

De l'«Indo-Européen» au latin par ordinateur : quelques résultats et quelques perspectives

par

Albert MANIET

Université Laval, Québec - CANADA

Ma communication a pour objet la recherche que je mène par ordinateur depuis 1976 sur la phonologie diachronique du latin. Etant donné le peu de temps dont nous disposons, je ne retiendrai que la partie centrale de ce qui a déjà été réalisé et je ne mentionnerai que quelques-unes des perspectives de recherche qui s'ouvrent à partir de cette réalisation⁽¹⁾.

La partie centrale des étapes déjà réalisées est l'élaboration d'une série algorithmique de règles d'évolution devant permettre d'expliquer l'état phonique du latin ancien (en fait, celui de l'époque de Plaute) en regard de celui des périodes qui l'ont précédé, et qui est soit attesté (on peut remonter ainsi, dans une mesure variable, jusqu'au 6ème siècle avant J.C.) soit reconstruit par comparaison (on peut alors remonter, dans les cas les plus favorables, à l'indo-européen commun). Depuis environ cent cinquante ans que l'on traite scientifiquement de phonétique historique du latin et, par la suite, de phonologie diachronique, aucun des auteurs d'ouvrages ou d'articles publiés sur cette matière, y compris celui qui vous parle (la dernière édition de mon manuel de *Phonétique historique du latin* . . . ne remonte pourtant qu'à six ans d'ici) n'avait encore présenté systématiquement l'ensemble de ces règles d'évolution de façon algorithmique, c'est-à-dire en tenant compte de leur enchaînement logique, ce qui suppose, en l'occurrence -en principe, du moins- leur enchaînement chronologique. Stimulé par l'aide que peut apporter l'ordinateur, que j'ai appris à apprécier ici-même, au LÁSLA, il y a une dizaine d'années, j'ai commencé par élaborer une série de règles à titre provisoire, en m'aidant des formulations existantes, que j'ai adaptées au langage-machine. Je l'ai testée à partir d'un corpus de trois cents étymons que j'ai rassemblés sur la base de l'illustration qu'ils fournissaient des règles d'évolution. Ce corpus n'était factice qu'en partie. S'il est vrai que chacune des formes qu'il contient servait à illustrer une ou même plusieurs règles, d'abord cette illustration ne portait pas sur tous les phonèmes du mot, les autres phonèmes ressortissant à d'autres règles, fût-ce de conservation⁽²⁾, et donc une seule des formes de ce corpus constituait déjà un faisceau de preuves de la validité des règles. Mais surtout, si l'on tient compte du nombre de trois cents formes, ce faisceau atteignait une complexité considérable, que je laisse aux statisticiens le soin de calculer. Peu importe d'ailleurs, puisqu'il ne s'agissait que d'un test provisoire. Ce test m'a permis de modifier sensiblement ma première série de règles, soit par addition, soit par modification du contenu ou de la place des règles dans la série.

Après cette étape, j'ai procédé à l'étymologie d'un corpus authentique, constitué à partir de textes de l'époque de Plaute et qui compte deux mille lignes complètes de ces textes prises au hasard (soit environ 80 000 segments phoniques). J'ai soumis aux règles confiées à la mémoire de l'ordinateur des tranches de 45 lignes de formes ainsi reconstituées. Par exemple, si le texte latin comportait la forme *fictus*, mon *input* a été la forme *dhightos. Pour que l'*output* soit effectivement *fictus*, l'ordinateur aura dû faire intervenir cinq règles d'évolution placées, du point de vue logique, dans un ordre déterminé et, du point de vue chronologique, s'échelonnant d'au moins 800 avant J.C. jusqu'un peu avant l'époque de Plaute, soit une période d'au moins six cents ans. Quelle est la portée théorique de l'*output correct fictus* ? Cet *output* implique que tous les phonèmes que comprenait cette forme à date ancienne (p. ex., /*dh/) ont soit évolué soit persisté de la même façon dans tous les mots latins où ils se trouvaient dans la même position (p. ex., /*dh/ initial).

