

**INTONATION, STRUCTURE PROSODIQUE
ET VERSIFICATION**

**INTONATION, PROSODIC STRUCTURE
AND VERSIFICATION**

**ENTONACIÓN, ESTRUCTURA PROSÓDICA
Y VERSIFICACIÓN**

PHILIPPE MARTIN
Université de Paris

Résumé: Les contraintes rythmiques fortes de la versification classique ont évolué dans la versification contemporaine vers une totale liberté. Alors que la versification classique, par des contraintes rythmiques et des permutations syntaxiques, rend l'accès au sens parfois difficile, la versification contemporaine au contraire rend plus facile l'accès au sens poétique en optimisant la mise en jeu des mécanismes cognitifs des lecteurs et auditeurs par l'intonation et les pauses.

Mots-clés: versification, structure rythmique, structure prosodique, groupe accentuel.

Abstract: The strong rhythmic constraints of classical versification have evolved in contemporary versification towards total freedom. While classical versification, by rhythmic constraints and syntactic permutations, sometimes makes access to meaning difficult, contemporary versification on the contrary

makes access to poetic meaning easier by optimizing the use of cognitive mechanisms of readers and listeners by intonation and pauses.

Keywords: versification, rhythmic structure, prosodic structure, accent phrase.

Resumen: Las fuertes restricciones rítmicas de la versificación clásica han evolucionado en la versificación contemporánea hacia la libertad total. Mientras que la versificación clásica, por restricciones rítmicas y permutaciones sintácticas, a veces dificulta el acceso al significado, la versificación contemporánea, por el contrario, facilita el acceso al significado poético al optimizar el uso de mecanismos cognitivos de los lectores y oyentes por la entonación y las pausas.

Palabras clave: versificación, estructura rítmica, estructura prosódica, grupo acentual.

Introduction

A la question, qu'est-ce qu'un vers ? Qu'est-ce qui distingue *Chassez le naturel il revient au galop de Tout condamné à mort aura la tête tranchée* ? Alors que les structures rythmiques de ces deux exemples sont similaires et composées de deux segments de six syllabes, je dirai paraphrasant Jean-Claude Anscombe¹ que c'est l'intention du texte qui importe, intention qui se manifeste par la manière de dire ou de lire le vers. L'intention se manifeste dans la diction ou dans la lecture par le rythme et l'intonation telle que révélée par les contours mélodiques placés sur les voyelles accentuées terminant chaque groupe accentuel. Alors que le vers impose un débit de parole moyen conduisant à quatre groupes accentuels :

[ChassEZ] [le naturEl] [il reviENT] [au galOp]

une lecture rapide du second exemple peut mener à seulement deux groupes accentuels :

[Tout condamné à mOrt] [aura la tête tranchÉE]

Alors que pour le vers les mouvements mélodiques sont bien marqués par des variations de fréquence descendantes et montantes perçues comme telles et non comme tons statiques :

[ChassEZ ↘ le naturEl] ↗ [il reviENT → au galOp] ↓

ces mêmes variations dans l'extrait du code pénal seront bien moins marquées du fait de la neutralisation des contours dans leur fonction indicatrice de la structure prosodique, neutralisation manifestée par des variations mélodiques moins amples, moins marquées, plus plates :

¹ ANSCOMBE, Jean-Claude, « Parole proverbiale et structures métriques », *Langages*, 139 (2000), pp. 6-26.

[Tout condamné à mOrt] → [aura la tête tranchÉe] ↓

Se pourrait-il que des différences comparables puissent être observées dans la versification, effets qui différencieraient des styles et des époques différentes, comme le vers classique et la poésie contemporaine ?

Lecteurs et auditeurs

Avant l'accès au sens poétique, il y a pour le lecteur comme pour l'auditeur l'accès au sens littéral du texte. Pour y accéder, lecteurs comme auditeurs ont à leur disposition quantité de mécanismes syntaxiques, lexicaux, morphologiques, sémantiques, etc. Mais une caractéristique importante différencie les auditeurs des lecteurs : alors que ces derniers sont maîtres du temps et du chemin pris dans l'exploration du texte, ce n'est pas du tout le cas pour les auditeurs, pour lesquels le déroulement du temps est imposé de manière inexorable. La mise en œuvre des outils disponibles à loisir pour les lecteurs est pour les auditeurs impérativement limitée dans le temps.

C'est à partir de cette constatation que j'évoquerai quelques courts exemples de l'évolution de la versification du xvième au xxième siècle du point de vue de l'accès au sens et au sens poétique, en examinant comment le choix des ressources prosodiques disponibles dans la langue va évoluer pour aider l'auditeur à l'accès au sens et au sens poétique.

Parmi ces ressources, les plus importantes sont a) le phrasé, c'est-à-dire la segmentation en groupes accentuels ; b) la structure prosodique, qui assemble ces groupes accentuels hiérarchiquement en permettant une appréhension préliminaire de la structure syntaxique du texte ; et c) les pauses, essentielles on le verra pour donner assez de temps à l'auditeur pour le décodage d'un texte oral dans un continuum de parole.

Syllabes et voyelles accentuées

Les mouvements mélodiques caractéristiques évoqués dans l'introduction sont attachés aux syllabes accentuées, et essentiellement aux voyelles de ces syllabes accentuées. On ne s'en rend pas toujours compte, mais l'énonciation d'un vers, d'un proverbe, ou de n'importe quel texte, opère une segmentation du flux de parole non pas en mots, mais en groupes de mots, appelés groupes rythmiques

ou accentuels. Chaque groupe accentuel ne comporte qu'une seule syllabe accentuée (hors accent d'insistance) qui, en français, est toujours en position finale du groupe, et porte donc sur la dernière syllabe du dernier mot du groupe.

