Annie Desnoyers
1998/01/01

Orthographe

L’orthographe lexicale, une question d’organisation des données à mémoriser

Annie Desnoyers est linguiste. Elle a travaillé pendant trois ans chez Machina Sapiens, compagnie spécialisée dans le traitement informatique des langues naturelles et reconnue pour son logiciel *Le Correcteur 101*. Dans son article, elle fait ressortir le fait que l’habileté à détecter les fautes d’orthographe lexicale dépend de la « mémoire », et elle propose une avenue non pas pour « augmenter » la mémoire humaine, mais pour réduire le nombre d’éléments à mémoriser.

Au cours du processus d’amélioration de la langue écrite à l’ordre collégial, un aspect important, mais souvent oublié, est l’orthographe lexicale. Peut-être l’oublie-t-on parce que l’on juge que cette matière ne s’enseigne pas : « Il faut deux *n*, c’est comme ça, il faut le savoir par coeur, c’est tout ! » est souvent l’explication donnée à la faute **honeur*, par exemple. On se dit que ce sont les étudiants qui lisent ou qui cherchent dans les dictionnaires les mots qu’ils écrivent, qui réussiront en orthographe lexicale. Certains pédagogues, par

contre, ne sont pas satisfaits de cet « apprentissage », s'il en est un, et cherchent comment on pourrait ENSEIGNER l'orthographe.

Dans ce bref article, je tenterai de donner une piste à suivre pour arriver à la compréhension du système orthographique, puis à son enseignement. D'abord, j'expliquerai la place de l'orthographe lexicale dans le traitement informatique des langues naturelles, ce qui confirmera les intuitions sur le fait que sa connaissance dépend de la mémoire. Puis, je présenterai René Thimonnier, chercheur reconnu qui a cru à un système orthographique au même titre que les autres systèmes de la langue, en a étudié le fonctionnement, et a développé un programme pour l'enseigner.

L'orthographe lexicale dans le traitement informatique des langues

En informatique, la règle de base est la suivante : ce qui est difficile pour l'humain est facile pour la machine et ce qui est facile pour l'humain est difficile pour la machine. Des exemples ? Exécuter des calculs avec plusieurs chiffres élevés est difficile pour un humain sans papier ni crayon, mais vite fait par un ordinateur ; par contre, allonger son bras, saisir une tasse et la porter à ses lèvres est facilement exécuté par un enfant de deux ans, mais nécessite des années de programmation et d'essais pour un robot.

Dans le traitement informatique des langues naturelles, c'est la même chose. Voici une phrase produite par un enfant de quatre ans : « J'ai mis de l'eau pour pas que les poissons *mourissent*. » Cet enfant possède tellement la syntaxe de sa langue qu'il insiste pour donner des allures de subjonctif à son verbe *mourir*, comme dans *fassent* ou *finissent.* Pour un analyseur de langue, il suffit de quelques dixièmes de seconde pour constater que **mourissent* n'est pas une forme verbale possible en français, alors que l'enfant devra, plus vieux, travailler dur pour apprendre les conjugaisons. Par contre, il aura fallu une dizaine de linguistes et d'informaticiens, et plusieurs années de travail, pour programmer une grammaire qui puisse « savoir » que le verbe *mourir,* dans cette phrase, va au subjonctif, alors qu'un enfant de quatre ans le sait, sans même avoir reçu d'enseignement explicite à cet égard.

L'orthographe lexicale, comme la conjugaison du verbe *mourir,* fait partie de ce qui est difficile pour l'humain et facile pour la machine ; rares sont ceux qui connaissent l'orthographe de 70 000 mots, mais un ordinateur fouille en moins de deux une liste de cette envergure et peut vous dire rapidement si la forme **honeur* fait partie ou non de la liste. La connaissance de l'orthographe lexicale dépend donc de la mémoire. Sachant que la mémoire humaine n'est pas fiable au-delà d'un nombre *n* de données, la solution réside dans leur regroupement par classes, pour arriver à un nombre plus petit d'éléments à mémoriser. C'est ce qu'a fait René Thimonnier.

