

Ondes cérébrales et contraintes de la structure prosodique



Philippe Martin

LLF UMR 7110, UFRL, Université Paris Diderot, France
philippe.martin@linguist.univ-paris-diderot.fr

Reçu le 05-07-2014 / Évalué le 20-09-2014 / Accepté le 18/10/2014

Résumé

L'observation attentive des données de parole spontanée et lue portant sur les durées minimales et maximales des groupes accentuels (Accent Phrases) conduit à proposer un modèle explicatif qui rend également compte de l'eurythmie de groupes accentuels successifs (tendance à l'équilibrage des durées), ainsi que de l'alignement de ces groupes avec des constructions syntaxiques déterminées. Ce modèle suggère que la perception des syllabes est synchronisée par les ondes cérébrales thêta (variant de 100 ms à 250 ms) et celui des groupes temporel par les ondes cérébrales delta (variant de 250 ms à 1200 ms environ). De plus, la différenciation des réalisations mélodiques des syllabes accentuées conduit à proposer un mécanisme incrémental de reconstitution de la structure prosodique par l'auditeur.

Mots-clés : structure prosodique, groupe temporel, eurythmie, ondes cérébrales, ondes delta, ondes thêta

Brain waves and the prosodic structure constraints

Abstract

Accent Phrase (AP) minimum and maximum duration measured in spontaneous and read speech leads to an explanatory model, which also pertains to observed eurhythm in successive AP as well as alignment of these groups with specific syntagms. This model suggests 1) that the perception of syllables is synchronized with theta brain waves (ranging from 100 ms to 250 ms) and 2) that Accent Phrases are synchronized by delta brain waves (ranging from 250 ms to 1200 ms). Furthermore, the differentiation of melodic movements observed on stressed syllables leads to an incremental model taking into account the temporal aspects of the sentence prosodic structure perception by the listener.

Keywords: prosodic structure, accent phrase, eurhythm, delta waves, theta waves

Introduction

Que ce soit en parole spontanée ou en lecture orale ou silencieuse, la production de parole est nécessairement accompagnée d'une variation musicale qui constitue l'intonation de l'énoncé. Cette « musique » résulte des variations de hauteur mélodique, de durée et d'intensité des syllabes et est directement liée aux émotions et aux attitudes des locuteurs (Léon, 2005). Mais elle assure aussi plusieurs fonctions dans le système linguistique, telles que celle, bien connue, de l'indication de la modalité de l'énoncé, mais également et surtout de la segmentation du flot syllabique en unités appelées communément groupes accentuels.

En effet, outre la corrélation observée depuis longtemps avec la modalité des énoncés, modalité déclarative, et leurs variantes implicatives d'évidence et de commandement, et modalité interrogative et leurs variantes implicatives de doute et de surprise (Martin, 1987), l'intonation de l'énoncé manifeste une fonction importante, sinon essentielle, dans la communication linguistique : celle de segmenter les énoncés en unités plus petites et de hiérarchiser les segments obtenus de manière à constituer une structure dite prosodique. Sans cette fonction, le traitement de la parole humaine par les auditeurs serait très lent et laborieux, comme on le verra plus loin.

Du point de vue du locuteur, c'est l'intonation de l'énoncé qui, par la hiérarchisation des groupes accentuels, lui permet de structurer dynamiquement au cours du temps les groupes temporels en les rassemblant ou au contraire en les séparant de manière à fournir une première indication sur la structuration des unités syntaxiques et sémantiques, dont l'analyse pourra éventuellement être affinée par l'auditeur dans un processus ultérieur.

L'objectif de cet article est de montrer que des propriétés déjà établies (Martin, 1987) portant sur les contraintes associées à la structure prosodique de l'énoncé trouvent une explication à la fois qualitative et quantitative dans les découvertes récentes en neurolinguistique, et en particulier dans les caractéristiques des ondes cérébrales Thêta et Delta.

1. Structure prosodique

Le premier type de segmentation de l'énoncé opéré par le locuteur est dû à une contrainte physiologique. La production de parole ne peut opérer que lors de la phase d'expiration du cycle respiratoire puisque la phase d'inspiration essentielle à la survie du locuteur ne permet pas de produire des sons de parole et se traduit par une pause silencieuse. La production de parole est donc segmentée en premier lieu en groupes de souffle, séquences de phonation entrées deux phases d'inspiration successives du cycle

respiratoire. Pour faciliter la reconstruction de la structure syntaxique par l'auditeur, les pauses respiratoires sont normalement positionnées par le locuteur aux frontières importantes de la structure syntaxique, par exemple en fin d'énoncé ou de syntagmes. Ce positionnement résulte d'un compromis entre la durée du groupe de souffle et la capacité respiratoire (le volume d'air inspiré) du locuteur à un moment donné.

La deuxième type de segmentation produit des segments plus petits et hiérarchisés, segments appelés dans la littérature groupes accentuels, groupes rythmiques ou encore groupes temporels, ou même souvent, dans la littérature du FLE, groupes de sens. Ces groupes, séquences temporelles de syllabes, contiennent des catégories syntaxiques variées (et donc pas nécessairement un seul verbe, nom, adjectif ou adverbe comme cela a été affirmé entre autres par Selkirk, 2000). En français, ils contiennent normalement une seule syllabe accentuée en position finale. Un groupe temporel (c'est la dénomination que nous adoptons ici, suite à Gilbert et Boucher, 2007), est donc une séquence de syllabes terminée par une syllabe accentuée. Ces groupes sont classés et hiérarchisés dynamiquement par le locuteur et cette hiérarchie des groupes temporels constitue la structure prosodique de l'énoncé (Martin, 1975, 1987, 2009).

