

## *UN PROGRAMME D'ORDINATEUR POUR LA COUPE SYLLABIQUE EN LATIN*

La syllabe est un élément qui intéresse, à des titres divers, quiconque a choisi la langue ou le style comme sujet de ses recherches.

### *La syllabe en phonétique*

Les linguistes s'accordent pour considérer la syllabe comme une unité de caractère phonétique; elle est donc, pour les phonéticiens, un objet d'observation essentiel.

L'évolution différente qu'elle connaît souvent, selon qu'elle est ouverte ou fermée, prouve l'importance de cette distinction; or repérer ce caractère par l'examen simpliste du signe qui suit la voyelle de base ne suffit pas; s'il est lui-même voyelle, il peut être le second élément d'une diphtongue ou constituer une nouvelle syllabe; s'il est consonne, la réponse ne sera donnée que par le signe qui le suit; autrement dit, le recours à un véritable système de syllabation est indispensable. Il en va de même si l'on s'attache à l'observation et à la comparaison des groupes consonantiques particuliers que la juxtaposition des syllabes crée à l'intérieur des mots.

### *La syllabe en linguistique*

Chaque langue a son génie propre; il détermine son système de composition et de flexion des mots; si ces derniers comportent un nombre variable de

syllabes, toutes les longueurs ne se rencontrent pas avec une égale fréquence; la prédominance de certaines d'entre elles est, sans aucun doute, une caractéristique linguistique essentielle que le spécialiste est curieux de déceler et d'apprécier quantitativement.

Ces recherches, phonétiques ou linguistiques, relatives aux phénomènes généraux d'une langue, n'ont de valeur que si elles portent sur des textes de styles variés et de très longue étendue; elles se conçoivent mal sans l'aide d'un système de syllabation automatique dont elles seraient une des applications les plus simples et les plus utiles.

#### *La syllabe en stylistique*

Bien que moins directement concernée par la syllabation, la stylistique, à l'occasion, y a recours.

"Il n'y a de langue que parlée; en un sens même, on peut dire que la langue écrite est une langue parlée deux fois du fait qu'elle est soumise à l'épreuve de la lecture, c'est-à-dire à un énoncé au moins imaginaire, et par celui qui écrit et par celui qui lira." Ainsi débute le *Traité de Stylistique latine* de J. MAROUZEAU; le tiers de l'ouvrage (1) étudie l'aspect oral de la langue et prouve, par une multitude de témoignages, l'attention particulière, étonnante même à nos yeux, que lui donnaient les Latins. Il montre combien ils étaient sensibles à la sonorité plus ou moins agréable de certains phonèmes ou de certaines rencontres de phonèmes; l'aspect phonique était, en somme, un élément constitutif du style et il était fatal qu'il devienne, surtout chez les poètes et les orateurs, un moyen d'expression parmi tous les autres.

Voulus ou non, des effets apparents proviennent de la valeur expressive des sons, ou même plus simplement, en dehors de celle-ci, de la répétition ou de la rencontre de certains d'entre eux.

Marouzeau propose et commente une longue gamme d'exemples allant de l'hiatus le mieux accepté à la cacophonie, du simple accord entre la sonorité et le sens d'un mot à l'harmonie imitative la plus élaborée. Il note que l'homophonie en début de mot "a un rôle de première importance en latin, à cause peut-être de la valeur particulière attribuée à l'initiale." (2). Il ne néglige pas pour autant l'homéotéleute. Si l'allitération est particulièrement frappante à l'initiale, il n'est pas rare que l'effet soit renforcé par une allitération intérieure. On s'étonne que Marouzeau ne mentionne pas explicitement le phénomène, bien qu'il en ait conscience puisque, parfois, ses exemples le signalent par le truchement de la typographie.

De tels effets peuvent être repérés par une comparaison simpliste des signes, sans syllabation préalable. Mais si l'on ambitionne une appréciation exacte et nuancée des allitérations, qu'elles soient initiales, intérieures ou finales, on ne peut oublier que la similitude des signes n'implique pas nécessairement celle des valeurs phoniques qu'influencent la place dans la syllabe et l'environnement. Ainsi en va-t-il dans *lacto*, *lacrima* et *laceo*, *paupera* et *ultra*, *monstrum* et *donare*, *arctatus* et *arcola*, *coepi* et *coeptus*... L'examen du contexte immédiat (voyelle ou consonne) éclairera, mais il faut admettre que c'est là une très grande approche déjà de la syllabation, et que celle-ci seule empêchera de confondre les *p* de *templum* et *temptare*, les *c* de *arctus* et *inclino*, enserrés entre deux consonnes.