Jusqu'à présent, mille lignes de texte ont été soumises à l'application des règles par l'ordinateur et j'ai été amené, selon les résultats, à apporter bon nombre de modifications nouvelles à la série publiée dans la Revue de l'Organisation internationale pour l'Etude des Langues anciennes par ordinateur (voir note 1). Cette série compte maintenant quarante-sept règles de plus, soit deux cent six (236 lignes). C'est à partir de cette expérience et de ses résultats que je vais, dans cette seconde partie de mon exposé, commenter brièvement le but et les prolongements de ce type de recherche.

Le résultat brut est que, sur environ 40 000 segments phoniques en *input*, moins de 2500, soit 6,2%, ne sont pas "corrects", c'est-à-dire conformes au texte latin authentique. En ce qui concerne le nombre des transformations dues à l'application des règles, sur un total d'environ 17 400 modifications de l'*input* "indo-européen", environ 15 400 se sont réalisées "correctement", soit un peu moins de 89%. Ce qui paraît très satisfaisant.

Le but du processus de transformation n'est évidemment pas d'obtenir en *input* une forme correcte, que l'on connaît d'avance, de toute façon. Cet *output* n'est qu'un test de la valeur du processus : c'est celui-ci qui importe, car il représente la reconstitution des diverses étapes par lesquelles sont passées les formes qui ont abouti à l'état latin, et ces étapes sont le reflet de l'histoire. La conclusion à tirer du degré satisfaisant de réussite du processus est qu'il fournit une base, sinon parfaite, du moins solide, concrète, à toutes sortes de réflexions sur l'évolution des langues au plan phonique, la première étant qu'il existe effectivement une très grande régularité dans cette évolution. C'est là un point de départ fondamental, sur lequel doivent prendre appui toutes les théories sur cette matière. Des théories explicatives, bien sûr, on en a présenté déjà un certain nombre. Mais, jusqu'à présent, jamais ces théories n'ont été mises en rapport avec un corpus réel de quelque étendue. C'est cette lacune qui m'a incité à entreprendre ce travail.

Je mentionne maintenant, à titre d'exemple, deux des recherches auxquelles il pourra donner lieu.

- 1.- On sera en mesure de vérifier dans le détail le pourcentage et la nature des exceptions à une règle d'évolution. Ces exceptions sont au moins aussi instructives que les résultats corrects, puisqu'elles forcent à rechercher les conditions spéciales qui ont fait échapper tel phonème de tel mot à la règle commune. Par exemple, un des résultats "incorrects" des transformations est la forme **percuntarī, au lieu de *percontarī*. Or, on sait que *contus* "gaffe, perche" a été repris au grec *κόρυτος* et il suffit donc d'émettre l'hypothèse que le composé imagé "sonder à fond (*per-*) au moyen d'une gaffe (*contus*)", au sens de "s'informer" est relativement récent et postérieur à la loi de l'apophonie (en l'occurrence "/*o/ → u en syllabe intérieure fermée") qui, elle, est antérieure au 3ème siècle avant J.C. Mais c'est là un type d'explication bien connu. D'autres le sont moins ou, en tout cas, n'ont pas fait l'objet de recherches systématiques. Dans cette optique, je compte notamment comparer l'état des monèmes du corpus de l'époque de Plaute avec celui des monèmes d'époques plus anciennes et, d'autre part, comparer le traitement des phonèmes en tant que simples éléments de séquences phoniques et en tant qu'éléments de monèmes.