Or, cette segmentation dépend du débit de parole du diseur de vers, ou de la vitesse de lecture du lecteur. Ainsi, *Chassez le naturel il revient au galop* peut être dit ou lu à haute voix ou silencieusement avec 4 syllabes accentuées :

[ChassEZ] [le naturEI] [il reviENT] [au galOp]

et former 4 groupes accentuels, mais peut également être prononcé avec un débit ou une lecture plus rapide, avec seulement deux syllabes accentuées et donc deux groupes accentuels :

[Chassez le naturEI] [il revient au galOp]

À l'inverse, chaque mot ou même chaque syllabe peut être prononcée séparément lors d'une diction très lente, de manière à obtenir autant de groupes accentuels que de syllabes

[ChA] [ssEZ] [IE] [nA] [tU] [rEI] [II] [rE] [viENT] [AU] [gA] [IOp]

D'un autre côté, il semble très difficile de ne réaliser qu'une seule syllabe accentuée pour l'entièreté du proverbe, sans accentuer aucune autre syllabe que la dernière :

?[Chassez le naturel il revient au galOp],

en incluant les 12 syllabes dans un seule groupe accentuel, et ce même en lecture silencieuse. On atteint là la limite des durées syllabiques qui peuvent être non seulement prononcées, mais aussi perçues, et ce qui peut paraître plus étonnant, perçues en lecture silencieuse.

Comme on l'a vu plus haut, les deux modes rapide et lent « ne conviennent pas » au proverbe, et seule une vitesse moyenne semble appropriée. Mais on peut alors se demander si et comment la versification interagit avec cette liberté (relative) de débit. Serait-ce en utilisant une disposition graphique particulière du texte, comme dans le poème de Victor Hugo, où chaque groupe accentuel n'a qu'une seule syllabe :

Fort
 Belle
 Elle
 Dort ?

Je vais tenter de montrer que le passage de la versification classique dans la tragédie du 17^{ème} siècle à la poésie contemporaine correspond aussi à un passage d'un modèle fondé d'abord sur le rythme contraint par des règles, l'accès au sens ne venant qu'ensuite, à un modèle où la (ré)génération du sens, et surtout du sens poétique, est primordiale, le rythme, libre, ne se révélant que dans un second temps. La versification classique apparaît comme un corset autant pour le diseur que pour le lecteur, qui peut cependant jouer des libertés offertes par le système langagier, et en particulier par les règles prosodiques rythmiques en modulant éventuellement le débit de parole par le ralenti, l'accélération, et les pauses.

La versification moderne ne corsète pas la lecture en imposant des structures rigides en strophes, hémistiches, coupures, et rimes, assonances ou allitérations. Elle procède en imposant une élaboration du sens grâce à une disposition graphique du texte qui, instanciée à l'oral, conduit l'auditeur et le lecteur à un assemblage spécifique des groupes de mots par les pauses dont les durées sont déterminantes, assemblage qui ne serait pas obtenu par une disposition linéaire du même texte.

Pour démontrer ce point, je ferai appel à trois notions liées à l'accès au sens : le phrasé, la structure prosodique de la phrase, et la mémoire à court terme du son de la parole.

Le phrasé

On peut lire syllabe par syllabe, mot à mot, mais le plus souvent on lit groupe accentuel par groupe accentuel². Un groupe accentuel, comme son nom l'indique, est un groupe de mots ne contenant qu'une seule syllabe accentuée, hors accent d'insistance. En parlant très lentement, on peut prononcer chaque syllabe d'une phrase séparément, et chacune d'elles constitue alors un groupe accentuel. De la même manière, on peut prononcer chacun des mots d'une phrase séparément, et chacun d'eux constitue alors un groupe accentuel, avec la dernière syllabe de chaque mot accentué. Mais dans

² BLANCHE-BENVENISTE, Claire, « La naissance des syntagmes dans les hésitations et répétitions du parler », en J.-L. Araoui (ed.), *Le sens et la mesure. Hommages à Benoît de Cornulier*, Paris, Honoré Champion, 2003, pp. 40-55.

la plupart des cas, pour obtenir une meilleure efficacité, on parle et aussi on lit en traitant plusieurs mots ensemble comme s'il ne s'agissait que d'un seul. Cet ensemble constitue un groupe accentuel, dont la dernière syllabe (mais parfois l'avant dernière) est accentuée en français.

Cette segmentation du flot de parole en groupes accentuels est appelée phrasé, en référence au phrasé musical, découpage du flot de notes musicales en mesures rythmiques. Selon le débit de parole du locuteur, mais aussi de la vitesse de lecture du lecteur, on peut donc avoir un phrasé par syllabes, par mots ou par groupes de mots. Ce qui est remarquable vient de ce que ces différentes variations de phrasé sont encadrées par des limites de nature uniquement temporelle. En effet, pour détacher un mot en syllabes, il nous faut les séparer dans la prononciation par un silence, une courte pause, d'environ 250 ms, soit un quart de seconde, et cette durée est indépendante des mécanismes articulatoires.

De même, le regroupement de mots dans un seul groupe accentuel est limité par la durée d'énonciation à haute voix de ce groupe, durée qui ne peut dépasser 1250-1350 ms environ, soit une seconde et un quart³. Ceci implique qu'en parlant vite, on pourra inclure plus de syllabes dans un seul groupe accentuel, et donc plus de mots, qu'en parlant lentement. Alors que la valeur moyenne de débit de parole est de l'ordre de 4 à 5 syllabes par seconde, les débits les plus rapides atteignant 9 à 10 syllabes par seconde (parole des jeunes de banlieue), la limite de perception étant de 100 ms par syllabe⁴. Le nombre de syllabes des mots les plus fréquents étant de 3,5 (loi de Zipf), on peut s'attendre à ce qu'un groupe accentuel moyen contienne 2 ou 3 mots, soit 2 à 3 groupes accentuels par seconde.

