Corela

Cognition, représentation, langage

HS-9 | 2010

Parole

Vers une synthèse vocale du français classique à l'aide du logiciel ©KALI

Philippe Caron et Michel Morel

Édition électronique

URL : <http://journals.openedition.org/corela/1050>

DOI : 10.4000/corela.1050

ISSN : 1638-573X

Éditeur

Cercle linguistique du Centre et de l'Ouest - CerLICO

Référence électronique

Philippe Caron et Michel Morel, « Vers une synthèse vocale du français classique à l'aide du logiciel ©KALI », *Corela* [En ligne], HS-9 | 2010, mis en ligne le 05 mars 2010, consulté le 03 mai 2019. URL : <http://journals.openedition.org/corela/1050> ; DOI : 10.4000/corela.1050

Ce document a été généré automatiquement le 3 mai 2019.

Corela – cognition, représentation, langage est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Vers une synthèse vocale du français classique à l'aide du logiciel ©KALI

Philippe Caron et Michel Morel

1. Introduction

1.1. Objet

- 1 Les récentes expériences de reconstitution orale du français classique¹, à travers des pièces de Lully ou de Molière, font apparaître des erreurs, souvent spectaculaires, en contradiction avec les témoignages du temps, traduisant un manque de lien avec les chercheurs². De plus, il n'y a pas eu jusqu'ici, à notre connaissance, de tentative de reconstitution du français classique par synthèse de la parole. Cette technique présente pourtant l'intérêt d'étendre l'oralisation expérimentale de quelques pièces à l'ensemble des écrits d'une période donnée et d'aider les chanteurs et les comédiens de manière interactive.

1.2. Faisabilité

- 2 Tous nos témoins potentiels étant morts depuis longtemps, il n'y a pas de témoins directs. D'autre part la tradition orale de cette diction s'est perdue, tant au théâtre français qu'à l'opéra. Il y a donc totale solution de continuité. Nous n'avons plus qu'un accès très indirect à l'hypothétique vérité que nous traquons : à la façon d'un détective, nous disposons d'indices dont nous devons déchiffrer la pertinence et, en les croisant, nous pouvons tenter ce qu'en jargon judiciaire on appelle une reconstitution. Nous n'accédons donc pas à la chose, nous tentons de modéliser une sorte d'image-robot. Mais la période offre toutefois des atouts pour une opération périlleuse comme celle-ci. En effet autour de 1700, nous commençons à pouvoir croiser des données directes ou indirectes. Directes lorsqu'elles traitent explicitement de prononciation, indirectes lorsqu'à propos d'une

autre question (la réforme de l'orthographe par exemple) elles livrent des indices sur la prononciation. Croiser des données suppose des sources multiples et si possible indépendantes les unes des autres. Cette condition fonde la science historique dont un des principes s'énonce ainsi, comme on sait : *testis unus, testis nullus*. Un seul témoin, pas de témoin. Or un climat sociologique propice à la fixation des usages est à l'œuvre sous le règne personnel de Louis XIV. D'où des traités, des manuels, des discours savants, des tentatives de réforme de l'orthographe, des remarques de langue qui, recoupés, offrent des indices. Il s'agit seulement d'indices pour la raison que ce que nous glanons n'est pas toujours d'une interprétation univoque et cela pour au moins deux raisons : le lexique employé n'est pas un jargon de spécialité univoque et d'autre part ce que nous cherchons à comprendre n'est pas toujours écrit pour indiquer la juste prononciation. Ainsi la nouvelle graphie de Vaudelin [Vaudelin 1713] est destinée à optimiser l'acte d'écrire le français, non à nous renseigner sur sa prononciation. Nous détournons donc sa *Nouvelle façon d'écrire comme on parle en France* pour en faire une lecture symptomale de la prononciation du temps. Il faut garder toujours à l'esprit cette distorsion que nous faisons subir à la lecture.

- 3 Ces sources s'acheminent en outre vers une certaine saturation. On peut entendre par saturation le fait que l'énoncé devient de plus en plus auto-suffisant, ne présupposant pas obscurément des connaissances préalables ou une pratique naturelle de l'idiome. Il suffit de voir par exemple comment d'une édition à une autre de son œuvre maîtresse Jean Hindret [1687, 1696] enrichit son discours : il y a un pas considérable de l'une à l'autre au point qu'on a l'impression d'une autre œuvre, explicite, fine, très soigneusement discursive. Si l'on croise ce second témoignage avec *La nouvelle manière d'écrire comme on parle en France* de Vaudelin,³ qu'on y ajoute le témoignage très circonspect de l'abbé de Dangeau (1754, mais rédaction dans la dernière décennie du XVII^e siècle), lequel est ensuite authentifié avec admiration par l'abbé Régnier-Desmarais dans sa *Grammaire française* [1706 :88], alors on a un faisceau assez intéressant qui converge par exemple vers l'idée que les articulations vocaliques nasales du français sont vraiment des voyelles monophthongues vers 1680.
- 4 Enfin ces sources s'essayaient parfois à une certaine exhaustivité, c'est-à-dire que par rapport aux époques précédentes les zones aveugles de la prononciation sont moins béantes. Hindret [1696] à cet égard est un modèle mais la *Prosodie française* de d'Olivet écrite vers 1730 est aussi de ces travaux qui visent à mieux couvrir le champ de la longueur vocalique en français. Tout cela invite prudemment à l'audace.

Quel niveau de langue et sur quelle période ?

- 5 Même prudence là encore que pour le point précédent : il ne saurait être question d'offrir un aperçu général des habitus de prononciation sous le règne de Louis XIV.
- 6 Faisons tout de suite un trait sur les français régionaux : stigmatisés ici et là pour une prononciation défectueuse, ils ne font jamais l'objet d'une description exhaustive et saturée. La visée de ceux qui en parlent est toujours corrective. C'est dire qu'aucune systématisme n'est apportée à leur description. Nous n'aurons toujours que des bribes. Même observation pour les sociolectes réprouvés par la norme. Par exemple celui des bourgeois parisiens, drapiers de leur état comme le Monsieur Jourdain du Bourgeois Gentilhomme par exemple.

