Journal of Speech Sciences 4(1): 5-12.2014
Available at: www.journalofspeechsciences.org

Analyse automatique de la structure prosodique d'énoncés de styles variés

Martin, Philippe
University of Toronto

Abstract

The aim of this study is to characterize automatically the boundary tones in French, from the point of view of their categories, their acoustic realizations and their distribution according to a prosodic grammar. If the automatic labelling proves pertinent, the proposed method would lead to an automatic process to determine the prosodic structure of speech recordings. From the C-PROM corpus, four spontaneous recordings with various styles (oral reading, university lecture, narration, political discourse) were automatically analyzed based on 1) the identification of prominent syllables (from the information already encoded in the corpus transcription), and 2) on the melodic contours glissando values relatives to prominent syllables. From an existing definition of prosodic contours in French, C0, C1, C2 and Cn, the validity of a prosodic grammar has been established, as reflected by the results shown in Table 1, as only one occurrence (*C2C0) not predicted by the grammar has been observed.

Table 1. Distribution of sequences of prosodic contours according to the speaker's styles

	lec-fr	cnf-fr	nar-fr	pol-fr
C1C0	2	2	1	3
*C2C0	0	0	0	1
C1C1	19	6	15	7
C2C1	7	2	1	9

Furthermore, the analysis gave some information pertaining to the distribution of prosodic contours relatively to the different speech styles, as shown in Table 2..

Table 2. Distribution of sequences of prosodic contours according to speaker's styles

	lec-fr	cnf-fr	nar-fr	pol-fr
C0	7%	1%	1%	7%
C1	43%	25%	35%	38%
C2	9%	3%	1%	10%
Cn	30%	66%	53%	48%
Ci	2%	3%	4%	5%
creak	0.7%	0%	3%	0%
euh	0%	0.5%	4%	0%
Total	136	211	176	169

Martin, Philippe

Résumé

Le but de cette étude est d'examiner la distribution des tons de frontière des groupes prosodiques à la fois du point de vue de leurs catégories, de leurs réalisations acoustiques et de leur conformité avec une grammaire prosodique. Si cette catégorisation automatique de ces tons s'avère pertinente, on disposera alors d'un processus automatisable pour déterminer la structure prosodique. À partir du corpus C-PROM, quatre enregistrements de parole spontanée de styles différents (lecture orale, conférence universitaire, narration, récit de vie, discours politique), ont été analysés automatiquement en se basant sur 1) l'identification des proéminences syllabiques (information déjà présente dans C-PROM) et 2) sur les valeurs de glissando des contours mélodiques à l'endroit des syllabes proéminentes. À partir des définitions des contours prosodiques du français, C0, C1, C2 et Cn, la validité d'une grammaire prosodique a pu être établie. De plus, des indications relatives aux styles des locuteurs analysés ont également pu être obtenues.

Mots clés intonation, français, contour mélodique, structure prosodique.

1. Structures prosodiques

Le but de cette étude est d'examiner la distribution des tons de frontière des groupes prosodiques à la fois du point de vue de leurs catégories, de leurs réalisations acoustiques et de leur conformité avec une grammaire prosodique. Si cette catégorisation automatique de ces tons s'avère pertinente, on disposera alors d'un processus automatisable pour déterminer la structure prosodique.

1.1. La structure prosodique autosegmentale-métrique

L'acceptation dominante du concept de structure prosodique relevant de la théorie Autosegmentale-Métrique (AM) organise hiérarchiquement en un ou plusieurs niveaux les groupes accentuels (*Accent Phrase*, AP) censés contenir une unité lexicale de classe ouverte (verbe, adverbe, nom ou adjectif) autour duquel gravitent un ou plusieurs mots grammaticaux, de classe fermée (conjonctions, pronoms, prépositions, etc.). Ces AP sont pourvus d'un accent mélodique (*Pitch Accent* dans la terminologie AM).

Dans une structure prosodique complexe, un premier regroupement des AP constitue un syntagme intonatif intermédiaire (ip), terminé par un ton de frontière. Le regroupement de ces ip constitue un syntagme intonatif (IP), également terminé par un ton de frontière. Enfin, le regroupement des IP constitue l'entièreté de la structure prosodique (SP), terminée par un troisième type de ton de frontière, dès lors conclusif.