Mais le phrasé est aussi contraint, ou du moins encouragé, à segmenter la parole en groupes accentuels qui correspondent à des unités qu'on pourrait à première vue qualifier de syntaxiques ou sémantiques (pas vraiment des « groupes de sens », mais des parties de groupes de sens). En fait, ces groupes ont pour vocation dans la communication linguistique de correspondre à des séquences enregistrées dans la mémoire du locuteur, et idéalement aussi dans

³ MARTIN, Philippe, *Intonation, structure prosodique et ondes cérébrales*, London, ISTE, 2018.

⁴ GHITZA, Oded, «Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm», *Frontiers in Psychology*, 2, 130 (2011). [<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3127251/>]

la mémoire de l'auditeur. Les sujets parlants ont constitué par leur pratique du langage un lexique non seulement de mots et de syllabes (en particulier lors de l'apprentissage de la langue), mais aussi et surtout de groupes accentuels complets, pouvant comprendre plusieurs mots. Ce qui va également les définir, ce n'est pas le nombre total de syllabes, mais que la durée qu'il faut pour les prononcer. Cette durée est nécessairement supérieure à 250 ms et inférieure à quelque 1250-1350 ms. Ces valeurs temporelles sont apparemment liées à la plage de variation des ondes cérébrales delta (de 0,8 Hz à 4 Hz environ), impliquées dans la perception des groupes accentuels dans la parole⁵.

On comprend que le rythme d'énonciation d'un vers ne dépend pas uniquement du nombre de syllabes et de la position des syllabes accentuées. Elle dépend aussi et surtout du débit de parole du diseur ou du lecteur, en mode oral ou silencieux, et ce débit devra être tel que la structure rythmique attendue soit effectivement réalisée. Ainsi dans *Chassez le naturel il revient au galop*, si quatre syllabes accentuées sont attendues pour respecter la structure rythmique imposée par le style du proverbe, le débit de parole devra être compris entre 3 et 5 syllabes par seconde. Un débit plus lent de 2 syllabes par seconde aboutirait à un vers de 12 groupes accentuels, et un débit plus rapide à seulement deux groupes.

Petits accommodements avec les règles du phrasé

On sait que dans le vers classique, c'est la répétition des accents de mots, en réalité de l'accent final des groupes accentuels successifs, qui crée le sentiment du rythme. Cet effet est codé par une structure définie en nombre de syllabes, 12 syllabes pour l'alexandrin, chaque vers étant au départ supposé constituer une unité de diction (donc sans considérer un enjambement qui serait rendu nécessaire par la syntaxe). Or, des accommodements avec le respect des règles du phrasé sont fréquents. Ainsi, on trouve chez Georges Brassens :

[SupposEZ qu'un de vOUs puisse Être]
 [Comme le sINge obligé dÊ]
 [Violer un jUge ou une ancÊtre]

et dans une chanson de Françoise Hardy :

⁵ MARTIN, Philippe, *Intonation...*, cit, pp. 235-269.

[Ils s'en vONt]
 [amourEUx]
 [sans peur dU]
 [lendemAIN].

On a donc dans ces exemples deux cas de groupes accentuels non congruents avec la syntaxe : [*obligé de*] et [*sans peur du*].

Des cas différents de groupes accentuels dont la syllabe accentuée n'est pas en position finale sont caractéristiques dans le rap et aussi dans la chanson française contemporaine. Ainsi, BigFlo et Oli dans « C'est dommage » :

[ParfOIs]
 [elle s'imagIne]
 [sous la lumiEre]
 [des projEcteurs]

et

[Vaut mieux vIvre]
 [avec des remOrds]
 [c'est ça le sECret]

exemples dans lesquels on trouve deux groupes accentuels avec des mots accentués sur la pénultième [*des projEcteurs*] et [*c'est ça le sECret*], alors que la syllabe finale n'est pas un e muet.

Mais on rencontre aussi de tels cas dans le vers classique, par exemple dans un vers de Phèdre de Jean Racine, impliquant la prononciation du « e » orthographique (le « e » muet) qui n'est pas réalisé dans l'oral :

[Oui PrINce]
 [je lANguis]
 [je brUle]
 [pour ThésÉE]

Cette même structure rythmique en 3-3-3 impose un découpage tel qu'on y trouve des groupes accentuels surprenants comme [*De ses jeu*] et [*nes erreurs*] en plus d'une syllabe accentuée pénultième dans [*par un indigne*] [*obstacle*], un effet habituel de la prononciation du « e » graphique :

[De ses jEU][nes errEURs] [désormAIs][revenU]
 [par un indIgne] [obstAcle] [il n'est pOInt] [retenU]

On le voit, le respect des règles de l'alexandrin impose toutes sorte de petites entorses faites en se justifiant par un rapport supposé à l'étymologie⁶. Ainsi le « e » graphique apparaît seulement sur demande pour obtenir le nombre de syllabes désirées, 6 dans chaque hémistiche, et si possible 3 dans chaque coupure.

De même, diérèse et synérèse sont des procédés bien pratiques pour ajuster un nombre de syllabes en plus ou en moins a priori vu comme inadéquat. Le prix de ces ajustements est souvent la non congruence de l'énonciation avec la pratique de l'oral, établissant en définitive la prépondérance du rythme sur l'accès au sens.

La diérèse et la synérèse procèdent également à une sorte de trucage faisant ressortir une prononciation soi-disant ancienne basée sur l'étymologie latine, où des semivoyelles + voyelle deviennent des séquences de deux voyelles, ce qui permet de faire surgir une syllabe supplémentaire, ou au contraire des séquences de voyelles, donc de deux syllabes n'en deviennent qu'une seule par le changement de la première en semivoyelle. Outre l'ajustement au nombre de syllabes imposé par le vers, le trop ou trop peu devient nombre exact de syllabes, la diction résultante exhale un parfum archaïque, ce qu'ont bien compris le groupe Les Goguettes⁷ dans un pastiche récent de la tragédie classique, « Le postillon maudit » où les mots *distanciation* et *épicerie* sont prononcés avec 5 syllabes dans *dis.tan.ci.a.tion* et *é.pi.ce.ri.e*, dans les vers

[à toujours respectER] [la distanciatIOn]

et

[qui se ruait d'assAUt] [dans une épicerIe].

