

# STRUCTURE PROSODIQUE ET DICTION POÉTIQUE

---

*Philippe* MARTIN

**ABSTRACT • Prosodic structure and poetic diction.** Speech styles are particularly characterized by prosodic features, such as rhythm, melody, intensity or vowel quality, which indicate to the listener emotions, attitudes, sociogeographic origin of the speaker. Whatever the speech style, there are some unsuspected rules that are always applied, linked to specific cognitive properties of the brain. Among these, the necessity to segment speech units in chunks called stress groups and assemble these groups into a prosodic structure. Both stress groups duration and their hierarchical organization into a structure are constrained by theta and delta brain waves timing properties. Examples of these limits are illustrated in acoustic analysis of poetic diction showing that whatever the style adopted by readers, cognitive constraints are always observed.

**KEYWORDS •** Intonation; Prosodic Structure; Stress Group; Brain Waves

## 1. Introduction

Il existe différentes façons de marcher : en mettant un pied devant l'autre, mais aussi en crabe, un pied à côté de l'autre, à reculons, un pied en arrière de l'autre, en dansant, etc. Observé dans le détail, chacun des mouvements impliqués dans le déplacement du corps est différent. Un observateur attentif peut facilement identifier l'origine sinon géographique du moins culturelle de chaque marcheur, même lors d'une marche régulière. Ainsi les marcheurs anglais se démarquent des marcheurs français qui eux-mêmes se distinguent des italiens, ce que les enfants grecs apprennent à deviner par jeu sur les plages, en observant les estivants ne pouvant difficilement se distinguer par leur seul maillot de bain.

Un seul but est cependant commun à tous les marcheurs : se déplacer, et une seule contrainte : ne pas tomber. Il en va de même pour la parole : un seul but est commun à tous les locuteurs, transmettre ou demander une information. Pour autant, de nombreuses variations sont possibles, liées à l'origine sociogéographique affichée ou simulée de chacun, à son émotion, son attitude. Mais quelles que soient ces variations, il existe une contrainte connue de tous, celle d'arrêter de parler de temps en temps pour pouvoir respirer. Cette contrainte détermine la durée des groupes de souffle, segments temporels pendant lesquels la phonation est possible, à l'inverse de la phase d'inspiration du cycle de respiration, pendant laquelle parler est impossible. Dès lors, pendant que nous parlons, nous devons optimiser la durée d'expiration du cycle respiratoire, et minimiser la durée d'inspiration qui ne produit que silence dans la parole continue.

Pour autant, il existe une autre contrainte, beaucoup moins évidente mais pourtant ressentie par chacun, une contrainte directement liée à nos capacités cognitives. Cette contrainte découle des limites de notre mémoire auditive à court terme, et à l'obligation que nous avons de segmenter la parole que nous entendons, ou les séquences de mots que nous lisons, en groupes de mots. C'est cette contrainte qui donne rend compte de la constitution de groupes accentuels,

---

des groupes de mots contenant une seule syllabe accentuée, et à la structure prosodique qui structure ces groupes accentuels en plusieurs niveaux.

## **2. Groupes accentuels et structure prosodique**

### **2.1. La voix intérieure**

En effet, on ne lit pas mot à mot (sauf lorsqu'on apprend à lire une langue), mais par groupes de mots. Ceci est aussi vrai en lecture orale, avec la voix, qu'en lecture silencieuse, intérieure, et c'est aussi vrai lorsqu'on monologue oralement ou silencieusement en parlant à soi-même. En réalité, même si on n'en a pas toujours conscience, ni la lecture d'un texte ni un monologue intérieur ne peut se faire sans sous-vocalisation, c'est-à-dire sans restituer par une voix intérieure une voix, la nôtre ou une autre voix connue ou inconnue, qui nous parlerait « dans notre tête ». Les rares exceptions à ce processus concernent des représentations ou des mots courts et fréquents, comme par exemple « 1789 » pour un lecteur français, ou un panneau « STOP » à l'intersection de deux rues. Ces cas relativement peu fréquents mis à part, il ne semble pas possible de lire ou même de réfléchir à un problème ou à une question sans passer par la sous-vocalisation, donc en entendant et en écoutant une voix intérieure.

