



Un algorithme de segmentation en phrasé

Philippe Martin¹

(1) LLF, UFRL, ODG, Place Paul Ricoeur, 75013 Paris, France

philippe.martin@linguist.univ-paris-diderot.fr

RESUME

En lisant à voix haute ou silencieusement, nous segmentons la parole en groupes accentuels, ne contenant qu'une seule syllabe accentuée (hors accent d'insistance). Dans les langues à accent lexical comme l'anglais, les groupes accentuels contiennent un mot lexical (nom, adverbe, verbe ou adjectif) dans lequel la position de l'accent est définie dans le lexique. En français, langue sans accent lexical, les groupes accentuels sont définis non par la catégorie de mots qu'ils contiennent mais par le temps qu'il faut pour les lire ou les prononcer. Le phrasé, c'est-à-dire la segmentation en groupes accentuels, dépend donc du débit de parole. Avec un débit de parole lent, toutes les syllabes finales de mots lexicaux sont accentuées, alors qu'un débit de parole rapide peut rassembler plusieurs mots lexicaux dans un seul groupe accentuel. Un algorithme de phrasé basé sur cette propriété est présenté et illustré sur deux exemples de parole lue et spontanée.

ABSTRACT

Automatic phrasing in French.

Whether we read aloud or silently, we segment speech in accent phrases, containing only one stressed syllable (excluding emphatic stress). In lexically stressed languages such as English, the location of stress in a noun, an adverb, a verb or an adjective (content words) is defined in the lexicon, and accent phrases include one single content word. In French, a language without lexical stress, accent phrases are defined by the time it takes to read or pronounce them. Therefore, actual phrasing, i.e. the segmentation into accent phrases, depends on the speech rate. With a slow speech rate, all content words final syllables are stressed, whereas a fast speech rate could merge more than one content word in a single accent phrase. Based on this observation, a computer algorithm for automatic phrasing operating in a top-down fashion is presented and applied to two examples of read and spontaneous speech.

MOTS-CLES : Groupe accentuel, français, accent lexical, phrasé, syllabe accentuée

KEYWORDS: Accent phrase, French, lexical stress, phrasing, syllabic stress

1 Introduction

Lorsque nous lisons un texte en anglais, à voix haute ou silencieusement, nous pouvons procéder mot à mot ou même syllabe par syllabe, mais si nous maîtrisons la langue et identifions tous les mots, nous procédons généralement par groupe de mots. Il est facile d'observer dans une

transcription orthographique où tous les mots seraient terminés par un point final que nous ne lisons pas mot à mot, comme ce serait le cas dans l'exemple en anglais : *In. The. Orthographic. Representation. Of. Speech. Of. Most. Written. Languages. Segmentation. Is. Defined. By. Spaces. Between. Words.* En réalité, si nous sommes suffisamment familiers avec la langue en question, nous lisons normalement en regroupant les mots en unités contenant un nom, un adverbe, un verbe ou un adjectif (i.e. un mot lexical, de classe ouverte), accompagnés chacun par des mots grammaticaux (pronoms, conjonctions, prépositions, déterminants..., mots de classe fermée) qui leur sont associés : [*in the orthographic*] [*representation*] [*of speech*] [*of most*] [*written*] [*languages*] [*segmentation*] [*is defined*] [*by spaces*] [*between*] [*words*]. En phonologie, de tels groupes de mots sont appelés groupes accentuels, et définissent les unités prosodiques minimales, qui organisés en une hiérarchie, constituent la structure prosodique de la phrase (Martin, 1975, Selkirk, 1978).

Pour tous les locuteurs de l'anglais, la position des syllabes accentuées dans les mots lexicaux est prévisible et résulte de l'acquisition du lexique de la langue. D'autres syllabes accentuées peuvent également apparaître, mais contrairement à l'accent lexical, elles ne sont pas prévisibles car elles résultent d'un choix particulier du locuteur pour indiquer une emphase. Ce type d'accentuation emphatique peut se produire sur une syllabe différente de l'accent lexical ou sur la même syllabe. Dans ce dernier cas, le locuteur utilisera une réalisation acoustique différente car l'accent emphatique doit être perçu par les auditeurs comme différent et imprévisible par rapport à l'accent lexical prévisible.