Notons que la segmentation respiratoire et la segmentation en groupes temporels ne sont pas nécessairement dans une relation de congruence, dans laquelle les structures prosodiques s'emboîteraient complètement à l'intérieur des groupes respiratoires. D'ailleurs, lors de l'acquisition du langage et de la lecture par les enfants, des groupes temporels peuvent se trouver interrompus par une phase d'inspiration d'air, perturbant parfois ainsi la compréhension du message voulu par les jeunes locuteurs. Il en va de même lors d'un effort physique important consommant plus d'air pulmonaire que prévu lors de la production d'un groupe de souffle (nage, course à pied,...).

2. Contraintes de la structuration prosodique

La structure prosodique, et donc la segmentation de l'énoncé en séquences de groupes temporels successifs, est *a priori* indépendante des autres structures organisant d'autres unités composant l'énoncé, et en particulier la structure syntaxique (Ceci s'explique par le fait que la structuration prosodique préexiste à la structure syntaxique, sauf évidemment lors de la lecture). Toutefois, si elle est indépendante par nature, le locuteur placé dans des conditions de production du discours favorable, peut s'efforcer de faire correspondre, du moins en partie, les regroupements des groupes temporels avec les syntagmes de la structure syntaxique, et ce à différents niveaux dans la structure. Dans une adéquation optimale, dans un style de production langagière particulier rencontré dans la lecture, le locuteur s'efforcera de réaliser une congruence entre les structures prosodique et syntaxique, c'est-à-dire que la seconde

pourra s'emboîter dans la première. Cette correspondance entre les deux structures ne va pas de soi, elle résulte souvent d'un apprentissage scolaire qui laissera toujours des traces ne fût-ce que dans la lecture silencieuse.

Depuis une vingtaine d'années, l'approche théorique autosegmentale-métrique est dominante dans l'étude des structures prosodiques d'un système linguistique donné, et utilise la notation ToBI pour représenter les données prosodiques. J'utiliserai plutôt ici une autre approche, qui me semble mieux rendre compte de phénomènes spécifiques au français. En effet, si la structure prosodique est *a priori* indépendante de la structure syntaxique, d'une part plus d'une structure prosodique peut être associée à un texte donné et donc à une structure syntaxique donnée, et d'autre part toutes les combinaisons de hiérarchies de toutes les séquences syllabiques des groupes temporels ne sont pas admissibles (ce dont ne rendent pas compte, ou très partiellement, les descriptions autosegmentales-métriques).

On a observé (cf. par exemple Martin, 2009), aussi bien sur des corpus lus qu'en production spontanée, qu'il existe des contraintes spécifiques limitant le nombre de configurations des structures prosodiques. Je décrirai brièvement ces contraintes pour ensuite proposer un principe d'explication s'appuyant sur des découvertes récentes en neurolinguistique.

2.1 Nombre de syllabes des groupes temporels

Une certaine régularité dans les intervalles entre syllabes accentuées successives a été observée depuis longtemps. Déjà, Louis Meigret en 1550 avait remarqué en fabriquant des néologismes apparaissant dans des énoncés tels que *Les Mégapolitains surreparlementeront quoique nous surreparlementassions* qu'on ne pouvait prononcer cette phrase sans accentuer au moins une syllabe toutes les 7 syllabes (voir aussi Miller, 1956). Outre le mot bien connu *anticonstitutionnellement*, un exemple plus récent est donné par le mot *paraskevidekatriaphobie* (la peur du vendredi 13). Ce mot ne peut être prononcé qu'avec une ou deux syllabes accentuées en plus de la syllabe finale. On remarquera au passage que les accents supplémentaires, soit *paraskevi dekatrìa phobie* ou *para skevidekatriaphobie* sont alors positionnés aux frontières morphologiques de ces mots : *para*, *paraskevi* (« vendredi » en grec moderne) *dekatrìa* (« treize ») ou *trià* (« trois »).

En réalité, ce n'est pas le nombre magique 7 de syllabes qui est en cause ici, mais la durée d'énonciation du groupe de syllabes terminé par une syllabe accentuée. Ainsi, un débit rapide de *Le frère de Max* permet de n'accentuer que la syllabe finale, alors qu'un débit plus lent nécessitera deux syllabes accentuées *Le frère de Max*. C'est

pour cette raison qu'il est préférable d'utiliser le terme *groupe temporel* plutôt que *groupe accentuel* ou *groupe rythmique*, l'emploi de l'adjectif *temporel* trouvant sa justification dans le traitement neuronal dans la perception de la parole, dont il est question ci-dessous.

D'autre part, le terme classique *groupe de sens*, encore fréquemment utilisé dans des manuels de FLE, n'a en réalité pas de... sens. Pour s'en convaincre, il suffit de rapprocher des exemples tels que *la table*, *la petite table*, *la petite table ancienne* et *la petite table ancienne vert-bouteille*. Les trois premiers exemples pourront être prononcés avec une seule syllabe accentuée finale, alors qu'il en faudra deux dans le dernier (sauf à adopter un débit très rapide). Il s'agit pourtant dans les trois cas d'un seul « groupe de sens ».

