L'analyse informatique du rythme de la parole: Pertinence pour l'étude de la césure en métrique

Brigitte Zellner Keller
Olivier Bianchi
Eric Keller
LAIP, Lettres, Université de Lausanne

Introduction

À l'interface entre la langue et le locuteur, le fait de *parler* est au coeur de la communication humaine. L'acte de parole peut se définir comme une "incorporation" du langage, véhiculant à la fois des propriétés linguistiques (sociales) et extralinguistiques (en particulier individuelles). Cette activité de parole se caractérise par un paradoxe important : d'une part, la parole utilise un code linguistique partagé par une communauté et donc relativement stable. D'autre part, elle est affectée par une propriété fondamentale partagée par toute gestuelle, à savoir son caractère unique. Il est impossible de redire la même chose exactement à l'identique.

Deux raisons majeures expliquent cette impossibilité. Tout d'abord, un très grand nombre de ressources cognitives et neuromusculaires sont activées en production de parole. Ces ressources sont requises pour articuler correctement entre douze et quinze phones[1] à la seconde, avec toutes les modulations adéquates de hauteur de voix - intonation -, de rythme temporel et d'intensité pour apporter les nuances requises à la communication orale. Une deuxième explication vient de ce que cette gestuelle est ancrée dans des situations de communication uniques. Il est à peu près impossible de reproduire le même acte de parole dans exactement les mêmes conditions environnementales, avec la même position du corps et de la tête, le même niveau de tonicité musculaire, la même qualité d'interaction avec un(e) autre, le même niveau de bruit ambiant, etc.

L'analyse scientifique de la parole cherche à mettre en évidence les aspects stables du code tout en respectant la variabilité importante des gestes sous-jacents. Si de telles analyses ont été lourdes et difficiles à effectuer dans le passé, elles sont devenues plus faciles récemment, grâce à l'analyse informatique et statistique d'enregistrements acoustiques. L'augmentation actuelle de l'utilisation d'instruments autrefois considérés rebutants, motive la présence de cette contribution dans ce volume consacré essentiellement à l'analyse métrique traditionnelle : par une considération des faits empiriques entourant l'organisation temporelle de la parole, nous montrerons que l'analyse instrumentale permet de relier les phénomènes métriques traditionnels à un important éventail de phénomènes acoustiques et psycholinguistiques bien documentés.

En effet, à partir d'enregistrements de la parole, un ensemble d'outils et de méthodes développés en phonétique expérimentale permet de quantifier et d'analyser le signal acoustique de la parole. L'informatique a permis de réduire la taille et le coût des instruments d'analyse de la parole et d'augmenter la précision, la fiabilité ainsi que la vitesse des analyses. Ces possibilités de mieux traiter un plus grand nombre d'informations ont eu pour effet d'améliorer considérablement notre compréhension des aspects stables tout comme variables de la communication orale. Finalement, il est maintenant possible de *simuler* sur ordinateur l'activité de parole, sur la base de "modèles" de la parole. Cette dernière procédure nous permet d'apprécier rapidement et intuitivement la qualité de nos modèles linguistiques (Keller, à paraître).

[1] Phone: production sonore d'une voyelle ou d'une consonne.

Dans cette contribution, il est proposé tout d'abord de présenter l'apport aux études du rythme de la parole de quelques analyses acoustiques sélectionnées. Nous placerons ensuite ces événements de parole (avant tout, les pauses et les regroupements spontanés de mots, c'est-à-dire les « structures de performance ») dans la perspective de recherches menées sur les langues anciennes, afin d'entrevoir les bénéfices à tenter de faire revivre - simuler sur ordinateur - ces langues d'autrefois et de tester une série d'hypothèses spéculatives. Nous finirons sur quelques conclusions thématiques et méthodologiques découlant de ces considérations.

I. Des silences acoustiques aux pauses perçues

Le rythme de la parole est une dimension perceptive complexe, inférée à partir d'un certain nombre d'indices acoustiques, dont celui des pauses (Zellner Keller, 2002). Dans cette section, la question du rapport entre un phénomène physique, le silence, et la perception d'une pause sera posée.

La parole est jalonnée de sons, de bruits et de silences, comme illustré en figure 1. Nous allons montrer que les silences sont constitutifs de la parole. Différents types de silences peuvent être distingués. Tout d'abord, certains silences sont à peine audibles et se produisent au cours même de l'émission d'un phone, comme par exemple dans la production d'un [k]. La figure 2 montre la structure acoustique de [k] dans le mot "mannequin".

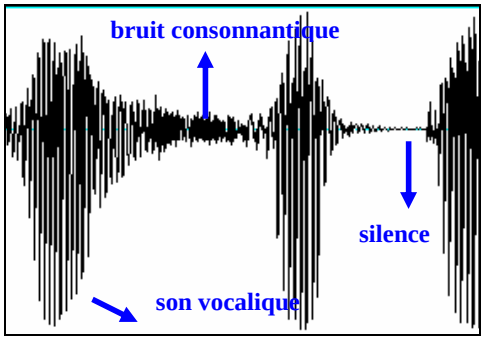

Figure 1. Signal acoustique de parole comprenant des sons, des bruits et des silences.

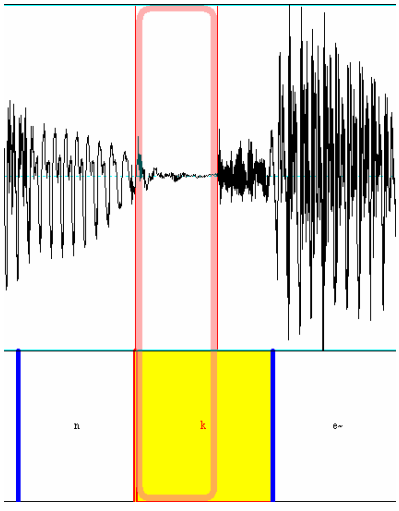

Figure 2. Structure acoustique d'un [k] comprenant un silence.

Ce “silence” acoustique - appelé “Voice Onset Time”- dure dans cet exemple 47 ms. Il correspond à l'intervalle de temps où le conduit vocal est intentionnellement fermé, empêchant l'air de s'échapper. A la fin de ce silence, le signal acoustique témoigne de l'ouverture brusque des cordes vocales, permettant tout à coup l'écoulement d'air qui était retenu derrière les cordes vocales, en même temps que la langue entre rapidement en contact avec le palais.

Cet ensemble de gestes articulatoires permet de produire l'événement acoustique [k]. Remarquons que ce silence qui constitue une portion de la structure acoustique du phone n'est pas perçu en tant que silence. Il s'agit d'un silence intrasegmental perceptivement intégré dans une entité plus grande, un bruit consonantique, ici le [k]. Ce silence est requis pour la perception normale de cette entité plus grande. Si par manipulation acoustique, le silence est coupé, le phone [k] est structurellement modifié, avec pour effet potentiel d'altérer la perception de cette entité.

Au niveau interlexical, le hiatus entre présence/absence d'un événement acoustique et perception se retrouve fréquemment. Soit l'exemple extrait de la lecture à haute voix d'un titre de journal “Primakov, en émissaire pour déjouer la crise au Kosovo”. La figure 3 montre un silence acoustique de 110 ms, entre les deux mots “Primakov” et “en”; ce silence est tout à fait congruent avec la ponctuation du texte. Dans ce cas, le silence acoustique est perçu comme une pause.

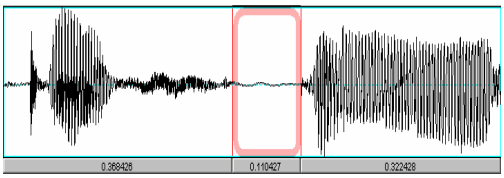

Figure 3. Signal acoustique représentant la portion “kov, en émi”, extrait de la phrase “Primakov, en émissaire pour déjouer la crise au Kosovo”. La pause interlexicale est signalée dans le rectangle gris.

Si, par manipulation acoustique, on supprime ce silence, la perception d'une pause demeure paradoxalement inchangée. Cette illusion auditive est bâtie sur la perception d'autres propriétés acoustiques contenues dans la portion avoisinante du signal.