La prévisibilité de l'accent lexical suggère que la perception des syllabes accentuées ne dérive pas directement du traitement des caractéristiques acoustiques spécifiques du discours, telles que la durée de la voyelle, le changement de fréquence fondamentale ou la modulation d'intensité, les paramètres prosodiques classiques souvent mentionnés dans la littérature comme paramètres de l'accent. En fait, la perception des syllabes accentuées peut être considérée comme le résultat d'un mécanisme d'identification comparant les caractéristiques acoustiques réelles des syllabes avec une position prévisible dérivée de la connaissance de la langue.

On peut citer à ce sujet l'expérience sur la perception des syllabes accentuées du berbère et de l'hébreu par des sujets qui n'ont aucune notion de ces langues (Mettouchi et al., 2007). Les caractéristiques acoustiques de l'accent syllabique sont présentes dans le signal vocal, mais dans cette expérience, les auditeurs n'ont identifié que très peu de positions correctes de l'accent lexical, puisqu'ils ne disposaient d'aucun lexique approprié permettant de prépositionner les syllabes accentuées lors de l'audition des mots correspondants, contrairement aux locuteurs du berbère ou de l'hébreu.

2 Les groupes accentuels en français

Le français est une langue où la position de l'accent lexical a évolué progressivement vers la dernière syllabe des mots lexicaux (et même sur la dernière syllabe de tous les mots prononcés isolément) en perdant progressivement toutes les syllabes placées à l'origine après l'accent (Väänänen, 1995). La fonction de l'accent lexical comme marqueur de frontière morphologique existant dans les autres langues romanes a été progressivement perdue puisque devenue redondante. Il est alors devenu possible pour les locuteurs d'ignorer certains accents syllabiques portés par les mots lexicaux, comme dans *la petite armoire violette*, qui peut recevoir une, deux ou même trois syllabes accentuées : *la petite armoire violette*, *la petite armoire violette*, *la petite armoire violette* ou *la petite armoire violette*. Pour un locuteur francophone, il est facile de se rendre compte que la différence d'accentuation de ces exemples est liée au débit de parole. Pour

prononcer (ou même lire silencieusement) *la petite armoire violette* avec une seule syllabe finale accentuée sur *violette*, il faut utiliser un débit de parole (très) rapide, alors qu'un débit plus lent conduit à la prononciation de trois syllabes accentuées dans le même exemple.

On pourrait peut-être conclure qu'il n'y a pas de limite au nombre de syllabes et donc de mots qui peuvent être prononcés en français avec une seule syllabe accentuée finale, et qui peut être contenue dans un seul groupe accentuel. La prononciation des mots longs permet toutefois de déterminer une limite. Des mots tels que *l'anticonstitutionnalité* ("contre la constitution"), (8 syllabes) ou *intergouvernementalisation* ("inter-gouvernemental") (10 syllabes) semblent difficiles voire impossibles à prononcer ou même à lire silencieusement avec une seule syllabe accentuée finale. Déjà au XV^{ème} siècle, le grammairien Louis Meigret (1550) concluait que le mot le plus long qui pouvait être prononcé avec un seul accent final est formé d'un maximum de 7 syllabes. Bien plus tard, Martin (2014) a montré que ce n'est pas le nombre de syllabes qui importe, mais le temps qu'il faut pour les prononcer. Les données expérimentales montrent en effet que l'intervalle maximum entre deux syllabes accentuées consécutives (dans la parole continue) ne peut pas dépasser 1250 ms environ. Dans le style « parole de jeunes », on trouve des séquences de 10 ou 11 syllabes avec seulement une seule syllabe accentuée finale. Cette valeur est proche de la limite théorique, dérivée de la durée moyenne minimale des syllabes qui pourraient être perçues dans une séquence, soit 100 ms (Ghitza & Greenberg, 2009). Ces observations situent la durée maximale des phrases accentuées en français à environ 1250 ms à 1400 ms, le débit le plus rapide atteignant 8 à 9 syllabes par seconde (Lekha et Le Gac, 2004).

