

La réforme de l'orthographe

L'informatique linguistique teste les qualités des réformes.

Comme le Phénix (fénix?) l'idée d'une réforme de l'orthographe renaît périodiquement ; ces temps derniers elle a été suscitée par les enseignants. Aussi la controverse séculaire sur son opportunité, son sens et son étendue recommence-t-elle à occuper les esprits, controverse d'autant plus riche que les contributeurs feignent d'ignorer ce qui a été dit et redit sur le sujet dans les ouvrages fondamentaux de F. Brunot (sur les aspects philologiques) et de C. Beaulieu (sur les aspects historiques, des Capétiens à nos jours). Nous contribuerons au débat en présentant quelques-uns des faits nouveaux révélés par les progrès de la linguistique au cours des dernières décennies, et surtout en montrant comment les moyens informatiques permettent de quantifier les faits et de transformer une illustration anecdotique en une argumentation mieux fondée.

Deux thèmes dominent le débat : le thème de l'orthographe témoin de sa propre histoire, elle-même reflet déformé de celle de la langue et de la culture, et le thème de l'orthographe comme codage d'une langue parlée, avec tous les objectifs qui peuvent et doivent être invoqués, depuis l'efficacité

de la transmission des idées jusqu'à la facilité d'apprentissage de la lecture ou de la prononciation.

Examinons ces deux thèmes sur l'orthographe du mot désignant « un des cinq prolongements de la main » (Petit Robert), le mot *doigt*. Le premier explique et justifie sa graphie par la théorie qui avait cours, il y a quelques siècles, sur son origine (*digitus*) ; cette position, fréquemment invoquée pour d'autres monstres comme le mot *poids*, justifierait que l'on écrive *onvle* avec un *v* muet pour rappeler l'affection avunculaire. Le second thème insiste, d'une part, sur l'utilité de marquer une parenté entre *doigt* et *digital* (comme dans les *empreintes digitales* et non pas comme dans le *calcul digital*, que l'Académie rejette) et d'autre part sur le besoin de distinguer *doigt* de *doit*, ce qui pourrait néanmoins être réalisé en attribuant à l'un d'eux la graphie *doua* ou *doie*.

Nous choisirons de nous limiter ici au problème de l'ambiguïté. Puissamment exploitée par les sciences de la psychanalyse et les arts de la contrepétrie, l'ambiguïté constitue cependant un obstacle pour les constructeurs futuristes de machines à écrire à commande vocale.

Précisons le problème sur des exem-

ples. La conjonction de coordination *ou* et le pronom, relatif ou interrogatif, *où*, se prononcent de la même façon et une ambiguïté existe dans la langue parlée. Des grammairiens ont décidé de la lever en distinguant les deux formes écrites. La situation est identique pour la forme *a* du verbe *avoir* et la préposition *à*. Dans la même crainte de confusion potentielle, on a voulu distinguer par un accent, circonflexe cette fois, le mot *du* – contraction de *de le* – et le mot *dû* – dérivé du verbe *devoir* ; comme il n'y a pas d'ambiguïté pour le féminin et le pluriel de *dû*, on écrit *due*, *dus*, *dues* sans accent.

Toutes les conséquences de telles décisions n'ont certainement pas été prévues par leurs auteurs :

- le *ù* est un caractère d'imprimerie supplémentaire du français (et une touche du clavier) qui ne s'emploie que dans le seul mot *où*,

- l'emploi de *à* est limité à sept mots que nous citons pour que le lecteur puisse partager l'effroi que causerait sa suppression : *à, là, déjà, déjà, delà, holà, et voilà* ! On connaît les incohérences qui opposent *cela* et *celui-là*...

- on n'a pas considéré que la forme *où* (le relatif et l'interrogatif) reste ambiguë et on n'a pas non plus distingué la forme verbale de *dû* comme dans *Jo réclame son dû* et *Jo a dû partir*. Dans ce cas on aurait pu tout aussi bien introduire, pour l'une quelconque de ces deux formes, l'orthographe *dú* – avec accent aigu – ou *dub* pour rappeler le *b* de *débiteur*. Le mot *tu* étant également ambigu, pourquoi ne le distinguerait-on pas, et n'écrirait-on pas : *tu ne t'es pas tû* ?