Il reste donc deux objets oraux complexes que nous pouvons espérer modéliser comme deux objets distincts qu'ils sont :

- L'honnête conversation, c'est-à-dire la diction familière mais conforme au bon usage, un standard en cours d'élaboration permanente et sans cesse évolutif. Ce pourrait être celle d'un aigrefin mais homme de qualité comme Dorante dans le *Bourgeois*. La diction de celui-ci est donc largement moins conjecturale que celle de Monsieur Jourdain.
- La diction haute, qui se déploie sur un continuum qui va de la lecture publique au récitatif en passant par la grande parole publique et le théâtre grave. On peut sans risque d'incompréhension parler là de déclamation. Cet habitus est un artéfact qui s'élabore et qui est plus ou moins bien intériorisé par les acteurs, les orateurs, les chanteurs ou ceux qui pratiquent la lecture publique.

- 7 Sur ces deux habitus, nous disposons d'un ensemble assez riche de sources pour pouvoir espérer procéder à une reconstitution plausible.
- 8 Il faut maintenant se poser la question de savoir si nous pouvons toujours discriminer ces deux dictions. Cette question n'est pas du tout indifférente : les contemporains font couramment la différence car il y a là deux 'dialectes' oraux, dont le deuxième, la diction qu'on peut qualifier de haute, est une diction plus artificielle, plus contrôlée encore que la première. Voici un passage bien connu de Bacilly [1688 : 248] qui effectue nettement la différenciation :

Je diray donc qu'il y a de deux sortes de Prononciations en general qui font naistre bien des doutes & des difficultez dans le Chant. Il y a vne Prononciation simple, qui est pour faire entendre nettement les Paroles, en sorte que l'Auditeur les puisse comprendre distinctement & sans peine ; Mais il y en a vne autre plus forte & plus énergique, qui consiste à donner le poids aux Paroles que l'on recite, & qui a vn grand rapport avec celle qui se fait sur le Theatre & lors qu'il est question de parler en Public, que l'on nomme d'ordinaire *Declamation*.

- 9 Evidemment cette réduction de la complexité, en réalité plus feuilletée, de l'usage est déjà un artéfact. Les documents indiquent plus ou moins précisément un standard, émaillé de zones de variation, ils ne nous renseignent qu'accessoirement sur les pratiques idiolectales qui, toutes, ont leurs idiotismes. Mais il existe quand même une frontière manifeste entre les usages de conversation plus ou moins familiers et ceux de la parole publique. La manière de nommer ces deux « dialectes » varie mais l'existence de désignations distinctes en cooccurrence, dans les ouvrages de remarques par exemple, ne laisse aucun doute sur la conscience claire que les observateurs ont d'un double registre de la diction. Peut-on pour autant distinguer systématiquement les deux ? La réponse doit être mesurée : il n'est pas simple de décoller les deux images sonores car si les commentateurs sont conscients des particularités des deux, ils ne spécifient pas toujours si ce qu'ils disent de l'honnête conversation vaut aussi pour la diction haute. Nous sommes assez heureux lorsque Tallemant [1698 : 88] commente la prononciation du pronom personnel *ils* en contexte aval vocalique : 'La seule [prononciation] où l'on s'accorde est la prononciation entiere [ilz] dans le style soutenu & dans les vers [...] aussi est-elle entierement bannie du discours ordinaire.'
- 10 Ici la distinction est très nettement marquée dans le texte mais le chercheur est loin d'avoir toujours ce même confort de conscience. Aussi sommes-nous souvent conduits à procéder par inférence : si les commentateurs ne stipulent pas qu'il existe une variante haute de la prononciation d'un item, alors nous en déduisons (avec une marge de risque qu'il nous faut prendre si nous voulons aller à l'exhaustivité) que ce qui vaut pour l'honnête conversation est transposable vers la diction haute.

- 11 Enfin il reste la délicate question de la période de validité de nos sources. Là encore il convient de rester très prudent. Les limites de pertinence des ouvrages que nous compulsions sont évidentes en amont et en aval. Nous manquons d'informations assez exhaustives en amont pour vérifier la permanence. En aval, nous savons que des réorganisations de grande ampleur sont en cours dans la seconde moitié du dix-huitième siècle dans le dialecte oral dominant. Une tranche chronologique de cinquante ans, à peu près coextensive du règne personnel de Louis XIV (1661-1715) semble être une zone de validité raisonnable.

1.3. Principes de la mise en œuvre

- 12 Une fois définies les règles de prononciation et de prosodie, avec une part intuitive inévitable que l'on tentera de réduire par la suite en approfondissant la recherche, il est possible à un locuteur averti et entraîné de simuler la lecture de textes de l'époque choisie. L'intérêt de cette simulation, de cette reconstitution de corpus oral, est double : (1) proposer une diction plausible pour le chant et le théâtre, à destination des amateurs et des professionnels ; (2) présenter un objet concret, à la portée du public, illustrant l'élocution de cette époque.
- 13 Mais comme les textes disponibles représentent une quantité de corpus importante (plusieurs milliers d'heures de lecture), il est exclu d'enregistrer la lecture humaine d'une part significative de ceux-ci. La synthèse vocale est alors un moyen d'étendre cette simulation à l'ensemble des textes par la création d'un modèle à partir de morceaux choisis (250000 mots) et de seulement 30 mn de lecture d'un corpus fabriqué pour la circonstance. Bien qu'imparfait, le résultat doit être suffisamment performant pour fournir une prononciation correcte de toutes les phrases et reproduire pour l'essentiel l'intonation du locuteur en respectant au maximum le rythme, le naturel et l'expressivité.

2. Ressources linguistiques pour la synthèse de parole

2.1. Ressources de base

- 14 Pour modéliser une langue parlée – actuelle ou ancienne, nous devons d'abord sélectionner un corpus représentatif, puis l'oraliser, le corpus oral étant la base de la synthèse vocale. En pratique, nous devons disposer des ressources suivantes :
- un corpus écrit, numérisé en format texte, d'environ 250000 mots, constitué d'extraits représentatifs des productions écrites de l'époque concernée,
 - un dictionnaire grammatical de la plupart des mots du corpus écrit,
 - un dictionnaire de prononciation de ces mêmes mots,
 - un corpus oral d'environ 30 mn, enregistré par un locuteur simulant la diction de l'époque sur des extraits sélectionnés pour la variété des prononciations d'une part et des unités prosodiques d'autre part.
- 15 Ces ressources de base sont à fabriquer à partir des corpus dont nous disposons, de nos connaissances et de nos hypothèses. Il s'agit plus d'un travail de recherche que d'un travail de développement.
- 16 Nous avons décidé de nous appuyer pour l'oralisation du vers (qui constitue notre objectif numéro un) sur un corpus de passages les plus nettement récitatifs de deux tragédies

lyriques de Lully : Atys et Armide, dont les livrets sont dus à la collaboration de Philippe Quinault.