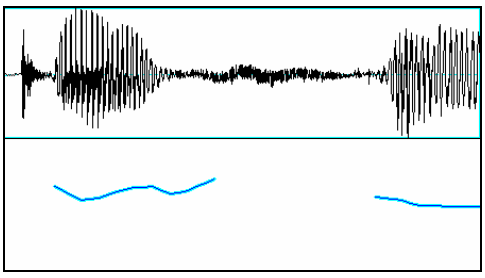

Figure 4. Signal acoustique représentant la portion “kov en émi”, extrait de la phrase “Primakov, en émissaire pour déjouer la crise au Kosovo”, après suppression du silence acoustique. La courbe mélodique extraite de ce signal acoustique est représentée sous le signal. Entre les syllabes “kov” et “en”, on observe une diminuation de 15 Hz, qui concourt à la perception d'une pause, même en l'absence d'un silence acoustique. La baisse mélodique se poursuit sur la syllabe suivante.

En figure 4, la courbe mélodique (en bleu) de ce signal montre par exemple que la perception d'une pause peut être inférée à partir de l'augmentation suivie de la diminution de la hauteur de voix, et cette variation peut suffire à créer l'illusion perceptive d'une pause pourtant physiquement absente.

Ainsi, la perception d'un silence acoustique varie, selon la durée d'un événement, selon le moment d'occurrence de cet événement et selon les valeurs des autres paramètres acoustiques présents dans le signal acoustique, en particulier courbe mélodique (hauteur de voix), courbe d'amplitude (intensité de la voix) et accélérations/décélérations du rythme.

Lorsqu'un silence est perçu en tant qu'interruption, cet événement est qualifié de pause *silencieuse*. Un autre type de pause est la pause *sonore*. Une pause sonore n'est jamais associée à un silence acoustique; elle se manifeste par exemple par un allongement vocalique, ou une interjection vide de type [euh], ou par la répétition du dernier mot; elle est perçue comme un ralentissement de parole, par exemple une hésitation.

La motivation pour appréhender le domaine de la parole selon des méthodes de phonétique expérimentale vient de ce hiatus entre événements physiques et perception, en particulier au niveau des pauses. La section suivante explore les aspects fonctionnels des pauses.

II. Les fonctions des pauses

Les pauses remplissent des fonctions multiples qui peuvent être considérées des points de vue physiologique et cognitif (Zellner, 1994).

II.1. Fonction physiologique

La production de parole est rendue possible par la modulation de petits flux d'air au niveau des cordes vocales qui s'ouvrent et se ferment très rapidement (environ 100-120 fois par seconde pour un homme, 200-240 fois par seconde pour une femme). L'air phonatoire provient donc des poumons. Lorsque la source d'air se tarit, le besoin de respirer s'impose, tant pour satisfaire un besoin physiologique d'oxygénation que pour fournir le support nécessaire à la phonation. Au repos, l'adulte prend en moyenne une quinzaine d'inspirations[2] par minute.

Chaque bouffée d'air émise au niveau des cordes vocales est ensuite “sculptée” à l'intérieur de notre appareil phonatoire (pharynx, bouche), selon différentes formes et au moyen de muscles divers, dont la langue et les lèvres. Les positions respectives de ces muscles modifient la forme du conduit vocal, faisant subir à ces bouffées d'air des effets de résonances et amplifications, sources de bruits et de sons. La vitesse d'articulation de ces sons et bruits est ultimement limitée par nos capacités motrices, soit environ 5 à 8 syllabes par seconde - deux à trois mots. Les contraintes cognitives qui s'exercent au niveau du système de production de la parole déclenchent alors le besoin de produire d'autres pauses.

II. 2. Fonction cognitive

Comme toute gestuelle, la parole ne se déroule pas de manière uniformément continue. Elle se déroule dans le temps par groupes de mots.

Les pauses respectent généralement ces groupes de mots. Il existe toute une littérature démontrant que les pauses entre les mots ne sont pas produites au hasard n'importe où dans la chaîne parlée, y compris lors d'une hésitation (cf. par exemple: Boomer et al, 1962; Butcher, 1980; Levelt, 1989; Zellner, 1992). Selon Goldman-Eisler (1972), une pause témoigne des traitements cognitifs impliqués dans la parole. Du point de vue du locuteur, une pause fournit par exemple du temps additionnel pendant lequel le message peut être planifié et programmé.

[2] http://www.lung.ch/dielunge/aufbauundfunktion/lungenfluegel3.asp?script=false&sprache=fr; http://www.ulg.ac.be/physioan/chapitre/ch3s3b.htm

Grosjean et Deschamps (1975) ont par exemple observé que plus la tâche communicative est complexe, plus le nombre et la durée des pauses tendent à augmenter. Il est aussi bien décrit dans la littérature qu'un locuteur en situation de parole spontanée produit plus de pauses qu'en situation de lecture. Du point de vue de l'auditeur, les pauses permettent la subdivision de la chaîne parlée en éléments plus petits, ce qui facilite le traitement perceptif et donc la compréhension de parole.

Les pauses tendent à être produites *entre* des groupes de mots. Ces groupes de mots constituent des unités cohérentes des points de vue prosodique, sémantique et potentiellement phonosyntaxique. Exemple (cf. Levelt, 1989):

/Attila believed / the world to be flat/

Grosjean et al. (1979), en particulier ont montré que les occurrences et la longueur des pauses sont fortement corrélées avec le degré de cohésion inter-lexicale: les pauses tendent à être plus longues et plus fréquentes entre les mots qui sont peu liés entre eux, et elles tendent à être plus courtes et moins fréquentes entre les mots qui sont fortement interdépendants. Par exemple, dans la phrase précédente, il y a davantage de cohésion entre les mots “plus” et “longues” qu'entre les mots “longues” et “et”. Si une pause devait survenir, elle se produirait alors plus probablement entre “longues” et “et”. Les groupes de mots sont également peu sensibles aux contraintes de respiration, ***i.e.***, la respiration s'adapte et intervient plutôt *entre* ces entités.

Différents types d'expériences, telles que des tâches de mémorisation ou de segmentation intuitive d'un énoncé, aussi bien que des mesures empiriques des pauses et des durées syllabiques valident cette hypothèse d'organisation en petits groupes de mots présentant une forte cohésion (cf. Levelt, 1989).

Les pauses se produisant plutôt entre des groupes de mots, beaucoup d'études ont exploré la possibilité d'une interdépendance entre structures syntaxiques et pauses. Il s'avère que les occurrences des pauses dans la chaîne parlée ne sont que partiellement dépendantes des structures syntaxiques, comme souligné par Levelt (1989). Intuitivement, cette non-congruence se comprend par le fait que la parole est soumise à beaucoup d'autres contraintes, comme par exemple les variations de débit ou les interactions entre locuteurs, etc. (voir section 1). Dans sa revue de littérature, Levelt conclut qu'une relation entre les structures syntaxiques et les occurrences des pauses peut être observée, en particulier avec les pauses sonores (remplies), sans que cette relation soit systématique. Il note aussi qu'il y a peu d'évidences d'une relation entre la complexité des opérations syntaxiques et le phénomène pausal, hypothèse qui avait souvent été postulée du point de vue de la hiérarchie des structures syntaxique et la longueur des pauses.

De manière universelle, la production de parole est délivrée en groupes de mots, les pauses ayant tendance à intervenir entre ces groupes de mots. Cette organisation fondamentale est à la base du rythme de la parole. La question de la prédictibilité d'une telle organisation est examinée dans la section suivante.

III. La prédiction des pauses sur la base des structures de performance

D'un point de vue psycholinguistique, l'agencement de la parole en groupes de mots peut se définir en termes de *structures de performance*. Fondée sur cette notion de cohésion inter-lexicale (cf. section précédente), l'organisation de la parole en groupes de performance (ou groupes temporels) présente un certain nombre de propriétés qui sont intéressantes dans la perspective d'une modélisation du rythme de la parole.

Grosjean et Dommergues (1983) et Monnin et Grosjean (1993) ont montré que pour le français et l'anglais, ces structures de performance présentent trois caractéristiques. Tout d'abord, les groupes de mots tendent à s'équilibrer en longueur, ce qui produit un effet rythmique équilibré. Ensuite, cette organisation est hiérarchisée: les petits groupes de mots sont encapsulés dans des unités plus grandes. Enfin, dans un énoncé, les structures de performance tendent à s'équilibrer entre elles - ex: la structure de la première moitié de l'énoncé ressemble à celle de la seconde moitié; cet équilibrage se réalise en particulier grâce aux pauses, la pause la plus importante ayant tendance à se situer vers le milieu de l'énoncé.