On peut aussi établir une durée minimale séparant deux syllabes accentuées successives, dans une configuration dite de « collision accentuelle ». Sa valeur s'évalue en sélectionnant des occurrences telles que *par le fait que* ou *le travail de nuit nuit*, sans déplacement ou suppression possible du premier accent. Il est souvent mentionné dans la littérature que ces cas nécessitent un écart acoustique entre syllabes accentuées consécutives (par exemple Di Cristo, 2016), habituellement mis en œuvre par la présence de consonnes après la première syllabe accentuée. En réduisant progressivement avec un éditeur de signal l'écart jusqu'à ce que la première syllabe cesse d'être perçue comme accentuée (sans modifier leur structure acoustique, i.e. en supprimant seulement la partie silencieuse), on obtient une limite d'environ 250 ms, ce qui détermine la durée minimale d'un groupe accentuel constitué de la deuxième syllabe de la séquence et précédé d'un silence suffisant. La désaccentuation perçue de la première syllabe [*par le fait*] va alors restructurer le groupe accentuel en [*par le fait que*].

3 Syllabes finales

Il est également facile de démontrer expérimentalement que toute syllabe suivie d'au moins 250 ms de silence est perçue comme accentuée en français. En insérant un silence de 250 ms, les syllabes finales de n'importe quelle catégorie de mots sont perçues comme accentuées, quelle que soit leur durée réelle ou leur mouvement de hauteur mélodique. Dans les langues à accent lexical, la perception d'une syllabe finale d'accent comme accentuée est préemptée par la position de l'accent lexical (s'il n'est pas en position finale). En italien par exemple, l'accent lexical de l'avant-dernière syllabe de *Marco* dans *la Sorella di Marco à partita* ("la sœur de Marco est partie") empêche un auditeur connaissant la langue de percevoir la dernière syllabe de *Marco* comme accentuée, même si elle est réalisée avec une forte montée mélodique, alors que pour un locuteur de français ne connaissant pas l'italien, cette syllabe sera probablement perçue comme accentuée, puisque c'est sa position attendue.

4 Glissando

Selon le modèle de Martin (1975, 2018), la structure prosodique résulte d'une organisation hiérarchique des groupes accentuels. En référence à un contour terminal attendu, perçu comme un marqueur de non-continuation de la phrase, deux autres contours mélodiques, les uns montants, les autres descendants, indiquent respectivement une continuité majeure et une continuité mineure (pour reprendre la terminologie déjà proposée par Delattre en 1966, mais avec une autre définition puisque portant ici sur les seules voyelles accentuées).

Les contours de continuité indiquent une relation de dépendance, de la continuation mineure vers la continuation majeure et de la continuation majeure vers le contour terminal, par un contraste de pente mélodique, où un contour descendant indique une dépendance en vers un contour montant situé plus loin dans le temps. Ce modèle implique que les mouvements mélodiques descendants et montants soient effectivement perçus comme tels, c'est-à-dire que la vitesse du changement mélodique dans le temps soit supérieure au seuil de glissando. Ce seuil est évalué à partir de la différence entre le début et la fin de la variation mélodique évaluée en demi-tons par rapport à la durée du contour (en supposant une variation linéaire, voir Rossi, 1971). Toute syllabe de mot portant une variation mélodique supérieure au seuil de glissando est donc accentuée selon le modèle de la structure prosodique retenu.

5 Pronoms toniques et démonstratifs

Les pronoms toniques en français (*moi, toi, lui, elle, nous, vous, eux, elles*) n'appartiennent pas à la catégorie des mots lexicaux, mais partagent leurs caractéristiques en termes d'accentuation, en particulier en position postverbale. En particulier, ils seront accentués s'ils sont suivis d'au moins 250 ms de silence ou si leur variation mélodique est supérieure au seuil de glissando. Ainsi dans l'exemple *moi ma mère le salon c'est de la moquette*, le pronom tonique *moi* est accentué s'il est suivi de 250 ms de silence, ***moi*** # *ma mère le salon c'est de la moquette*, mais ne l'est pas s'il n'y a pas de silence suffisant après *moi* : *moi ma mère* ... Il en va de même pour les pronoms démonstratifs (*celui, ceux, celles, celui-ci, etc.*) accentuables même non suivis d'un silence de 250 ms.