De façon moins apparente, l'aspect phonique de la langue nuance à tout moment l'expression. Le volume, la longueur des mots peut déterminer

l'allure de la phrase (3), mais ce serait un leurre de ne tenir compte que du nombre de signes; comparez *pulchrae* et *amicitia*, deux termes de huit lettres chacun : le second, avec ses syllabes plus nombreuses mais plus courtes, paraît plus long mais plus alerte que le premier; compter le nombre de voyelles ne suffirait pas non plus; le problème, nous le verrons, est moins simple (4) et ici encore sa résolution ramène inmanquablement à la syllabation.

Au moment où l'on envisage de mesurer le mot d'après le nombre de syllabes, on est pris de scrupule. Le français qui, dans sa métrique traditionnelle, a adopté le principe de l'égalité des syllabes autorise le procédé, mais le latin ? Jusqu'à quel point l'opposition des longues et des brèves que la métrique latine implique, rendra-t-elle nos mensurations caduques ? La métrique est une convention déterminée par les réalités phonétiques d'une langue, mais qu'elle a dû simplifier à son profit. La française feint de croire que toutes les syllabes sont égales en durée, ce qui est faux, et l'on ne peut même penser qu'à l'intérieur de l'alexandrin les différences se compensent au point que tous les alexandrins, lus de la voix la plus neutre, aient une longueur égale. La latine, de son côté, oublie que chaque syllabe correspond à une articulation distincte; les auditeurs devaient avoir conscience de ce nombre plus élevé d'articulations là où les brèves dominaient. En sorte qu'il me semble que si la durée de la syllabe est incontestablement une notion beaucoup plus importante en latin, choisir la syllabe comme unité de mesure, en faisant abstraction de sa durée, n'est cependant pas sans fondement et c'est d'ailleurs le seul procédé pratique.

On peut, nous l'avons vu, considérer des effets stylistiques isolés (à la condition, bien entendu, d'avoir une norme de référence); on peut aussi étudier des textes dans toute leur étendue et caractériser les styles selon, par exemple, la longueur moyenne des mots qu'ils utilisent. Dans chacun de

ses deux ouvrages de statistique lexicale (5), Ch. MULLER consacre un chapitre à cette notion. Il y prouve que les recherches quantitatives sur cet aspect du style sont fructueuses puisqu'elles lui ont fait découvrir, entre autres choses, que la longueur moyenne des mots est plus élevée dans les tragédies de Corneille que dans ses comédies, et a tendance à se réduire au cours de la carrière de cet écrivain; ce sont là sûrement deux faits dignes d'intérêt.

Or, pour calculer cette longueur moyenne, Muller a pris la syllabe comme unité de mesure de l'étendue du texte; en effet, il a appliqué ce principe sans devoir recourir à la syllabation parce qu'il a bénéficié de la régularité de l'alexandrin qui représente un groupe fixe de douze syllabes. Il a donc pu se contenter de diviser le nombre total de mots du passage étudié par le nombre de vers. Le procédé est sûrement fort pratique, mais ses possibilités restreintes. Il ne s'adapte qu'à des ensembles de vers uniformes et qui ont un nombre fixe de syllabes, comme c'est le cas des alexandrins (6). Il n'est d'aucun secours pour les textes latins où les vers présentent un nombre variable de syllabes. Il n'est pas non plus applicable à la prose, qu'elle soit latine ou française. On ne peut remplacer l'unité stable qu'est le vers régulier par l'unité trop fluctuante qu'est la ligne, variable selon les typographies, et incomplète en tête et en fin de paragraphe.

Dans ces cas, on pourrait penser à diviser par le nombre de mots le total de tous les signes du texte. Mais on n'obtiendrait que la longueur moyenne graphique des mots, ce qui est une notion qui satisfera plus le typographe que le linguiste, et bien différente de celle que Muller utilise dans ses travaux, puisqu'il se base sur la syllabe.