2.2 Collision accentuelle / Collision temporelle

La contrainte de collision accentuelle (en anglais *stress clash*) est également connue depuis longtemps (Passy, 1887). Elle concerne l'évitement de deux syllabes accentuées successives (contigües dans l'analyse de la transcription écrite). Cet évitement se traduit d'après de nombreuses études théoriques a) par un recul du premier accent pris dans la collision sur la syllabe précédente (ou sa disparition s'il n'y a qu'une seule syllabe porteuse du premier accent), ou b) sur une syllabe normalement porteuse d'un accent secondaire, comme l'accent dit « intellectuel » sur la première syllabe des mots lexicaux, cf. Léon et Martin, 1980). Dès lors, en français, du fait de la position finale de l'accent dans les groupes temporels, cette configuration s'observe lorsque le deuxième groupe impliqué n'a qu'une seule syllabe. Ainsi, la prononciation de *soulier noir* entraîne un recul de l'accent final de *soulier* sur la première syllabe de ce mot : *soulier noir*, constituant ainsi un seul groupe temporel, pourvu d'un accent secondaire (ou emphatique) sur la première syllabe et d'un accent de groupe temporel sur la dernière (un « arc accentuel »).

En fait des contre exemples se trouvent facilement, et le déplacement ou le non déplacement du premier accent dans un cas de collision peut même différencier des constructions syntaxiques distinctes. Un exemple comme *chocolat chaud* est éclairant de ce point de vue. Lorsqu'il y a un recul d'accent, donc lorsque l'exemple est prononcé *chocolat chaud* ou *chocolat chaud* (réalisation d'un accent secondaire) comme dans *Max aime le chocolat chaud*, le déplacement d'accent indique le regroupement des séquences *chocolat* et *chaud* en un seul groupe temporel (effectivement terminé par une syllabe accentuée et porteur d'un accent secondaire sur *chocolat*), on peut suggérer que l'énoncé répond à une question du type *qu'est-ce que Max aime ?* Par contre, si le déplacement du premier accent n'est pas réalisé, *Max aime le chocolat # chaud*, peut

constituer une réponse à *comment est-ce que Max aime le chocolat ?* Notons que dans ce dernier cas une pause est introduite entre les deux syllabes accentuées successives. C'est également le cas dans des exemples tels que celui présenté Fig. 1.

Indépendamment d'une différenciation syntaxique, l'exemple *un soulier # noir* peut également être réalisé avec deux groupes temporels, impliquant nécessairement la présence d'une pause intercalée.

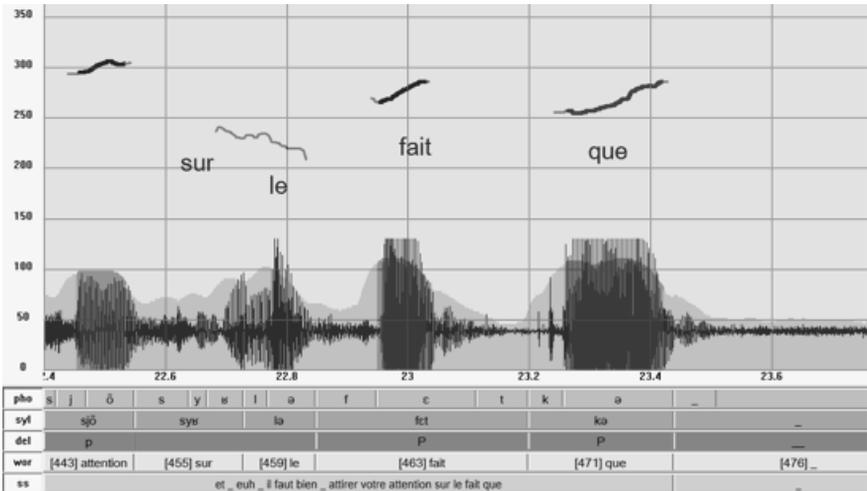


Fig. 1. *et euh il faut bien attirer votre attention sur le fait que la description que non qu'on en fait* (cnf-be, Corpus C-PROM). Les deux syllabes accentuées successives sont séparées d'environ 300 ms.

La contrainte de collision d'accent n'est donc pas une contrainte portant sur la présence de syllabes accentuées successives, mais bien sur la temporalité dans l'énonciation des deux syllabes en question : il faut (et il suffit) qu'une distance temporelle suffisante existe dans la prononciation des syllabes accentuées successives. Des cas soi-disant remarquables trouvés dans la parole spontanée (Avanzi et al. 2010) s'expliquent alors aisément, et au lieu de *collision accentuelle* on devrait parler de *collision temporelle*, accentuées. Mais il reste à trouver la raison de cette distance temporelle nécessaire dans l'énonciation de syllabes accentuées successives, même en l'absence de pause entre les deux syllabes. C'est l'objet de ce qui suit.