2.2. Ressources modélisées

- 17 Le synthétiseur de parole est un programme – un moteur en quelque sorte – dont le fonctionnement ne dépend pas de la langue à synthétiser. Il utilise pour chaque langue des ressources linguistiques et phonétiques spécifiques sous une forme qu'il est capable d'interpréter et de généraliser. Le travail de développement consiste à fabriquer ces ressources modélisées à partir des ressources de base, soit :
- un dictionnaire pragmatique (probabilité de mise en relief),
 - un dictionnaire des catégories grammaticales,
 - un dictionnaire des racines verbales,
 - un dictionnaire des flexions verbales,
 - un jeu de règles de désambiguïsation des catégories grammaticales,
 - un jeu de règles de groupement et de mise en relation des unités prosodiques,
 - un jeu de règles de transcription graphème-phonème,
 - une base de diphtongues,
 - un dictionnaire de contours prosodiques.
- 18 Une fois ces ressources modélisées intégrées au logiciel Kali, celui-ci est capable de prononcer toute phrase du corpus complet. La lecture du corpus écrit par la synthèse permet de cerner les erreurs, les omissions, les défauts et de les corriger en amont. La synthèse de la parole permet aussi, par sa souplesse, de faire avancer la recherche en stimulant une démarche fondamentale et expérimentale riche (essais, comparaisons, possibilité d'options, etc.)

3. Architecture de la synthèse vocale Kali

- 19 Le logiciel Kali est utilisé dans l'industrie, les applications multimédia, les serveurs vocaux, le handicap de la vue ou de la parole. Il reçoit un texte en entrée et fournit de la parole en sortie [Morel 2001]. Celle-ci doit être intelligible et dans la mesure du possible agréable, naturelle et expressive. Pour la restitution d'une langue ancienne, la parole doit être aussi fidèle que possible au modèle, notamment sur les plans de l'articulation et du rythme.

Le texte à prononcer traverse successivement les modules suivants :

3.1. Analyse syntactico-prosodique et pragmatique

- 20 L'analyse syntactico-prosodique fournit un découpage du texte en constituants prosodiques ainsi que leurs relations de dépendance [Vergne 1999]. Pour les besoins de la prosodie, l'analyse syntactico-prosodique fusionne en un même constituant prosodique les syntagmes élémentaires ou tronçons qui sont fortement liés entre eux, par exemple *pronom personnel + verbe* ou *"être"* ou *"avoir" + groupe nominal*. L'analyse pragmatique du texte (organisation informationnelle et expressivité) complète le modèle prosodique en repérant les constituants prosodiques à mettre en relief ainsi que leur modalité affective, améliorant ainsi le naturel et facilitant l'accès au sens.

3.2. Transcription graphème-phonème

- 21 Pour prononcer correctement un texte, il faut appliquer de nombreuses règles de prononciation. Pour s'en faire une idée, il suffit de comparer les prononciations des mots : *chocolat* vs. *chorégraphe*, *les options* vs. *nous options*, *aliment* vs. *animé*, *écarter* vs. *carter*, etc. Le module de transcription est chargé de fournir un texte phonétique à partir du texte alphabétique [Morel 1998]. Le français comporte environ mille règles, l'anglais plusieurs milliers. Les mots d'emprunt et les noms propres nécessitent plusieurs milliers de règles supplémentaires.

3.3. Base de diphones

- 22 A partir de la voix d'un locuteur, on extrait une base de diphones, c'est-à-dire 1000 à 2000 segments de signal qui seront ensuite concaténés pour former le signal acoustique de parole. Chaque segment va de la moitié d'un phonème à la moitié du suivant. L'interface entre les segments est le phonème, ce qui limite le nombre de segments nécessaires à la fabrication de n'importe quel énoncé.
- 23 Les diphones étant prélevés dans des contextes différents, il est nécessaire de les modifier pour que leur raccordement ne présente pas de discontinuités. Des traitements complexes à base de fenêtres de Hanning et de transformées de Fourier entrent en jeu.

3.4. Prosodie

- 24 Le rôle de la prosodie est de donner à la lecture un rythme et une intonation naturels, agréables, expressifs, et de faciliter l'accès au sens en découpant l'énoncé en groupes de mots fortement liés entre eux, plus ou moins mis en relief. Pour l'heure, la prosodie de Kali a été schématisée en traits principaux : déclinaison, pauses, allongements, accents. Un modèle utilisant un dictionnaire de contours prosodiques naturels est en cours de réalisation. En partant du principe que la forme de chaque patron prosodique émerge de la combinaison de contraintes multiniveaux (rythmiques, syntaxiques, pragmatiques) [Lacheret 2002], la méthode utilisée consiste à associer une clé à chaque contour entré dans le dictionnaire, c'est-à-dire une représentation mathématique de ces contraintes. Le principe général consiste à coder des domaines (de taille et de nature linguistique variable, du groupe accentuel au paragraphe), des distributions (initiale, médiane, finale du domaine) et des fonctions (modale, stylistique) associées à ces domaines. Il suffit qu'un des paramètres change pour qu'une autre clé soit produite. La phase de génération quant à elle consiste à calculer, sur la base de l'analyse automatique du texte (cf. *supra*), la clé associée à chaque segment, puis à la comparer aux clés en mémoire grâce à une fonction de coût (distance entre les deux clés). Cette méthode permettra ainsi de sélectionner dans le dictionnaire la forme prosodique la plus pertinente et de l'appliquer à l'unité en cours de traitement.