6 Eurythmie

Selon Wioland (1985), l'eurythmie de la parole spontanée procède en ajustant la durée moyenne des syllabes accentuées pour atteindre une durée comparable des groupes accentuels successifs. En lecture, les locuteurs utilisent le plus souvent une stratégie visant à équilibrer le nombre de syllabes des groupes accentuels successifs, au détriment éventuel de la congruence avec la structure syntaxique. Un exemple classique est donné par la phrase *Marie adore les chocolats*, où le locuteur a tendance en parole spontanée à réaliser un phrasé congruent avec la syntaxe [*Marie*] [*adore les chocolats*] et éventuellement à atteindre l'eurythmie en ralentissant le débit syllabique de [*Marie*] et en accélérant sur [*adore les chocolats*]. Au contraire, les lecteurs de cette même phrase montrent une tendance à regrouper les mots pour équilibrer le nombre de syllabes en phrases d'accent consécutives, au détriment de la congruence avec la syntaxe [*Marie adore*] [*les chocolats*].

7 Syllabes accentuables et syllabes accentuées

Le phrasé détermine une étape essentielle dans la compréhension de la parole. La segmentation en groupes accentuels constitue la première phase de reconstruction de la structure prosodique voulue par le locuteur, indispensable et incontournable pour accéder à la structure syntaxique dans une étape ultérieure. La structure prosodique résultante ne correspond pas nécessairement à la structure prosodique voulue par l'auteur du texte lu, car le phrasé dépend de la vitesse de lecture choisie, à haute voix ou silencieusement.

Le simple fait que nous connaissons la position possible des syllabes accentuées lorsque nous lisons à haute voix ou silencieusement suggère que nous n'avons pas vraiment besoin de données acoustiques pour percevoir des syllabes accentuées (non-emphatiques). Non seulement la lecture à voix haute ou silencieuse du même texte peut conduire à des segmentations différentes, mais à l'écoute, on ne peut empêcher d'avoir des attentes vis-à-vis d'une localisation des syllabes accentuées différente de celle effectivement réalisée par le locuteur. En d'autres termes, nous pouvons "entendre" des syllabes accentuées qui peuvent ne pas être présentes acoustiquement. Cette illusion apparente se retrouve dans de nombreux processus impliqués dans la perception de la parole (Amal & Giraud, 2017), et suggère non pas un traitement direct d'une entrée physique, mais la validation d'une entrée attendue par comparaison entre ce qui est attendu et ce qui est physiquement réalisé.

Puisque la réalisation effective de l'accentuation dépend de la vitesse de prononciation, la seule manière d'éviter la perception d'un accent syllabique virtuel qui ne serait pas réalisé effectivement serait d'adapter constamment le débit à celui utilisé par le locuteur. Cette adaptation n'est pas toujours facile ni même possible. Des exemples avec un débit de parole très rapide dépassant 7 ou 8 syllabes par seconde sont difficiles à suivre pour la plupart des auditeurs, au point que certains auront du mal à comprendre les énoncés, et avoir tendance à entendre des syllabes accentuées là où elles n'existent pas acoustiquement.

L'écart possible entre les syllabes accentuées perçues et celles effectivement réalisées en français conduit à différencier les syllabes accentuables des syllabes effectivement accentuées, selon la terminologie de Paul Garde (1968, 2013).

8 Annotation des syllabes accentuées : mission impossible ?

Le problème pour un annotateur de syllabes accentuées en français est donc de s'adapter au débit de parole de l'enregistrement. La perception sera influencée par le processus de prédiction de l'annotateur, tendant à détecter les syllabes accentuées aux endroits où il les aurait placées en lisant ou en parlant non pas avec le débit du locuteur mais avec le sien. On peut aussi imaginer pour la même raison qu'un annotateur ne perçoive pas une syllabe effectivement accentuée.