2.3 Alignement syntaxique

Outre la collision accentuelle, une autre contrainte limitant la combinatoire des structures prosodiques porte sur l'assemblage de catégories syntaxiques permises dans les groupes temporels, dont, rappelons-le, la durée d'énonciation est limitée. Cette contrainte, qui est la seule qui ait un lien avec le regroupement d'unités syntaxiques (Martin, 1987), stipule qu'un groupe temporel ne peut contenir des séquences syllabiques correspondant à des unités syntaxiques dominées directement par des nœuds distincts dans la structure syntaxique, donc au niveau le plus bas de la structure.

Plus simplement, on ne peut, par la position finale de la syllabe accentuée, avoir dans un même groupe temporel un nom suivi d'un déterminant, comme par exemple **adore les* ou **chocolat de chez* dans *Max adore les chocolats de chez Machin*, les groupes *adore les* et *chocolat de chez* contenant des mots grammaticaux se rapportant aux groupes syntaxiques suivants. S'il est facile de se convaincre de la validité de cette contrainte, qui trouve une explication syntaxique évidente (les déterminants *les* et *de* suivent des unités qu'ils ne déterminent pas), une explication rationnelle qui ne soit pas circulaire (i.e. qui ne renvoie directement pas à la syntaxe) serait bienvenue. On trouve aisément des contrexemples dans la parole spontanée, lors d'hésitations et de répétitions, comme dans l'exemple de la Fig. 2. La reformulation décidée par le locuteur après le déterminant *le* entraîne son accentuation et la formation du groupe temporel *que le*. Des exemples semblables s'entendent souvent dans la production orale journalistique au point de devenir un style (comme dans *nous verrons demain une hausse de la # température*, France 2, 1/7/14).

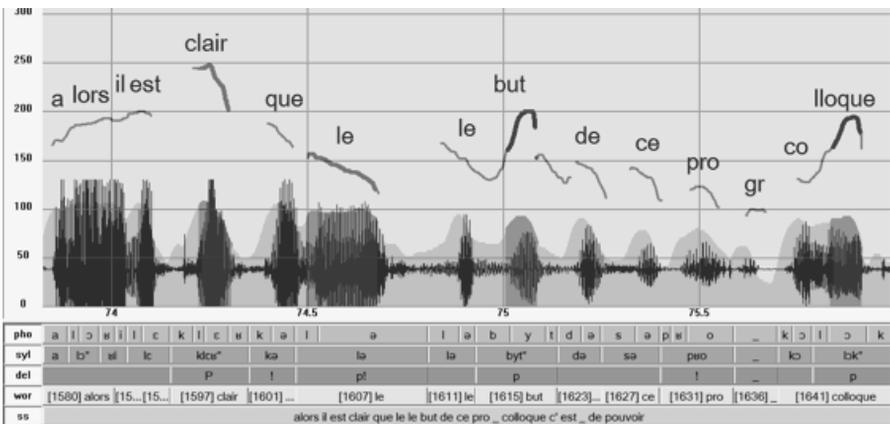


Fig. 2. *alors il est clair que le le but de ce pro _ colloque* (cnf-ch, C-PROM).

Redoublement du déterminant *le* impliquant l'accentuation de la première occurrence de *le* et reformulation (*le but*), formant ainsi le groupe temporel *que le*. L'accentuation du déterminant ainsi que la courte pause qui suit permet de vider la mémoire syllabique de l'auditeur et laisser la place pour le groupe temporel corrigé *le but*.

2.4 Eurythmie

Il s'agit ici aussi d'une contrainte temporelle qui a trait à la préférence observée aussi bien en parole lue qu'en parole spontanée. L'analyse des données montre en effet (Wioland, 1983, Martin, 1986) une tendance à l'eurythmie dans la réalisation des groupes temporels ; des groupes avec beaucoup de syllabes (*a fortiori* ceux comprenant 7 syllabes ou plus) sont naturellement énoncés avec un débit plus rapide que ceux comprenant peu de syllabes, au point que lorsqu'un groupe ne comprend qu'une syllabe, une pause est souvent introduite pour tendre à équilibrer les durées d'énonciation de chaque syntagme prosodique, c'est-à-dire de regroupements de groupes temporels.

Mais d'autres stratégies tendant à l'eurythmie sont possibles. L'énoncé *Max aime les chocolats suisses* peut être prononcé avec une pause après *Max*, [*Max #*] [*aime les chocolats suisses*] ou par le regroupement temporel de *Max* et *aime*, [*Max aime*] [*les chocolats suisses*]. La première solution rend la structure prosodique congruente avec la syntaxe, puisque le groupe nominal *Max* correspond à un groupe temporel, alors que le second réalise une structure prosodique eurythmique sans l'insertion d'une pause, au prix d'une structuration non congruente avec la syntaxe (quelques chercheurs d'orientation autosegmentaliste parlent de restructuration prosodique, révélant par là la primauté de la syntaxe (Post, 1999). Pourtant, une explication simple peut être trouvée par l'examen des processus neuronaux de la perception des séquences syllabiques par l'auditeur.

En parole spontanée, on observe le plus souvent un ajustement des durées syllabiques à l'intérieur d'un même groupe temporel selon le nombre de syllabes que ce groupe contient. La Fig. 3 montre le caractère linéaire de cet ajustement, avec des durées syllabiques moyennes allant de 250 ms pour des groupes d'une syllabe, à 100 ms pour des groupes de 7 syllabes.