Une structure prosodique donnée ne comprend pas nécessairement tous ces niveaux de regroupement. On peut par exemple avoir une SP "plate", regroupant une énumération de AP en un seul niveau. Une SP à deux niveaux est également possible, regroupant des AP en IP, et des IP en SP. La SP à trois niveaux regroupe alors des AP en ip, des ip en IP, et finalement des IP en SP. C'est le cas général mentionné plus haut.

1.2 Structure prosodique en français

En toute généralité, les groupes accentuels portent un accent mélodique lexical dont la position dans la séquence de syllabes est définie par la morphologie ou la syntaxe, ou encore par une règle rythmique. En français par contre, il n'y a pas d'accent lexical mais seulement un accent de groupe. On est donc conduit à admettre l'existence d'un troisième type de ton de frontière, placé sur la dernière syllabe prononcée des AP comme les tons de frontières de la SP, des IP et des ip.

De ce fait, il ne peut donc pas y avoir de AP en français, mais seulement des mots prosodiques (*Prosodic Words*, PW). Ceux-ci n'étant pas des AP ne contiennent pas non plus nécessairement un mot lexical (Verbe, Adverbe, Adjectif ou Nom), mais peuvent en contenir plusieurs, ou encore se limiter à une seule syllabe.

Il y aurait donc en français 4 types de tons de frontières (et aucun accent lexical). Appelons les C0 (frontière de SP), C1 (frontière de IP), C2 (frontière de ip), et Cn (frontière des AP). Dans ce système, les mots prosodiques sont donc dotés en français d'un seul ton de frontière, qui peut être C0, C1, C2 ou Cn. Il n'y a pas de coexistence d'accent lexical et de ton de frontière comme cela peut être le cas en Italien par exemple. En fait, on rejoint par ce raisonnement la définition de la structure prosodique donnée il y a longtemps déjà par Martin (1975) et par Mertens (1987). Cette structure est clairement récursive en français, assemblant des groupes prosodiques de plus en plus grands.

1.3 Nature des tons de frontière

Ayant adopté le système de transcription ToBI, les descriptions de la SP autosegmentale-métrique utilisent des tons Haut et Bas (et leur variantes) pour décrire les différents tons de frontière. Cependant, l'existence manifeste de contrastes d'empan mélodique (contrastant les tons C0 et C2 par exemple) conduisent à utiliser plutôt des contours, dont les traits descriptifs impliquent par exemple la durée, la fréquence moyenne, et la variation mélodique. Outre des propositions de révision de la notation ToBI adaptée au français (Post & Delais, 2011), on trouve ainsi dans la littérature des définitions de tons de frontière par des contours -Haut, - Montant, -Ample pour C0, +Haut, +Montant, +Ample pour C1, +Haut, -Montant, -Ample pour C2.

Pas plus que la notation ToBI par tons Haut et Bas, ce dernier type de notation ne permet de rendre compte efficacement des réalités acoustiques manifestant ces différentes réalisations des tons de frontière, et en particulier du mécanisme dit du *contraste de pente*. Ce mécanisme prévoit essentiellement la réalisation pour C2 d'une variation mélodique de sens opposé à celle instanciée pour C1, dont C2 dépend (la relation de dépendance résulte du regroupement de ip terminés par C2 en un IP terminé par C1, la présence de C2 requérant celle de C1 à sa droite, i.e. apparaissant après C2).

1.4 Domaines des tons de frontière

Si l'analyse de corpus lus et fabriqués permet de valider expérimentalement les caractéristiques attendues des tons de frontière, il n'en va pas a priori de même avec la parole spontanée (i.e. non préparée). Toutefois, on peut s'attendre que, du point de vue de l'auditeur, les tons de frontière de même niveau présentent dans leurs réalisations suffisamment de traits communs pour qu'ils puissent être identifiés comme appartenant à une même classe, et ce au cours du déroulement du temps où surviennent les évènements prosodiques.