3.5. Générateur de parole

- 25 Le module générateur de parole convertit la chaîne phonétique en un signal de parole par concaténation des diphones de la base. Les contours prosodiques sélectionnés sont

concaténés et les valeurs acoustiques sont interpolées sur toute la phrase puis appliquées en temps réel. Le signal de parole ainsi enrichi est envoyé en continu vers les haut-parleurs à fréquence d'échantillonnage constante.

4. Plan de travail

26 Le développement de ce projet basé sur la synthèse vocale Kali, implique la réalisation des points suivants (les points déjà réalisés sont en italique) :

- Collecte et numérisation d'un corpus écrit, aussi varié et exhaustif que possible.
- Définition d'un jeu de phones (réalisations réelles des phonèmes) du français aux environs de 1700, à partir des connaissances acquises.
- Fabrication d'un jeu de règles de transcription graphème-phonème (environ 1500 règles), en mode "déclamation" (réalisation des schwas, des liaisons et de la synérèse selon les règles en vigueur).
- Phonétisation automatique d'un corpus écrit aussi large et varié que possible ("grand corpus").
- Fabrication d'une liste de phrases à partir de ce corpus, par sélection des parties les plus productives en termes de diphtongues différentes. Des phrases complémentaires seront fabriquées, afin de réaliser des diphtongues rares risquant d'être rencontrés dans des textes non encore numérisés. Il va de soi que ces 150 à 200 phrases hors contexte ne constituent qu'un outil de travail, en aucun cas un corpus de démonstration.
- Enregistrement d'un corpus oral à partir de ces phrases, pour le prélèvement des diphtongues. L'oralisation sera effectuée avec le souci d'une diction phonétique soignée, prioritairement à la prosodie.
- Prélèvement des diphtongues du corpus oral (environ 1000 à 1500) et fabrication de la voix.
- Construction de dictionnaires et de règles pour l'analyse syntaxique et pragmatique. Toutes les variantes orthographiques pouvant être rencontrées dans les textes devront être prévues (nécessité d'un prétraitement).
- Construction d'un petit corpus écrit pour le prélèvement de contours prosodiques, par sélection des parties les plus productives du grand corpus en termes de variabilité syntaxique et rythmique.
- Oralisation de ce petit corpus (corpus prosodie) avec le souci d'une prosodie soignée, prioritairement à la diction phonétique.
- Prélèvement des contours prosodiques de plusieurs centaines d'unités prosodiques du corpus prosodie.
- Intégration dans Kali d'une voix en français classique.

5. Dialogue entre développement et recherche

27 Le développement des ressources linguistiques soulève de nombreux problèmes en lien avec la recherche fondamentale. La nécessaire exhaustivité du travail de développement pousse le chercheur à affiner ses hypothèses et parfois à prendre parti sur des points qui sont en discussion. Nous traiterons ici de quelques problèmes liés à la transcription graphème-phonème.

5.1. Variation orthographique

- 28 L'orthographe du français n'est pas encore stabilisée à la fin du XVII^e siècle. De plus, des corrections ont pu être apportées ultérieurement et de nombreux textes ont été transcrits en orthographe contemporaine. Il fallait pouvoir exploiter tous les textes, y compris ces derniers. Plutôt que de multiplier les entrées des dictionnaires grammaticaux et de prévoir de nombreux cas particuliers dans la transcription graphème-phonème, nous avons préféré mettre en place un prétraitement, une réécriture des textes sous une forme orthographique stable, qui facilite l'analyse du texte et sa transcription phonétique. Il s'agit bien sûr d'un code interne, sans relation avec un éventuel état de la langue ayant pu exister.
- 29 Dans ce code interne, nous avons conservé – ou ajouté – les accents afin d'éviter les ambiguïtés de prononciation et de reprendre avec peu de modifications les règles d'analyse syntaxiques du français contemporain. Dans le même esprit, certains "i" ont été transformés en "y" (*aiant* → *ayant*). En revanche, les "oi" qui se prononcent /wɛ/ ont été conservés, même si leur équivalent actuel est "ai", afin de faciliter la phonétisation. Nous avons donc retranscrit en "oi" les "ai" des textes transcrits en français contemporain lorsque leur prononciation est /wɛ/.
- 30 Ce prétraitement par dictionnaire est inopérant dans les cas d'homographie. Par exemple, les participes passé pluriels et les noms qui en dérivent sont homographes des formes verbales à la deuxième personne du pluriel (*les blessez, ils sont blessez, vous blessez*). Les règles d'étiquetage par déduction contextuelle permettent néanmoins de résoudre la majorité des cas.

5.2. Allongement phonologique vs. prosodique

- 31 A qui se rend familier des manuels de poésie, de grammaire ou d'orthoépie du XVII^e siècle, il ne fait pas de doute, au bout d'un certain temps, que le vocalisme du français classique diffère sensiblement, en tant que système, de celui du français contemporain. Mais comme il n'existe pas qu'un seul français à cette époque (tout comme aujourd'hui), il convient d'être plus précis en spécifiant que nous parlons exclusivement dans cet article de cette forme de diction qu'est la diction haute du français. Plus généralement, de cette variété normée de français que l'on peut appeler à la suite de Vaugelas le bon usage, en sachant très bien que certains français régionaux de cette époque (celui de Provence notamment, comme le font remarquer Morin et Dagenais [1988]) ne présentent pas de longueurs phonologiques.⁴

5.2.1. La longueur vocalique est-elle phonologique en français classique ?

- 32 Que nous prenions à cet égard les ouvrages de Jean Hindret [1687 et 1696], de Mourgues [1697 pour la seconde édition] ou de d'Olivet [1736 pour la première édition], force est de nous rendre à l'évidence suivante : tous entendent très nettement des longueurs vocaliques différentes dans la diction mais, chose plus troublante pour un locuteur du XXI^e siècle, ces longueurs ont, en bien des cas, un statut phonologique, c'est à dire qu'elles sont distinctives. On parle alors, pour adopter un jargon plus précis, de 'quantités vocaliques'. Ainsi, selon que la voyelle [e] est longue [e:] ou brève [e], on obtient une paire minimale pour le mot 'pied'. En effet le singulier est bref /e/ et le pluriel est long /