Si le principe de séparation orthographique des formes ambiguës avait été appliqué avec cohérence, il aurait fallu introduire un accent (lequel d'ailleurs ?) dans *car*, pour distinguer la conjonction de l'autobus, dans *or* pour distinguer la conjonction du métal précieux, ou bien écrire *aur* ou *aurre* (comme dans *Aurillac*) en souvenir de *aureus*, car il n'est pas plus difficile dans une conversation de choisir la bonne signification de *ou/où, a/à* que celle de *car* ou de *or*.

Au contraire, les adverbes *dûment* et *indûment* ne sont pas ambigus, bien que leur accent soit exigible... On pourrait multiplier les exemples mais notre propos est plutôt d'examiner comment les travaux de linguistique informatique font avancer ce type d'études.

En effet les travaux sur la formalisation des structures lexicales et syntaxiques pour la reconnaissance d'unités de sens dans les textes (en vue de la traduction et de l'indexation de textes) mettent en lumière les formidables ambiguïtés des langues naturelles. Pour les

EXERCICE DE SENSIBILISATION : LE CŒUR A SES RAISONS

Dans le texte qui suit, on remplacera chaque occurrence du signe \$ par un synonyme du mot raison que l'on peut choisir ad libitum. On a, par exemple :

raison = la vérité + tête + occasion + une bénédiction + position + comprendre + la cadence + les comptes + la cadence + les comptes, etc.

Il n'y a pas de \$ pour que cela dure. \$ de plus pour ne pas lui donner \$. Pourtant il a l'âge de \$, et il a une \$ sociale. Mais il fait des mariages de \$ à \$ d'un ou deux par an, cela n'a ni rime ni \$. Pour des \$ de santé évidentes, on ne peut lui en demander \$. Il faut donc se faire une \$ et espérer qu'il aura \$ de ses problèmes et retrouvera une \$ de vivre.

linguistes, la notion d'ambiguïté constitue d'ailleurs l'un des principaux critères d'adéquation d'un modèle linguistique aux faits observés. Les mots d'un dictionnaire électronique doivent être décrits de façon entièrement explicite ; en particulier il est nécessaire de représenter toutes les structures où ils entrent ; ces structures syntaxiques constituent alors le contexte que l'ordinateur utilise pour lever l'ambiguïté du mot.

Considérons le mot *raison* : nous en présentons une trentaine de sens différents, dans leur structure de base ou dans un contexte sans aucune ambiguïté. S'il est certes possible et légitime de regrouper – de bien des manières – toutes ces constructions, il n'en reste pas moins que deux quelconques de ces acceptions, comme dans *le triomphe de la raison* ou *la raison du plus fort*, sont traduites par deux « mots » différents dans la plupart des langues étrangères.

Notre dictionnaire comprend donc une trentaine de mots *raison* ; autrement dit, une trentaine de mots de la forme *x raison y* présentent ces ambiguïtés. Plusieurs dizaines de milliers de mots du français ou de n'importe quelle autre langue sont dans cette situation et le locuteur ou le programme d'analyse automatique de texte doit explorer le contexte pour lever les ambiguïtés : les contextes sont également explorés pour les mots *car* et *or* de la même façon qu'ils le sont pour les sons des mots *a/à* et *ou/où*. En conclusion, il n'y avait pas de raison raisonnable de privilégier par une distinction orthographique quelques cas dans la masse immense du vocabulaire qui requière une analyse contextuelle.

Les ambiguïtés peuvent être réelles et plus ou moins difficiles à éviter, mais certaines d'entre elles sont le fait des représentations orthographiques choisies pour les langues. L'arabe et l'hébreu peuvent être écrits sans voyelles mais la lecture de textes non voyellés demande un apprentissage difficile. L'analyse du contexte permet de rétablir la plupart des voyelles : ce qui importe n'est pas tant l'ambiguïté, mais le fait qu'elle oblige une exploration plus ou moins difficile du contexte : toute modification de l'orthographe devra prendre en compte ce paramètre.