En français, ces structures de performance correspondent, à quelques exceptions près, à des distributions de un ou plusieurs mots grammaticaux (articles, pronoms, prépositions, etc) suivis de un ou plusieurs mots lexicaux (noms, verbes, adjectifs, etc). Cette disposition est simple, robuste quel que soit le style de parole, et assez bien prévisible. Elle est à la base de nos algorithmes de prédiction de la structure temporelle de la parole pour la lecture neutre (Keller et al., 1993; Zellner, 1996) et la lecture lente et rapide (Zellner 1998).

La simulation de la parole sur ordinateur à partir d'un tel algorithme montre la plausibilité de l'hypothèse qu'une composante majeure du rythme de la parole repose sur une subdivision adéquate de la chaîne parlée en groupes de mots, même si la création du rythme de la parole implique beaucoup d'autres paramètres avec différents niveaux d'interaction entre ces paramètres (Zellner Keller, 2002.a.b). La section suivante montre comment ces principes de groupements de mots peuvent être appliqués à la prédiction du rythme poétique.

IV. Structures rythmiques de la prose et de l'alexandrin[3]

IV. 1 Introduction

Les philologues et les métriciens se sont tous heurtés, au moins une fois, à la question de savoir comment les Anciens récitaient leurs oeuvres en vers. Cette question, simple en apparence, se révèle en fait fort épineuse et n'a pas encore trouvé, malgré l'abondante littérature qu'elle a suscitée, de réponse unanimement acceptée. On observe cependant, depuis une vingtaine d'années, une certaine convergence des regards que les spécialistes portent sur cette question : la recherche tend à découvrir les facteurs susceptibles de jouer un rôle rythmique dans un vers grec ou latin non plus seulement dans la structure même du vers, mais aussi dans l'observation d'un contexte énonciatif plus large qui inclut la prose[4].

Cette mise en parallèle de la structure rythmique de la poésie et de la prose, deux types d'énonciation *à priori* différents, va servir de fil conducteur à notre propos.

Notre étude porte sur le français. Les avis divergent sur une définition du « rythme poétique » du français. Ces divergences proviennent du fait que ces définitions découlent essentiellement d'une approche perceptive[5]. L'approche que nous proposons ici est de type empirique, basée sur

[3] L'étude pilote présentée lors du colloque Damon X a donné lieu à une seconde étude, plus vaste, dont les résultats viennent confirmer et compléter les conclusions initiales. Ce sont donc les résultats de cette seconde étude que nous livrons ici.

[4] Une illustration de cette nouvelle dimension de la recherche sur la métrique, notamment latine, nous est donnée par BOLDRINI (1992, 36) : *« i Latini leggevano i versi esattamente come la prosa… »*.

[5] Jean-Claude MILNER et François REGNAULT s'en amusent d'ailleurs : « Mais le fait est que, dans l'ordre de la diction, on ne part pas de rien. La rumeur, sans doute, va répétant que les acteurs disent mal aujourd'hui. Comme la plupart des déplorations, celle-là manque l'essentiel. Car, malgré tout, il est des acteurs qui disent bien. Mais ceux qui disent bien et ceux qui disent mal ont néanmoins un point commun : ils se sont formés des conceptions fausses sur ce qu'est, ce que peut être et ce que doit être la diction. On peut même soutenir que, littéralement, ils ne savent pas ce qu'ils font : ils croient dire des choses qu'ils ne disent pas, ils croient ne pas dire des choses qu'ils disent, ils croient devoir dire des

une expérience effectuée en laboratoire et avec l'objectif de tenter de définir ce qu'est le « rythme poétique » des alexandrins. Le choix de l'alexandrin comme type de vers nous a paru le plus raisonnable, dans la mesure où c'est un vers de théâtre, conçu pour être récité à haute voix. Nous montrerons d'abord que les structures rythmiques que l'on peut dégager de l'analyse d'une récitation de textes en prose et ceux que l'on peut dégager de l'analyse d'une récitation de textes en vers reposent en fait sur une base rythmique unique, inhérente à la langue française. Puis, nous tenterons d'isoler les éléments constitutifs de cette base rythmique.

IV. 2 Expérience

L'expérience avait pour objectif d'analyser les structures rythmiques produites par plusieurs locuteurs lors d'une lecture à haute voix non préparée d'un texte comprenant des passages versifiés et des passages en prose, et de les comparer aux structures rythmiques produites lors d'une lecture préparée du même texte.

IV. 2.1 Méthode

Sujets — Quatre sujets (ou « locuteurs ») ont été enregistrés : un étudiant de troisième année du Conservatoire de Lausanne, section d'art dramatique ; un professeur de français à l'Université de Lausanne ; un maître d'enseignement et de recherche à l'Université de Lausanne (section autre que français) ; un membre du personnel d'exploitation de l'Université de Lausanne n'ayant pas suivi d'études littéraires. Si tous les sujets sont de langue maternelle française, leur connaissance de la poésie classique française est, en revanche, variable. Un sujet « hors catégorie » a également été pris en compte : le système informatique text-to-speech (TTS) développé par le Laboratoire d'Analyse Informatique de la Parole (LAIP), intitulé MoulinAParole (version 1998) et entraîné sur la lecture de textes avec intonation neutre.

Matériel — Deux textes ont été soumis aux locuteurs :

TEXTE A
[a Seigneur, dans cet aveu dépouillé d'artifice,]
[b J'aime à voir que du moins vous vous rendiez justice,]
[c Et que voulant bien rompre un noeud si solennel,]
[d vous vous abandonniez au crime en criminel.]
[e Est-il juste après tout, qu'un conquérant s'abaisse]
[f Sous la servile loi de garder sa promesse ?]
[g Non, non, la perfidie a de quoi vous tenter ;]
[h Et vous ne me cherchez, que pour vous en vanter.]
1 Quoi ? sans que ni serment ni devoir vous retienne,
2 Rechercher une Grecque, amant d'une Troyenne ?
3 Me quitter, me reprendre, et retourner encor
4 De la fille d'Hélène à la veuve d'Hector ?
5 Je suis l'Empire à la fin de la décadence
6 Qui regarde passer les grands barbares blancs
7 En composant des acrostiches indolents
8 D'un style d'or où la langueur du soleil danse
9 De n'importe quelle hauteur de la ville et tout autour,
10 On peut toujours apercevoir les montagnes,
11 Le fleuve ou bien la campagne. Cette nature,
12 Souvent sauvage, n'est pas qu'à la portée du regard
13 Mais à portée de la main, si l'on peut dire.

L'ensemble du texte A est présenté visuellement comme s'il s'agissait d'un passage continu en vers. Les douze premières lignes sont des alexandrins de type classique (tirés de Racine), avec

---

choses qu'il ne faut pas dire, et qui sont parfois matériellement impossibles à dire, ils croient ne pas devoir dire des choses qu'il faut dire et que parfois on ne peut pas matériellement ne pas dire. » (MILNER/REGNAULT 1987, ??)

une césure après la sixième syllabe. Parmi ces douze vers, seuls les quatre derniers (numérotés 1-4) ont été retenus pour l'analyse, les huit premiers permettant au locuteur de se détendre et de prendre ses marques. Les quatre vers suivants (5-8) sont des alexandrins non classiques[6] (tirés de Verlaine) dont seul le vers 6 présente une césure après la sixième syllabe. Les cinq dernières lignes (9-13) sont tirées d'un ouvrage en prose sur la ville de Québec. Ces dernières lignes ont été artificiellement mises en forme pour donner une impression visuelle de continuité avec les vers qui précèdent.

> TEXTE B
> [i] Malgré son charme romantique, Québec, avec ses fortifications, sa citadelle et ses canons pointant du haut du promontoire vers le fleuve, [14] conserve ce panache dramatique et guerrier [15] qui lui permit d'affronter les aléas de l'histoire. [16] Lieu stratégique de toutes les convergences, [17] elle fut en effet sans cesse convoitée. [18] L'Europe qui vous hait, vous regarde en riant. [19] Comme si votre roi n'était plus qu'un fantôme, [20] la Hollande et l'Anglais partagent ce royaume.