e :/. En d'autres termes, le locuteur natif du XVII^e siècle n'a pas besoin d'un déterminant pour savoir s'il s'agit d'un pied ou de deux pieds. Les exemples en sont si nombreux en français classique qu'il devient absolument clair, à la lecture des auteurs, que la longueur a vraiment un statut phonologique, puisqu'elle distingue des paires minimales comme 'faîte' /ε:/ et 'faite' /ε/⁵. Quiconque s'aventure dans la diction du français classique soutenu ne saurait viser juste s'il ne prend conscience de cette réalité incontournable. D'ailleurs les remarques désobligeantes sont nombreuses contre les normands par exemple qui abrègent ce qui est long et allongent ce qui est bref. Certes la réalisation de ces longueurs dans la substance sonore varie selon la cadence générale de la diction (selon qu'on est en pure citation conversationnelle, en lecture publique ou au théâtre) mais en valeur relative, ces contrastes font partie du système de la langue, ils demeurent actualisés de façon distinctive puisqu'ils jouent un rôle fondamental dans la forme de l'expression⁶. Tous les timbres vocaliques⁷ du temps (sauf le schwa toujours très bref) comportent donc deux voyelles, une brève et une longue. Voici quelques exemples de paires minimales :

- [a] : Parti vs partis
- [e] : Aimé vs aimés
- [ε] : Faîte vs faite
- [a] : Tache vs tasche
- [y] : Fut vs fût
- [œ] : Feu vs feux
- [u] : Clou vs clous
- [O] : Pot vs pots

5.2.2. Le domaine labile de la longueur positionnelle en prose

- 33 Une fois ce premier niveau systémique admis, il existe un deuxième niveau à explorer avant d'entrer enfin dans ce qui est spécifique au vers : celui de la prosodie générale du français. La prosodie comporte dans sa description les pauses, les longueurs, les hauteurs et le volume. Sur la question des longueurs, et corrélativement des pauses, la situation est plus confuse que pour le premier. En effet les grammairiens du temps, même fins observateurs de la prononciation, déclarent le plus souvent forfait, perplexes devant la complexité du phénomène auquel ils sont confrontés. Et la métrique latine, qui déclare longue une pénultième qui se termine par deux consonnes consécutives, joue encore un rôle parasite dans la juste appréhension du mécanisme accentuel du français. Même l'abbé d'Olivet [1771 : 67], pourtant fin observateur, déclare forfait dans sa *Prosodie française* rédigée dans les années 1730 :

Ainsi à parler exactement, nous aurions cinq temps syllabiques, puisqu'on pourroit diviser nos syllabes en muettes, breves, moins breves, longues, & plus longues. Mais il est inutile de tant anatomiser les sons : & nous n'avons qu'à suivre l'exemple des Grecs & des Latins, qui ne connoissoient que breves, longues, & douteuses.

- 34 On voit dans cet extrait qu'après avoir nettement pris conscience du phénomène prosodique en général dans le domaine des longueurs, l'auteur préfère s'en tenir au socle phonologique de la quantité dans sa variété de français : ce niveau qui permet d'opposer, même en dehors de la syllabe accentuable du mot deux lexèmes par la seule quantité vocalique : matin vs matin, i.e. /ma'tē/ vs /ma:'tē/ (le chien). Pourquoi cette perplexité et ce sentiment d'impuissance ? C'est que la question du rythme est encore dans les limbes et qu'elle est perturbée par les essais de poésie mesurée à l'antique. La langue française est-elle nombreuse, se demande-t-on ? L'échec de la poésie en vers mesurés semble avoir

beaucoup troublé les esprits. De fait la rigidité de la syntaxe du français interdit de faire des vers « à l'antique » qui ne soient pas contournés, bizarres, voire carrément amphigouriques. Les auteurs du XVII^e siècle en tirent de là une conclusion qui est erronée : la langue française n'est pas nombreuse. Or il faut distinguer les possibilités offertes par le système d'un côté et ce qu'on a essayé d'en faire de l'autre. L'échec des vers mesurés n'est pas pour autant la preuve qu'il n'existe pas de prosodie esthétiquement mobilisable en français. Il indique simplement qu'elle ne peut se prêter à l'usage qu'en faisaient la langue grecque ancienne et la langue latine.

- 35 Essayons toutefois de mettre, dans ce lieu si confus, un peu de netteté. Quels sont, en bref, les différents niveaux d'organisation rythmique de la diction du français au-delà de la quantité vocalique ?

La place et l'incidence de l'accent de mot

- 36 En français, la saillance acoustique de l'accent de mot est faible relativement à d'autres langues néo-latines mais ce n'est pas une raison pour le déclarer nul. Cet accent se loge toujours, s'il est réalisé dans la substance sonore, sur la dernière syllabe (non muette) d'un mot non clitique. Il suffit de prononcer « du chocolat » en mettant l'accent de la façon suivante : [dy 'ʃokola] et on se rendra compte qu'on a tendance à comprendre quelque chose comme ' du chaud cola'. L'accent juste est bien sûr sur la dernière syllabe du syntagme [dy ʃoko'la]. Donc, même discret, l'accent en français est absolument présent. Mais sa réalisation va varier selon plusieurs critères et règles.
- 37 La règle majeure ici est que, parmi les effets acoustiques possibles d'un accent, figure un allongement du segment vocalique. Mais cet allongement n'est pas mécanique car tout laisse à penser qu'il est en français classique d'une durée proportionnelle à la place du mot dans l'organisation hiérarchique du syntagme auquel il appartient. Quand on dit « tout laisse à penser », c'est qu'il y a une sorte de saut périlleux qui consiste à postuler que de ce point de vue, le français classique fonctionne comme le français contemporain. Il faut prendre ce risque en l'absence d'indications formelles du temps⁸ et sous réserve d'inventaires de plus en plus serrés.
- 38 En français contemporain, la tête prosodique d'un syntagme étant toujours à droite, l'accent le plus saillant d'un syntagme (donc l'allongement le plus perceptible) reposera sur la dernière syllabe (non muette) du dernier mot de celui-ci. Ainsi dans cet hémistiche du *Saint-Genest* de Rotrou, 'verser d'indignes larmes', le syntagme nominal 'd'indignes larmes' comprendra deux accents mais la longueur vocalique réalisée a toute chance d'être plus grande sur 'larmes' que sur 'indignes' car, la tête prosodique du syntagme étant à droite, l'accent prosodique sur 'larmes' est plus saillant que sur le /i/ d' 'indignes'. Cette première considération, toute simple, fait que la longueur vocalique réalisée va désormais reposer sur des considérations non seulement phonologiques mais sur des critères positionnels complexes. Le résultat en termes d'allongement acoustique sera évidemment varié selon que la voyelle accentuée se trouvera dans une position plus ou moins saillante à l'intérieur du syntagme où elle s'insère.
- 39 Les règles de réalisation de l'accent de mot prosodique selon sa position (neutralisation, déplacement en fonction des clitiques droits, saillance selon la position de celui-ci dans le syntagme ou l'unité de souffle) sont d'une grande complexité et elles ne font pas l'unanimité selon les modèles théoriques. Renvoyons sur cette question à la synthèse de Lacheret et Beaugendre [1999]. Disons simplement qu'à cette lumière la prosodie du