### **2.2. La mémoire à court terme**

Lorsqu'on écoute quelqu'un parler, on peut se remémorer ses paroles pendant 20 à 30 secondes, mais seulement si cette parole est suivie d'un silence. Par contre, si le locuteur poursuit son discours, et si on continue à l'écouter, on ne dispose que de 2 à 3 secondes pour se remémorer les dernières paroles prononcées (Acheson & MacDonald, 2009). Dans ce dernier cas en effet, notre mémoire auditive à court-terme doit laisser la place à la suite des paroles prononcées par le locuteur que nous écoutons. Pendant ces 2 ou 3 secondes, il nous faut donc décoder la parole entendue, c'est-à-dire identifier les syllabes successives constituant des mots de manière à avoir accès au sens, tout en poursuivant notre écoute pour la suite.

Pour assurer ce décodage dans ce laps de temps réduit, il nous faut donc segmenter le flot de parole en unités temporelles de durée inférieure aux 2 ou 3 secondes fatidiques, faute de quoi le message sera perdu. Pour nous seconder dans cette opération de segmentation, nous disposons dans le système de la langue d'une balise de nature acoustique, qui nous signale régulièrement qu'une segmentation doit être faite et démarrer ainsi l'opération d'identification des unités linguistiques. Ce sont les traits d'accentuation, faisant apparaître certaines syllabes comme particulières par rapport à certaines autres, qui instancient cette balise nécessaire pour avoir accès au sens des phrases que nous entendons.

Dans les langues dites à accent lexical, comme l'italien ou l'anglais, les syllabes ainsi « surlignées », i.e. accentuées, sont réalisées sur des mots appartenant à la classe ouverte dite de « contenu », par opposition à la classe fermée dite « grammaticale ». La classe de contenu contient les noms, adjectifs, verbes et adverbes, alors que la classe grammaticale rassemble les conjonctions, prépositions et pronoms. La classe du contenu est dite ouverte parce qu'on peut toujours y ajouter des mots au cours de l'évolution de la langue, ce qui est très rarement le cas pour la classe fermée.

Dans les langues dites à accent rythmique, dépourvues d'accent lexical comme le français ou le coréen, la réalisation d'une syllabe accentuée est tout aussi nécessaire, mais n'est pas obligatoirement déterminée par la classe des mots, mais seulement par une contrainte

temporelle. Cette contrainte limite la durée séparant des syllabes accentuées successives à ne pas être inférieure à 250 ms ni d'être supérieure à quelque 1250 ms, donc à avoir une valeur située entre un quart de seconde et une seconde et un quart à peu près.

La réalisation des syllabes accentuées présente un caractère obligatoire dans le sens où on ne peut prononcer une suite de mots sans accentuer une syllabe au minimum toutes les secondes et un quart, même en lecture silencieuse, et ce indépendamment de la catégorie des mots prononcés. D'autres réalisations de syllabes accentuées sont possibles dans la production de parole, par exemple caractérisant une insistance ou une emphase mise sur certains mots, mais leur réalisation n'a rien d'obligatoire et résulte d'un choix du locuteur. Ce type d'accentuation est généralement qualifié d'emphatique ou d'insistance, par opposition aux accents rythmiques ou lexicaux évoqués plus haut.

### 2.3. Les groupes accentuels

Les balises temporelles réalisées par l'accentuation de certaines syllabes guident l'auditeur dans la segmentation du flot de parole (d'où la terminologie ancienne d'accent démarcatif). Pour les langues sans accent lexical, ces balises accentuelles sont toujours placées sur la syllabe finale d'un mot a priori quelconque. On définit alors le groupe accentuel comme une séquence de mots dont la dernière syllabe est accentuée. Ainsi, *la banlieue de Paris* constitue un groupe accentuel si la dernière syllabe de *Paris* est accentuée et que la syllabe finale de *banlieue* ne l'est pas. Par contre, si cette dernière est accentuée, deux groupes accentuels seront formés : *la banlieue de Paris*. On note facilement que la réalisation de l'une ou l'autre de ces solutions ne dépend que de la vitesse de lecture, une lecture ou une élocution rapide, qu'elle soit orale ou silencieuse, permettant d'insérer plus de syllabes, donc plus de mots dans un seul groupe accentuel qu'une lecture ou une élocution lente. On peut aussi en déduire, sachant que l'intervalle maximum entre deux syllabes accentuées successives est de l'ordre de 1250 ms, et que la durée minimale d'une syllabe est de 100 ms, qu'un groupe accentuel ne peut pas contenir plus de 11 syllabes (en comptant la dernière comme plus longue, par exemple de 250 ms).