Contrairement à ce que peuvent suggérer les représentations graphiques des structures prosodiques, cette identification se fait séquentiellement dans l'axe temporel. Étant donné les limitations de la mémoire à court terme de l'auditeur, l'appartenance à une classe de tons de frontière donnée ne pourra se faire que relativement aux tons précédant et suivant le ton considéré. L'identification d'un ton est donc un processus local, par lequel l'auditeur doit décider si un ton (i.e. le contour mélodique qui le réalise) appartient à la même classe que le précédent ou non, ou éventuellement de plusieurs tons précédents.

Une exception à cette règle est donnée par le contour conclusif C0, dont les réalisations successives peuvent être temporellement éloignées, mais qui peuvent toujours être identifiées par les auditeurs quelles que soient ses variantes de réalisation (l'auditeur peut toujours savoir si l'énoncé est terminé ou pas).

Du point de vue phonologique, cet aspect temporel revient à considérer l'existence de domaines locaux, définis comme des séquences de tons de frontière terminés par un ton de niveau supérieur. Dans chaque séquence de SP, l'auditeur devrait donc pouvoir identifier des contours terminaux C0 par des caractéristiques acoustiques similaires. De même, dans chaque séquence de IP, l'auditeur devrait pouvoir identifier des contours terminaux C1 par des caractéristiques acoustiques similaires, etc. Il en résulte que des contours de même classe C1, C2, Cn ne présenteront des traits semblables qu'à l'intérieur de chaque domaine, et ne seront pas nécessairement similaires du point de vue acoustique d'un domaine à l'autre dans l'énoncé.

Font également exception à ce processus les contours relatifs à l'accent secondaire (aussi appelé accent d'insistance), identifiés par des traits acoustiques similaires (montée mélodique), et qui, n'étant pas des tons de frontière, sont placés sur la première syllabe des mots lexicaux. Une ambiguïté apparait alors lorsque le mot lexical impliqué ne possède qu'une seule syllabe : s'agit-il alors d'un ton de frontière (éventuellement neutralisé) ou d'un accent d'insistance ?

Les catégories et les définitions fonctionnelles et perceptives des tons de frontière suivants :

C0 : ton de frontière conclusif, facile à identifier par simple écoute, au cours de laquelle l'auditeur n'attend pas de continuation de l'énoncé ;

C1, C2, Cn : catégories jugées perceptivement comme non conclusives. À l'écoute de segments extraits de l'énoncé et se terminant par un de ces contours, l'auditeur s'attend à une continuation de l'énoncé ;

Ci : correspondant à l'accent d'insistance, identifiable par sa position s'il n'est pas placé sur la syllabe finale des mots lexicaux ;

Creak : généralement utilisé par certains locuteurs comme ton conclusif, ou comme contour C1.

2. Analyse expérimentale

2.1. Corpus d'analyse

Si le mécanisme de neutralisation et la notion de domaines permet de mieux comprendre les variations de réalisations de tons de frontière observées, il pose aussi un problème difficile quant à la catégorisation des tons des données expérimentales. De plus, bien des auteurs ont remarqué que des tons apparemment classés de catégorie C1 étaient réalisés de manière très diverses.

Le corpus utilisé pour les tests est C-PROM (2010). C-PROM est un corpus aligné et annoté, développé pour l'étude des proéminences syllabiques en français. Il inclut 24 enregistrements échantillonnés en 7 genres (ou styles) de parole et produits par des locuteurs francophones (issus de Belgique, de France et de Suisse). Dans ce corpus, on n'a retenu que les locuteurs hexagonaux.

Les enregistrements du corpus C-PROM analysés appartiennent aux genres suivants :

lec-fr : lecture orale (149s.);

cnf-fr : conférence universitaire (224s.);

nar-fr : narration, récit de vie (197s.);

pol-fr : discours politique (217s.).

Le détail des formats de transcription peut être consulté en ligne.

2.2. Méthode d'analyse

À partir des transcriptions et des annotations de proéminences disponibles dans le corpus C-PROM, on a analysé plusieurs extraits de parole de locuteurs hexagonaux de styles différents (lecture orale, conférence universitaire, narration, discours politique). La durée totale des extraits est de 787s.