L'ensemble du texte B est présenté visuellement comme s'il s'agissait d'un passage continu en prose. La première partie du texte (i-17) est effectivement en prose. Le membre de phrase [i] sert d'introduction et n'a pas été retenu dans l'analyse. La seconde partie du texte (18-20) est constituée de trois alexandrins classiques (tirés de V. Hugo).

Procédure — Chaque sujet a reçu un exemplaire du texte A (qu'il n'avait jamais vu auparavant) – sans numérotation ni crochets carrés – avec pour consigne d'en faire une lecture à voix haute la plus naturelle possible. La consigne insistait particulièrement sur le fait que la lecture ne devait pas être interrompue, même en cas de lapsus ou d'hésitation. Cette première lecture a été enregistrée et constitue l'échantillon de lecture « spontanée ». Les sujets ont ensuite reçu pour consigne de relire ce texte pour eux, autant de fois que nécessaire, afin de procéder à un second enregistrement du même texte. Ce second enregistrement constitue l'échantillon de lecture « préparée ». On a procédé de la même façon pour le texte B.

Une fois les enregistrements effectués[7], les signaux de parole ont été segmentés[8] (niveau syllabique) et analysés. Dans un premier temps, les durées syllabiques mesurées pour les lectures spontanée et préparée d'un même locuteur ont été comparées : c'est ce que nous appellons ci-dessous la comparaison « intra-locuteur ». Puis on a comparé entre elles les durées mesurées pour chaque locuteur : c'est la comparaison « inter-locuteurs ». La comparaison intra-locuteur permet de mettre en relief le rôle joué par le contexte, et notamment par le passage d'un type d'énoncé à un autre, dans le cas des lectures spontanées et préparées. La comparaison inter-locuteurs permet d'identifier, entre autres, le rôle potentiel joué par l'expérience littéraire des quatre locuteurs.

IV. 2.2 Résultats

Comparaison intra-locuteur
L'objectif étant d'analyser les variations de la structure rythmique produite par chaque locuteur lorsqu'il est confronté de manière inattendue à des types d'énoncés différents, nous n'exposerons ici que les résultats de l'analyse portant sur les passages de *transition* entre deux types d'énoncés (alexandrins classiques/non classiques [texte A, v. 5], alexandrins non classiques/prose [texte A, l. 9] et prose/alexandrins classiques [texte B, v. 18]).

La figure 1 montre un exemple de courbe de durées syllabiques obtenues pour l'un des quatre locuteurs en lecture spontanée et en lecture préparée, pour le vers 5 du texte A (transition alexandrins classiques/ non classiques) :

[6] Alexandrins sans césure obligatoire après la sixième syllabe.
[7] Les enregistrements se sont déroulés au Centre audio-visuel (CAV) de l'Université de Lausanne, dans une cabine insonorisée, puis ont été numérisés (fréquence d'échantillonage: 44'100 Hz, 16 Bit, Stéréo).
[8] A l'aide du logiciel Praat (www.praat.org).

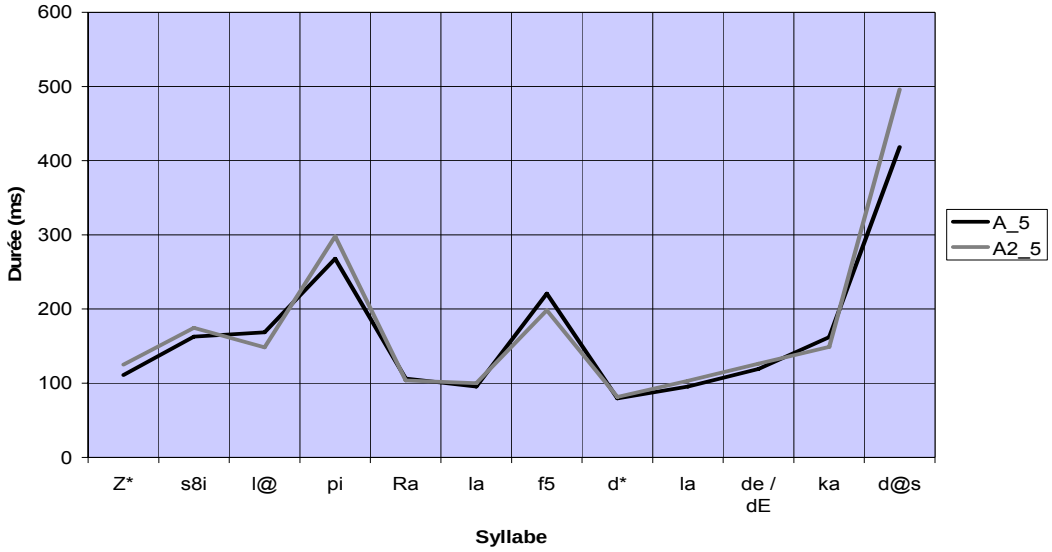

Figure 1 — Durées syllabiques pour le locuteur 1, texte A, v. 5 (« Je suis l'Empire à la fin de la décadence »), lecture spontanée (A_5) et préparée (A2_5). L'axe horizontal représente la découpe syllabique, l'axe vertical la durée (en millisecondes). L'alphabet utilisé pour représenter les syllabes est celui utilisé par le LAIP ; les « - » sont mis pour les pauses, les « ◊ » pour des syllabes n'existant pas dans l'une des deux lectures. La similitude des deux courbes indique que les durées syllabiques mesurées pour la lecture spontanée et pour la lecture préparée sont très proches (écart moyen : 17,8 ms).

On constate que les deux courbes de durées syllabiques se superposent presque exactement, ce qui signifie que le locuteur a temporellement structuré l'oralisation de ce vers de manière semblable dans les deux modes de lecture.

Cette similarité est illustrée par la table 1, où l'on constate que la différence de durée syllabique, en pourcentage de la durée totale de l'énoncé, est souvent très proche de zéro :

		A_5	**A2_5**	**écart-%**
1	Z*	5,53	5,94	0,41
2	s8i	8,11	8,30	0,19
3	l@	8,40	7,06	1,35
4	pi	13,35	14,16	0,81
5	Ra	5,30	4,92	0,38
6	la	4,76	4,75	0,01
7	f5	11,00	9,43	1,58
8	d*	3,97	3,87	0,10
9	la	4,76	4,91	0,15
10	de / dE	5,93	6,00	0,07
11	ka	8,08	7,09	1,00
12	d@s	20,82	23,58	2,76
	Total	100 %	100 %	0,73

Table 1 — Durées syllabiques exprimées en pourcentage de la durée complète de la lecture du texte dans les deux modes de lecture, locuteur 1, texte A, v. 5, lecture spontanée (A_5) et préparée (A2_5). L'écart, exprimé en pourcent, de la durée syllabique d'une lecture à l'autre varie de 0,01 pour le plus faible à 2,76 pour le plus important, pour une valeur moyenne de 0,73%.