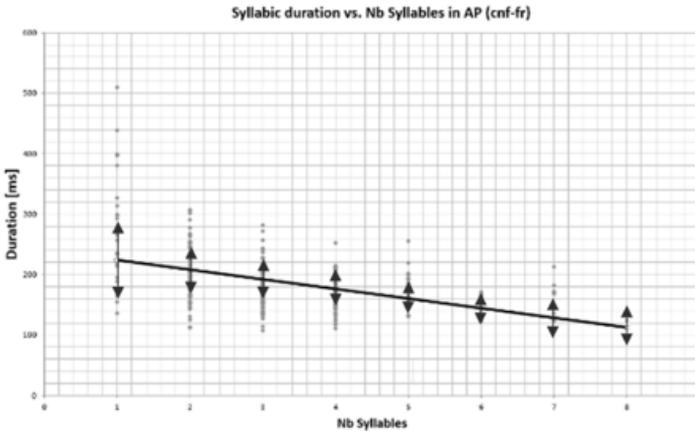


Fig. 3. Durée syllabique en fonction du nombre de syllabes dans un mot prosodique (C-PROM, *nar-fr*). L'écart type est indiqué par des doubles flèches

3. Origine des contraintes de structuration prosodique

3.1 Les ondes Thêta et Delta

On a observé depuis longtemps que le cerveau est constitué d'un nombre considérable de neurones (de l'ordre de 100 milliards) interconnectés par groupes dans des régions spécifiques de la masse cérébrale. On sait aussi que ces interconnexions permettent un transfert de l'information stockée chimiquement dans chaque neurone. Or ce transfert induit des variations de potentiel électrique faible (de l'ordre du μV) mais cependant observables par des capteurs positionnés sur le crâne des sujets auditeurs (électro encéphalographie, EEG). Ces variations sont appelées potentiels évoqués car résultant d'une stimulation sensorielle, auditive ou autre. L'activité électrique de transfert de groupes de neurones à d'autres groupes de neurones ne se fait pas de manière désordonnée. D'une part ils s'opèrent selon des gammes de fréquences spécifiques liées aux activités cognitives déterminées, et d'autre part ils sont synchronisés en phase dans chaque gamme de variation. Chaque gamme d'onde est désignée par une lettre grecque : ondes Alpha, Bêta, Delta, etc. Les gammes d'onde cérébrales qui nous intéressent ici sont les ondes Delta, variant de 1 à 4 Hz, et les ondes Thêta, variant de 4 à 10 Hz.

Ces expériences ont montré, et c'est un premier point, la préférence pour l'auditeur du traitement de la structure prosodique par rapport à celui de la syntaxe (Steinhauer & al. 1999). D'autres chercheurs (Gilbert & Boucher, 2007, Obrig & al. 2010) ont établi

que la segmentation du flux sonore syllabique se faisait à la fois grâce à des balises prosodiques et par identification directe d'unités déjà connues et mémorisées.

Deux résultats relativement récents en recherche par potentiels évoqués s'appliquent à la perception de la structure prosodique :

1. La synchronisation de la perception des syllabes par les ondes Thêta (Gilbert & Boucher, 2007, Giraud, 2013) ;
2. Le traitement neuronal totalement indépendant des frontières prosodiques (Boundary Tones, tons de frontière terminant les syntagmes prosodiques IP dans la terminologie autosegmentale-métrique) (Pannekamp & al. 2005).

3.2 L'hypothèse de synchronisation

La clé de l'explication pourrait résider dans le paramètre temps lié aux contraintes de la structuration prosodique. Ces différentes observations conduisent à formuler l'hypothèse qui suit. On sait que les ondes du cortex Delta et Thêta (entre autres) gouvernent le flux de transfert d'information d'ensembles neuronaux. Ces valeurs des gammes de fréquence suggèrent alors que les ondes Delta seraient responsables de la synchronisation du transfert des séquences syllabiques, synchronisées elles par les ondes Thêta (Martin, 2013). Ceci rendrait compte d'une part de l'étendue de variation des durées des groupes accentuels, variant de 250 ms à 1200 ms environ (gamme de variation des périodes des ondes Delta) d'une part, et d'autre part des plages de durées syllabiques, contraintes par la propriétés mécaniques des organes articulatoires, mais aussi par la gamme de variation des périodes des ondes Thêta, de 100 ms à 250 ms, ondes synchronisant la perception des syllabes par les auditeurs.

La synchronisation des ondes Thêta par les ondes Delta s'observe par exemple pour des stimuli composés de sons purs (Fig. 4 et 5). La présence des ondes Delta est donc indispensable pour le traitement des ondes Thêta dans la perception des syllabes (Gilbert, Boucher, Martin, 2014).

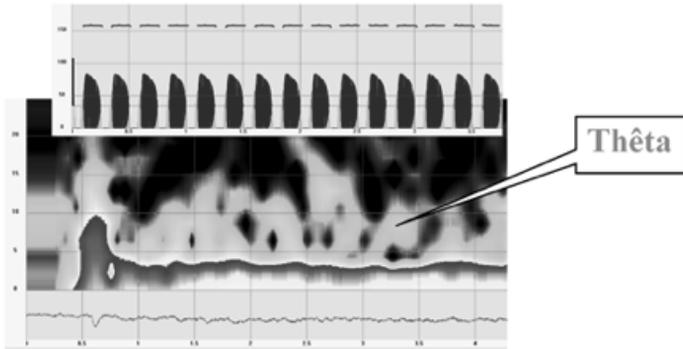


Fig. 4. Ondes cérébrales Thêta non structurées résultant de l'écoute d'une séquence de sons purs de même durée.