Pour les langues à accent lexical par contre, cette syllabe accentuée n'est pas nécessairement en position finale, et est souvent déterminée par des règles de composition morphologiques des mots de contenu, en se plaçant souvent à une limite entre racine, suffixe et flexion. Ainsi dans le mot italien *repubblica*, la syllabe accentuée antépénultième se trouve à la limite de la racine *pub* accentuable et du suffixe inaccentuable *-ic*. Dans ce cas, le groupe accentuel est défini comme une séquence de mots contenant un seul mot de contenu (i.e. verbe, adverbe, adjectif ou nom), auquel s'adjoignent éventuellement les mots grammaticaux qui en dépendent, par exemple *la repubblica*.

La limite de 2 ou 3 secondes pour effectuer l'identification des groupes accentuels successifs implique que l'identification de leur contenu ne se fait pas mot à mot, c'est-à-dire mot orthographique par mot orthographique, mais bien groupe accentuel par groupe accentuel, c'est-à-dire des groupes de mots contenant une seule syllabe accentuée (non emphatique). L'auditeur n'a littéralement pas le temps cognitif de réaliser l'identification dans sa mémoire de chaque mot lexical ou grammatical contenu dans chaque groupe. On est donc conduit à considérer que le lexique mémorisé par l'auditeur n'est pas organisé comme un dictionnaire, quand bien même il contiendrait toutes les formes fléchies de chaque mot lexical, mais par groupes de mots, contenant les mots lexicaux et les mots grammaticaux qui en dépendent. C'est alors l'unique syllabe accentuée du groupe (donc en l'absence d'accent emphatique) qui va synchroniser comme balise l'identification de chaque groupe accentuel, non pas dans un intervalle de temps de 2 ou 3 secondes, mais dans celui qui sépare deux syllabes accentuées

successives, soit au minimum 250 ms et au maximum 1250 ms, c'est-à-dire dans une limite confortable inférieure aux 2 ou 3 secondes de mémorisation maximales. Ces valeurs limites temporelles ont été obtenues expérimentalement par l'analyse acoustique de nombreuses données de parole lue et spontanée (Martin, 2014).

#### **2.4. La structure prosodique**

Les limites de la mémoire auditive à court terme imposent non seulement une segmentation du flot de parole en groupes accentuels, mais aussi la quasi obligation de structurer les séquences de ces groupes, c'est-à-dire de les organiser, de les rassembler en plusieurs niveaux. En effet, sans cette structuration, les groupes accentuels successifs apparaîtraient successivement en formant une chaîne, sans relation les uns avec les autres, comme une longue énumération, dont il serait souvent difficile d'en extraire le sens. Certains procédés dits de lecture rapide procèdent de cette façon, en présentant au lecteur une suite de mots clés extraits du texte et censés être importants (Nowak, 2012). La durée de sous vocalisation est ainsi réduite, voire annulée pour certains mots, mais le sens résulte alors d'une longue suite de mots dont le lecteur ne peut, ou très difficilement, reconstituer les relations syntaxiques ou sémantiques qui les lient entre elles.

C'est aussi ce qui se passe lorsqu'on écoute un locuteur d'une oreille distraite ou distante, sans bien percevoir la structure des phrases énoncées. On alors pourrait penser que la structure des énoncés entendus se trouve restituée par l'auditeur grâce à une analyse syntaxique effectuée au fil du temps, groupe accentuel après groupe accentuel. Cette analyse devrait se faire à partir des marques grammaticales éventuellement présentes dans chaque groupe accentuel, et ce en moins de 2 ou 3 secondes, cette durée de rétention de notre mémoire auditive à court terme, laps de temps duquel il faut déduire la durée d'identification de chaque groupe accentuel dans le lexique.