Le plus souvent, la détection automatique des syllabes accentuées en français opère de bas en haut (bottom-up) à partir de l'enregistrement, recherchant des variations acoustiques significatives entre syllabes consécutives en durée, fréquence fondamentale et intensité (pour les exemples récents, voir Goldman et al., 2013 ; Mertens & Simon, 2013). Dans ces processus, la qualité des voyelles n'apparaît pas comme un paramètre significatif en français.

Dans un article publié en 2013, M. Avanzi, confronté à l'incertitude dans l'annotation des syllabes accentuées, décrit en détail une procédure complexe impliquant deux experts, éventuellement aidés

d'un troisième en cas de désaccord entre les deux premiers. Même avec ce protocole, l'accord entre les annotateurs varie entre 60% et 80%.

Dans un autre article sur le même sujet, Christodoulides et Avanzi (2014) ont mis en œuvre un détecteur automatique de prééminence (c'est-à-dire également de l'accent d'emphase) par des méthodes d'apprentissage automatique appliquées à un grand corpus d'une durée de 11 heures. Les auteurs utilisent un ensemble complet de paramètres acoustiques censés être appropriés pour différencier les syllabes prééminentes des autres syllabes (durée syllabique minimale et maximale, fréquence fondamentale moyenne, minimale et maximale, intensité maximale, équilibre spectral, partie de l'étiquette vocale, présence et durée des pauses, structure syllabique, position de la syllabe dans le mot, etc.). Leurs meilleurs résultats, évalués par rapport à la référence déterminée par des experts, atteignent un niveau d'identification correct de 90%.

Compte tenu de ces difficultés, il semble que la détection des syllabes accentuées ne devrait pas procéder directement de l'analyse du signal de parole, mais plutôt indirectement par validation des hypothèses relatives aux positions potentielles basées sur les observations faites plus haut.

9 Un algorithme de segmentation descendant (top-down)

Pour exploiter les données acoustiques, et innover par rapport aux approches *bottom-up* opérant à partir des données acoustiques, on propose ici un algorithme *top-bottom* basé sur les mécanismes cités plus haut, en retenant les règles suivantes :

1. Toute syllabe suivie d'un silence de plus de 250 ms est accentuée
2. Toute syllabe finale d'un nom, adjectif, verbe, adverbe ou pronom (tonique ou démonstratif) est accentuable (définition classique du groupe accentuel contenant un seul mot lexical)
3. Si deux syllabes accentuables ou accentuées successives sont séparées par moins de 250 ms, la première n'est pas accentuée (durée minimale du groupe accentuel)
4. Toute syllabe accentuable avec changement de F_0 au-dessus du seuil de glissando est accentuée (définition de la structure prosodique)
5. Si 2 syllabes accentuées consécutives sont séparées de plus de 1250 ms en parole continue, au moins une syllabe accentuable dans cet intervalle est accentuée (durée maximale du groupe accentuel). Celle ayant avec la plus haute valeur de glissando est accentuée (caractère approximatif du calcul du glissando)
6. Une syllabe accentuable doit exister dans n'importe quelle durée de fenêtre égale à la durée moyenne de l'accentuation (eurythmie). L'eurythmie est mise en œuvre par le calcul incrémental de la moyenne des durées des groupes accentuels successifs à partir du début de l'enregistrement. Un test de cohérence eurythmique est ensuite appliqué par une fenêtre temporelle glissante, censée correspondre à un débit de parole supposé constant dans tout l'enregistrement. La durée de cette fenêtre résulte de la moyenne cumulative des groupes accentuels considérés successivement. Une syllabe accentuée supplémentaire sélectionnée selon sa valeur de glissando est ajoutée en cas d'absence d'accent dans une fenêtre donnée.

10 Un exemple de parole lue

Premier exemple de parole lue : *il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en randonnée car il avait du mal à suivre le rythme de ses compagnons.* Dans les étapes détaillées ci-dessous, les syllabes accentuables sont soulignées et les syllabes retenues comme effectivement accentuées sont soulignées et en gras.