Dans chacun des extraits, les voyelles (et seulement les voyelles) des segments annotés comme proéminents dans les annotations d'origine (proéminence forte P ou faible p, ce jugement étant éventuellement revu pour tenir compte de la contrainte des 7 syllabes qui limite le nombre de syllabes non proéminentes successives), ont été annotés et surlignés automatiquement.

Pour différencier C0, C1, C2 et Cn, on utilise comme critère la valeur correspondante de glissando en demi-tons par seconde (Mertens, 2004), du reste également affichée automatiquement par le logiciel d'analyse pour chaque segment. Si le glissando et supérieur au seuil de perception (n demi-tons par seconde au carré, avec une valeur du paramètre multiplicatif égale à 32), un contour mélodique montant réalisant le ton de frontière est catégorisé comme C1 ; S'il est descendant comme C2. Si le glissando du contour est inférieur au seuil, il est noté Cn.

Les segments ainsi annotés apparaissent sur l'écran d'analyse acoustique dans une couleur dénotant leur catégorie (Fig. 1)

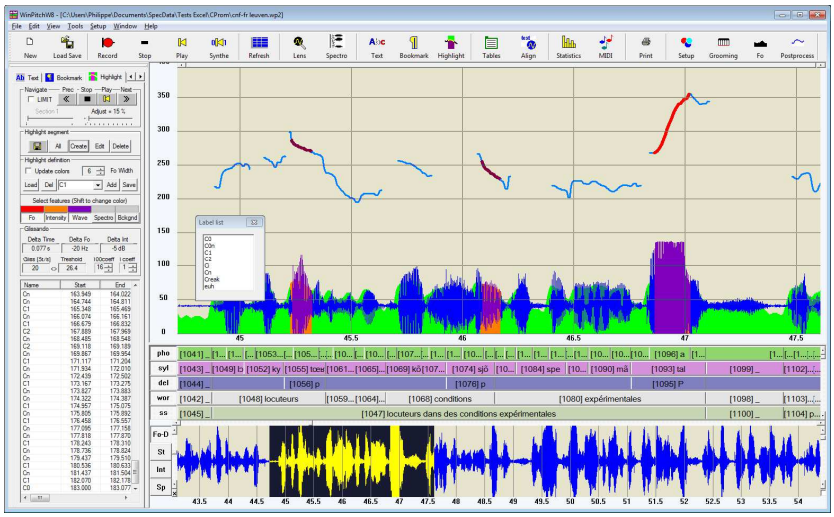

Fig. 1. Exemple de surlignage de segments en couleur différentes selon les catégories de contours prosodiques

Une fonction du logiciel d'analyse permet alors de mesurer automatiquement pour chaque segment correspondant à des tons de frontière annotés comme proéminents plusieurs paramètres acoustiques tels que la différence de fréquence fondamentale F0, la durée, la fréquence fondamentale moyenne, la valeur du glissando ainsi que le seuil correspondant. Ces valeurs sont automatiquement transférées dans un tableau Excel aux fins d'analyse. Ce logiciel dispose de plusieurs fonctions permettant l'analyse statistiques et des représentations efficaces des données, en particulier par l'utilisation des fonctions *powerview*. Par manque de place, ces résultats ne sont pas reportés ici.

2.3. Vérification de la grammaire prosodique

On a vu plus haut que les limitations de la mémoire à court terme de l'auditeur permettent de rendre compte du caractère local de la catégorisation des évènements prosodiques. Lors du déroulement temporel des évènements prosodiques, l'identification des contours prosodiques se fait domaine par domaine, chaque

domaine étant défini par un contour de rang supérieur placé en fin de domaine : domaine des C0, contenant une séquence de C1, lesquels définissent des séquences de C1, les C1 terminant des séquences de C2, et les C2 des séquences de contours neutralisés Cn.

Cette hiérarchie de domaines permet de définit des séquences de classes de contours bien formées et mal formées. Ainsi, les séquences C1 C1 C0, C2 C1 C0, C1 Cn C0 sont bien formées, mais C1 C2 C0 ne l'est pas. De même C2 C2 C1, Cn Cn C1, Cn Cn C2 sont bien formées, mais pas C2 Cn C0.

