

« Tapez le mot qui vous crée une difficulté » Orthographe et Minitel

Charles MULLER

Il ne s'agit pas ici de faire jouer de subtiles opérations statistiques, mais seulement de présenter quelques données quantitatives inédites.

Leur intérêt, s'il existe, est de concerner une démarche linguistique qui, jusqu'alors, échappait à toute observation : la consultation d'un répertoire lexical, le geste, fugace et généralement solitaire, par lequel un usager ouvre un dictionnaire (le dictionnaire ?, son dictionnaire ?), non pour le feuilleter au hasard, mais pour le consulter, l'interroger sur un point précis; le consultant s'oriente rapidement, dans l'ordre alphabétique, vers le mot qui le préoccupe, atteint l'article qu'il cherchait, pour y cueillir le renseignement qui lui faisait défaut.

Quel renseignement ? une signification, une graphie, un fait de flexion, un choix syntaxique ? La diversité est grande, suivant le type de l'ouvrage consulté (encyclopédique ou linguistique), suivant aussi le degré de culture du consultant ... ; mais ne nous attardons pas à de faciles énumérations ! Passons aux faits.

L'expérience « ORTHOTEL » nous place dans un cas particulier, assez limité, ce qui du reste rend ses données interprétables sans trop d'hypothèses. La richesse de cette banque de données est avant tout orthographique, morphologique si l'on préfère; la sémantique y est la parente pauvre; si on l'interroge sur SYNALLAGMATIQUE ou EMPHYTÉOTIQUE, ce sera pour la bonne graphie de ces adjectifs; on y lira une définition très sommaire (un « repère sémantique »), mais qu'on n'attende pas un commentaire juridique.

Le public d'ORTHOTEL se recrute surtout parmi les Français de culture moyenne, et souvent inférieure à la moyenne; parmi ceux (et celles), en tout

✉ Avenue de la Forêt-Noire, 12; F-67000 Strasbourg (France).

MOTS-CLÉS : Orthographe, minitel, statistique lexicale

cas, dont l'écriture n'est pas l'occupation première, qui ne sont guère à l'aise dans le code écrit et devant ses pièges, mais qui, à l'occasion, demandent à leur Minitel un secours, un « dépannage » que le dictionnaire usuel ne leur livre pas toujours, ou pas aussi rapidement, ou moins commodément. Je dirai plus loin ce qui m'autorise, malgré le parfait anonymat de ce public, à en donner une certaine image.

Chaque fois que l'utilisateur du Minitel appelle « 36 15 ORTHOTEL » et ouvre le choix 2 : « LEXIQUE, dépannage immédiat », son écran lui répond : « tapez le mot qui vous crée une difficulté », mais en même temps l'avertit : « attention : un seul mot à la fois; les verbes à l'infinitif, les noms au singulier, les adjectifs au masculin singulier, et pas de noms propres »; en somme, les mêmes limites que celles que requiert la consultation d'un Petit Robert ou d'un Petit Larousse Illustré (première partie). La différence, c'est qu'ORTHOTEL peut répondre même si le questionneur ignore la bonne graphie : qu'il ait tapé IMANENT, ou IMMANANT, ou IMANANT, il obtiendra : *immanent, -te, adj.*, et quelques exemples; pour IGROMETTRE, l'écran répondra *hygromètre, n. masc.*; enfin, en interrogeant ACQUERIRE ou AQUERIR, on lira toutes les formes simples de ce verbe; quant aux adjectifs juridiques cités plus haut, il suffira d'appeler SINALAGMATIC ou ANFITEOTIC pour être servi.