Les figures 3 et 4 montrent les graphiques obtenus après analyse du locuteur 1 (très expérimenté) sur le vers 9 du texte A et le vers 18 du texte B, figurant respectivement une transition alexandrin/prose et prose/alexandrin :

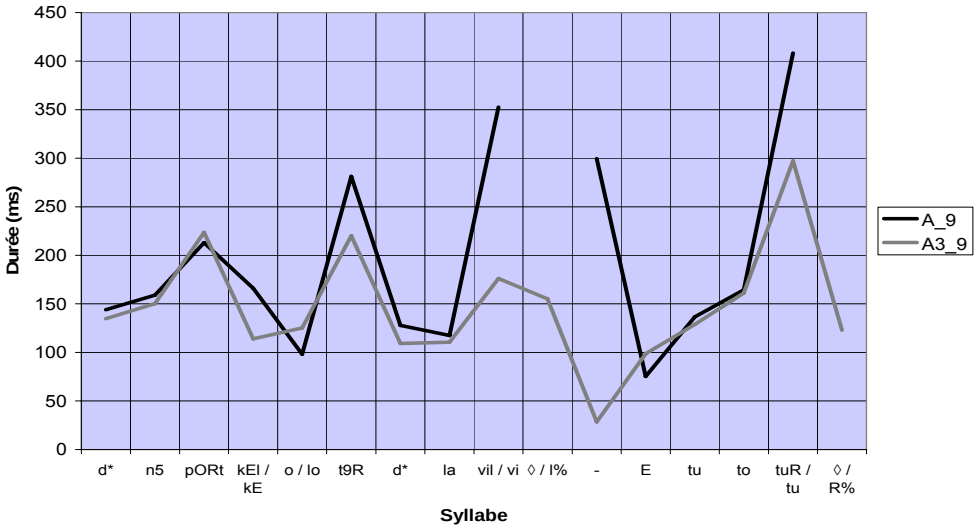

Figure 2 — Durées syllabiques pour le locuteur 1, texte A, l. 9 (« De n'importe quelle hauteur de la ville et tout autour »), lecture spontanée (A_9) et préparée (A3_9). La coupure que l'on observe pour la courbe A_9 est due à une différence dans la structuration syllabique produite par ce locuteur lors des deux lectures : en lecture préparée (A3_9) le locuteur énonçait le mot « ville » avec deux syllabes, en prononçant le *e* muet ; en lecture spontanée, en revanche, il énonçait ce mot de façon monosyllabique, en ne prononçant pas le *e* muet. La syllabe /l%/ n'existant pas pour A_9, elle n'a naturellement pas de valeur de durée. Cette différence de syllabation explique aussi une partie de la différence de la hauteur du « pic » de durée pour la syllabe 9, puisque en lecture spontanée, cette syllabe compte trois segments (/vil/) alors qu'elle n'en compte que deux (/vi/) en lecture préparée.

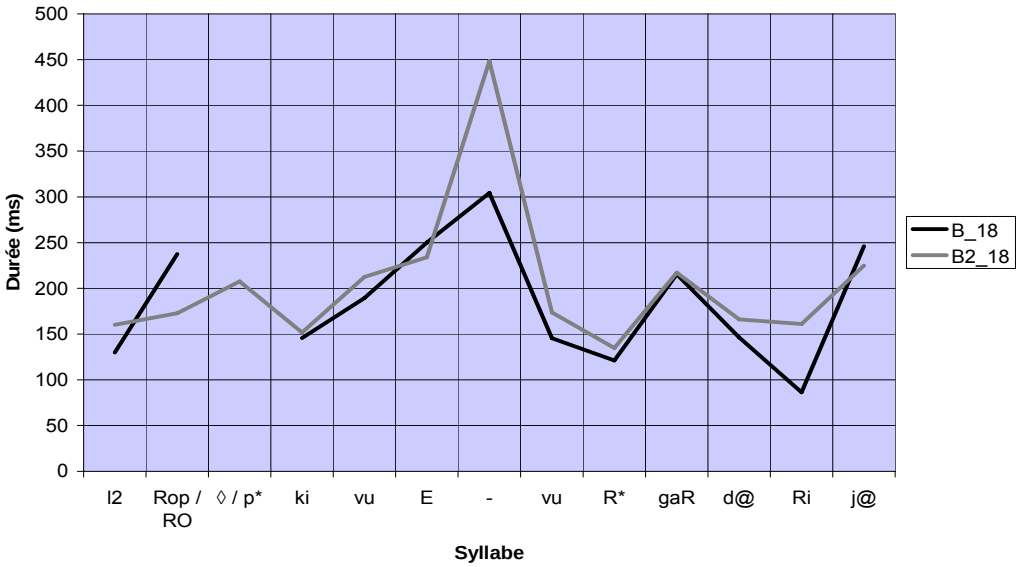

Figure 3 – Durées syllabiques pour le locuteur 1, texte B, v. 18 (« L'Europe qui vous hait, vous regarde en riant. »), lecture spontanée (B_18) et préparée (B2_18). La courbe B_18 s'interrompt entre la deuxième et la quatrième syllabe pour des raisons de syllabation différente (pour des explications plus complètes, voir ci-dessus la légende de la figure 2).

L'analyse des transitions alexandrin/prose (texte A) ou prose/alexandrin (texte B), basée sur la comparaison des lectures spontanées et préparées, fait apparaître une grande ressemblance des structures rythmiques produites. Cette similarité est toutefois ponctuée par des micro-variations de durée touchant principalement les pauses et les finales de groupes temporels ou *structures de performance*[9], comme on le voit pour les vers 9 du texte A (cf. ci-dessus figure 2) et 18 du texte B (cf. ci-dessus figure 3) qui présentent en effet les groupes temporels suivants :

A9 [[[[de n'importe] [quelle hauteur]] [de la ville]] *(pause)* [et tout autour]]
B18 [[[L'Europe] [qui vous hait]], *(pause)* [[vous regarde] [en riant]].]

Les syllabes finales (/-t9R/, /vil/ ou /-l%/, /-tuR/ ou /-R%/ pour le vers A9 et /-Rop/ ou /-p*/, /-E/, /-j@/ pour le vers B18) des groupes temporels ainsi que les pauses sont effectivement les endroits où l'on observe les micro-variations les plus importantes. L'analyse comparative entre les locuteurs montre que ces variations sont d'autant plus marquées que le locuteur a plus d'expérience littéraire.

Comparaison inter-locuteurs
Etablir une comparaison entre tous les locuteurs pour un type d'énoncé permet d'estimer entre autres le rôle joué par l'expérience littéraire de ces locuteurs dans la production de structures rythmiques.

La comparaison inter-locuteurs montre que les quatre sujets accordent la même proportion aux mêmes syllabes, comme l'illustrent les figures 4 (alexandrin) et 5 (prose), indépendamment du niveau littéraire ou de l'expérience des locuteurs (l'axe vertical exprime la durée d'une syllabe en pourcentage de la durée totale du vers) :

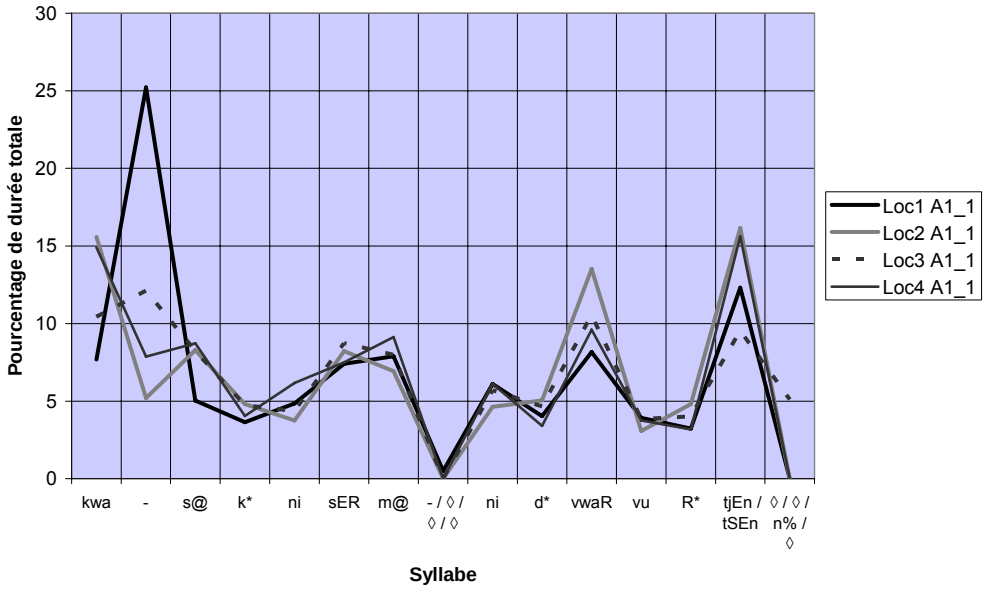

Figure 4 — Durées syllabiques en pourcentage de la durée de tout l'énoncé, pour les locuteurs 1-4, texte A, v. 1 (« Quoi ? sans que ni serment ni devoir vous retienne »), lecture préparée. La pause qui suit le mot « quoi » génère des différences plus marquée dans la proportion de durée que lui accordent les locuteurs. Ces différences s'expliquent peut-être par le fait que le mot *quoi* peut être compris,

[9] Pour une définition des structures de performance (ou groupes temporels), voir ci-dessus p. 5. On observe également des variations de structuration syllabique pour certains mots, notamment des mots comportant des *e* muets. Ces différences de structuration, bien qu'elles altèrent la lisibilité des graphiques, n'ont que peu d'influence sur l'élaboration des patterns rythmiques dans la mesure où un système de compensation des durées des syllabes précédentes et suivantes se met en place avec pour effet de maintenir le pattern rythmique inchangé.

selon le locuteur (et malgré la présence du signe de ponctuation), soit comme un interrogatif, soit comme un exclamatif. Le statut grammatical qu'on lui confère et les différences d'interprétation que ce choix engendre peuvent entraîner des variations substantielles dans la durée de la pause subséquente.