La table 1 donne les occurrences de quelques-unes de ces séquences pour les 4 locuteurs. On constate qu'il n'y a qu'un seul cas de séquence mal formée ne correspondant pas aux principes donnés plus haut.

Table 1. Distribution des séquences de contour selon le genre des locuteurs

	lec-fr	cnf-fr	nar-fr	pol-fr
C1C0	2	2	1	3
*C2C0	0	0	0	1
C1C1	19	6	15	7
C2C1	7	2	1	9

La seule occurrence mal formée C2C0 apparait dans pol-fr, et à l'écoute s'interprète comme un accent d'insistance placé sur la syllabe finale du mot prosodique ce contour étant inattendu sur cette syllabe.

Le contraste de pente C2C1 s'observe relativement fréquemment dans cnf-fr, mais dans des séquences CnCn…CnC2C1, indiquant une structure prosodique [[[Cn Cn Cn] C2] C1] non congruente avec la structure syntaxique associée, alors qu'on s'attendrait à une SP [[Cn Cn Cn] C1] ou [[C2 C2 C2] C1]. La locutrice réalise donc un contraste de pente, mais seulement en précédant directement un contour C1, dénotant ainsi une moindre planification de la structure prosodique.

Table 2 Répartition des contours réalisés par genre

	lec-fr	cnf-fr	nar-fr	pol-fr
C0	7%	1%	1%	7%
C1	43%	25%	35%	38%
C2	9%	3%	1%	10%
Cn	30%	66%	53%	48%
Ci	2%	3%	4%	5%
creak	0.7%	0%	3%	0%
euh	0%	0.5%	4%	0%
Total	136	211	176	169

La table 2 présente les pourcentages d'emploi des différents contours, reflétant l'utilisant de structures prosodiques plus ou moins complexes. Ainsi les enregistrements lec-fr et pol-fr présentent un plus grand nombre d'occurrences de contours C2, donc de SP à 3 niveaux, caractéristiques de la parole lue (le locuteur pol-fr lit son discours). À l'inverse, la narratrice nar-fr utilise des structures plus simples, accompagnée d'un plus grand nombre d'hésitations.

3. Conclusion et perspectives

Le processus d'analyse présenté constitue une technique simple pour établir automatiquement la structure prosodique à partir de l'identification des syllabes effectivement accentuées. L'identification de ces syllabes utilise comme vérification la contrainte des 7 syllabes et la position finale des proéminences sur les mots lexicaux. Ce type d'analyse a déjà été tenté par des processus comme Analor (2013), mais seulement pour l'analyse de SP à un seul niveau.

L'identification des catégories de contour suppose, on l'a vu, la quasi linéarité de la variation mélodique des contours implicite dans le calcul du glissando. Elle dépend aussi de la valeur des paramètres retenus pour ce calcul (coefficient de variation d'intensité et coefficient de la différence de demi-tons). L'analyse de variations régionales ou idiosyncratiques présentant des contours convexes ou concaves impliquera l'élaboration de critères d'identification des contours plus élaborés, mettant en jeu les propriétés de contrastes locaux décrits plus haut.

Références

Analor (2013) Logiciel d'étiquetage et séquençage basée sur l'analyse Prosodique du discours, http://www.lattice.cnrs.fr/Analor,70

C-PROM (2010). Corpus libre de parole multigenre,

https://sites.google.com/site/corpusprom/

Martin, Ph. (1975). Analyse phonologique de la phrase française *Linguistics*, (146) Fév. 1975, 35-68.

Martin, Ph. (1987). Prosodic and Rhythmic Structures in French *Linguistics*, 1987, 25-5, 925-949.

Mertens, P. (1987) *L'intonation du français. De la description linguistique à la reconnaissance automatique.* Unpublished Ph.D. (Univ. Leuven, Belgium).

Mertens, P. (2004). Le prosogramme : une transcription semi-automatique de la prosodie *Cahiers de l'Institut de Linguistique de Louvain* 30, 1-3, 7-25.

Post, B. and Delais-Roussarie, E. French ToBI, Workshop on Romance ToBI, Satellite workshop PaPI 2011, Universitat Rovira i Virgili (Tarragona), June 23, 2011.