Pour avoir abordé le problème, je suis d'ailleurs parfaitement convaincue qu'il est extrêmement difficile de concevoir, pour le français, un système de syllabation automatique qui soit pratique et satisfaisant (7). Mais si on possède l'outil - et nous avons tenté de le créer pour le latin - non seulement nous mesurons sans difficulté la longueur moyenne des mots, quelle que soit la forme particulière du texte, mais nous pouvons aller plus loin dans l'appréciation des effets stylistiques qui découlent de la longueur des mots.

On peut, en effet, connaître la longueur moyenne des mots et ignorer les contrastes entre mots courts et mots longs, les accumulations, en certains endroits, de mots brefs ou longs, l'usage de mots dont la longueur excède la norme, leur fréquence et leur répartition. Ainsi, même si nous savons que ces deux vers de Lucrèce (8) comptent chacun 5 mots et 15 syllabes, qu'est-ce qui nous ferait deviner combien leurs structures et donc leurs valeurs stylistiques diffèrent ?

*" Membra movere videbantur vocesque superbas "*

(2,3,4,3,3)

*" Subpeditabatur facies et forma manebat "*

(6,3,1,2,3)

De tels faits caractérisent un style et sont importants à déceler; il faut donc chercher à mesurer la longueur individuelle des mots et c'est par la syllabation qu'on y arrive.

*La syllabe dans les études de statistique linguistique*

Qui dit statistique dit comptage et mensuration.

Pour assurer l'exactitude des résultats, il faut définir avec précision l'étendue du texte considéré. Selon la nature des recherches, l'unité de mesure sera le signe, la syllabe, le mot, le vers., mais il faut savoir que ce choix n'est pas sans conséquence; nous venons de le voir à l'instant à propos du calcul de la longueur moyenne des mots.

On pense souvent à traduire l'étendue d'un texte par le nombre de mots qu'il contient, mais la longueur moyenne des mots étant variable, des extraits comportant un nombre égal de mots peuvent être d'étendues différentes.

Puisqu' "il n'est de langue que parlée", le linguiste pourra préférer à toute autre, une unité de caractère phonique comme la syllabe. Muller a constaté et utilisé l'incidence du choix de la mesure sur les résultats (9). En outre, dans une note de son *Initiation à la statistique linguistique*, il conseille une grande prudence et, tout en regrettant la difficulté du procédé, attribue à la syllabe un rôle privilégié : "Ces variations dans la densité en mots grammaticaux conduisent à réviser l'opinion suivant laquelle N, nombre de mots, est la meilleure mesure de l'étendue d'un texte; il est utile, dans certains cas, de doubler cette mesure par une autre : nombre de vers, si les textes le permettent, ou nombre de lignes, de pages etc., si leur réalisation typographique est uniforme; quant au nombre de syllabes, il donnerait sans doute une base excellente, mais c'est en général une donnée difficile à établir." (10).

### *Un programme de syllabation*

On peut conclure de toutes ces observations que la syllabe est un élément que phonéticiens, linguistes et spécialistes de la stylistique doivent, à l'occasion, isoler.

L'automatisation facilite les recherches systématiques, appliquées à des textes étendus et qui, sans elle, seraient ardues, fastidieuses, voire pratiquement impossibles.

En outre, si la nécessité crée l'outil, l'outil une fois créé ne peut-il, dans les mains d'un chercheur imaginatif, mener à des utilisations imprévues ?

Un programme qui réalise la coupe syllabique des mots latins a donc son utilité.

Dès qu'on aborde le problème, on se heurte à deux ordres de difficultés. L'un théorique : il y a plusieurs manières de concevoir la coupe des mots; il faudra donc prendre des options; celles que j'ai prises l'ont été sur l'avis de M. J. DENOOZ, beaucoup mieux qualifié que moi pour établir ce choix, et à qui revient d'ailleurs l'initiative de ce travail.

L'autre, plus technique : on rencontre des homographies qui exigent des coupes différentes; c'est là un problème pratique de programmation.

### *Les principes*

Tout latiniste se rebiffera de prime abord contre la coupe *abire - a-bi-re*; pourtant, nous l'avons adoptée.



En effet, nous devons le répéter ici, la syllabe est une unité de caractère phonétique. Décomposer le mot d'après le radical est aussi légitime, mais c'est faire de la sémantique, ce qui est tout autre chose.

Notre décision s'appuie sur d'autres arguments que cette position de principe.