On conviendra que ces petits arrangements avec les *règles* d'accentuation et de syllabification du français ne favorisent pas la compréhension facile et rapide des vers dans leur version orale.

La structure prosodique

L'accent dit « tonique », dont la position détermine la frontière « droite » des groupes accentuels en français, n'est pas comme on

⁶ MAZALEYRAT, Jean, *Éléments de métrique française*, Paris, Armand Colin, 1995.

⁷ LES GOGUETTES (consulté le 22-07-2020), *Ô Postillons Maudits*, [https://www.youtube.com/watch?v=Nf8nA8_p5Y8]

l'a dit souvent un accent d'intensité,⁸ c'est même le contraire pour l'accent final de la phrase, le plus souvent d'une intensité acoustique bien inférieure à celle des autres syllabes de la phrase (de l'ordre de -6 dB à parfois -30 dB chez Emmanuel Macron !). En fait, à la différence de l'accent d'insistance résultant d'un libre choix du locuteur, la réalisation de l'accent du groupe accentuel résulte de la combinaison d'une durée et d'une variation mélodique plus importante que les syllabes voisines. On peut donc avoir des syllabes accentuées du fait de leur seule durée accompagnée d'une variation mélodique faible, ou au contraire une durée peu ou pas allongée mais accompagnée d'une forte variation mélodique. Un paramètre relativement commode et pertinent combinant ces deux paramètres en un seul est le glissando, décrivant la vitesse de variation mélodique, c'est-à-dire l'amplitude de variation de fréquence rapportée à la durée de cette variation.

La valeur du glissando pour un segment de parole donné, en l'occurrence une voyelle d'une syllabe accentuée, nous donne une estimation de la perception du changement de fréquence. Une variation mélodique est perçue comme telle, c'est à dire comme un changement de fréquence, lorsque la valeur du glissando dépasse un certain seuil appelé seuil de glissando. En dessous de ce seuil la variation mélodique est perçue comme un changement tonal statique (donc sans variation) situé en fréquence à environ 2/3 de la variation⁹.

Une fois la complexité de l'intonation de la phrase, réduite uniquement aux variations mélodiques à l'endroit des seules voyelles accentuées, un jeu de relations de dépendance entre groupes accentuels apparaît clairement. Ces relations sont indiquées par les contours mélodiques et déterminent en définitive la structure prosodique de la phrase.

Munis d'un critère de perception de la variation mélodique, au-dessus ou en dessous du seuil de glissando, variations montantes ou descendantes, on est tout naturellement conduit à envisager les classes de contours mélodiques suivantes :

Un contour montant de glissando supérieur au seuil donc perçu comme variation mélodique montante, noté $C_{ris} \nearrow$.

⁸ *Ibid.*, pp. 109-110.

⁹ Rossi, Mario, « Le seuil de glissando ou seuil de perception des variations tonales pour la parole », *Phonetica*, 23 (1971), pp. 1-33.

Un contour descendant de glissando supérieur au seuil donc perçu comme variation mélodique descendante, noté $C_{fal} \searrow$.

Un contour descendant de glissando supérieur au seuil donc perçu comme variation mélodique descendante, noté $C_{fal\#} \searrow$, similaire à C_{fal} , mais apparaissant devant une pause d'au moins 250 ms. C'est le contour caractéristique de fin de segments d'une diction lente de phrases longues, utilisées souvent dans le discours politique et dans la dictée scolaire.

Un contour de glissando inférieur au seuil donc perçu comme ton statique et non comme variation mélodique, noté $C_{neu} \rightarrow$. Les variations montantes ou descendantes inférieures au seuil de glissando ne sont pas différenciées, d'où le qualificatif de contour « neutre ».

À ce jeu de 4 contours mélodiques C_{ris} , C_{fal} , $C_{fal\#}$, C_{neu} s'ajoutent les contours dits terminaux conclusifs, qui terminent la phrase sur sa dernière syllabe accentuée (sauf dans des configurations appelée classiquement « propos-thème »). On note ces contours terminaux conclusifs par $C_{dec} \downarrow$, indiquant une modalité déclarative, et $C_{int} \uparrow$, indiquant une modalité interrogative. Le contour terminal déclaratif est caractérisé par sa fréquence finale la plus basse de la phrase (comparée aux autres contours, et non à la courbe mélodique globale). Le contour terminal interrogatif est caractérisé par sa fréquence finale la plus haute de la phrase (toujours comparée aux autres contours). Dans les deux cas, ces variations s'appliquent aux seules voyelles accentuées (et non à une éventuelle consonne voisée finale de la syllabe).

Relations de dépendance

Les contours mélodiques placés sur les voyelles accentuées ne sont pas réalisés au hasard par le locuteur. Ils servent à indiquer des relations de dépendance entre les groupes accentuels, qui en déterminent les regroupements successifs pour former la structure prosodique de la phrase. Dans ce processus, opérant de manière incrémentale au cours du déroulement de la phrase dans le temps, que ce soit oralement ou en lecture silencieuse, les regroupements des groupes accentuels établis par le phrasé sont déterminés par les relations de dépendance entre les groupes et entre les groupes de groupes.