1 : Toute syllabe finale de mot suivie d'un silence de plus de 250 ms est accentuée, cas du mot final *randonnée* :

Il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en randonnée

2 : Toute syllabe finale d'un nom, adjectif, verbe, adverbe ou pronom est accentuable :

Il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en randonnée

3 : Si 2 syllabes accentuées successives sont séparées par moins de 250 ms, la première n'est pas accentuée : l'écart entre *chaque* et *fois* est de 180 ms, sous la limite de 250 ms, *chaque* ne peut être accentué :

Il était une fois un pauvre escargot qui souffrait beaucoup à chaque |180 ms| fois qu'il partait en randonnée

4 : Toute syllabe accentuable avec changement de F0 au-dessus du seuil de glissando est accentuée (notation {valeur de glissando / seuil de glissando avec coefficient 0,16}). Les syllabes accentuables en dessous du seuil ne sont pas accentuées :

Il était {35/76} une fois {36/17} un pauvre {44/66} escargot {32/12} qui souffrait {54/144} beaucoup {79/66} à chaque fois {46/106} qu'il partait {32/51} en randonnée

5 : Deux syllabes accentuées successives séparées de plus de 1250 ms, comme dans le cas de [*à chaque fois qu'il partait en randonnée*] de durée 1367 ms. En sélectionnant la valeur glissando la plus haute, sur *fois* : [*à chaque fois qu'il partait en randonnée*].

6 : Cohérence eurythmique. Les durées moyennes cumulées pour chaque groupe accentuel déterminé au stade précédent sont successivement : *il était une fois* 726 ms, *Il était une fois un pauvre escargot* 706 ms, *Il était une fois un pauvre escargot qui souffrait beaucoup* 606 ms, le dernier groupe *qu'il partait en randonnée* a une durée de 1033 ms, supérieure à la moyenne cumulée de 606 ms. La syllabe finale de *partait* ayant la plus grande valeur de glissando est ajoutée à la liste des syllabes accentuées. Le résultat final est alors : *Il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en randonnée.*

11 Un exemple de parole spontanée

Le deuxième exemple présente un débit rapide caractéristique du « parler jeune » : *Juste pour une carte d'identité tu n'as pas ta carte tu fais tes vingt-quatre heures tu en ressors t'as la haine encore plus ça augmente.*

1 : La dernière syllabe de l'énoncé est suivie de plus de 250 ms de silence :

*Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre heures tu ressorts t'as la haine encore **plus***

2 : Toute syllabe finale d'un nom, d'un adjectif, d'un verbe ou d'un adverbe est accentuable :

*Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre heures tu ressors t'as la haine encore **plus***

3 : Si deux syllabes accentuables ou accentuées successives sont séparées par moins de 250 ms, la première n'est pas accentuée : les écarts entre *vingt-quatre* et *heures* (230 ms) et entre *encore* et **plus** (240 ms) sont en dessous de la limite de 250 ms :

*Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre | 230 ms | heures tu ressors t'as la haine encore | 240 ms | **plus***

4: Toute syllabe accentuable avec changement de F0 au-dessus du seuil de glissando est accentuée. Les syllabes accentuables sous le seuil ne sont pas accentuées : **Juste** {64/36} pour une **carte** {44/38} d'**identité** {54/45} t'as pas {18/35} ta **carte** {44/38} tu fais {54/142} tes vingt-quatre | 230 ms | **heures** {49/37} tu **ressors** {38/32} t'as la **haine** {25/22} encore | 240 ms | **plus** {38/23}

L'étape 5 ne s'applique pas, et le test d'eurythmie d'ajoute pas de syllabes accentuées supplémentaires.