Dès l'époque classique, la conscience de la valeur sémantique des éléments constitutifs des mots devait être fort atténuée, même souvent disparue dans l'usage normal de la langue (11).

Les phénomènes phonétiques qui résultent de l'agglutination d'éléments distincts le prouvent et, par un choc en retour, ils ont rendu les usagers de la langue moins sensibles à l'analyse étymologique. On trouve dans le vocabulaire le plus fréquent maints exemples de contraction, d'assimilation, d'insertion ou d'altération de sons.

Si la contraction *amatust* apparaît encore clairement comme la forme abrégée de *amatus est*, *nemo* et *cogo*, issus de *ne-homo* et *co-ago*, sont vraiment de nouveaux mots. L'assimilation travestit les préfixes - *ad-fluo* et *sub-gero* deviennent *affluo* et *suggero* -, de même que le son consonantique qui s'intercale - *re-eo* et *pro-eo* deviennent *redeo* et *prodeo*. Le radical lui-même est parfois atteint et, comme cette altération se combine souvent avec une modification du préfixe, l'identification spontanée devient plus aléatoire encore : ainsi *coerceo*, *illido*, *accipio*, *redimo*, issus de *co-arceo*, *in-laedo*, *ad-capio*, *re-emo*.

Un autre fait justifie notre choix. Si l'on fonde la coupe des mots sur l'étymologie, où s'arrêter ? Il sera logique de reconnaître *neg-otium*, *neg-ligo*, *sed-itio*, *se-ebrius* dans *negotium*, *negligo*, *seditio*, *sobrus*. Même si l'attention se porte plus souvent sur les préfixes que sur les suffixes, on n'aura



deux consonnes; il sera donc traité de la même façon et coupé après *s* - *nos-trum, as-tra*.

Les groupes *bsc, bst, nsc, nst, nsp, nstr* doivent, semble-t-il, recevoir une seule et même solution. La phonétique, je l'ai dit, n'impose aucune règle; même si, en désespoir de cause, on appelle la sémantique à l'aide, elle ne peut être d'aucun secours, puisqu'elle exige des coupes opposées selon les mots : *ob-stare, abs-terreo; in-specto, trans-pono*. Comme les groupes *sc, st, sp, str* se rencontrent fréquemment en début de mot et que la coupe devant *s* nous a peut-être paru la plus naturelle ici, c'est elle qui fut choisie et le programme fournit donc : *ob-stare, ob-scurus, in-specto, mon-strum, ab-sterreo, ab-scondit, in-stillo, in-struit*.

Cependant *mon-strum* étonne en regard de *nos-trum*. La présence de l' *n* dans le premier mot, modifiant le groupe phonétique, peut expliquer la différence de traitement, mais s'il est vrai que cet *n* était déjà amui à l'époque classique, l'argument perd toute valeur. A moins que d'adopter *no-strum, a-stra*, mais ce ne serait pas sans beaucoup de réticence.

Une autre contradiction apparaît dans le traitement du préfixe *trans*. La coupe phonétique étant notre principe de base, il n'y a aucune raison de réserver à ce préfixe un traitement spécial; nous aurons donc *tran-situs, tran-sactus*. Il n'en va plus de même quand sa jonction au radical aboutit à un groupe de deux ou trois consonnes. On pourrait adopter la solution choisie pour les groupes *nsc, nst, nsp* et obtenir *tran-spono, tran-scendo, tran-stineo*. Cette coupe pourtant ne peut convenir dans les autres cas, car les groupes *sf, sg, sl* ... n'existent pas en tête de mot et ne sont donc pas des phonèmes naturels au latin. Dès lors, chaque fois que *trans* est suivi de consonne, il semble plus logique, par souci d'uniformité, de scinder

le mot après lui, même si dans de rares cas (*trans-cribo, trans-cindo*) on empiète sur le radical. On obtiendra donc : *trans-pono, trans-tineo, trans-fugo, trans-latus, trans-gero, trans-sulto*...

Je me suis étendue sur ces cas pour montrer leur complexité et la difficulté de les résoudre de façon logique et satisfaisante; mais c'est le problème de l'utilisateur plutôt que celui du programmeur. Que plusieurs solutions soient, comme on l'a vu, subjectives, donc discutables, est de peu d'importance; en effet, à condition que l'on reste fidèle au principe de coupe phonétique, une légère et facile modification du programme peut renverser les options, non seulement selon l'opinion de l'utilisateur, mais selon les exigences de ses recherches.