Or, c'est un procédé bien plus efficace qui est mis à la disposition de l'auditeur, par la réalisation de la structure prosodique par le locuteur. Cette structure indique le regroupement en plusieurs niveaux (de l'ordre de 2 ou 3) des groupes accentuels, et s'opère au fil du temps, c'est-à-dire par des regroupements réalisés et révisés à chaque survenance d'un groupe accentuel au cours de l'écoute de la phrase. Ce sont alors les caractéristiques particulières des syllabes accentuées, modulables en termes de durée, de variation mélodique et d'intensité, qui vont permettre le codage nécessaire à la (re)construction dynamique, incrémentale, de la structure prosodique de la phrase voulue et codée par le locuteur. Ainsi, l'auditeur dispose d'un mécanisme bien plus efficace que l'analyse syntaxique qui serait effectuée à partir des mots contenus dans chaque groupe accentuel, et en particulier les mots grammaticaux. Ce mécanisme est plus efficace surtout parce qu'il est plus simple, puisqu'il se contente d'indiquer les relations de dépendance existantes entre les groupes accentuels, relations nécessaires et suffisante pour assurer l'indication de la structure prosodique, et qui de plus sont bien adaptée à un traitement au fil du temps. Il est aussi plus efficace car se contentant d'un codage par variation prosodique à l'endroit des syllabes accentuées (ou finales pour les langues à accent lexical), codage au nombre de symboles très restreint.

#### **2.5. Lecture et mouvements oculaires**

Une expérience facile à réaliser pour établir la durée minimale de traitement d'un groupe accentuel consiste à lire silencieusement un texte quelconque, par exemple d'environ 200 mots. Bien que le nombre de groupes accentuels en français dépende de la vitesse de lecture, on peut

en estimer le nombre maximal en comptant le nombre de mots de contenu. Un lecteur rapide aura tendance à regrouper plusieurs mots du contenu en un seul groupe, mais ce comptage nous donne un ordre de grandeur du nombre de groupes accentuels. On chronomètre ensuite le temps qu'il faut pour lire ce texte silencieusement, le plus rapidement possible. Quelle que soit cette vitesse de lecture, pour autant qu'on lise effectivement le texte sans négliger aucun mot, et qu'on l'entende en sous vocalisation « dans sa tête », on ne pourra pas dépasser un quart de seconde par groupe accentuel. Un exemple de segmentation en groupes accentuels est donné ci-dessous.

D'après François Richaudeau / la subvocalisation / est inutile / et ralentit beaucoup / la lecture / sans améliorer / la compréhension / En effet / d'après lui / le cerveau / n'a pas besoin / d'entendre le mot / pour en comprendre / le sens / Cependant / aucune expérience / n'a pu prouver / que supprimer / la subvocalisation / avait le moindre effet / sur la vitesse de lecture / supprimer / celle-ci / n'a donc pas d'effet / observable / Dans certaines expériences / le fait de subvocaliser / permet d'améliorer / la compréhension / du texte lu / La subvocalisation / a aussi / un effet facilitateur / sur la mémorisation / De nombreuses expériences / portant sur la boucle phonologique / et notamment / sur la suppression articulaire / l'ont montré /

Or la lecture d'un texte s'opère par saccades oculaires, passant d'un point de fixation au suivant en quelque 70 à 80 ms (Quercia, 2010). L'empan de focalisation ne dépasse en général pas 10 caractères en arrière (à gauche du point de fixation), et 20 à 30 caractères vers l'avant (à droite du point de fixation). On constate donc que la vitesse de lecture, opérant de 70 à 80 ms de mouvement oculaire d'un groupe accentuel au suivant, est ralentie par la vitesse de sous vocalisation, qui demande au moins 250 ms pour passer d'un groupe accentuel au suivant. Le ralentissement apparaît donc dû à la segmentation en groupes accentuels et à la reconstruction d'une structure prosodique. Le même test effectué sur une langue à accent lexical comme l'anglais ou l'italien peut se révéler encore plus convaincant, puisque le nombre de groupes accentuels ne varie pas avec la vitesse de lecture, chaque mot lexical devant (en principe) être accentué.

## **2.6. Les ondes cérébrales**

Pour savoir ce qui ralentit la lecture, il faut se tourner vers les ondes cérébrales, et en particulier les oscillations thêta et delta. On sait que le cerveau comporte des milliards de neurones, regroupés fonctionnellement en zones, et que ces zones communiquent par impulsions électriques entre elles. Pour différencier ces ondes cérébrales entre elles, on les désigne par leurs fréquences d'oscillations, par exemple de 4 Hz à 10 Hz pour les ondes thêta, et de 0,8 Hz à 4 Hz pour les ondes delta. Ces fréquences différentes reflètent des échanges entre des zones distinctes, plus ou moins facilement localisables expérimentalement.