Ainsi un contour $C_{fal\#} \searrow$, descendant et supérieur au seuil de glissando, indique une relation de dépendance envers un contour terminal conclusif $C_{dec} \downarrow$ qui le suit et donc un regroupement du

sous-groupe terminé par $C_{fal\#} \searrow$ avec le sous-groupe terminé par $C_{dec} \downarrow$.

On peut représenter cette relation par $C_{fal\#} \searrow \rightarrow C_{dec} \downarrow$. Les autres relations de dépendance sont alors

$$\begin{array}{l} C_{ris} \nearrow \rightarrow \{C_{fal\#} \searrow, C_{dec} \downarrow\} \\ C_{ris} \searrow \rightarrow C_{fal\#} \nearrow \\ C_{neu} \rightarrow \rightarrow \{C_{fal} \searrow, C_{ris} \nearrow, C_{fal\#} \searrow, C_{dec} \downarrow\} \end{array}$$

Il faut également tenir compte des mécanismes de neutralisation, souvent utilisés en phonologie pour rendre compte de la non réalisation d'un trait phonologique dans des conditions où la réalisation de ce trait devient redondante, et donc non nécessaire pour le locuteur.

Mais pourquoi la structure prosodique existe-t-elle ? On conçoit sans problème que l'intonation puisse indiquer la modalité de la phrase, déclarative ou interrogative par exemple. Ou que l'intonation signale à l'auditeur des émotions et des attitudes particulières, joie, lassitude, énervement ou mépris, arrogance, empathie... Et pourtant l'intonation de la phrase assure une autre fonction, souvent insoupçonnée, et cependant indispensable à tel point qu'on ne peut jamais s'en passer. Même en lecture silencieuse, où aucun son n'est émis, on ne peut appréhender un texte sans reconstituer une parole, une diction du texte qu'on est en train de lire, « dans sa tête ». Et cette reconstitution s'accompagne toujours d'une intonation de phrase. Mais pourquoi ?

L'explication tient essentiellement aux limitations de notre mémoire des sons de la parole à court terme. En effet, l'image sonore des sons de parole que l'on écoute ou que l'on perçoit en lecture silencieuse s'évanouit après très peu de temps, 2 ou 3 secondes au maximum lorsque nous écoutons un locuteur de manière continue, c'est-à-dire que les pauses entre les phrases ou parties de phrases sont très courtes ou inexistantes. Toutefois, lorsqu'une phrase est suivie d'un long silence, nous pouvons garder l'image sonore en mémoire plus longtemps jusqu'à 20 ou 30 secondes, et beaucoup plus longtemps si nous nous répétons cette phrase un certain nombre de fois de manière à l'enregistrer dans notre mémoire à long terme¹⁰. La versification moderne fait largement

¹⁰ VOIR FRANKISH, Clive, « Auditory short-term memory and the perception of speech », *Models of Short-term Memory*, Psychology Press, S. E. Gathercole, 1996, pp. 179-207 et ROLL, Mikael, LINDGREN, Magnus, ALTER, Kai, and HORNE, Merle, « Time-driven effects on parsing during reading », *Brain and Language*, 121, 3 (June 2012), pp. 267-272.

usage de ce fait par l'insertion de pauses dans le texte, suggérées par la ponctuation et surtout par la disposition graphique des vers.

Dans la parole continue, nous sommes donc contraints de procéder à une analyse du contenu d'un segment de parole à intervalles réguliers. Pour des phrases courtes, énoncées en moins de 2 ou 3 secondes, le locuteur ménage souvent une pause, permettant à l'auditeur après la réalisation du contour terminal conclusif d'effectuer l'identification et l'analyse linguistique du segment de parole précédent. Cette pause correspond idéalement à la phase d'inspiration du cycle respiratoire du locuteur, mais en lecture silencieuse, cette synchronisation n'est pas nécessaire et la phase d'inspiration du lecteur n'influe en rien sur le processus d'accès au sens. Pour des phrases longues par contre, il est possible de n'avoir aucune pause à l'intérieur de la phrase, et c'est alors la construction incrémentale de la structure prosodique qui va permettre à l'auditeur d'avoir accès au sens (partiel) des segments précédents. Les positionnements des contours prosodiques, et en particulier celui de $C_{fal\#}$ et de C_{ris} , sont donc très importants, car ils vont déterminer la manière dont les groupes accentuels sont regroupés, et donc dont se fera l'accès au sens du dernier segment entendu.

Les pauses

On a vu que le décodage par l'auditeur ou le lecteur du flot de parole implique d'abord la segmentation en groupes accentuels, simultanément au regroupement hiérarchique de ces groupes pour former par incrémentation la structure prosodique de la phrase. On peut montrer¹¹ que cette opération basée en premier lieu sur l'identification des syllabes accentuées et des contours mélodiques portés par les voyelles de ces syllabes précède l'identification des mots du texte inclus dans chaque groupe accentuels, suivi de l'identification des relations syntaxiques qui les lient le cas échéant. Dans ce mécanisme, ce sont les contours terminaux C_{dec} , descendants devant pause $C_{fal\#}$, montants C_{ris} et éventuellement descendants C_{fal} qui sont susceptibles d'alarmer l'auditeur quant à l'opportunité ou la nécessité de réaliser sans tarder l'accès au sens des segments terminés par ces contours (les contours neutralisés C_{neu} quant à eux jouent plus rarement ce rôle, à moins d'avoir été réalisés comme tels pour précisément éviter d'attirer l'attention de l'auditeur).

¹¹ BLANCHE-BENVENISTE, Claire, « La naissance des syntagmes... », *cit.*

Dans cette opération qui se déroule sur l'axe temporel, le temps laissé à l'auditeur pour effectuer le décodage et la hiérarchisation des groupes accentuels est évidemment très important, et des locuteurs professionnels ont appris à ménager des espaces après des séquences de parole, qui ne devraient pas de toute manière dépasser une durée d'énonciation de 2 ou 3 secondes, au risque que l'auditeur ne puisse plus synchroniser sa compréhension avec le débit de parole qu'on lui impose dans une parole continue. Le rôle des pauses dans la versification a été abondamment étudié dans une perspective littéraire¹².

