pause + consonne + voyelle ; voyelle + pause ; voyelle + consonne + pause). Ces valeurs de formants plus périphériques sont également corrélées à un allongement vocalique important.

L'objectif de cette étude est de déterminer une catégorie prosodique intermédiaire aux deux catégories analysées ci-dessus. Cette catégorie correspondrait peu ou prou aux groupes accentuels (Jun et Fougeron, 2000) ou aux groupes de sens (Grammont, 1931). Ces groupes de sens révèlent une cohésion sémantique des différents mots à l'intérieur de chaque constituant. Il a fréquemment été observé dans la littérature à quel point il est difficile de distinguer les catégories prosodiques intermédiaires (groupes accentuels) sur la base des mouvements de f_0 et des valeurs de durée. En effet, les groupes accentuels ne sont pas nécessairement conclus par une montée de f_0 (Vaissière, 2002) et il est indispensable d'effectuer une normalisation de la durée de par les importantes variations intrinsèques de durée, or cette dernière s'avère délicate.

La méthode utilisée ici est une segmentation syntaxique sur la base de catégories grammaticales établies automatiquement, combinée à quelques règles.

- Etiquetage de chaque mot avec toutes les catégories disponibles dans le lexique de formes fléchies du français (Leff : Clément et al. 2004), légèrement corrigé pour nos besoins.
- Utilisation du chunker (http://nltk.org/index.php/Main_Page) du Natural Language Toolkit afin de créer deux types de segments :
 - les segments nominaux, prépositionnels et verbaux avec leur entourage le plus proche (clitiques, déterminants, prépositions, adjectifs, etc.)
 - toutes les suites de mots qui ne sont pas définies par la première règle
 - Ensuite, trois règles de combinaisons ('merging') sont appliquées :
 - Combinaison de tout segment terminant sur un auxiliaire ou modal avec le segment suivant.
 - Combinaison de tout segment verbal avec le segment suivant si la combinaison fait moins de 7 syllabes.
 - Combinaison de tout autre suite de segments qui fait moins de 7 syllabes.
 - Les segments découpés par cet algorithme peuvent avoir plus de 7 syllabes, si les règles précédentes le permettent, par exemple « avec qui j' ai pu m' entretenir », qui forme un groupe très naturel et difficilement à découper. La règle des 7 syllabes (Wioland, 1985) pourra être supplantée ultérieurement par une durée plus respectueuse du style de parole)

Cette étude ne vise pas à proposer un découpage prosodique unique sur des critères syntaxiques mais plutôt à comparer les différentes possibilités sur la base des mesures spectrales. Ainsi, il sera possible de moduler les différents paramètres prosodiques par des critères phonétiques tels que la durée maximale d'un groupe prosodique avant d'en démarrer un suivant. Ces différents critères devraient permettre à terme d'aider à la modélisation des constituants prosodiques intermédiaires et systématiser la recherche sur le lien entre prosodie et syntaxe. Il est envisageable qu'une bonne disposition des règles de segmentation permettra de déterminer les positions et types de groupes accentuels, essentiel pour la synthèse de la parole. Les premiers résultats obtenus sur la base de ces règles montrent des résultats en accord avec les hypothèses, à savoir des voyelles en position finale de constituants prosodiques intermédiaires (groupes accentuels) hyperarticulées par rapport à des voyelles en

Bibliographie :

- Clément, L., Sagot, B. et Lang, B. (2004) Morphology based automatic acquisition of large coverage lexica LREC'04, Lisbonne.
- Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.-F. & Gravier G. (2005). ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News, Proceedings of Interspeech, Lisboa, Portugal, 1149-1152.
- Gendrot, C. & Adda, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. Proceedings of Eurospeech – Lisbon (Portugal), September 2005, pp. 2453-2456.
- Gendrot, C. & Adda-Decker, M. (to appear). Impact of duration and consonantal context on formant values of oral vowels: An automatic analysis of large broadcast news corpora in French. In M. Embarki & C. Dodane (Eds.), Patterns of speech coarticulation: Between physics and metaphysics. Cambridge: CUP.
- Gendrot, C. & Adda-Decker, M. (2006) Analyses formantiques automatiques en français : périphéralité des voyelles orales en fonction de la position prosodique. XXVIèmes Journée d'Etude de la Parole, 12-16 juin 2006.
- Grammont M., 1933, Traité de phonétique, Paris: Librairie Delagrave.
- Jun S.-A. & Fougeron C. (2000), A Phonological model of French intonation. In A. Botinis (ed.) Intonation: Analysis, Modeling and Technology. Dordrecht : Kluwer Academic Publishers. pp.209-242.
- Vaissière J., 2002, Cross-linguistic prosodic transcription: French versus English, in Problems and methods in experimental phonetics, N. B. Volskaya, N. D. Svetozarova et P. A. Skrelin, St.-Petersburg, St Petersburg State University, pp. 147-164.
- Wioland, F. Les structures syllabiques du français : Fréquence et distribution des phonèmes consonantiques, Contraintes idiomatiques dans les séquences consonantiques. Slatkine-Champion, Genève-Paris, 1985.