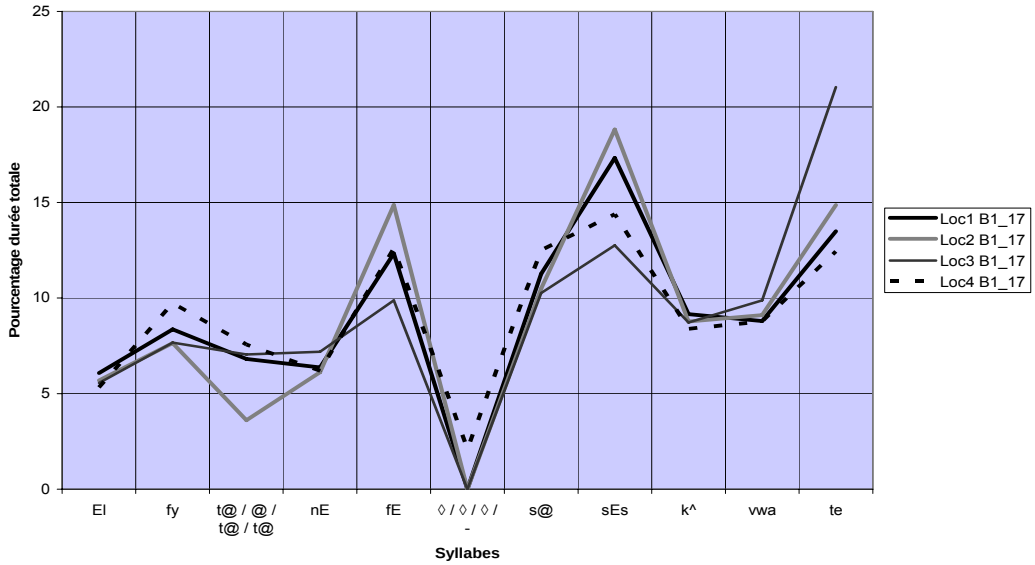

Figure 5– Durées syllabiques en pourcentage de la durée de tout l'énoncé, pour les locuteurs 1-4, texte B, l. 17 (« elle fut en effet sans cesse convoitée »), lecture préparée.

Des micro-variations (et même des variations plus importantes) peuvent également être observées entre les différents locuteurs. Ces variations apparaissent aux mêmes endroits que ceux révélés par la comparaison intra-locuteur, soit aux pauses et en finale de groupes temporels. Nous observons par exemple des micro-variations au vers 1 du texte A (cf. ci-dessus figure 4) pour les syllabes finales des groupes temporels : [quoi], [ni devoir] et [vous retienne], ainsi qu'à la ligne 17 du texte B (cf. ci-dessus figure 5) : [elle fut en effet] et [[sans cesse] [convoitée]].

IV. 3 Discussion

Les résultats de l'expérience montrent clairement que ni le contexte dans lequel apparaissent les énoncés testés, ni l'expérience qu'ont les sujets de ce type d'énoncé n'ont d'influence sur les structures rythmiques produites.
L'analyse montre aussi que des micro-variations de durée, sont observables tant entre la lecture spontanée et la lecture préparée, qu'entre les différents locuteurs sur un même énoncé. Deux facteurs contribuent principalement à l'apparition de ces micro-variations : 1) la présence de pauses ; 2) la position relative dans la chaîne parlée (finale de structure de performance).

Les pauses et les finales de structures de performance sont proportionnellement les endroits les plus longs de la chaîne parlée. La figure 7 fait apparaître clairement que ces endroits clés sont bien déterminés pour les alexandrins classiques :

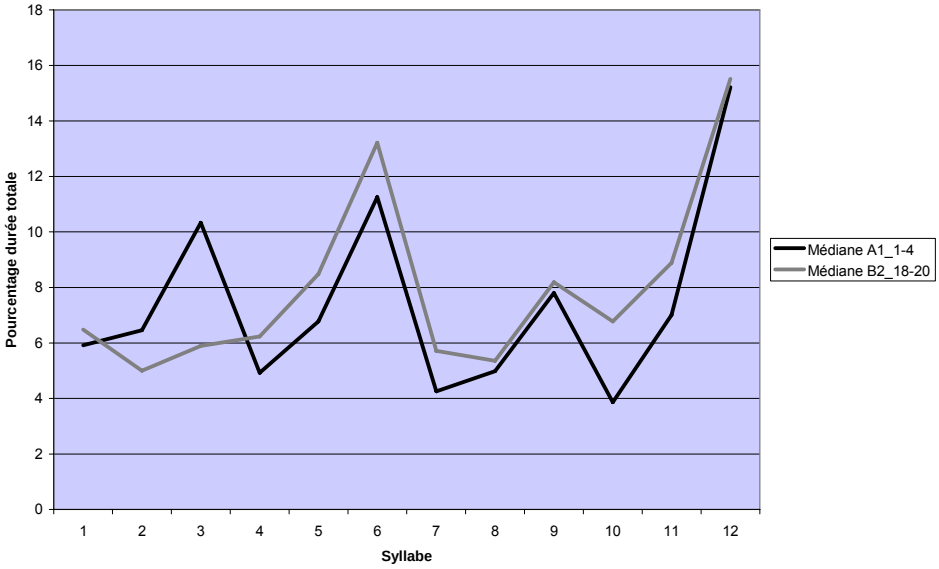

Figure 6 — Durées syllabiques en pourcentage de la durée de tout l'énoncé, pour les locuteurs 1-4, texte A, v. 1-4 et texte B, v. 18-20 (alexandrins classiques), lecture préparée

Le profil rythmique des alexandrins classiques comprend généralement trois ou quatre pics principaux[10] découpant l'énoncé, selon les vers, en trois ou quatre unités rythmiques de trois ou quatre syllabes, soit une moyenne de 3,5 syllabes par unité[11]. L'analyse du profil rythmique des alexandrins non classiques révèle le même nombre d'unités (trois ou quatre) pour douze syllabes ; seuls la position des syllabes longues dans la chaîne parlée varie par rapport aux alexandrins classiques (pics aux syllabes 12, 8 et 4), comme le montre la figure 7 :

[10] Dans les cas où quatre pics apparaissent, ils correspondent aux syllabes 3, 6, 9 et 12 (cf. vers A1_1-4, figure 6) ; dans les cas où trois pics seulement apparaissent, ils correspondent aux syllabes 6, 9 et 12 (cf. vers B2_18-20, figure 6).

[11] Cette moyenne de 3,5 syllabes par unité calculée pour les alexandrins rejoint la moyenne calculée par MONNIN et GROSJEAN (1993) pour une lecture à haute voix de phrases simples en prose. Beaucoup d'études phonétiques suggèrent que cette longueur correspond à une sorte d'empan rythmqiue (LEVELT, 1989).