Dans la parole spontanée, et dans la parole lue d'un texte dépourvu de ponctuation, les pauses n'ont rien d'obligatoire (hors bien sûr les pauses de respiration) et sont donc l'objet d'un choix du locuteur, tout comme les accents d'insistance. Elles ne sont pas obligatoires, mais elles ont une utilité : donner du temps à l'auditeur (ou au lecteur) pour l'accès au sens, et en particulier ici au sens poétique.

On a vu que la mémoire à court-terme ne retenait le son de la voix que pendant deux ou trois secondes en flot de parole continu. L'insertion d'un silence suffisant à la suite d'un certain nombre de groupes accentuels facilite le travail de l'auditeur et du lecteur. La durée laissée aux pauses de respiration (d'inhalation de l'air lors du cycle de respiration est d'environ 500 ms¹³. Dans l'hypothèse d'un débit oral moyen de 4 syllabes par seconde, et un nombre de syllabes moyen par groupe accentuel de 6, on peut estimer le nombre maximum de groupes accentuels qui peuvent être mémorisés pendant 3 secondes, toujours en parole continue, de l'ordre de seulement 2 ou 3 !

Heureusement le locuteur est contraint par son cycle de respiration à interrompre sa production de parole après des durées équivalentes, souvent plus courtes. Par contre en lecture, la vitesse maximale est d'environ 250 ms par groupe accentuel. On peut donc en 3 secondes mémoriser 12 groupes accentuels, ce qui constitue la limite pratique de l'espacement entre des signes de ponctuation successifs, invitant chacun à une courte pause dans la lecture.

¹² Voir par exemple MARTÍNEZ CANTÓN, Clara Isabel, « El silencio del verso. La pausa y sus implicaciones métrico-estilísticas », *Rhythmica*, XVII (2019), pp. 55-81.

¹³ WŁODARCZAK, Marcin, & HELDNER, Mattias, « Respiratory Constraints in Verbal and Non-verbal Communication », *Front Psychol*, 8 (2017), p. 708.

On peut raisonnablement faire l'hypothèse que l'accès au sens poétique d'un vers (ou d'un texte en général) demande plus de temps que le simple accès au sens lexical, c'est-à-dire sans toutes les « figures de style » telles que métaphores, plus ou moins bien répertoriées dans l'analyse littéraire.

Les déplacements et les permutations syntaxiques

Les déplacements syntaxiques fréquents dans la versification classique ne favorisent pas non plus la compréhension du texte à l'oral. Des vers bien connus comme *Contre nous de la tyrannie l'étendard sanglant est levé* ou *Qui va du dieu des morts déshonorer la couche* (*Phèdre*, J. Racine) demandent un recalcul syntaxique pour rétablir la forme attendue, la structure prosodique étant d'un secours limité :

L'étendard sanglant de la tyrannie est levé contre nous

et

Qui va déshonorer la couche du dieu des morts.

Ce recalcul prend environ au moins une demi-seconde après la fin du vers (« Closure Positive Shift »). Dans une production de parole continue, ce recalcul est pratiquement impossible.

En résumé, l'accès au sens et au sens poétique en particulier procède par le phrasé et la reconstruction de la structure prosodique, permettant l'identification des groupes accentuels (durée minimale 250 ms) et le (re)calcul de la structure syntaxique à partir de la hiérarchie des groupes accentuels indiqués par les contours mélodiques placés sur les voyelles accentuées (durée de 500 ms environ). On comprend que des pauses sont bienvenues pour permettre à l'auditeur d'avoir accès au sens du texte.

Phrasé, structure prosodique et pauses dans la versification

Les exemples brièvement traités ici portent successivement sur un court extrait de *Phèdre*, de Racine, interprété par Sarah Bernhardt¹⁴ et par Marie Bell¹⁵, la première strophe de « L'Albatros »

¹⁴ BERNARDT, Sarah (1902), Extrait de *Phèdre*, enregistré par Thomas Edison. [https://www.youtube.com/watch?v=FWGjd39dPg8].

¹⁵ BELL, Marie (1975), Extrait du film français réalisé par Pierre Jourdan. [https://www.youtube.com/watch?v=FIxUPxYNz5M].

de Charles Baudelaire interprété par André Nerman¹⁶, et finalement un poème contemporain de Jean-Pierre Lemaire¹⁷, dit par l'auteur.

Un premier exemple porte sur les 4 premiers vers d'un enregistrement de *Phèdre* (Jean Racine, 1677) par Sarah Bernardt (1844-1923), datant de 1902 (*Phèdre*, Acte 1 Scène 1). Sarah Bernardt a rencontré Thomas Edison lors d'une tournée à New York et y a enregistré cet extrait sur cylindre :

Oui, Prince, je languis, je brûle pour Thésée.
Je l'aime, non point tel que l'ont vu les Enfers,
Volage adorateur de mille objets divers,
Qui va du dieu des morts déshonorer la couche ;

La fig. 1 montre l'analyse acoustique (obtenue par le logiciel WinPitch¹⁸) représentant l'onde de parole et la courbe mélodique, sur laquelle on a ajouté une annotation graphique des contours mélodiques. Ces contours permettent de déterminer la structure prosodique indiquée par les relations de dépendance entre groupes accentuels. La structure prosodique est représentée par des arbres aux branches orthogonales, qui indiquent également la direction de la dépendance (ici toujours « à droite », c'est-à-dire vers un contour situé plus loin dans la phrase). Outre le signal de parole, la figure montre l'évolution de la courbe mélodique, mesure de la fréquence de vibration laryngé des plis vocaux, variant ici de 250 Hz à 360 Hz environ.