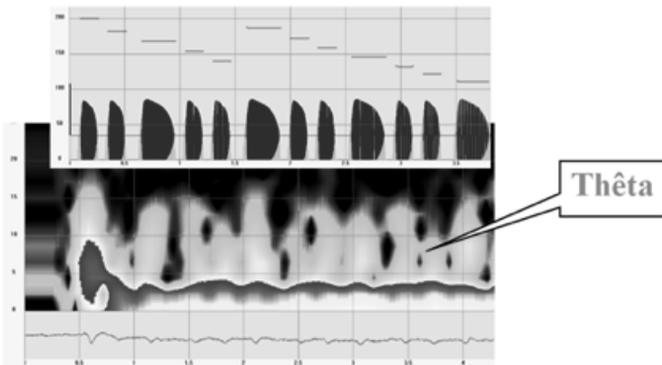


Fig. 5. Ondes cérébrales Thêta structurées par groupe de trois résultant de l'écoute d'une séquence de groupes de trois sons purs dont le dernier présente une durée plus grande.

Pour valider cette hypothèse, il faudrait donc pouvoir observer une activité par potentiel évoqué qui résulterait de la perception des syllabes accentuées dans les séquences de groupes temporels. Les expériences de Pannekamp & al. (2005) n'ont pas permis de valider ce point, mais les travaux récents de Giraud et al. (2013) vont tout à fait dans ce sens.

De plus, notre hypothèse permet de proposer une explication pour chacune des contraintes limitant la combinatoire des structures prosodiques :

La collision temporelle, présence de deux syllabes accentuées successives, est en fait permise pourvu qu'il existe un espace de temps suffisant entre les deux syllabes

successives impliquées dans la collision. Cet espace correspondrait à la période minimale des ondes Delta, de l'ordre de 250 ms. Une valeur plus petite ne permettrait pas la synchronisation des périodes d'onde Thêta par les ondes Delta, bien que la durée de la syllabe puisse être inférieure à 250 ms (Martin, 2014) ;

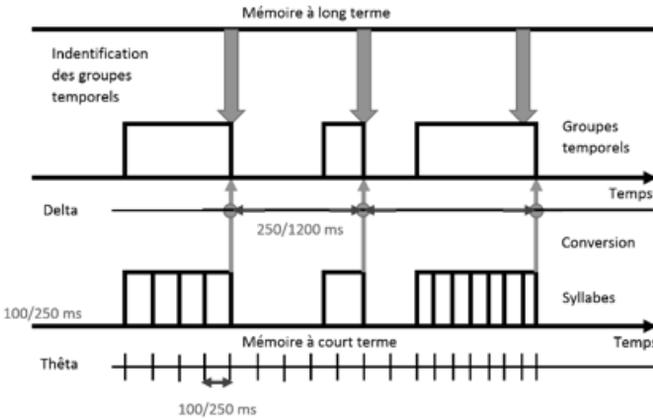


Fig. 6. Modèle de conversion série-parallel des séquences dans la mémoire syllabique à court terme, montrant les durées minimales (250 ms) et maximales (1200 ms) des groupes temporels

La règle des 7 syllabes, qui résulte d'une contrainte sur la durée maximale d'une séquence syllabique ne contenant qu'une seule syllabe accentuée, s'expliquerait par la durée maximale d'une période des ondes Delta, de l'ordre de 1200 ms. Une durée plus longue ne permettrait plus cette synchronisation (Martin, 2014) ;

La contrainte de collision syntaxique rend compte de la durée d'identification par l'auditeur de séquences syllabiques erronées ou inconnues. Le non-respect de cette contrainte entraîne la nécessité de révision du découpage initial réalisé par l'auditeur, révision toujours possible mais coûteuse en temps (cf. ci-dessous). C'est pour la même raison que les reprises de groupes accentués faites après une interruption ou un abandon se font toujours avec l'énonciation d'un groupe accentuel complet (cf. l'exemple de la Fig. 2) ;

L'eurythmie, préférence pour que les structures prosodiques soient équilibrées soit en nombre de syllabes à chaque niveau de la structure, soit à une modulation du débit syllabique de syntagmes prosodiques avec des nombres de syllabes très différents, s'expliquerait par la difficulté de modifier les périodes successives des ondes de synchronisation Delta par des valeurs extrêmes, par exemple de 250 ms à 1200 ms environ pour la période suivante. L'analyse de données de parole spontanée révèle que

la compression des durées syllabiques pour les groupes temporels à grand nombre de syllabes est préférée par la plupart des locuteurs (Martin, 2014).

La non-congruence entraîne de la part de l'auditeur une ré-analyse des regroupements définis par la structure prosodique. De nombreuses expériences montrent que ces « incongruïtés » prosodiques se traduisent par des potentiels évoqués retardés par rapport aux stimuli (Steinhauer & al. 1999), c'est-à-dire aux séquences de groupes temporels. On parle alors de phénomènes P600, variation de potentiel positive survenant 600 ms après le stimulus, ou de N400, variation négative après 400 ms).