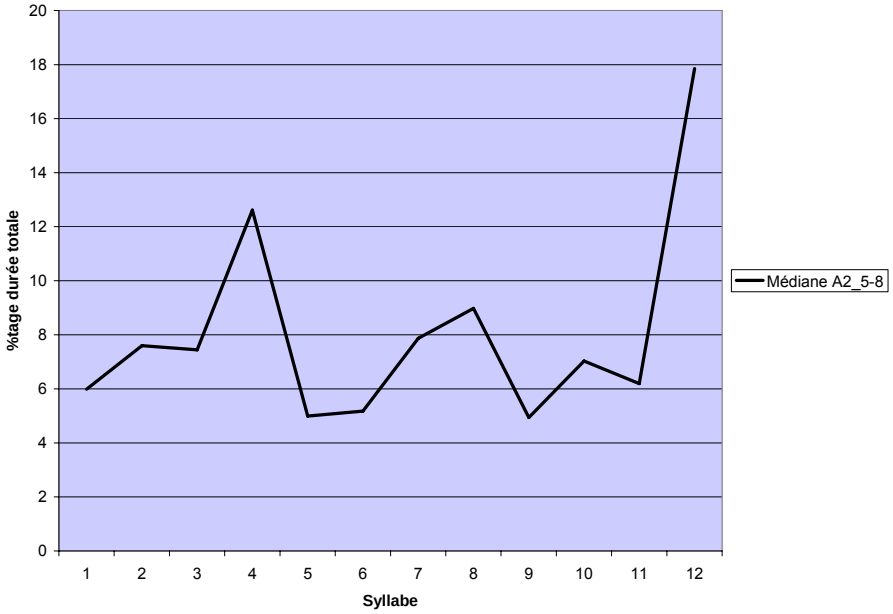

Figure 7 — Durées syllabiques en pourcentage de la durée de tout l'énoncé, pour les locuteurs 1-4, texte A, v. 5-8 (alexandrins non classiques), lecture préparée

Cette structuration de l'énoncé en groupes temporels (approche psycholinguistique) est à la base des modèles temporels développés pour le Moulin-à-parole au LAIP[12]. Cette approche permet, selon nous, de prédire non seulement les structures rythmiques de la parole lue et de la parole spontanée, mais aussi de la récitation d'alexandrins et, par extension, d'autres types de vers[13]. Nous avons vu en effet que les structures rythmiques étaient produites sans que ni le contexte d'énonciation, ni l'expérience littéraire des sujets ne jouent de rôle déterminant, mais que les frontières rythmiques étaient balisées autour des éléments les plus susceptibles d'allongements (pauses, syllabes en fin de groupe temporel). Ces éléments fonctionnent comme des marqueurs de l'agencement desquels découle le profil rythmique[14]. Ces marqueurs peuvent être :

- de type syntaxique (ponctuation) ;
- de type morphologique (présence de syllabes longues « par nature » : diphtongues ou syllabes formées de trois phonèmes) ;
- de type lexical (noms, verbes, adjectifs et adverbes, par opposition aux mots grammaticaux) ;
- de type sémantique (présence d'un mot important pour le sens du passage) ;

Ces marqueurs étant des éléments constitutifs de la langue française, ils sont les mêmes pour les énoncés en prose que pour les alexandrins. Comme le montre la figure 9, le système "text-to-speech" du LAIP, bien que reposant sur un système composé d'algorithmes calculés à partir d'énoncés en prose, est capable de produire des structures rythmiques, pour l'alexandrin classique, qui sont similaires à celles produites par des locuteurs humains :

[12] Cf. ZELLNER 1997, 64 *sq.*

[13] Il serait intéressant à cet égard de poursuivre cette étude en y incluant les « vers libres » de la poésie française.

[14] La liste des marqueurs que nous avons repérés sur la base de nos échantillons enregistrés recoupe la liste des prédicteurs de frontières de groupes temporels telle qu'établie par ZELLNER (1997, 68).

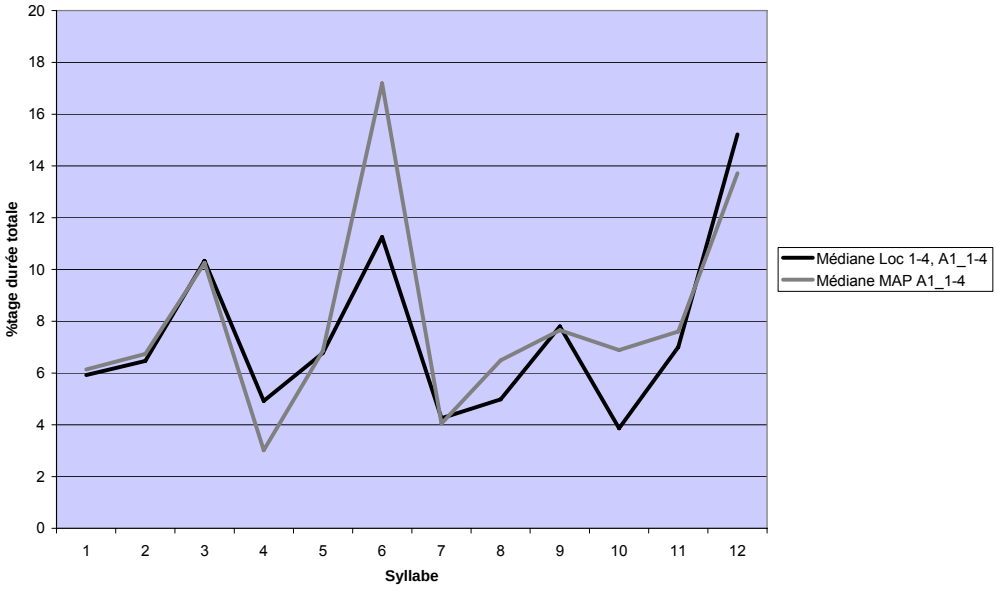

Figure 8 — Durées syllabiques en pourcentage de la durée de tout l'énoncé, pour les locuteurs 1-4 + MoulinAParole, texte A, v. 1-4 (alexandrins classiques), lecture préparée

Les patterns rythmiques de la prose et de l'alexandrin présentent donc, dans cette étude, une base temporelle commune. C'est de la présence ou de l'absence de certains marqueurs à des positions données (pour l'alexandrin classique, par exemple, la présence de marqueurs aux 3^{e}, 6^{e}, 9^{e} et 12^{e} syllabes), ou de l'agencement particulier de ces marqueurs (par exemple l'inversion de l'ordre traditionnel substantif-adjectif que l'on trouve généralement en prose au profit d'un ordre adjectif-substantif) que provient la différence de structures rythmiques observée entre ces deux types d'énoncés. Les micro-variations de durées syllabiques, constatées entre la lecture spontanée et la lecture préparée pour les passages de transition d'un type d'énoncé vers un autre, peuvent ainsi s'expliquer par le fait que les sujets ont été légèrement surpris – lors de la lecture spontanée – par l'agencement différent des marqueurs rythmiques entre la prose et les alexandrins. Cet instant de surprise a entraîné un ajustement que l'on ne retrouve pas lors de la lecture préparée. D'autre part, les micro-variations observées d'un locuteur à l'autre peuvent s'expliquer par la reconnaissance, de la part des locuteurs avec une expérience littéraire plus grande, des agencements de marqueurs constitutifs du rythme alexandrin, reconnaissance qui a entraîné une emphase prosodique, emphase qui est absente chez les locuteurs moins expérimentés, car ils n'ont pas reconnu cet agencement.

Conclusion

Plusieurs conclusions thématiques et méthodologiques émanent de cette discussion. Par divers examens des phénomènes acoustiques sous-jacents au *problème de la césure*, nous sommes arrivés aux constats suivants :

1. Les pauses empiriquement mesurables n'atteignent leur signification linguistique que dans leur contexte phonétique et linguistique.
2. Les pauses empiriquement mesurables ne sont pas les seuls indices acoustiques de la césure. D'autres indices prosodiques (intonation, intensité, accélérations/décélérations) contribuent de manière importante au percept de la césure.
3. Il n'y a que peu de différences mesurables entre les structures temporelles de la poésie et de la prose. Ce sont avant tout les contextes linguistique et sémantique ainsi que la répétition de schémas métriques particuliers, qui nous renseignent sur la nature, la présence ou l'absence d'une structure métrique. En général nous pouvons

émettre l'hypothèse qu'un poète n'impose pas une structure métrique particulière à un texte. Il exerce plutôt un choix diligent de mots et de structures grammaticales dont l'effet ultime et presque inévitable[15] sera de laisser surgir et resurgir une structure temporelle particulière, de manière à évoquer chez l'auditeur la perception d'une structure métrique donnée.