français classique est d'une grande sophistication (par rapport à celle du français contemporain) dans la mesure où la longueur vocalique combine à la fois :

- la quantité phonologique de base,
- la place, inaccentuée ou accentuable, de la voyelle,
- si elle est accentuée, la position de cet accent à l'intérieur du syntagme et de l'unité de souffle.
- Ceci sous réserve de réorganisations accentuelles pour des raisons d'équilibre ou de trop grande proximité.

40 Cela a pour conséquence évidente qu'une voyelle phonologiquement longue en français classique comme le /ɛ:/ de *faîte/faîtage* va recevoir une traduction dans la substance sonore variable selon que :

- la voyelle est sous l'accent potentiel : elle l'est dans 'faîte' [fɛ:t⁰], elle ne l'est plus dans 'faîtage' [fɛ•'ta⁰]. Du coup la quantité phonologique longue sera d'autant plus saillante acoustiquement qu'elle est sous l'accent possible dans le premier cas et au contraire moins saillante dans le second cas.
- La voyelle accentuée est ou pas en position forte dans son environnement syntaxique.

On contrastera par exemple le même mot 'faîte' selon qu'il est en finale absolue d'une unité de souffle dans (1) ou en position secondaire dans le syntagme de (2) :

Qu'est-ce qui est le plus fragile dans la construction ?

(1) - Le *faîte*.

(2) - Le *faîte* en tuiles.

41 Cette matière, on le voit, est complexe mais il ne faut pas renoncer à la mettre en règle. Tout en sachant que nous sommes, pour le français classique, dans une position très délicate puisque les théoriciens eux-mêmes peinent à élaborer les règles qui régissent la prosodie. Toutefois nous pouvons, avec d'innombrables précautions, émettre l'hypothèse qu'en travaillant sur les récitatifs des tragédies lyriques nous pourrions nous aider de la ligne musicale en prenant bien garde aux effets locaux dictés par les exigences strictement musicales ou les affects mimés par le chanteur [Voir Caron-Morel 2006 : 19]. Ce chapitre constitue à lui seul la matière d'un article. Qu'il nous suffise de renvoyer ici à Bettens [Bettens 200] qui, sur des dépouillements importants distingue plusieurs phases dans l'histoire du vers chanté au XVII^e siècle et conclut de façon méthodique sur le parti prudent que l'on peut tirer d'une « rétroconversion » des propriétés musicales d'un récitatif sur la prosodie de la déclamation.

42 Pour l'heure, nous avons défini le symbole ' : ' qui provoque un allongement vocalique de 20 % dans certaines configurations de milieu de mot (*manger, lundi, abondant, gauche, râble, naître, feste, etc.*), dans les fins de mot phonétiquement en voyelle-r (*canard, heure, enviers, empire, remords, amour, allure, etc.*) et les terminaisons en voyelle-s, voyelle-x et voyelle-z (*repas, heureux, après, repris, ennuis, etc.*). Il s'agit ici de l'allongement phonologique. La valeur de cet allongement pourra être modifiée selon le cas et selon l'évolution de nos connaissances. De son côté, la prosodie se superposera à ce phénomène phonologique. Elle agira sur la durée – des longues comme des brèves – en fonction des contraintes syntaxiques et pragmatiques qui résulteront de l'analyse du texte et du modèle prosodique associé qui reste à créer.

5.3. Diérèse / synérèse

- 43 Pour définir les cas de diérèse et de synérèse dans la déclamation, nous avons choisi de passer au crible plusieurs pièces de Corneille, Racine et Molière en alexandrins afin de vérifier la régularité des règles en vigueur. Le comptage des syllabes permet de classer les mots concernés dans chacune des deux catégories. Nous avons pu constater que les règles étaient stables, en synchronie comme en diachronie. Nous n'avons rencontré que deux problèmes :
- un faux alexandrin de 11 syllabes dans *Cinna* : *Vous pourriez m'opposer tant et tant d'obstacles*,
 - une variation sur le mot *paysan/paysanne* dans "l'école des femmes" : 2 syllabes pour *paysanne* et 3 syllabes pour *paysan*.
- 44 Lors de la mise en place des règles de diérèse/synérèse, un point s'est avéré délicat, nécessitant de nombreuses règles sur les mots entiers : les terminaisons en "iez" et "ions" selon qu'il s'agissait de verbes à l'imparfait (synérèse) ou de verbes au présent ou de noms (diérèse).

5.4. Schwas

- 45 Comme pour la diérèse/synérèse, nous avons utilisé les alexandrins de notre corpus. Là aussi les règles sont stables : les schwas sont toujours conservés (*Qu'on se contentera de s'en rire sous main*) sauf dans deux cas :
- schwas de fin de mot, sans consonne finale, suivis d'un mot commençant par une voyelle (*En un mot, qu'elle soit d'un(e) ignoranc(e) extrême*),
 - fin de vers (*Je reçus et je vois le jour que je respir(e) / Sans que mere...*)
- 46 Encore dans ce dernier cas le schwa est-il parfois faiblement articulé pour ne pas apparaître comme une syllabe surnuméraire.
- 47 En prose, l'élision du schwa se fait normalement devant une pause, mais en poésie, c'est la fin du vers qui compte comme pause (même sans ponctuation) et les virgules placées à l'intérieur des vers sont ignorées à l'oral (*Tout cede, tout ressent ses funestes efforts*). La synthèse vocale devra donc dans ce cas effectuer un prétraitement particulier sur la ponctuation afin de restituer le rythme et la prononciation adéquats.
- 48 Dans le chant, la réalisation du schwa est encore plus marquée : même en fin de vers, celui-ci est conservé et chanté :