Pendant toute notre vie, ces ondes oscillent constamment, même pendant notre sommeil, mais peuvent être synchronisées par des stimuli extérieurs à nous, tel le son de la parole par exemple. Or, on a montré il y a quelques années que les ondes thêta sont synchronisées, c'est-à-dire que leur fréquence d'oscillation s'aligne sur celles des syllabes de la phrase perçue par l'auditeur. En considérant non pas les fréquences des ondes thêta de 4 Hz à 10 Hz, traditionnellement données comme première caractéristique descriptive, mais plutôt les périodes correspondantes, c'est-à-dire de 100 ms à 250 ms, on trouve une analogie troublante, mais maintenant démontrée, avec la gamme de variation de durée des syllabes, soit également de 100

ms à 250 ms. Une durée de 100 ms constitue effectivement la limite inférieure, en deçà de laquelle une syllabe ne peut plus être identifiée par l'auditeur (Ghitza, 2011).

Il s'avère qu'une même correspondance peut être établie entre les variations de durée des groupes accentuels, ou plus exactement entre les variations de durée entre deux syllabes accentuées successives (en parole continue). En effet, comme on l'a vu, cette gamme s'étend de 250 ms (durée minimale) à 1250 ms (durée maximale), ce qui correspond à la gamme de variation des oscillations delta, une fois les valeurs de fréquence converties en valeurs temporelles, soit de 250 ms (4 Hz) à 1250 ms (0,8 Hz). De là est apparue l'hypothèse que les ondes delta s'alignent sur les syllabes accentuées. De là aussi l'explication voulant que la lecture, orale ou silencieuse, est ralentie essentiellement par la synchronisation nécessaire avec les oscillations delta (Martin, 2018).

Il en découle un modèle de la perception de la parole, dans lequel les oscillations thêta plus rapides, synchronisées par les sommets d'intensité syllabiques (donc les voyelles), optimisent la perception des syllabes par le transfert de l'information acoustique vers l'identification des groupes accentuels, identification synchronisée par les ondes delta. Ceci rend compte de l'impossibilité de percevoir comme accentuée la première de deux syllabes pourvues toutes deux des paramètres acoustiques de l'accent si elles sont séparées de moins de 250 ms (Martin, 2018). De même, ce modèle rend aussi compte de la nécessité d'accentuer au moins une syllabe toutes les 1250 ms (environ) dans la parole, sans quoi la première partie d'une telle séquence de pourrait pas être traitée, c'est-à-dire sans que l'identification du groupe puisse se faire dans le lexique de l'auditeur. Du reste, si une telle syllabe accentuée est absente dans des longues séquences de mots, par exemple par synthèse texte parole, les auditeurs auront une forte tendance à percevoir dans cet intervalle trop grand entre syllabes accentuées pourtant aucunement marquées acoustiquement (Martin, 2018).

## 2.7. Contours mélodiques

Les syllabes accentuées terminant les groupes prosodiques sont porteuses, en français, mais aussi dans les autres langues romances, de contours mélodiques divers, montants, descendants ou plutôt plats. Certains de ces contours présentent des changements mélodiques supérieurs au seuil de glissando, c'est-à-dire que leur variation mélodique est perçue comme telle, alors que d'autres varient en dessous du seuil de glissando, de sorte que leur variation mélodique est perçue comme un ton statique. La valeur du glissando peut être estimée par la formule  $(st_2 - st_1) / (t_2 - t_1)$ , c'est-à-dire la différence de fréquence en demi-tons divisée par la durée du contour, le seuil valant coef  $/(t_2 - t_1)^2$ . Le coefficient coef varie de 0.16 à 0.32, et constitue une valeur d'ajustement espérant tenir compte des approximations relatives à cette formule par rapport à la perception réelle (Rossi, 1971).

On désigne alors par C1↗ un contour montant supérieur au seuil de glissando, C2↘ un contour descendant supérieur au seuil de glissando, Cn→ un contour montant, descendant ou plat mais inférieur au seuil de glissando, et finalement C0↓ le contour terminal conclusif déclaratif terminant la phrase, atteignant le niveau mélodique le plus bas et descendant. C0 peut ne pas être en position finale dans les cas décrits par la macrosyntaxe de *propos-thème* et de *complément différé* (Bally, 1944), cas qui ne seront pas décrits ici. Le contour terminal conclusif interrogatif, noté Ci↑, est montant et atteint la fréquence la plus haute de la mélodie de la phrase.