Étudier le système de l'orthographe lexicale afin de l'enseigner

L'orthographe française n'a jamais été complètement phonétique. Dès que l'on a essayé de la fixer, on l'a fait avec des lettres latines, qui représentaient donc des sons latins, alors que de nouveaux sons français naissaient depuis 1 000 ans déjà. De plus, son histoire est parsemée de changements : des lettres ont été ajoutées ou supprimées selon les époques, pour des raisons étymologiques (vraies ou fausses), sociales ou analogiques. Devant ce qui semble être un chaos et que plusieurs voudraient réformer, les pédagogues sont souvent dépourvus et tentent d'enseigner une matière mal définie, les linguistes ne se souciant généralement pas du problème, et les psychologues étudiant plutôt les procédés d'acquisition.

René Thimonnier s'intéresse à l'orthographe française non pas pour ce qu'elle devrait être, mais pour ce qu'elle est. Selon lui, ce qu'on appelle « chaos orthographique » n'est pas un vrai problème ; si l'orthographe est observée à la loupe, des règles générales surgissent et un système apparaît. L'élève, une fois les mots classés, n'aurait pas à mémoriser l'orthographe par mots isolés, mais par séries analogiques. Si le mot *honneur* prend deux *n* et *honorer* seulement un, ce n'est pas de façon arbitraire, mais bien parce que les mots dont le radical est *honn* prennent deux *n* et ceux dont le radical est *honor* n'en prennent qu'un : honn*ête,* honn*êteté,* honn*êtement,* honor*able,* honor*abilité,* honor*aire,* honor*ifique,* etc. Aussi, il y a des règles selon les contextes : la consonne nasale (*m* ou *n*) n'est jamais doublée après *i* ou *u* dans le radical du mot, comme dans *finesse, finement, jardinier, brunette, parfumeur* (*immigrant* est formé du préfixe *in-* et du radical *migrant*, donc la règle ne s'applique pas). Un dernier exemple

du système de Thimonnier, les homonymies suffixales : pourquoi *méchamment* avec *a* et *patiemment* avec *e* alors qu'on prononce les deux de la même façon ? Parce que *méchamment* vient de *méch*a*nt* et *patiemment*, de *pati*e*nt*.

René Thimonnier publiait les premiers résultats de ses études en 1967 et, dès 1977, le ministre français de l'Éducation reconnaissait son travail dans la circulaire ministérielle exposant les principes et les modalités d'une nouvelle pédagogie de l'orthographe, basée sur un enseignement méthodique et non plus empirique. Au Québec, son programme ne fait pas officiellement partie du système scolaire, mais les enseignants intéressés à consulter ses travaux pourront nourrir leur curiosité et aussi commencer à ENSEIGNER l'orthographe. Les nombreuses publications de Thimonnier font le tour de la question, allant de l'exposition théorique du système graphique du français aux manuels pratiques suivant les années scolaires.

En résumé, une pédagogie de l'orthographe lexicale est d'abord une classification des données à mémoriser (les mots), afin de faire ressortir le système à enseigner. Comme les professeurs (ou les personnes-ressources des centres d'aide en français) n'ont pas de temps à consacrer à l'étude complète des graphies du lexique français et à l'élaboration d'un programme d'apprentissage, peut-être peuvent-ils faire connaissance avec des auteurs qui l'ont déjà fait pour eux, que ce soit René Thimonnier, présenté dans cet article, ou d'autres chercheurs reconnus, comme Nina Catach.

Pour finir, une petite note terminologique : tout au long de cet article, il a été question d'*orthographe lexicale*, alors que le terme utilisé normalement dans le réseau collégial québécois est *orthographe d'usage*… Cette différence est voulue, le terme *orthographe d'usage* connotant fortement l'arbitraire, alors que le terme *orthographe lexicale* suggère l'idée d'un sous-système dépendant du système lexical. « Tatillonnement » de linguiste ou choix du mot juste ? À vous de choisir !

RÉFÉRENCES

CATACH, Nina, *L'Orthographe française : traité théorique et pratique*, Paris, Éditions Fernand Nathan, 1986.

CATACH, Nina, Daniel DUPRIEZ et Michel LEGRIS, *L'enseignement de l'orthographe : l'alphabet phonétique international, la typologie des fautes, la typologie des exercices*, Paris, Éditions Fernand Nathan, 1980.

DESMEUZES, Jean et René THIMONNIER, *Les 30 problèmes de l'orthographe, méthode Thimonnier, cours complet*, Paris, Hachette, 1979.

THIMONNIER, René, *Le système graphique du français*, Paris, Plon Éditeur, 1967.



Annie Desnoyers

Étudiante au troisième cycle en linguistique, UQAM