5.5. Oralisation des consonnes terminales, liaisons

- 49 Ce secteur est sans doute l'un des plus complexes et il est hors de question de le traiter dans les limites d'un tel article. Le cas particulier de l'oralisation des marques de pluriel à la pause que nous ne ferons qu'évoquer ici est révélateur de cette complexité. A la différence des disciples d'Eugène Green dans la diction du *Bourgeois Gentilhomme* en

DVD (voir ci-dessus note 1 et aussi Caron 2008 pour une critique suivie), nous pensons qu'il n'y a pas d'argument réel pour en soutenir l'oralisation marquée vers 1700 : le *Traité de la poésie françoise* de Mourgues qui date des années 1680 [1724 : 72] apporte nettement sa contribution au diagnostic :

Quatrièmement, Que la licence [en poésie] ne doit jamais aller jusqu'à faire rimer le Singulier avec le Plurier, sous prétexte que l's ou le z qui distingue les deux Nombres ne se fait point entendre dans la prononciation, quant les mots terminent par ces deux lettres coupent ou finissent un sens ; comme font tous ceux qui sont placez à la fin du Vers, où celui qui lit ou qui récite est obligé de faire une pause. Ainsi il n'est point d'exemple de ces Rimes.

Larme : charmes, bonté : beautez, desir : plaisirs, desert : concerts. [Les mots sont soulignés par moi et les accolades qui conjoignent les paires de mots sont remplacées par les deux points].

- 50 Il faut donc bien lire que malgré l'évolution phonétique qui gomme l'oralisation du pluriel hors liaison, les vers proscrivent les rimes singulier-pluriel. Sans doute parce que la longueur, elle, continue à varier. Une deuxième source, convergente, est Hindret. Dans l'édition de 1696 [II, 743], Hindret pose de façon basique une première règle concernant les consonnes finales en général :

La premiere Regle que nous avons à donner touchant les consonnes finales qui ne se prononcent point, est que la dernière consone des mots qui finissent une phrase ou une periode, ne se prononce ni dans le discours ordinaire, ni dans celui qui se fait en public. Prononcez donc, *Il aïait le premié ; Araité vou ; doné man dè plû bô ; Tout ce kila de plû maïchan ;* pour dire il étoit <sic> le premier ; Arrêtez vous ; Donnez m'en des plus beaux ; Tout ce qu'il a de plus méchant.

- 51 Il est intéressant que l'abbé d'Olivet [1771 : 46-47], secrétaire perpétuel de l'Académie française, sanctionne (tardivement il est vrai) cette zone de consensus, y compris pour les vers. Et c'est de la façon la plus incidente qu'il nous renseigne, à propos d'un autre phénomène, l'oralisation des schwas :

Allons plus loin. Puisque l'E muet écrit, ou non écrit, ne fait qu'une différence oculaire, voyons de conséquence en conséquence, où ceci nous conduira. Voici des paroles à mettre en chant.

Esprits, qui portez le tonnerre,
Impétueux tyrans des airs,
Qui faites le péril des mers,
Et les ravages de la terre,
Vents, &c. Ode du P. de la Rue

J'avoue que mon oreille n'en sait point assez pour distinguer le son de ces quatre rimes. Je n'entends qu'*erre* par-tout, en supposant qu'on ne fera pas mal-à-propos, & contre l'usage, sonner des s, d'airs & de mers, où elles ne sont que signes du pluriel. [les mots sont soulignés par nous]

Que nous dit exactement cet ajout tardif de la *Prosodie françoise* à propos des marques de pluriel ?

- Que d'Olivet considère ici la frange haute de la diction haute, puisque nous sommes dans l'ode.
- Que dans ce dialecte il est hors de question d'oraliser l's d' *airs* et de *mers* en fin de vers et qu'un observateur averti n'entend qu'une même terminaison à la fin des quatre vers. Les S ne sont que "signes du pluriel".

- 52 Tout autre est le cas des liaisons, qui nécessiterait un autre développement. D'après les témoignages du temps, toutes les consonnes terminales non prononcées sont des liaisons potentielles en vers, y compris à l'hémistiche, bien que dans cette position, la liaison soit

considérée comme moins évidente, voire disgracieuse ou facultative selon les auteurs. Mais la nature syntaxique de la frontière entre les mots génère des cas de figure différents qui, de surcroît, ont un traitement différencié selon que l'on est dans la parole oratoire, le vers déclamé ou le chant.

6. Résultats attendus

- 53 Plusieurs types de résultats sont attendus : d'une part, ce travail permet de concrétiser l'état d'avancement de la recherche en lui offrant une visibilité nouvelle ; ce faisant, il fournit un matériel expérimental souple en généralisant l'oralisation d'un corpus de travail à celui de l'ensemble des corpus écrits. D'autre part, au-delà de la recherche, il permet de conseiller les professionnels du théâtre ou de la musique baroque de façon conviviale. Enfin, il peut jouer un rôle pédagogique auprès du public scolaire, universitaire, enseignant ou autre.

BIBLIOGRAPHIE

Bacilly, 1679. *L'Art de bien chanter de M. de Bacilly. Augmenté d'un discours qui sert de réponse à la critique de ce traité et d'une plus ample instruction pour ceux qui aspirent à la perfection de cet Art*. Paris, Chez l'auteur. (Genève, Minkoff, 1971).

Caron Philippe & Morel Michel, 2006. « Note récapitulative sur le programme de reconstitution de la diction haute du français vers 1700 ». *Annales de l' Association pour un Centre de Recherche sur les Arts du Spectacle aux XVII^e et XVIII^e siècles*, Vol 1 (2007), pp. 17-24.

Caron, Philippe 2008. « Pouvons-nous reconstituer la diction haute du français vers 1700 ? A propos du Bourgeois Gentilhomme en DVD ». Maney, *17th century French studies*, volume 30-2 pp. 182-195.

Dangeau, Louis Courcillon de, 1700. *Essais de Grammaire contenus en trois discours*. Paris, Coignard. dans *Opuscles sur la langue française par divers académiciens* (Paris : Brunet, 1754) (Genève : Slatkine, 1969).