Ceci montre bien l'importance de la structure prosodique dans le décodage du flux syllabique. S'il y a congruence avec la structure syntaxique, il n'y a pas ou peu de retard de traitement de l'information syntaxique. Le même phénomène s'observe pour des incongruïtés syntaxiques ou sémantiques (du type *Max aime maximum* ou *Max aime le café avec imaginaire*).

4. Codage et décodage dynamique de la structure prosodique

L'accentuation sur les syllabes finales des groupes temporels ne sert pas seulement à délimiter ces groupes et à déclencher le transfert des séquences syllabiques d'une mémoire sérielle à court terme vers un traitement d'unités linguistiques de rang supérieur. Si on observe leurs réalisations phonétiques manifestées par des paramètres acoustiques tels que hauteur mélodique, durée, intensité, ces variations constituant la mélodie de l'énoncé, on constate que ces accents présentent des différences remarquables, notamment par leurs mouvements mélodiques. Le modèle dit de « stockage-concaténation » et de « structure prosodique incrémentale » (Martin, 2009) permet de rendre compte de l'utilisation de ces différences accentuelles dans le mécanisme de reconstruction de la structure prosodique par l'auditeur.

Brièvement résumé, le processus décrit par ce modèle présuppose que l'auditeur ait acquis une connaissance des classes d'évènements prosodiques instanciés sur les syllabes accentuées. Ainsi la syllabe finale de l'énoncé (en dehors des cas de postnoyaux et de complément rapporté analysés en macrosyntaxe) est reconnue comme appartenant à une classe corrélatrice de fin d'énoncé, et est manifestée par des traits acoustiques bien connus (variation mélodique descendante et basse pour une modalité déclarative de l'énoncé, montante pour une modalité interrogative, etc.). Appelons cette classe C0.

Deux instanciations successives de C0 définissent un domaine prosodique, dans lequel peuvent s'insérer des évènements prosodiques de rang inférieur, de classe C1. Deux instanciations successives de C1 définissent alors un domaine prosodique plus petit, dans lequel peuvent s'insérer des évènements prosodiques de rang inférieur, cette fois

de classe C2. Et ainsi de suite, quoique en pratique des réalisations de classe C3 ou au-delà sont rares et s’observent seulement en lecture ou dans la parole professionnelle (de politiciens par exemple).

L’exemple simple qui suit (corpus C-PROM) permet d’illustrer ce processus (les symboles de contours sont alignés sur les syllabes accentuées). Avec C0 bas long et descendant, C1 montant, C2 descendant ou plat, Ci contour montant d’accent secondaire (emphatique). Les différentes phases d’assemblage des groupes syllabiques est alors le suivant :

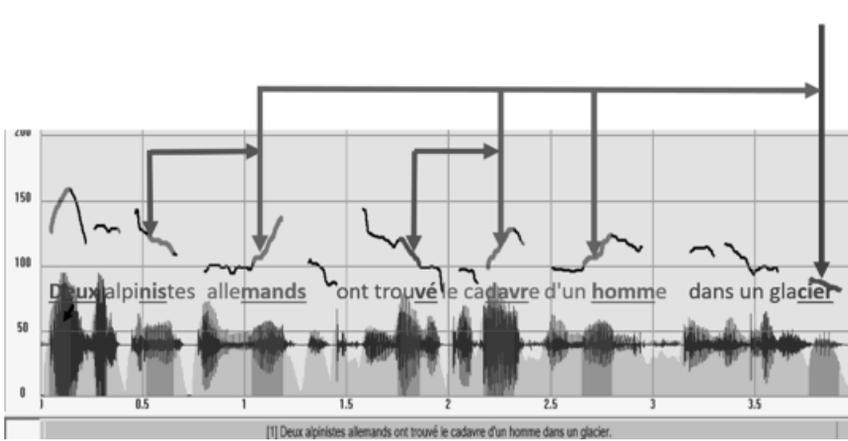


Fig. 7. Mécanisme d’assemblage de la structure prosodique à partir du décodage des classes de contours mélodiques successifs alignés sur les syllabes accentuées (Corpus Eurom 4). Les flèches présentes indiquent les relations de dépendance « à droite » de chacun des contours, c’est-à-dire envers les contours survenant après eux.

Le mécanisme d’assemblage de la structure prosodique de l’exemple est alors le suivant : la première séquence de groupes temporels terminée par un contour de classe C1 est mise en mémoire, en attendant d’être, lors de l’apparition de C0, concaténée avec les autres séquences terminées par C1. Les séquences terminées par C2 sont ensuite stockées avant d’être concaténée avec celles terminées par C1. La séquence résultante est mise en mémoire en parallèle avec celles terminées par C1. Ce mécanisme se répète jusqu’à l’apparition du contour de classe C0. Le processus restitue donc au fur et à mesure dans son déroulement temporel la hiérarchie prosodique censée aider l’auditeur à reconstituer partiellement ou totalement la structure syntaxique.

Il est intéressant de remarquer que les réalisations des contours d’une même classe, par exemple C2, ne doivent être semblables qu’à l’intérieur de chaque domaine, l’important est qu’ils soient identifiés par l’auditeur comme appartenant à une même

classe. Ainsi dans l'exemple précédent, C2 dans le domaine *Deux alpinistes allemands* pourra être réalisé différemment du C2 dans le domaine *ont trouvé le cadavre*, les réalisations phonétiques ne devant être semblables qu'à l'intérieur d'un même domaine.