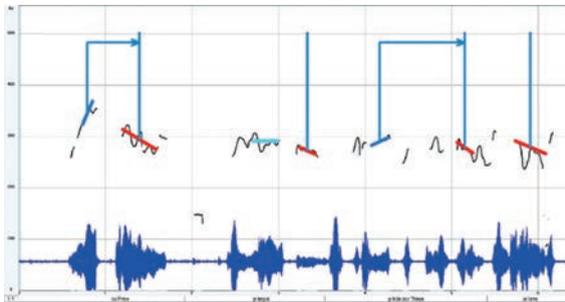


Figure 1. Extrait d'une lecture d'un vers de *Phèdre* de Jean Racine, enregistré en 1902 par Sarah Bernardt. La courbe mélodique variant d'environ 250 Hz à 360 Hz est caractérisée par des vibratos importants.

¹⁶ NERMAN, André (consulté le 21-07-2020) [<https://wheatoncollege.edu/vive-voix/titres/lalbatros/>].

¹⁷ LEMAIRE, Jean-Pierre (consulté le 21-07-2020), *La ménagère*, extrait de *Meubler le silence*. [<https://www.youtube.com/watch?v=wB7FxYzxNQM>].

¹⁸ WINPITCH : logiciel d'analyse acoustique de la parole. [www.winpitch.com].

Sarah Bernardt parvient à réaliser toutes les syllabes, même non-accentuées, avec des vibratos importants oscillant autour des contours mélodiques montants et descendants. Ce type de réalisation était à l'époque la norme dans le discours politique tel enregistré à l'Assemblée Nationale avant 1914¹⁹. On peut toutefois estimer les contours mélodiques des voyelles accentuées à partir des syllabes attendues comme accentuées. Ces contours définissent quatre structures prosodiques indépendantes indiquées entre crochets :

[OuI, PrINce],
 [je lANguis],
 [je brûle pour ThésÉE]
 [je l'AIme]. (Acte 1, scène IV).

Il y a donc une segmentation du texte différente de celle suggérée par la ponctuation (*je languis, je brûle pour Thésée*) et un regroupement prosodique.

Dans la suite de la strophe par contre, [*non point tel que l'ont vu les Enfers, Volage adorateur de mille objets divers*], les contours prosodiques sont tous neutralisés, donnant une structure prosodique « plate » d'énumération, rendant la compréhension de l'interprétation peu aisée.

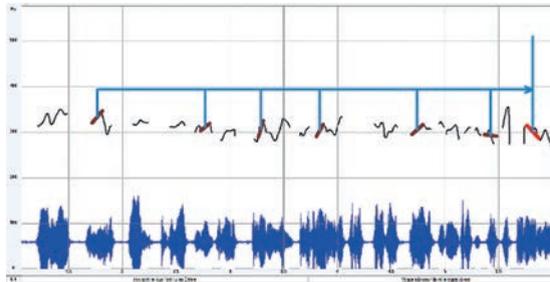


Figure 2. Lecture des vers de *Phèdre* [*non point tel que l'ont vu les Enfers, Volage adorateur de mille objets divers*] par Sarah Bernardt. Outre les vibratos réalisés sur chaque voyelle, on observe que les syllabes accentuées portent toutes un contour neutralisé, résultant en une structure prosodique d'énumération.

¹⁹ MARTIN, Philippe, *Intonation du français*, Paris, Armand-Colin, 2009.

La fig. 3 montre une interprétation plus moderne.

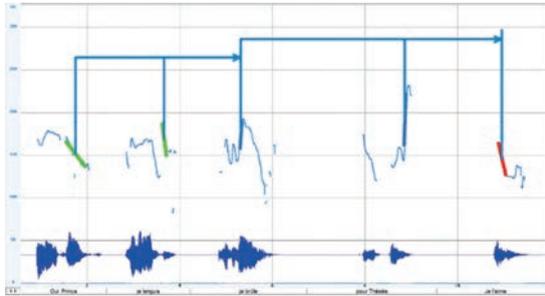


Figure 3. Extrait d'une lecture d'un vers de *Phèdre* de Jean Racine, enregistré en 1975 par Marie Bell.

Dans l'interprétation de Marie Bell (1900-1985), enregistrée en 1975, les vibratos ont disparu, les contours mélodiques sont nettement marqués et supérieurs au seuil de glissando, menant à la structure prosodique regroupant

[Oui, PrINce], [je IANguis], [je brÛle] [pour ThésÉE] [Je l'Aime]

avec un enjambement prosodique et une séparation du groupe syntaxique *je brûle pour Thésée* par une pause importante.

Dans la poésie classique, une correspondance évidente apparaît entre la phrase et la strophe, le syntagme de premier niveau et le vers, la coupure séparant le vers en syntagmes prosodiques et l'hémistiche, ainsi que la césure de l'hémistiche en groupes accentuels. Pour illustrer ce point, j'ai choisi la première strophe d'un poème très connu de Charles Baudelaire, « L'albatros », paru dans le recueil *Les fleurs du mal* en 1857 :

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

La ponctuation, en particulier dans le premier vers, impose un phrasé résultant d'un débit de parole relativement lent :

[SouVENt], [pour s'amusER], [les hOMmes] [d'équIAge]
 [PrEnnent] [des albatrOS], [vastes oisEAUx] [des mERs],
 [Qui suIvent], [indolENTs] [compagnONs] [de voyAge],
 [Le navIre] [glissANT] [sur les gOUffres] [amERs].

Le phrasé et la structure prosodique réalisés par André Nerman sont représentés Fig. 4.

Ici l'effet poétique ne paraît être pleinement accessible qu'en donnant du temps à l'auditeur. Mais comment s'assurer d'un débit de lecture suffisamment lent avec les limites typographiques ?