Chaque frappe se traduira par un succès (lecture d'une page concernant le mot appelé) ou un échec (« Notre LEXIQUE ne connaît pas cette forme »). Ce que l'utilisateur ignore, c'est que sa frappe est allée s'inscrire dans une des mémoires de l'ordinateur, soit celle des « présents » (formes ayant abouti à une consultation), soit celle des « absents » (formes non reconnues). C'est la lecture de ces deux mémoires qui nous renseigne, chaque fois que nous le désirons, sur les mots consultés, mais aussi sur les lacunes de notre LEXIQUE. Et, après lecture et parfois listage de ces relevés, nous en purgeons la mémoire de l'ordinateur; mais ils nous ont livré des renseignements qui peuvent intéresser le linguiste.

L'idée m'est venue, l'an dernier, de ne plus anéantir, après une rapide exploitation, ces curieuses données, ces traces quotidiennes des inquiétudes orthographiques d'un public de plus en plus nombreux, mais toujours aussi anonyme et discret. Du 1^{er} août 1991, à 0 heure, au 31 juillet 1992, à 24 heures, le contenu des deux mémoires a été stocké, puis listé, ce qui nous livre maintenant une année de consultation de notre « LEXIQUE ». Voyons ce qu'elle nous enseigne.

Pendant ces douze mois, ORTHOTEL a enregistré 26 912 appels (la glorieuse année de la « réforme » nous avait, en 1990, fait franchir le seuil des 40 000). Le LEXIQUE en a accueilli 14 151, un appel pouvant comporter plusieurs frappes;

le nombre de ces frappes a été de 23 183, dont 8 924 dans les « absents », et 14 259 dans les « présents ».

Réglons d'abord le cas des « absents ». Leur effectif se compose essentiellement (à plus de 80 %) de négligences des usagers : des étourderies surtout, par oubli des conseils : suites de mots [le seul cas des articles ou pronoms, suivis d'un ou plusieurs mots : « A LAQUEL (*sic*) — LES DIMANCHES — JE VOUS SERAIS — IL VA DE SOIT (*sic*) —, etc. » nous fournit 124 occurrences]; noms propres; innombrables formes fléchies. Un(e) acharné(e), ayant égaré un code de cinq lettres commençant par SIR et terminé par I, ne fait pas moins de 32 essais pour identifier la 4^e lettre (sans aucune chance de trouver ce code dans le LEXIQUE).

Outre ces distraits et ces égarés, il y a les maladroits : de simples fautes de frappe peuvent rendre la forme opaque : dans un mot comme STRUCTURE, l'omission d'une seule lettre (sauf l'*E* final) ou son échange fausse la prononciation et empêche l'affichage de la page demandée. On peut certes admettre que dans ce cas l'échec est aussitôt suivi d'une frappe rectifiée, qui va tomber dans les « présents ». Mais, perdus dans la foule de ces infirmes, parmi ce fatras d'erreurs, on découvre quelques « absents » de bon aloi, qui révèlent effectivement des lacunes de notre LEXIQUE; celles-ci sont alors réparées. C'est là le rôle utile de cette mémoire, qui nous permet de dire qu'un mot qui manque dans le LEXIQUE est généralement un mot qui ne nous a pas encore été demandé, ne serait-ce que parce qu'il ne crée aucune difficulté graphique.

Nous porterons donc notre attention sur les « présents ».

Pendant cette période de 12 mois, les 14 259 interrogations réussies ont concerné 5 993 mots différents (en fait plus de 6 000, car de nombreux appels concernent plusieurs unités : FIER conduit à l'adjectif et au verbe, BOUCHER au nom et au verbe, COU ouvre sur « cou », « coup », « coût », etc.); notre LEXIQUE comptant environ 25 000 vedettes, ce serait un sixième de la nomenclature qui a été exploité pendant ces douze mois.

Cet effectif n'est en aucune façon une limite, car ce sous-ensemble n'est pas stable. En effet, si l'on divise l'année en deux semestres, on observe que chacun d'eux a appelé plusieurs milliers de mots ignorés par l'autre; d'où l'on peut conclure que les périodes précédentes (dont l'enregistrement n'a pas été conservé) ou suivantes (que nous étudierons) augmenteraient largement cette proportion.