Conclusion

Loin d'être marginaux comme les linguistiques structurale et plus tard générative transformationnelle l'ont prétendu, les faits prosodiques comptent parmi les plus importants dans le codage et le décodage de la parole. De fait, ce sont les événements prosodiques indicateurs de la structure prosodique qui apportent les toutes premières informations à l'auditeur, ils sont donc loin de constituer des béquilles prosodiques qui viendraient au secours de structures syntaxiques localement défailtantes. Ce sont eux qui sont traités en premier et qui vont permettre de faire une première ébauche de la reconstitution de la structure syntaxique de l'énoncé, quitte, en cas de non congruence prosodie-syntaxe, à opérer la restructuration nécessaire, mais au prix de temps de traitement cognitif plus long.

C'est pourquoi cette congruence joue un rôle important pour des styles d'énonciation avantageant la compréhension du destinataire des énoncés oralisés. Elle doit être privilégiée lorsque les conditions de production du discours sont défavorables ou difficiles (bruit important, public nombreux et varié, etc.). Une réalisation correcte des structures prosodiques doit également et peut-être en premier lieu faire partie de l'acquisition d'une langue seconde, en permettant de suppléer si nécessaire à des réalisations imparfaites des autres éléments du discours, prononciation des syllabes, approximations morphologiques ou syntaxiques incluses.

Bibliographie

- Avanzi, M., Lacheret, A., Obin, N., Victorri, B. 2010. « Vers une modélisation continue de la structure prosodique : le cas des proéminences syllabiques ». *French Language Studies*, 21 (2011) 53-71.
- Friederici, A., Wartenburger, I. 2010, Language and brain, *Cognitive Science*, (10) 150-159.
- Gilbert, A., Boucher, V. 2007, What do listeners attend to in hearing prosodic structures? Investigating the human speech-parser using short-term recall, *Proc. Interspeech 2007* : 430-433.
- Giraud, A-L, Poeppel, D. 2012. *Cortical oscillations and speech processing: emerging computational principles*. Nature neuroscience E-pub, doi: 10.1038/nn.3063.
- Giraud, A-L. 2013. Are cortical oscillations a useful ingredient of speech perception? *Proc. Interspeech 2013*, to appear.
- Isel, F. Alter, K., Friederici, A. D. 2005. Influence of prosodic information on the processing of split particles: ERP evidence from spoken German, *Journal of Cognitive Neuroscience*, 17(1) 154-167.
- Léon, P., Martin, P. 1980. Des accents, *The Melody of Language*, Linda R. Waugh & Dwight Bolinger ed., University Park Press.

- Léon, P. 2005. *Précis de phonostylistique - Parole et expressivité*. Paris : Armand Colin.
- Martin, P. 1975. Analyse phonologique de la phrase française. *Linguistics*, (146) 3568.
- Martin, P. 1986. Structures prosodiques et structures rythmiques, Actes des 13ème JEP, Aix-en-Provence.
- Martin, P. 1987. Prosodic and Rhythmic Structures in French. *Linguistics*, (25-5) 925-949.
- Martin, P. 2009. *Intonation du français*, Paris : Armand Colin.
- Martin, P. 2013. Contraintes phonologiques de l'intonation de la phrase réinterprétées à la lumière des recherches récentes en neurophysiologie, *La Linguistique*, 2013/1, M. Mahmoudian éd.
- Martin, P. 2014. Spontaneous speech corpus validation of prosodic constraints, *Proc. Speech Prosody 2014*, Dublin.
- Meigret, L. 1550. *Le tretté de la grammaire française*, 1550, (réédité par Slatkine, Genève, et disponible sur *Gallica* <http://gallica.bnf.fr/ark:/12148/bpt6k507854/f1.image>) [Consulté le 01-07-2014].
- Miller, George A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *Trends in Cognitive Sciences* 7, (3) 141-4.
- Obrig, H., Rossi, S., Telkemeyer, S., Watenburger, I. 2010. From acoustic segmentation to language processing: evidence from optical imaging, *Front. Neuroenerg*, 2:13.
- Pannekamp, A., Toepel, U., Alter, K., Hahne, A., Friederici, A. D. 2005, Prosody-driven sentence processing: an event-related brain potential study. *J Cogn Neurosci* 2005, 17(3) 407-421.
- Passy, P. 1887. *Les sons du français : leur formation, leur combinaison, leur représentation*, Paris : Firmin-Didot.
- Post, Brechtje, 1999, Restructured phonological phrases in French: Evidence from clash resolution, *Linguistics*, 37 (1) 41-63.
- Selkirk, L. 2000. The Interaction of Constraints on Prosodic Phrasing. In: *Prosody: Theory and Experiment, Text, Speech and Language Technology, Volume 14*, p. 231-261.
- Steinhauer, Karsten, Alter, Kai & Friedrici & Angela D., 1999, Brain potentials indicate immediate use of prosodic cues in natural speech processing, *Nature Neuroscience*, 2(2) 191-196.
- Wioland, F. 1983, *La rythmique du français parlé*, Collection Foreign Language Study, Université de Strasbourg.
- C-PROM (2010) Corpus libre de parole multigenre, <https://sites.google.com/site/corpusprom/> [consulté le 15 mars 2014].