Ces contours mélodiques ne sont pas distribués au hasard, ils indiquent une relation de dépendance entre les groupes accentuels qui les portent. Ces dépendances opèrent toutes « à droite », c'est-à-dire qu'ils marquent la dépendance, et donc le regroupement dans la structure prosodique, par rapport à un groupe accentuel situé plus loin dans la phrase. Ainsi, dans ce

réseau de dépendances, le contour terminal conclusif, qu'il soit déclaratif ou interrogatif, constitue la racine de la structure prosodique ainsi définie. Un contour montant C1↗ signale la dépendance du groupe accentuel qui le porte, ainsi que tous les groupes qui en dépendent, par rapport à la racine de la structure prosodique C0↓, c'est-à-dire à la structure tout entière. Le contour C2↘ indique une relation de dépendance envers C1↗, et le contour neutralisé Cn une dépendance par rapport à tout contour C2↘, C1↗ ou C0↓. La relation de dépendance de C2↘ par rapport à C1↗ se manifeste par un contraste de pente mélodique, le contour descendant C2↘ de pente opposée à celle du contour montant C1↗. Ces relations opèrent donc « à droite », c'est à dire envers un groupe accentuel situé plus loin dans la phrase. Elles peuvent se formaliser comme suit, les flèches symbolisant la dépendance d'un contour Cx par rapport à un autre contour Cy « à droite » :

Cn -> {C2, C1, C0}  
 C2 -> C1  
 C1 -> C0

Ainsi, dans l'exemple les hommes d'équipage prennent des albatros, segmenté en 4 groupes accentuels [les hommes] [d'équipage] [prennent] [des albatros] portant les contours C2, C1, Cn et C0 [les hommes C2↘] [d'équipage C1↗] [prennent Cn →] [des albatros C0↓], les regroupements successifs indiqués au cours du déroulement temporel de la phrase par les contours mélodiques sont :

[les **hommes** C2↘]  
 [[les **hommes** C2↘] [d'**équipage** C1↗]]  
 [[les **hommes** C2↘] [d'**équipage** C1↗]] [**prennent** Cn → ]  
 [[les **hommes** C2↘] [d'**équipage** C1↗]] [[**prennent** Cn → ] [des **albatros** C0↓]]

Pour recycler une terminologie ancienne, le contour C1↗ est appelé de continuation majeure et C2↘ de continuation mineure, bien que leurs définitions soient assez différentes de celles proposées à l'origine par Delattre (1966).

### 3. Diction poétique

L'objectif ici est de montrer que, quel que soit le style de diction, et en particulier de diction poétique, les locuteurs respectent le système phonologique de la langue, non seulement en ce qui concerne la prononciation des voyelles et des consonnes, mais aussi quant à la réalisation des contours mélodiques dans leur fonction indicatrice de la structure prosodique. Souvent, dans les études phonostylistiques consacrées au langage poétique (par ex. Léon, 1993), faute de consensus sur un modèle phonologique de l'intonation, les chercheurs se contentent d'analyse acoustique mettant en relief les différences phonétiques des différentes réalisations orales étudiées, comme par exemple le débit syllabique, la hauteur mélodique moyenne, l'empan des variations mélodiques, etc.

Ce faisant, ils méconnaissent et sous estiment l'effet des contraintes cognitives évoquée plus haut, contraintes se manifestant en particulier par les durées des groupes accentuels et les variations mélodiques à l'endroit des syllabes accentuées. On peut monter en effet (Martin, 2013), que, à l'exception de colères extrêmes, les contrastes de pente caractéristiques des contours mélodiques dans leur fonction indicatrice de la structure prosodique sont toujours respectés.

Pour illustrer ce point, on a choisi la première strophe d'un poème très connu de Charles Baudelaire, l'albatros, paru dans le recueil *les fleurs du mal* en 1857.

Souvent, pour s'amuser, les hommes d'équipage  
 Prennent des albatros, vastes oiseaux des mers,  
 Qui suivent, indolents compagnons de voyage,  
 Le navire glissant sur les gouffres amers.

Deux enregistrements ont été analysés avec le logiciel WinPitch (2019), de manière à obtenir d'une part les contours mélodiques placés sur les syllabes accentuées, ainsi que les différences phonétiques caractéristiques des deux interprétations, l'une par Henri Rollan (1888-1967), l'autre, plus récent, par André Nerman (Wheaton collègue, 2019).