Cette double comparaison permettra, je crois, de dégager sur une base statistique dans quelle mesure les faits phoniques peuvent modifier le signifiant des unités significatives et, à l'inverse, jusqu'à quel point la reconnaissance d'un élément signifiant peut contrecarrer l'évolution de son signifiant. Soit, p. ex., les formes reconstituées /*sisō/ et /*disēmaj/, qui deviendront respectivement *serō* et *dirēmī*. Après le rhotacisme, le /i/ de /*sīrō/ et celui de *dirēmī* se trouvaient tous deux devant /r/ récent à la même époque. Or, le premier s'est ouvert en /e/ sous l'influence de ce /r/, tandis que le second a conservé son timbre. L'explication qui semble s'imposer est que les variantes du morphème *dis-*, conservé dans *distulī*: *dī-* (dans *dīlātus*), *dif-* (dans *differo*) etc. présentaient toutes normalement le timbre /i/. C'est ce modèle qui a dû provoquer la résistance au changement de /i/ en /e/. Tandis que la séquence /si/, perçue à juste titre comme un redoublement aussi longtemps que la forme était /*sisō/, a cessé d'être perçue comme telle dans la forme évoluée /*sīrō/. /i/ faisait dès lors partie d'une pure séquence phonique et non plus d'un morphème, et il n'y avait aucune raison de freiner, dans ce cas, le processus inconscient qui a abouti à /e/.

- 2.- On sera en mesure de vérifier le degré d'application des théories explicatives, par exemple la théorie du rendement fonctionnel proposée -avec des nuances- par A. Martinet et selon laquelle les phonèmes ou les séquences servant à opposer entre elles un nombre important de paires de mots ou de paires de mots importants résistent davantage à l'évolution. Dans ce but, j'ai déjà établi par ordinateur, à partir du corpus de deux mille lignes, les particularités du système phonique du latin ancien sous l'angle quantitatif. Plus précisément, j'ai calculé par rapport au total de 80 000 segments la fréquence des éléments vocaliques et consonantiques, à la fois globalement (tant d'occurrences de voyelles etc.) et à l'intérieur de chaque catégorie (tant d'occurrences de /a/ etc.), la fréquence des syllabes comprenant 1, 2, 3 ... *n* sons et, plus particulièrement, tel son ou telle séquence etc. En confrontant ces données avec les données correspondantes d'autres langues ou de tranches synchroniques de périodes antérieures à l'époque de Plaute, on pourra placer sous un éclairage nouveau un certain nombre de faits. Par exemple, en latin, diverses séquences biphonématiques anciennes : /*ln, *ld, *ls, *rs/ se sont réduites à un seul phonème, géminé : //, *rr*. Or, pendant la période précédant ce changement, la langue ne contenait pratiquement pas de géminées. Cette évolution normalise donc la proportion des géminées dans le système (j'en ai dénombré 1108 occurrences dans le corpus). Mais, en même temps, elle prive cette langue de certains types de séquences. Eh bien, nous constatons que d'autres séquences : /*len, *lid, *lgs, *rks/ etc. ont abouti à *ln, ld, ls, rs*, qui ont subsisté tels quels. Voilà donc réintroduites dans le système, les séquences fonctionnelles disparues.

On voit le genre d'études qui pourront être menées, à la fois sur le plan qualitatif et sur le plan quantitatif, à partir des résultats obtenus par l'application des règles d'évolution au corpus.

NOTES

- (1) Je signale que j'ai déjà publié deux articles sur cette recherche ("Recherche sur la phonologie diachronique du latin ancien", dans "Langues et linguistique", 4/5, 1978/1979, p. 39 à 64; "De l' "indo-européen" au latin par ordinateur : un programme de phonologie diachronique", dans "Revue de l'Organisation internationale pour l'Etude des Langues anciennes par Ordinateur", 1/2, 1979, p. 1 à 23) et présenté une communication au XVIème Congrès International de linguistique et philologie romanes (Palma de Mallorca, avril 1980). Je compare, dans cette dernière, mon propre travail avec celui du Prof. S. Burton, de la State Univ. of North Carolina, sur l'ancien français, et celui de M. Ch. Eastlack, de l'Université de Californie, sur l'ancien espagnol à partir du latin.
- (2) Par exemple, l'étymon *dhightos, d'où lat. *fictus*, servait à illustrer le passage progressif à l'initiale de /*dh/ à *f*. Mais il supposait en même temps les évolutions /*ght/ → /*gt/ → *ct*, /*o/ en finale devant /s/ → *u*, et le maintien de /*i/, /*t/ et /*s/.