#### *La programmation*

Une voyelle est suffisante mais nécessaire à l'existence d'une syllabe; aussi le principe de base du programme est-il le repérage des voyelles; la nature de ce qui les suit détermine les points de scission.

Le latin a une lexicologie trop complexe pour qu'on puisse espérer créer un programme répondant aux exigences de tous les mots de la langue.

A ce niveau pratique, de nouvelles options sont à prendre, des insuffisances à admettre; mais ici, le choix est plus facile; en cas de concurrence, ce seront évidemment les formes les plus rares qui seront sacrifiées.

Les difficultés proviennent des cas où le signe vocalique n'est pas le fondement d'une syllabe.

Il faut repérer les groupes *qu* et *gu* (celui-ci uniquement devant voyelle), de même que les diphtongues *ae* et *au*. *Oe* est déjà moins simple à traiter puisqu'il y a diérèse devant deux consonnes : *coe-lum*, *coe-pi* en face de *co-eptus*, *co-erceo*.

Mais la grosse difficulté vient de l'ambivalence du signe *i*, voyelle ou consonne (12).

*i* entre voyelles sera toujours consonne : *cu-ius*, *pe-ior*.

*i* en tête de mot et suivi de voyelle sera aussi toujours traité comme une consonne : *jam*, *ia-ceo*.

Les formes de *ire*, comportant à l'initiale *i* suivi de voyelle (*ii*, *iens*, *iero*, *ieram*, *ierim*) et le démonstratif *ii*, recevront donc des coupes incorrectes.

Le même inconvénient réside dans l'identité de graphie que présentent les mots composés d'un préfixe et d'un radical commençant par *i* consonne (*iaceo*, *iiceo*, *iuro*, *iudico*, *iungo*, *iuvo*...) et ceux qui accolent aux mêmes préfixes les formes verbales de *ire* qui viennent d'être citées; comment l'ordinateur pourrait-il distinguer l'*i* consonne de *subiectum* de l'*i* voyelle de *subieram* ?

On peut estimer que *ire* et ses composés se rencontrent fréquemment, mais si l'on considère que seules entrent en ligne de compte leurs formes commençant par *i* suivi de voyelle, que les verbes du premier groupe et leurs dérivés sont très nombreux et très usuels (et toutes leurs formes concernées), on comprendra qu'ils devaient emporter la décision en leur faveur. Ce traitement préférentiel a été réservé cependant aux préfixes qui ont paru, peut-être à tort, les plus fréquents : *ab*, *ad*, *con*, *dis*, *in*, *ob*, *per*, *sub*, et cela

uniquement pour une raison pratique; on pouvait aussi facilement l'étendre à *e, de, prae, pro, trans*, mais on a craint l'alourdissement de l'exécution. Le hasard des textes fait accepter ou regretter ce choix. Dans la forme actuelle du programme, il faut donc prévoir des coupes erronées comme *de-i-ectus, pro-i-e-ci, tran-si-i-cio, tran-si-aceo* (au lieu de *de-iectus, pro-ieci, trans-ii-cio, trans-ia-ceo*) aussi bien que les fautes inverses comme *sub-ie-ram, per-ie-runt* (pour *sub-i-eram, per-i-erunt*).

On pourrait, à la rigueur, repérer toutes les formes de *ire* qui échappent à la règle, mais un tel alourdissement de la procédure ne se justifierait que pour une recherche tout à fait spécifique. Pour la même raison pratique, nous avons sacrifié les composés des préfixes *inter* et *super* (*superiaceo, interiungo* devenant *supe-ri-aceo* et *inte-ri-ungo*), assez rares, au profit de *superius, superior, interius, interior*.

Les groupes de consonnes, en général, sont plus dociles et ne requièrent qu'un peu d'attention.

Il faut distinguer, parmi les groupes de deux consonnes, ceux qui sont insécables de ceux qui se défont : *ter-ra, mul-tus*, mais *pa-tria, de-clino*.

Certaines combinaisons consonantiques qui s'écartent des règles générales sont trop fréquentes pour qu'on puisse négliger de les traiter correctement.