Si les deux enregistrements se distinguent par quelques détails de réalisation phonétique, comme la prononciation de e muet chez Nerman, un débit plus lent chez Rollan (3,4 syl/sec vs. 4,4 syl/sec en moyenne), et un empan mélodique plus large chez Nerman (80-120 Hz vs. 80-150 Hz), la distribution des contours phonologiques présente peu de différences, et s'inscrit totalement dans le cadre de la grammaire distribuant les contours dans la phrase et définissant la structure prosodique.

Les deux dictions du poème se caractérisent par une segmentation assurée non seulement par les contours mélodiques mais surtout par des pauses importantes, notées # ci-dessous, et variant de 350 ms à 1200 ms chez Rollan, et de 400 ms à 900 ms chez Nerman. L'auditeur est donc conduit à traiter les informations linguistiques des segments alignés non pas sur les grandes unités de la structure prosodique, mais sur celles déterminées par les pauses. C'est à l'intérieur de chacun de ces segments que s'opère le décodage de la structure prosodique.

Henri Rollan

[souvent C1↗#]  
 [pour s'amuser C1↗#]  
 [les hommes d'équipage C2↘ prennent des albatros C2↘ #  
 vastes oiseaux des mers C2↘ qui suivent C1↗#]  
 [indolents compagnons de voyage #  
 le navire glissant sur les gouffres amers C0↓#]

André Nerman

[souvent C1↗#]  
 [pour s'amuser C1↗#]  
 [les hommes d'équipage C1↗] [prennent des albatros C2↘#]  
 [(vastes oiseaux des mers C0↓#)]  
 qui suivent C1↗#]  
 [indolents C1↗ compagnons de voyage C1↗#]  
 [le navire glissant sur les gouffres amers C0↓#]

Ces deux réalisations sont en fait très semblables en ce qu'elles encodent deux parenthèses par des pauses, guidés évidemment en cela par la ponctuation et la disposition graphique du texte : *vastes oiseaux des mers* et *indolents compagnons de voyage*. En retirant ces parenthèses, on obtient les séquences suivantes :



Pour H. Rollan

souvent C1↗# pour s'amuser C1↗# les hommes d'équipage C2↘ prennent des albatros C2↘# qui suivent C1↗# le navire glissant sur les gouffres amers C0↓

Pour A. Nerman

souvent C1↗# pour s'amuser C1↗# les hommes d'équipage C1↗ prennent des albatros C2↘# qui suivent C1↗# le navire glissant sur les gouffres amers C0↓

La seule différence porte sur le contour terminant les hommes d'équipage, C2↘ pour Rollan et C1↗ pour Nerman. Dans le premier cas les hommes d'équipage est intégré prosodiquement dans le groupe les hommes d'équipage C2↘ prennent des albatros C2↘# qui suivent C1↗# terminé par le contour C1↗, alors que pour Nerman, les hommes d'équipage terminé par le contour C1↗, constitue un groupe de premier niveau.

Une autre différence porte sur l'incise vastes oiseaux des mers. Dans réalisation de Rollan, cette parenthèse terminée par un contour descendant C2↘ est intégrée à la structure prosodique, alors que pour Nerman, la même séquence est associée à une structure prosodique indépendante, terminée par un contour terminal conclusif C0 ↓.

On voit donc que les différences prosodiques apparentes ne sont dues qu'à l'insertion de deux parenthèses, insertion signalée par des pauses qui précèdent et qui suivent (à une exception pour H. Rollan). Les différents groupes prosodiques, alignés sur les mêmes groupes syntaxiques dans les deux réalisations, sont indiqués selon la même grammaire prosodique faisant intervenir le principe de contraste de pente mélodique.

Le respect des règles de la grammaire prosodique du français s'observe également dans des cas apparemment extrêmes, comme dans l'enregistrement d'Antonin Artaud en 1947 de la lecture du texte *j'ai appris hier* (Artaud, 1947). L'analyse de la première phrase fait apparaître les contrastes de pente mélodique attendus, indiquant une structure prosodique congruente avec l'organisation syntaxique de l'énoncé :

J'ai appris hier C1↗ faut croire C2↘ que je retarde C2↘ peut-être C2↘ n'est-ce qu'un faux C2↘ bruit C1↗ l'un de ces sales ragots C0↓

Il s'agit donc de la phrase *J'ai appris UB C1↗ l'un de ces sales ragots C0↓*, dans laquelle s'insère les parenthèses texte faut croire C2↘ que je retarde C2↘ et peut-être C2↘ n'est-ce qu'un faux C2↘ bruit C1↗, associée à un seul syntagme prosodique terminé par C1 et donc intégré dans la structure prosodique de l'ensemble. A. Artaud a prononcé la première partie de la phrase, de *j'ai appris hier* à *peut-être* avec une voix de falsetto, et une voix normale par la suite de *n'est-ce qu'un faux bruit* à *ragots*.