Ces différentes possibilités laissent donc un choix, Dans cette version l'interprète s'est permis quelques variantes à l'intérieur de contraintes imposées par la ponctuation et la disposition graphique des vers. Ces pauses importantes incitent l'auditeur à une compréhension plus attentive du texte, la parole de l'interprète restant plus longtemps dans la mémoire à court terme.

[SouvENT pour s'amusER,]
 [les hOmmes d'équipAge PrEnnent des albatrOs,]
 [vAstes oisEAUx des mErs,]
 [Qui suIvent,]
 [indolENTs compagnONs de voyAge,]
 [Le navIre glissANT sur les gOUffres amErs]

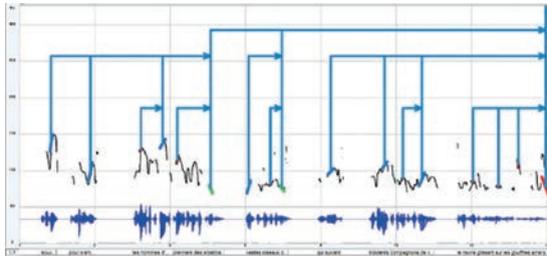


Figure 4. Première strophe de « L'Albatros » de Charles Baudelaire, dit par André Nerman. Analyse acoustique des contours mélodiques et structure prosodique correspondante.

L'accès au sens poétique n'est plus ralenti par le décodage des déplacements syntaxiques caractéristiques de l'époque classique, la ponctuation et la disposition graphique des vers suggérant des pauses bienvenues. Le seul avantage de cet arrangement est de démontrer le respect du format par les rimes. Le même texte avec la seule ponctuation n'invite pas à une lecture avec des pauses aussi importantes : *Souvent, pour s'amuser, les hommes d'équipage prennent des albatros, vastes oiseaux des mers, qui suivent, indolents compagnons de voyage, le navire glissant sur les gouffres amers.*

Un exemple de poésie contemporaine est donné par Jean-Pierre Lemaire (1948-). Il s'agit de « La ménagère », extrait du recueil *Le pays derrière les larmes* publié en 2006. Sans disposition graphique particulière, ce court texte ne semble pas offrir d'interprétation poétique remarquable :

La ménagère. Quand elle a fini de cirer les meubles, d'essuyer les vases, le dos des vieux livres, elle s'assied, la tête vide. Les grains de lumière ont partout remplacé les grains de poussière mais qui verra la différence ? Le soleil seul la félicite.

Mais sa disposition graphique dans le recueil imprime un tout autre rythme :

La ménagère.
 Quand elle a fini de cirer les meubles,
 d'essuyer les vases, le dos des vieux livres,
 elle s'assied, la tête vide.
 Les grains de lumière ont partout remplacé
 les grains de poussière
 mais qui verra la différence ?
 Le soleil seul
 la félicite

La figure 5 permet de visualiser les effets réalisés par l'auteur dans sa lecture :

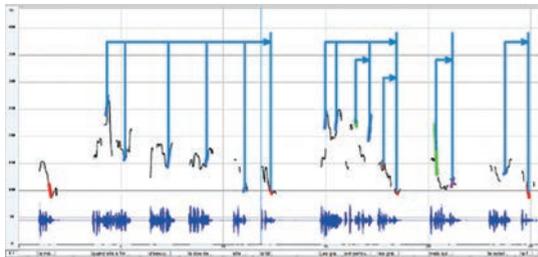


Figure 5. Poème « La ménagère » de Jean-Pierre Lemaire, dit par l'auteur. Analyse acoustique des contours mélodiques et structure prosodique correspondante.

On constate que ce sont essentiellement les pauses et non les contours mélodiques, dont l'alternance montant-descendant est réduite, qui conduisent l'auditeur ou le lecteur à réaliser un accès au sens poétique qui dépasse la simple identification des groupes accentuels successifs, en particulier par l'accélération dans les strophes successives.

[*Quand elle a finI*] [*de cirer les mEUbles*] [*d'essuyer les vAses*]
 [*le dos des vieux lIvres*] [*elle s'assiEd*] [*la tête vIde*] durée 9,2 s. pause 2 s.
 [*Les graINs*] [*de lumiEre*] [*ont partout remplacE*] [*les grains de poussieRE*]
 durée 4 s. pause 1,4 s.
 [*mais quI*] [*verra la différENce*] ? 1,7 s. pause 1,2 s.
 [*Le soleil sEUl*] durée 1,1 s. pause 3 0,5 s.
 [*la félicite*] durée 0,7 s.

Les strophes ont des durées décroissantes jusqu'à l'avant dernière, qui pose une question. La réponse vient après 1,2 s, en deux parties séparées par une demi seconde de silence.

Conclusion

J'ai essayé d'analyser les effets cognitifs dans la lecture des vers classiques et contemporains de l'abandon du corset rythmique en vigueur dans la période classique, et un remplacement par un système de pauses placées à des endroits stratégiques du texte.

D'un certain point de vue, la versification classique brime l'accès au sens, et en particulier l'accès au sens poétique, par un rythme de diction imposé, par un jeu d'ajustement du nombre de syllabes des vers (« e » graphique, diérèse, synérèse), enjambements, permutations syntaxiques, et un usage parcimonieux des pauses, semant une possible confusion chez l'auditeur et rendant la compréhension du texte malaisée. Cependant, contrairement à l'auditeur, le lecteur aura bien moins de difficultés à accéder au sens du texte, ayant tout le loisir d'en déjouer les pièges rythmiques et syntaxiques.

La versification contemporaine par contre, ayant jeté les contraintes classiques aux oubliettes, laisse le texte intact en ne jouant que sur les pauses pour en moduler le rythme, donnant ainsi à l'auditeur comme au lecteur un bien meilleur accès au sens qu'il soit littéral ou poétique.