Hindret, Jean, 1687. *L'art de bien prononcer et de bien parler la langue française*. Paris, Laurent d'Houry. (Genève : Slatkine, 1973).

Hindret, Jean, 1696. *L'art de prononcer parfaitement la langue française*. Paris, Laurent d'Houry. (Paris, Bibliothèque de l'Association pour un Centre de Recherche sur les Arts du Spectacle aux XVII^e et XVIII^e siècles, 2007).

Lacheret-Dujour, Anne & Beaugendre, Frédéric, 1999. *La prosodie du français*, CNRS Editions.

Lacheret-Dujour, Anne, 2002. *Modélisation prosodique du français parlé : analyse de la substance, représentation formelle, interprétation symbolique*, Dossier d'habilitation à diriger des recherches, Université de Paris X, Nanterre, juillet 2002.

Morel, Michel & Lacheret-Dujour Anne, 1998. « Utilisation d'une structure arborescente pour une hiérarchisation fine des règles de transcription graphème-phonème », *Actes des XXIIèmes Journées d'Etudes sur la Parole*, 15-19 juin 1998, Martigny, Suisse, pp. 151-154.

Morel, Michel & Lacheret-Dujour, Anne, 2001. « Le logiciel de synthèse vocale Kali : de la conception à la mise en œuvre », *Traitement Automatique des Langues*, Ch. D'Alessandro (éd.), Paris : Hermès, vol. 42, pp. 193-221.

Mourgues, Michel, 1724. *Traité de la poésie française*. Paris : Jacques Vincent, 2^e édition.

Olivet, Pierre-Joseph Thoulhier, abbé d', 1738. « Prosodie française ». dans *Remarques sur la langue française*. Paris, Barbou, 1771. (Genève : Slatkine, 1968).

Régnier-Desmarais François Séraphin, 1706. *Traité de la grammaire française*. Paris, Coignard Paul Tallemant, *Remarques et décisions de l'Académie française*. Recueillies par M. L. T (Paris : Coignard, 1698) (Genève : Slatkine, 1972), p. 88.

Vaudelin Giles, 1713. *Nouvelle maniere d'écrire comme on parle en France*. Paris, Vve Jean Cot et J.B. Lamesle.

Vaudelin Giles, 1715. Instructions cretiennes mises en orthographe naturelle, pour faciliter au peuple la lecture de la science du salut. Paris, J.B. Lamesle.

NOTES

1. Voir notamment *Cadmus et Hermione*, opéra de Lully dirigé par Vincent Dumestre et produit en DVD par Alpha et Amiral LDA ainsi que *Le Bourgeois Gentilhomme* produit en DVD par les mêmes producteurs. La diction est supervisée par Benjamin Lazar.
2. Voir notamment Caron 2008 pour une critique partielle de l'habitus articulatoire dans le DVD du *Bourgeois-Gentilhomme*
3. Vaudelin, p. 3.
4. Pour cette notion de diction haute, voir Caron [2006 et à paraître]. Cette distinction est fondamentale pour savoir ce que l'on restitue et comment on le restitue.
5. A quoi on ajoutera, dans le cas précis, que le timbre de la longue /ε:/ se réalise vraiment ouvert d'après les témoignages, assez proche du [æ] de l'anglais actuel 'cat'. Ce qui fait dire à certains grammairiens qu'il y a trois timbres de cavité moyenne dans la série des voyelles d'avant non labiales : fermé, moyen et ouvert.
6. Rappelons à titre d'information secondaire que dans les années 1970, Henriette Walter [Walter 2001] et ses informateurs enregistraient encore nettement, dans certaines variétés de français et chez des personnes âgées de plus de 60 ans des oppositions de longueur manifestement phonologiques. Il s'agit de traces (désormais presque totalement perdues) de l'ancien régime des longueurs en français classique central.
7. Le cas des voyelles nasales est un cas à part car selon d'Olivet [1771 : 73 sv AIM, AIN], la nasale "rend longue" la syllabe dans laquelle elle est inscrite. Il n'y aurait donc pas de nasale intrinsèquement brève.
8. On verra le rôle qu'éventuellement la musique vocale du temps peut avoir pour nous sortir d'un embarras évident

RÉSUMÉS

La reconstitution orale du français classique est un projet qui intéresse à la fois les historiens, les linguistes, les musiciens, les comédiens et le public. La synthèse vocale Kali permet d'automatiser l'oralisation de textes d'une époque donnée à partir de 30 mn de parole d'un lecteur expert. Le synthétiseur de parole est un logiciel qui utilise pour chaque langue des ressources linguistiques et phonétiques allant de l'analyse syntaxique et pragmatique à la fabrication d'une base de diphtongues en passant par un dispositif de transcription graphème-phonème et une bibliothèque de contours prosodiques. Moins naturel que la lecture réelle, le résultat doit cependant être suffisamment performant pour fournir une prononciation correcte de toutes les phrases et reproduire pour l'essentiel l'intonation du locuteur en respectant au mieux le rythme, le naturel et l'expressivité.

This paper deals with the complex issue of the pronunciation of French during the period of King Louis XIV. That question is a keypoint for actors and singers, who need to know how the French language was pronounced in that period. A joint programme between FORELL (University of Poitiers) and CRISCO (University of Caen) research teams was launched in 2004 to collect data from a very large range of sources coming from French speaking countries but also from countries such as Poland or the Low countries or Britain where a good practice of French was a main concern. Not all varieties of French pronunciation are concerned, except for the favoured pronunciation of verses and public occasions. As no direct witness of that period remains any more, scholars have to scrupulously compare those data in order to build up a hypothetical model of its phonetic and prosodic features. A large part of this paper is dedicated to ©KALI computer-aided device which is being implemented by Michel Morel to give an online overview of that pronunciation to any actor or singer needing to assess his/her own practice by 2012.

INDEX

Mots-clés : langue française 17^e siècle, Phonétique et prosodie, synthèse de parole

Keywords : 17th century French language, phonetics and prosody, speech synthesis.

AUTEURS

PHILIPPE CARON

Universités de Poitiers et Cambridge

MICHEL MOREL

Université de Caen