Grâce à la construction de dictionnaires électroniques, il est possible, depuis quelques années, de quantifier des effets d'ambiguïtés comme ceux que nous venons de présenter et, plus généralement, de simuler les effets d'une réforme orthographique sur la totalité du lexique, donc sur la forme des textes.

Nous avons compilé, au LADL, un dictionnaire électronique de plus de 60 000 entrées (B. Courtois, *Rapport*

technique du LADL, CNRS-Université Paris 7, 1986) ; il comporte les informations nécessaires à la flexion automatique de chaque mot (mise au féminin, au pluriel, conjugaison), et l'un de nous (E. Laporte, *Thèse de doctorat*) l'a complété en y ajoutant les éléments requis pour obtenir la transduction phonétique des mots, fléchis ou non. Ces programmes engendrent un lexique de plus de 500 000 formes qui donne l'ensemble des formes rencontrées dans les textes écrits ou oraux. Des traitements automatiques permettent alors de répondre à des questions variées sur l'ambiguïté. Voici quelques exemples de ces traitements.

C. Leclère et A. Guillet (*Rapport technique du LADL*) ont examiné par curiosité l'effet de la suppression des voyelles dans les mots français. Certains mots, dont les mots courts, sont évidem-

ment mutilés et non reconnaissables tant leur degré d'ambiguïté est élevé : *peau* = *api* = *p* ; *initié* = *aient* = *ont* = *ente* = *note* = *nt*. Mais pour les mots plus longs, il y a très souvent correspondance unique du mot et de son squelette consonantique.

Plus réaliste, B. Courtois a étudié la suppression des accents du français, autrement dit l'écriture en majuscules non accentuées. Pour 500 000 formes, 50 000 ambiguïtés nouvelles apparaissent (*MANGE* = *mange* + *mangé*). Comme on le sait, cet effet ne rend pas les textes illisibles, mais il est difficile d'évaluer l'accroissement de la difficulté de lecture, tant les situations sont variées. Par exemple, si l'ambiguïté de *MANGE* est difficile à lever, celle de *MANGES* = *manges* + *mangés* est plus aisée, car la forme *manges* comporte obligatoirement le pronom *tu* (ou *toi*)

LA SIGNIFICATION DES DIFFÉRENTS MOTS RAISON :

- (1) Jo a raison de partir
- (2) Jo a (une + des) raisons de partir
Jo a ses raisons de partir
Il n'y a pas de raison que Jo parte
Le départ de Max a (une + plusieurs) raisons
Le départ de Max est sans raison(s)
Une des raisons de son départ est que Max reste
Max a toutes (E + les) raisons de penser que Jo restera
- (3) Que Jo parte n'est pas une raison pour que Max reste
- (4) Jo a toute sa raison
Jo (a perdu + est revenu à) la raison
Max ramène Jo à la raison
- (5) Max s'est rendu aux raisons de Jo
- (6) Max s'est fait une raison, il ira travailler
- (7) Ce texte (n'a ni + est sans) rime ni raison
- (8) Jo a eu raison de toutes les difficultés
- (9) Max donne raison à Jo dans ce conflit
- (10) Max a demandé raison à Jo de sa conduite
Jo rendra raison
- (11) Max entendra raison
- (12) La raison de cette progression (arithmétique + géométrique) est
- (13) Max écouterait la voix de la raison
- (14) Jo a fait un mariage de raison
- (15) Jo a l'âge de raison
- (16) Jo a une raison (d'être + de vivre)
- (17) Jo a une raison sociale
- (18) La raison d'état impose cette mesure injuste
- (19) Max invoque (la raison pure + la raison du plus fort)
- (20) (partir) (à + avec + non sans) raison
- (21) (partir) à juste raison
- (22) (partir) comme de raison
- (23) (dormir) plus que de raison
- (24) (agir) à plus forte raison
- (25) (partir) pour des raisons de santé
- (26) (partir) à tort ou à raison
- (27) (partir) en raison du brouillard
- (28) (payer les gens) en raison de leurs efforts
- (29) (arriver) à raison de six à l'heure
- (30) (agir) en raison (directe + inverse) des interventions
- (31) (agir) pour la bonne raison que Jo part
- (32) (agir) raison de plus pour que Jo parte

Orthogonal, e — or, ...ore etc tôt, taux goth n'al..., orthogone al..., goune, gau n'as le, n'al..., goth, Gowan (angl. gone).