En basant ces conclusions sur des analyses acoustiques détaillées, nous avons pu mettre en évidence le grand intérêt de la *méthodologie empirique phonétique* pour l'étude des structures métriques. De fait, nous considérons qu'une partie de l'avenir des études métriques se fondera sur une telle méthodologie, car cette approche permet de construire une argumentation à partir d'évidences physiques, qui sont quasiment indépendantes d'une perception humaine particulière et assez indépendantes d'une interprétation théorique préconçue. Elle permet enfin d'examiner des hypothèses au moyen de statistiques de plus en plus adaptées aux besoins de l'analyse linguistique et phonétique.

Comme cette contribution l'illustre partiellement, une analyse conduite selon cette méthodologie procède comme suit:

- *Identification des paramètres.* Lors d'une première étape, les paramètres acoustiques pertinents pour une analyse donnée sont identifiés. Par exemple, pour des études ayant trait à la césure dans le contexte de structures métriques poétiques, nous identifions les pauses silencieuses, les modifications de l'intonation et de l'intensité ainsi que les accélérations/décélérations de la parole comme objets à analyser.
- *Création du parseur-analyseur.* Puis un parseur-analyseur (semi-) automatique pour le matériel linguistique à traiter est créé. Pour l'étude de la césure, il peut s'avérer utile, par exemple, d'assembler plusieurs centaines de textes pertinents et de constituer des logiciels de segmentation (des « parseurs ») automatiques permettant d'identifier les endroits dans le texte où les différents types de découpage risquent de se produire. Comme nous avons vu dans cette contribution, ces parseurs ont intérêt à s'inspirer des travaux sur les structures de performance (ou groupes temporels), telles qu'identifiées par diverses études psycholinguistiques.
- *Enregistrement de locuteurs.* Ensuite, une étude pilote sert à vérifier le bon fonctionnement de ce parseur. Il s'agit d'enregistrer un ou deux locuteurs en lecture ou en reproduction semi-spontanée de ces textes et de vérifier auditivement si les découpages prédits par le parseur-analyseur correspondent aux césures effectuées par les locuteurs. Le parseur-analyseur est amendé au besoin.
- *Analyse acoustique.* Il s'en suit une analyse acoustique détaillée des enregistrements sonores. La plus grande partie de cette analyse peut être effectuée automatiquement, p.ex., à l'aide de logiciels comme Praat[16]. Lors de cette phase, on crée des signaux secondaires, tels des extractions de fréquence fondamentale (d'intonation) ou d'intensité. Une certaine partie de l'analyse doit obligatoirement être effectuée manuellement (ou semi-automatiquement), notamment la segmentation phonétique.
- *Système de prédiction.* A partir des mesures vérifiées de ces enregistrements, vient la phase de la prédiction quantitative (Par exemple, prédire des durées, les variations de hauteur de voix ou de niveau d'intensité). Les classifications symboliques produites par le parseur-analyseur sont alors mises en relation avec les mesures acoustiques obtenues

[15] Nos expérimentations portant sur les durées segmentales et syllabiques ont clairement montré que la sélection d'une chaîne donnée de segments phonétiques (phones) détermine en grande mesure la structure métrique d'un passage oral. Exprimé en termes statistiques, l'identité phonémique d'un segment, du segment précédent et du segment suivant explique la plus grande partie de la variance systématique de durée du segment en question. Les autres contributions à cette variation (position dans la syllabe, dans le lexème ou dans le syntagme, etc.) ne sont responsables que d'une partie circonscrite de la variation temporelle des segments et syllabes.

[16] Logiciel multi-plateforme rendu disponible gratuitement par l'Université d'Amsterdam, voir www.praat.org.

sur la parole produite. Ceci implique généralement la création d'un modèle de prédiction statistique (par exemple, un « modèle linéaire général ») ou neuro-mimétique (un modèle de découverte de relations quantitatives non linéaires). Cette étape finit sur la génération d'un système de prédiction prosodique relativement autonome. Idéalement, ce système est en mesure de générer de manière fiable les phénomènes acoustiques en question. Par exemple, ce système produira de manière crédible différents types de césure aux bons endroits dans le texte.

- *Intégration dans un système de synthèse de la parole.* Idéalement, ce système de prédiction est ensuite intégré dans un système de synthèse de la parole, à des fins de vérification. Par exemple, le système "lira à haute voix" les textes et produira des pauses aux endroits prédits. Dans le cas de langues anciennes pour lesquelles de telles synthèses n'existent pas encore, ceci suppose, bien sûr, un effort supplémentaire. Cependant, les expériences de notre laboratoire montrent qu'il est possible, dans le cadre d'un effort bien structuré s'étendant sur deux ou trois ans, d'arriver à constituer un tel système.

Ce type de procédure basée sur une méthodologie empirique est intéressante car elle offre l'avantage de l'objectivation et de la duplication des analyses dans différents laboratoires, selon des protocoles explicites. Cette méthodologie est de plus en plus accessible aux étudiants de la métrique classique. Dans notre laboratoire, deux systèmes de synthèse de la parole pour le français et pour l'allemand contemporain ont été crées selon cette approche. De plus, deux systèmes initiaux pour l'anglais britannique et le latin classique sont en cours de réalisation. Notre objectif à long terme est de documenter de manière de plus en plus détaillée le fonctionnement de tels systèmes d'analyse-synthèse, afin de permettre à de nombreux autres laboratoires d'expérimenter les concepts et les outils que nous avons développés pour ce type de recherches.

Références bibliographiques

BOLDRINI, S. (1992) - *La metrica e la prosodia dei Romani*, Roma.

BOOMER, D.S., & DITMANN, A.T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, *5*, 215.

BUTCHER, A. (1980). Pause and syntactic structure. In W. DECHERT & M. RAUPACH (Eds.), *Temporal variables in speech* (pp. 86-90). Mouton.

GOLDMAN–EISLER, F. (1972). Pauses, clauses, sentences. Language and Speech, 15, 103-113.

GROSJEAN, F. , & DESCHAMPS, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. Phonetica, 31. 144-184.

GROSJEAN, F., & DOMMERGUES, J.Y. (1983). Les structures de performance en psycholinguistique. L'Année psychologique, 83, 513-536.

GROSJEAN, F., GROSJEAN, L., & LANE, H. (1979). The patterns of silence: Performance structures in sentence production. Cognitive Psychology, 11, 58-81.

KELLER, E. (à paraître). La vérification d'hypothèses linguistiques au moyen de la synthèse de la parole. Cahiers de l'institut de linguistique 28, Université de Louvain.

KELLER, E., ZELLNER, B., WERNER, S., and BLANCHOUD, N. (1993). The prediction of prosodic timing: Rules for final syllable lenthening in French. Proceedings ESCA Workshop on Prosody, September 27-29. Lund, Sweden. 212-215.

LEVELT, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

MILNER J.-C. et REGNAULT F. (1987) - *Dire le vers*, Paris.

MONNIN, P., & GROSJEAN, F. (1993). Les structures de performance en français: caractérisation et prédiction. L'Année Psychologique, 93, 9-30.

ZELLNER, B. (1992). Le bé bégayage et euh... l'hésitation en français spontané. Actes des 19eme J.E.P. (pp. 481-487). Bruxelles.

ZELLNER, B. (1994). Pauses and the temporal structure of speech, in E. KELLER (Ed.) *Fundamentals of speech synthesis and speech recognition.* (pp. 41-62). Chichester: John Wiley.

ZELLNER, B. (1996). Structures temporelles et structures prosodiques en français lu. Revue Française de Linguistique Appliquée: La communication parlée. 1. (pp.7-23).Paris.

ZELLNER, B. (1997) - “La fluidité en synthèse de la parole”, dans KELLER E. et ZELLNER B. (éds.), *Les défis actuels en synthèse de la parole*, EL 1997/3, 47-78.

ZELLNER KELLER B. (2002.a). Revisiting the Status of Speech Rhythm. in Bernard BEL & Isabelle MARLIEN (eds.), 2002. Proceedings of the Speech Prosody 2002 conference, 11-13 April 2002.(pp. 727-730) Aix-en-Provence: Laboratoire Parole et Langage. ISBN 2-9518233-0-4.

ZELLNER KELLER B. (2002.b). La simulation du rythme de parole. Travaux de l'Institut de Phonétique de Strasbourg. TIPS 31 (pp. 139-165). ISDN 0750-1315.