Ainsi, un groupe de trois consonnes se scinde normalement après la première d'entre elles : *tem-plum, ef-fractus*; mais *nct, rct, mpt* et le groupe formé par la copule *que* adjointe à une finale de deux consonnes, exigent une coupe différente : *sanc-tus, arc-tus, temp-tare, hiems-que*. On pourrait y ajouter *mps* si on préférait *contemp-si* et *consump-si* à *contem-psi* et

*consum-psi.*

Il faut aussi prévoir pour le préfixe *post* les solutions variées qu'il réclame : *pos-terior, pos-tridie, post-scribo, post-pono.*

Avant de conclure, et pour prouver que nous ne nous sommes pas contentés d'un examen superficiel du lexique, j'exposerai un cas insoluble.

*ae* exigeant normalement la synérèse, le programme fera une seule syllabe du mot *aer*. Il est possible de lui appliquer un traitement spécial, mais il est logique alors d'apporter la même correction à ses dérivés. Or il se trouve que les membres de la famille de *aes* ont les mêmes signes à l'initiale; il faudrait donc un test portant sur les formes entières et encore ne vaudrait-il rien pour les homographes (*a-ere, ae-re; a-ereus, ae-reus*). On pourrait n'appliquer le procédé que lorsque le thème du texte fait attendre une abondance de formes d'*aer*, mais on peut être malchanceux et tomber sur les 120 premiers vers du premier livre des *Métamorphoses* d'Ovide où se rassemblent 9 formes d'*aer* (v. 12, 15, 17, 23, 28, 52, 58, 75, 119) et 3 d'*aes* (v. 92, 98, 115).

### *Conclusion*

Au terme de cette analyse, il faut un bilan.

Les options indispensables ont parfois acculé à des résultats déconcertants;

mais, d'une manière générale, ce n'est pas le manque de souplesse de l'automatisme qui en est responsable; avec une méthode traditionnelle, on éprouverait aussi souvent le besoin d'uniformiser les solutions appliquées à des cas similaires et on aboutirait aux mêmes impasses. Il n'empêche que nous sommes bien conscients des failles du programme réalisé; aucune cependant, c'est essentiel, n'affecte les phénomènes généraux de la langue.

Il laisse échapper des coupes incorrectes de trois types :

1. Les unes sont volontaires et peuvent être facilement évitées; on ne les a admises que pour une raison pratique; l'alourdissement considérable de la procédure que représenterait le cumul de leurs traitements particuliers n'a pas paru justifié pour l'usage général; elles concernent des mots isolés, très spécifiques, donc d'usage restreint; citons quelques termes où *h* est inclus dans un groupe consonantique intérieur : *drachma (drac-hma)*, *athleta (at-hleta)*, *anthropophagus (ant-hropophagus)*, *isthmus (ist-hmus)*. Il peut aussi s'agir de groupes : ainsi certains verbes à préfixe suivi de *i* consonne, du type *proiectus (pro-i-ectus)*, dont on a parlé plus haut.
2. D'autres sont peut-être également éliminables, mais au prix de procédés beaucoup plus compliqués. J'ai expliqué le problème des composés de *ire*; c'est sûrement la plus grande faiblesse du programme (les tests, d'ailleurs, le prouvent), mais le nombre des formes affectées reste tout de même assez limité. On se souvient aussi des termes du type *superiaceo*, *interiiceo*.
3. Il y a par contre des cas qui me semblent insolubles; cette fois encore, ils sont rares et isolés, tels *aer* et ses dérivés, *seu*, *arguo*, *poeta*, *iambus*, *periodus* et quelques autres. En cas de nécessité absolue, on pourrait



tout de même, surtout pour les moins rares et les plus brefs, comme *aer* et *seu*, procéder à un traitement absolument spécifique. Il faut mentionner encore dans cette catégorie, les phonèmes *eu*, *ei* de certains noms propres d'origine grecque qui, ne se distinguant pas des groupes habituels de deux voyelles, se scindent comme eux : *Perseu* devient *Perse-u* comme *me-um*, *Promethei* devient *Promethe-i* comme *me-i*.

Ce n'est pas l'endroit, dans cet exposé général, d'énumérer tous les termes rebelles; on peut cependant affirmer qu'il en reste peu à ajouter à ceux qui furent cités pour que l'inventaire soit complet.

Même ainsi limitée, la liste des exceptions peut inquiéter parce qu'elle les rassemble, mais on épuiserait sa patience à les rechercher, dispersées de loin en loin dans les textes.