12 Conclusion

Basé sur le fait que la perception des syllabes accentuées résulte d'un processus de validation comparant la position prédite avec des paramètres acoustiques effectivement mesurés, une segmentation automatique descendante du phrasé en français est décrite brièvement. L'algorithme incorpore les observations suivantes : 1) Durée minimale des groupes accentuels de 250 ms et maximale de 1250 ms ; 2) Toute syllabe finale de mot suivie d'au moins 250 ms de silence est perçue comme accentuée 3) la durée des groupes accentuels dépend du débit de parole choisi par le locuteur ou le lecteur ; 4) L'accent syllabique définissant le phrasé comporte un mouvement mélodique supérieur au seuil de glissando. Les données acoustiques n'interviennent que par les durées entre syllabes accentuables successives et la valeur de glissando de leur contour mélodique.

Remerciements

À Damien Lolive, lecteur convaincant du conte « Le petit escargot »

(https://www.iletaitunehistoire.com/genres/albums-histoires/lire/la-maison-de-l-escargot-biblihis_027).

Références

ARNAL, L. et GIRAUD, A-L. (2017). Neurophysiologie de la perception de la parole et multisensorialité. *Traité de neurolinguistique*, Serge Pinto et Marc Sato éd., Louvain-la-Neuve : De Boeck, 97-108.

AVANZI, M., LACHERET-DUJOUR, A., and VICTORRI, B. (2010). A Corpus based Learning Method for Prominence Detection in Spontaneous Speech. *Proc. of Prosodic Prominence*, Speech Prosody Workshop.

- AVANZI, M. (2013). Note de recherche sur l'accentuation et le phrasé à la lumière des corpus du français, *Tranel*, vol. 58, 5-24.
- CHRISTODOULIDES, G. & AVANZI, M. (2014). An Evaluation of Machine Learning Methods for Prominence Detection in French. *Proc. Interspeech 2014*, 116-119.
- DI CRISTO, A. (2016). *Les musiques du français parlé*, Berlin : De Gruyter Mouton. 513 p.
- GARDE, P. (1968). *L'accent*. Paris : Presses universitaires de France, collection SUP « Le linguiste », n° 5. 172 p.
- GARDE P. (2013). *L'accent*. Paris : Lambert-Lucas.
- GOLDMAN, J-P., AUCHLIN, A., ROEKHAUT, S., SIMON, A-C., AVANZI, M. (2013). Prominence perception and accent detection in French. A corpus-based account, *Language Science* (39), 95-106.
- GHITZA, O. & GREENBERG, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, vol. 66, n°s 1-2, 113-126.
- LEHKA, I. & LE GAC, D. (2004). Etude d'un marqueur prosodique de l'accent de banlieue, *Actes des XXIIIème Journées d'Etudes sur la Parole*, avril 2004, Fès, Maroc.
- MARTIN, P. (1975). Analyse phonologique de la phrase française. *Linguistics*, vol. 146, 35-68.
- MARTIN, P. (2014). Spontaneous speech corpus data validates prosodic constraints. N. CAMPBELL, D. GIBBON, D. HIRST (éd.), *Proceedings of the 6th conference on speech prosody*, 525-529.
- MARTIN, P. (2018). *Intonation, structure prosodique et ondes cérébrales*. London : ISTE.
- MEIGRET, L. (1550). *Le tretté de la grammère françoéze*. Paris, C. Wechel.
- MERTENS, P. & SIMON, A-C. (2013). Towards automatic detection of prosodic boundaries in spoken French, *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, 81-87.
- METTOUCHI, A., LACHERET-DUJOUR, A., SILBER-VAROD, V. *et al.* (2007). Only Prosody ? Perception of speech segmentation in Kabyle and Hebrew. *Cahiers de linguistique française*, vol. 28, A. AUCHLIN (éd.), *Actes du 2^e Symposium international IDP07 (Interfaces Discours Prosodie)*. 207-218.
- ROSSI, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour la parole. *Phonetica*. n° 23, 1-33.
- SELKIRK, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim, ed., *Nordic Prosody II*. Trondheim: TAPIR, 111-140.
- VÄÄNÄNEN, V. (1995). *Introducción al latin vulgar*, Madrid: Editorial Gredos.
- WIOLAND, F. (1985). *Les structures rythmiques du français*, Paris : Slatkine-Champion.