La structure quantitative de ce corpus rappelle ce que l'on observe chaque fois que l'on quantifie le vocabulaire d'un texte :

f_i	V_i
> 10	138
10	25
9	30
8	56
7	54
6	123
5	156
4	298
3	552
2	1 161
1	3 400
	5 993

$N = 14\ 259$
$V = 5\ 993$
$V_1 = 3\ 400$
$\bar{f} = 2,38$
$p_1 = 0,567$

(Notons cependant que cette distribution n'obéit pas au modèle de Waring.)

De la structure, passons au contenu. Les mots appelés le plus fréquemment ont été :

avoir	106	pouvoir	59	ci-joint	43	orthographe	30
tout	88	dysfonctionnement	49	gré	41	renouveler	30
être	74	permettre	47	quelque	30	compte-rendu	28
rappeler	73	cent	44	appeler	37	donner	26
faire	61	voir	44	joindre	32	dû	26

Ce début de la liste révèle déjà les tendances de la consultation. D'abord les verbes, les verbes fréquents et irréguliers, dont certaines formes créent l'hésitation (*ayons* ou *ayions* ?, *faisait*, mais *feruit* ?, je me *permet(s)* ?, je *join(d)s* ?, etc.), et ceux qui mettent en jeu des consonnes doubles (*rappeler*, *appeler*); quant à *donner*, c'est sa conjugaison qui sert de modèle pour tous les verbes réguliers. Les pronoms *tout*, *quelque* (ou *quel que*), le numéral *cent* tendent des pièges bien connus; le participe *dû*, avec son accent circonflexe à éclipses, crée l'inquiétude; *compte rendu*, son pluriel et son absence de trait d'union ouvrent la série des problèmes posés par les composés.

Gré donne accès à l'inusable « je vous saurai(s) gré de ... », où l'usager flotte entre un conditionnel et un futur, quand il ne confond pas *saurais* et ... *serais*, sans parler de la questionneuse qui s'interroge sur un éventuel accord de *gré(e)*. Ajoutons enfin l'apparition des termes à la mode avec *dysfonctionnement*. Un seul de ces mots relève du jeu, de l'absence de consultation pratique : *orthographe*, tapé avec des fantaisies graphiques (la forme correcte 4 fois seulement, les 26 autres occurrences se distribuant entre 9 formes sauvages...) dues au seul désir de tester, sans chercher bien loin, la souplesse de notre logiciel.

Si la syntaxe n'est pas absente de ce relevé, il est pratiquement impossible d'en évaluer l'importance dans les interrogations sur *avoir, être, faire, etc.*

On notera que ce vocabulaire est lié à la vie pratique, à la correspondance utilitaire, professionnelle, commerciale, administrative, contentieuse, ..., où abondent les « je vous *joins* (mon C.V.), je me *permets* de vous *rappeler* ..., les sommes *dues* ..., les *comptes rendus* de nos réunions... », etc.

Un trait remarquable de ces données est leur grande stabilité dans le cours de cette année. Parmi ces mots à haute fréquence, un seul a des effectifs très inégaux entre les deux semestres : *tout*, avec ses formes *tous* et *toute*; de 30 occurrences pour les six premiers mois, il passe à 58 pour les suivants; il semble que de nombreux usagers aient constaté, au cours de cette année, qu'une série de problèmes grammaticaux (*tout(e)* entière, en *tout/tous* temps, etc.) peuvent être résolus par la consultation du LEXIQUE. Quant aux autres mots de la liste, la comparaison des sous-effectifs semestriels, par une série de χ^2 , n'a donné aucun écart significatif.