La valeur d'un outil se révèle à l'usage. Tel quel, et appliqué à un texte quelconque, le programme donne satisfaction.

Trois essais ont été réalisés sur des textes suivis, pris au hasard et comportant chacun plus de mille mots plurisyllabiques. Appliqué à 200 vers de Virgile, le programme laisse échapper 2 coupes incorrectes (*ae-ria*, *ehe-u*), outre 7 dans des noms propres (*Dap-hnim*, *Dap-hnidis* et *Melibo-e-e* en 5 exemplaires), tandis qu'un extrait en prose de Sénèque révèle 5 fautes dans le vocabulaire courant (*iis* en triple exemplaire, *ab-ie-rit*, *per-ie-runt*) et la coupe erronée du nom propre grec *Théop-hrasti*. Dans un passage de 200 vers du même écrivain, on découvre 2 erreurs dans le lexique ordinaire (*aer*, *ad-ie-re*) et 6 dans des noms propres grecs dont 3 dans le même mot

(*E-uropae, Perse-us, E-uryst-he-us, E-umenides*); le sujet du texte, un récit mythologique, explique l'abondance de ceux-ci. Si l'on tient compte de cette particularité défavorable, on peut croire qu'une moyenne de 5 erreurs pour 1000 mots, soit 0,5 %, est une estimation assez exacte pour le lexique commun, qu'en prévoir 10 serait pessimiste.

N'est-ce pas là une proportion tout à fait acceptable dans un traitement automatique d'une langue aussi complexe (13) et avec un programme PL1 qui excède à peine 125 instructions ?

Université de Liège  
Laboratoire d'Analyse statistique  
des Langues anciennes

M. DUCHESNE-DEGEY

## NOTES

- (1) Toute la première partie et les deux premiers chapitres de la deuxième partie.
- (2) A. Meillet insiste aussi sur l'importance de la syllabe initiale en latin et la fréquence de l'allitération - *Esquisse d'une histoire de la langue latine*, pp. 129-130.
- (3) Quintilien notamment, et bien des critiques modernes, cités par Marouzeau dans son *Traité de stylistique latine* ont considéré cet aspect du style.
- (4) Problèmes des diphtongues, de l'i consonantique ou vocalique, de l'u derrière q et g.
- (5) Ch. MULLER, *Essai de statistique lexicale : L'illusion comique de P. Corneille* et *Etude de statistique lexicale : Le Vocabulaire du théâtre de P. Corneille*.
- (6) Muller explique qu'il a dû procéder à une correction, d'ailleurs facile, dit-il, pour les passages du théâtre de Corneille qui ne sont pas en alexandrins. Cf. *Etude de statistique lexicale*, p. 47.
- (7) J'ai réalisé le programme de syllabation nécessaire à la composition typographique automatique d'un texte (ajustement des

fins de ligne, répartition des blancs), mais le fait qu'en français on ne coupe pas un mot entre deux voyelles m'a écartée du principal écueil.

- (8) *De Natura Rerum*, Ed. Budé, V vers 1173 et 1176.
- (9) Ch. MULLER, *Etude de statistique lexicale*, pp. 47, 49, 100.
- (10) Ch. 15, note 1, p. 208. La réticence qu'il formule à l'égard de la syllabe porte non sur le principe mais sur l'application; elle provient sûrement du fait que l'auteur travaille sur des textes français dont la syllabation automatique est beaucoup plus complexe à réaliser.
- (11) Décivant la langue du IIIe siècle avant J.-C., Meillet affirme que déjà "les anciennes relations étymologiques des mots sont souvent obscurcies ... Le latin est une langue où les mots d'une même famille tendent ainsi à s'isoler les uns des autres, par suite des altérations et des voyelles, et des groupes de consonnes à l'intérieur du mot." Dans *Esquisse d'une histoire de la langue latine*, p. 145.
- (12) Il faut noter que le programme est prévu pour un texte qui utilise les deux lettres distinctes *u* et *v*.
- (13) Le grec, lui, semble avoir été créé, de toute éternité, pour l'informatique; l'italien sûrement aussi, alors que le français s'en défend encore plus jalousement que le latin.

Un programme, fondé sur les mêmes principes, a été réalisé pour le grec. Ce fut un jeu d'élaguer le programme latin et d'en adapter la base; il se trouve réduit à 75 instructions et le nombre de cas qui lui échappent est infime.