#### 4. Conclusion

L'analyse de ces courts exemples montre combien les contraintes prosodiques, instanciées par la segmentation en groupes accentuels et leur organisation en structure prosodique, sont impératives au point que le locuteur, dans une diction poétique, ne peut s'en affranchir. La

liberté de diction ne peut s'exercer qu'à l'intérieur de ces règles non écrites, dictées par les propriétés des ondes cérébrales thêta et delta liées à la perception des syllabes et des groupes accentuels. Ce sont les gammes de variation des oscillations thêta, de 100 ms à 250 ms, et de delta, de 250 ms à 1250 ms, qui déterminent les possibilités de variation de durée des syllabes et des groupes accentuels, ce qui apparait clairement en observant les caractéristiques rythmiques de la parole en français, langue dépourvue d'accent lexical. Pour les langues à accent lexical par contre, ces contraintes sont masquées par les règles phonologiques de placement de l'accent, et sont donc moins facilement mises en évidence.

## REFERENCES

- Acheson D. J. and Maryellen C. MacDonald (2009) *Verbal Working Memory and Language Production: Common Approaches to the Serial Ordering of Verbal Information*, *Psychol Bull.* 2009 January; 135(1): 50–68.
- Artaud, A. (1947) <https://www.youtube.com/watch?v=oiScQ2wG3WU>
- Bally, Ch. (1944) *Linguistique générale et linguistique française*, Berne : Francke,
- Delattre P. (1966) Les dix intonations de base du français, *French Review* 40, 1966, 1-14.
- Ghitza, Oded (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm, *Frontiers in Psychol.* 2, 130.
- Léon, Pierre (1993) *Précis de phonostylistique : parole et expressivité*, Paris : Nathan Université, Série « linguistique », 1993
- Martin Ph. (2014) *Spontaneous speech corpus data validates prosodic constraints*, *Proceedings of the 6th conference on speech prosody*, Campbell, Gibbon, and Hirst (eds.), 2014, 525-529.
- Martin, Ph. (2013) *Émotions et structure prosodique : qui domine qui ?*, *Cartographie des émotions*, F. Baider et G. Cislaru éd., Paris : Presses universitaires de la Sorbonne.
- Nowak, Paul (2012) *Speed reading tips: 5 ways to minimize subvocalization*, <http://www.irisreading.com/speed-reading/speed-reading-tips-5-ways-to-minimize-subvocalization/>.
- Martin Ph. (2018) *Intonation, structure prosodique et ondes cérébrales*, London : ISTE.
- Rossi, M. (1971) Le seuil de glissando ou seuil de perception des variations tonales pour la parole, *Phonetica*, 23, 1-33.
- Quercia P. (2010) Ocular movements and reading: a review, *J. Fr. Ophtalmologie*, 33 (6): 416-423.
- WinPitch (2019) *Logiciel d'analyse de la parole*, <http://www.winpitch.com/>.
- Wheaton College (2019) (<https://wheatoncollege.edu/vive-voix/titres/lalbatros/>).

**PHILIPPE MARTIN** • has been teaching at the University of Toronto, Aix-en-Provence and Paris Diderot and is presently *professeur émérite* in the Linguistic Department of the University of Paris (formerly Paris Diderot). He published numerous papers on intonation phonology, phonetics and acoustic analysis of speech and his last released books are *Phonétique Acoustique*, *Intonation du Français* (Armand Colin), *The Structure of Spoken Language*, *Intonation in Romance* (Cambridge) and *Intonation, structure prosodique et ondes cérébrales* (ISTE). He also developed speech analysis software (WinPitch), and is currently working on brain oscillations and linguistic speech units.

**E-MAIL** • [philippe.martin@utoronto.ca](mailto:philippe.martin@utoronto.ca)