Orthographe, le — or tôt que raffé (raffe), rat fit, ...orte au gras fit, fils, ogre a f..., fils, affi(quet), raffi(ner), Hogue ras.

Orthographe, lque — orthographe y est, hie, ...orte au gras fils que, fic, aux graphiques, or tôt que raffes ic..., y e..., rat, ras fier, taux, fic..., ic..., ...aure.

Orthographe, lque — ...orte (morte, etc) au logia, or tôt l'eau, l'os gtt, j'y, que, logique, töle, auge, Auge hic, ic..., Ich...

Orthopédie, lque, iste — or taupe est dit(e), tope (interj.) est dit, toper, paie dix

Reforger — ...re, ...reux forger, fort j'ai, jet, geai, gé..., forge et, é..., faux rejet, Fo, Foë regé..., ceuf, orge et.

Réforme, er, é, e, able, iste — forme E, haie, hâble, réforme (v.) é..., E..., raifort met, m'est, mé..., Mab le, miss (angl.) te, rets, rais etc former, forme able, fort myst(érieux), mit c'te, mise (s. v.) te, ormaie, O remet, (b)ref !, or me, (g)refte.

Réformation, ateur, trice — rets etc formation, formâtes eur..., risse, forts masions, ma trice, triasse, matrice (de la) réforme hâte eurent, or mat, format.

Reformer — reforme etc é..., ai..., aia, hale, E..., ...reux former, faux remet, fort

Le *Nouveau dictionnaire français* de Baillargé, publié à Québec en 1888, montre bien les diverses ambiguïtés phoniques qui fournissent de nombreuses solutions à la représentation du thème de l'article : *La raie forme de l'or tôt gras feu. L'arrêt fort me de Laure Togo raffie. La raifort me deux...* etc.

dans son voisinage immédiat. En fait, on n'observe guère *manges* que dans l'une des séquences :

tu (ne) (le + la + les) (lui + leur + y) manges

tu (ne) (lui + leur + y) (en) manges

manges-tu

ou dans ses variantes en *toi*.

Enfin, pour évaluer une réforme minimaliste de l'orthographe, S. Woznika (*Rapport technique du CERIL*) a examiné la réduction des trois accents et du tréma, à un seul accent (plat ou vertical). Le nombre des ambiguïtés introduites est inférieur à 50, chiffre incluant la confusion de variantes orthographiques pour des mots comme *chénevotter*-*chénevotter* ou *euréka*-*eurêka*.

En comparaison, le dictionnaire phonétique du LADL permet d'évaluer les ambiguïtés de l'orthographe actuelle, c'est-à-dire les différences de prononciation qui existent pour une même orthographe, par exemple *ti* dans cinq verbes comme *initier* et cinq comme *châtier*; plus généralement, il existe 44 mots comme *partie* et 57 comme *inertie*. L'un de nous (E. Laporte) a recensé 300 situations de ce type pour lesquelles il a dû marquer les mots du dictionnaire. Le programme de phonétisation consulte le système de dictionnaire orthographique pour chaque mot de texte et fournit la solution phonique correcte. Un tel outil permet de simuler à volonté les diverses réformes de l'orthographe qui ont été proposées et de les évaluer. Par exemple, les observations sur la perte d'information qu'occasionnerait la disparition de certaines lettres redondantes et de certains signes diacritiques peuvent aujourd'hui être quantifiées.

Il faut garder raison et de telles évaluations devraient contribuer à dépassionner le débat. Certes il y a longtemps que le Poète (qui ne doit sans doute qu'à son origine savante de ne pas bénéficier d'une graphie plus vernaculaire qui eût pu être *Peauhaitte*)

a dit que l'orthographe était le visage des mots. Nul(le) amant(e) n'accepte sans émoi qu'un esthète chirurgien prétende à changer celui de l'aimé(e). Mais après coup, tout le monde s'en trouve bien, et même y découvre les charmes d'un amour rénové.

M. GROSS, E. LAPORTE et
M.-P. SCHÜTZENBERGER