Beaucoup de scripteurs inquiets ne découvrent pas immédiatement les ressources du LEXIQUE. Car ORTHOTEL leur propose d'autres secours : le choix 3 permet de poser une question : « comment écrire "tout(e) entière" ? » — « quel est le futur d'acquérir ? », dont on lira la réponse le lendemain, au réveil; le choix 4 les invite à taper non le « mot qui crée une difficulté », mais un texte ou une phrase qu'on soumet à notre correction, effectuée dans la demi-journée. Ce sont les innombrables questions et les abondantes fautes des textes qui nous renseignent sur les défaillances les plus fréquentes et sur les contextes que les appels au LEXIQUE ne révèlent pas. Des centaines de fois, nous avons arbitré des phrases comme : « les difficultés que j'ai eu(es) », ou : « que j'ai eu(es) à vaincre », « les travaux que j'ai fait(s) exécuter », etc., et la plupart du temps notre réponse s'accompagnait d'une charitable indication : « Notre LEXIQUE vous l'aurait dit ». On ne compte plus les questions ou les phrases à corriger sur « je vous saurai(s) gré » (« faut-il un S ? »), ou « ci-joint les documents », etc. C'est cette partie interactive d'ORTHOTEL qui nous donne quelques lumières sur les motivations des usagers du LEXIQUE.

Ce n'est que dans les fréquences faibles, dispersées entre des milliers de mots différents, qu'apparaissent des intentions sémantiques. En somme, notre public, dans une majorité croissante, a compris que nous répondons volontiers à des questions sur le sens, le niveau stylistique, la légitimité (« est-ce français ? »), l'origine des mots, mais que le LEXIQUE, lui, est centré sur la forme : « comment dois-je écrire... ? »

La lecture de nos listes confirme que l'une des difficultés les plus mal maîtrisées, au niveau lexical, est le choix entre consonne simple et consonne

double; que les *Y* attirent les *H*, et réciproquement; que certains mots, plus souvent entendus que lus, sont victimes de prononciations approximatives (*dilemme, aréoport, frustrer, omnubilé, suggestion*, etc.); le champion de cette catégorie est la locution très médiatique *d'ores et déjà*, dont nous aurions pu collectionner les transcriptions les plus inattendues (*d'hors, ..., dors*, etc.). Quant à la syntaxe, c'est une statistique sur notre choix 3 (Questions) qui fournirait des données utiles, et parfois surprenantes; mais tel n'est pas notre sujet d'aujourd'hui.

Un coup d'œil sur les mémoires des semaines suivantes (du 1^{er} août au 15 septembre 1992) confirme les données recueillies pendant ces douze mois : les mots souvent appelés conservent leur forte présence (AVOIR, 16 appels, sans y compter les AVIEZ, AVAIT, etc., égarés dans les « absents »; ÊTRE : 10; PARAÎTRE : 5; VOIR : 4; etc.). À l'inverse, les demandes à faible fréquence manifestent leur mobilité; les deux tiers des appels concernent des mots qui, pendant la période précédente, n'avaient pas figuré dans nos listes de « présents ». C'est dire que, sur une longue durée, la quasi totalité de notre nomenclature est consultée. En outre, ces six semaines, par les « absents », nous dicent une vingtaine d'additions.

On compare volontiers notre activité à celle des écrivains publics de jadis, qui du reste renaissent dans quelques villes. Si nous avons comme eux le souci de « dépanner » ceux à qui l'écriture pose des problèmes sérieux, nous différons de ces professionnels quand, invités à fabriquer une quelconque lettre, nous répondons : « ORTHOTEL corrige, mais ne rédige pas ». Et pourtant, l'an dernier, cédant à des demandes diverses, nous nous sommes risqués à ouvrir un choix d'« Aide à la Correspondance ». Mais c'est là une autre affaire, où le rôle de la grammaire est purement négatif, et où la statistique n'a pas encore à se manifester.

Ces quelques données montrent que l'expérience télématique qui se poursuit sous l'étiquette d'ORTHOTEL ouvre des regards sur le maniement du code écrit, tant lexical que syntaxique, et sur les soucis qu'il crée à bon nombre de francophones. On pourra sans doute y puiser la matière de quelques enquêtes révélatrices.