

L'internet pour corpus de l'usage orthographique

Ange Bizet

Presque tous les dictionnaires, à commencer par celui de l'Académie française, se disent d'usage. Les utilisateurs les considèrent comme source de la norme. Cette fonction de référence normative est renforcée par les jeux de société ou télévisés. Le corpus de référence des dictionnaires est surtout composé de littérature et de publications périodiques, passées au crible des correcteurs, automatiques ou pas, qui ne font que suivre le dictionnaire. Ce cercle vicieux fige, fossilise l'orthographe.

Un corpus immense.

L'internet présente un gigantesque corpus, constamment alimenté et mis à jour, bien plus important que ceux ordinairement utilisés par les dictionnaires. La rapidité de la consultation est telle, qu'en une fraction de seconde, on trouve des centaines de milliers, voire des millions d'occurrences, qu'une vie entière n'aurait pas permis de relever par les moyens classiques, donnant un aperçu instantané de l'usage.

On peut gloser sans fin sur la définition de *l'usage*. La toile présente l'éventail le plus large de l'écrit, du discours très soutenu, à la langue sauvage, composé de textes savants, de textes très officiels politiques ou administratifs, de textes journalistiques, humoristiques, commerciaux, des catalogues publicitaires, de l'écrit non discursif, listes, énumérations, tableaux..., peu de littérature à proprement parler, les ouvrages disponibles sur l'internet ne le sont pas directement par les moteurs de recherche. Les auteurs de ces écrits en sont des universitaires, des chercheurs, des organismes gouvernementaux, des organisations internationales, des amateurs plus ou moins éclairés, particuliers qui cherchent à partager leur passion ou qui diffusent leur curriculum vitae, des sociétés commerciales ou industrielles, des agences immobilières ou hôtelières, des groupes ou associations de toute sorte, club de pétanque ou de scrabble. La représentation de ces différents types de sources varie selon le registre ou le domaine du vocable recherché.

Les sites sont souvent des vitrines de prestige. L'orthographe n'y est pas spontanée, les textes sont pour la plupart passés à la moulinette des correcteurs de traitements de textes, ou contrôlés dans un dictionnaire plus ou moins récent, les gens ont rarement le sentiment que l'orthographe change.

Les moteurs de recherche exercent un certain contrôle de qualité par leur sélection pour le référencement. Rien à voir avec le langage sauvage des forums, des *chats* (tchates), et des courriels, dans lesquels on s'applique à casser l'orthographe ; SMS, abréviations de tous genres, allographie, orthographe approximativement phonétique, frimousses¹ composées de signes de ponctuations et de lettres. La transgression de la norme devient la règle, pour mieux masquer souvent les carences du scripteur. La bonne tenue de l'écrit sur les sites donne d'autant plus de poids, dans l'analyse des résultats, aux formes s'écartant de la norme. « Fautes » involontaires ou choix conscients sont alors très significatifs.

Méthodologie

La recherche sur la toile est un outil qui, comme tout autre, demande une maîtrise méthodologique pour valider les résultats et en tirer des conclusions. Examinons quelques problèmes qui se présentent dans la recherche lexicologique à travers quelques exemples. D'abord deux cas simples, la proportion des variantes, et les comparaisons en séries. Il suffit d'interroger le moteur de recherche successivement avec chacune des formes.² Le nombre obtenu n'est pas celui des occurrences à proprement parler, ni celui des sites, mais le nombre de pages de sites contenant au moins une fois la forme recherchée.

Les cas le plus simple

imbécillité : 4660 *imbécilité* : 1870

On peut en conclure que la forme régulière recommandée par les rectifications de 1990, avec un seul *l* comme *imbécile*, est bien entrée dans l'usage.

Une série

Ripou, verlan de *pourri*, désigne les policiers corrompus. Régulièrement son pluriel devrait se faire en *-s*. Seulement, ce vocable argotique est entré dans la langue courante par un film de Claude Zidi en 1983 (suivi de deux autres). Or le titre du film « Les ripoux » a adopté un pluriel irrégulier en *-x*, certainement

¹ Nom français pour remplacer « *smiley* ».

² Tous les exemples traités le sont sur Google. Les raisons de ce choix sont exposées plus bas.

par jeu de mot pour le rapprocher de *poux* dont la connotation est proche de *pourri*. Les dictionnaires qui ont admis le vocable donnent généralement les deux pluriels. La recherche sur la toile donne la mesure (01-11-03) :

	<i>ripoux</i>	%	<i>ripous</i>	%
<i>ripou</i>	5 660	98,6	78	1,4

La figure de style, qui crée du sens par un écart à la norme, n'était évidemment pas explicite. A tort, l'usage a pris modèle sur l'exception. La proportion 1,3 %, reste du même ordre que pour :

	x	%	s	%
<i>hibou</i>	137 000	99,75	350	0,25
<i>bijou</i>	158 000	99,6	704	0,4
<i>caillou</i>	65 700	98,4	1 040	1,6
<i>genou</i>	94 300	98,3	1 640	1,7
<i>Joujou</i>	5 890	91,53	545	8,47

La proportion de l'écart à la norme est le même pour les réguliers :

<i>écrou</i>	164	0,7	21 400	99,3
<i>bambou</i>	282	1,1	25 900	98,9
<i>tabou</i>	635	1,5	40 900	98,5
<i>gourou</i>	278	2,2	12 800	97,8

Ces données statistiques éclairent autrement la question. Comment choisir le pluriel de *ripou*, selon la règle ou selon l'usage. L'attitude normative commanderait de dénoncer *ripoux* comme fautif. En le retenant le dictionnaire ne fait que conforter la « faute ». L'usage ainsi établi commanderait de ne garder que cette forme. Il faudrait alors allonger la liste des exceptions. La solution ne serait-elle pas de supprimer cette liste en régularisant tous les cas ? L'idée n'est pas nouvelle, cette « réforme » était déjà au programme au début du siècle. Les rectifications de 1990 n'abordent pas la question.

Les variantes graphiques peuvent être multiples.

(01-11-2003)			
<i>gas(-)oil</i>	5 490	<i>gaz(-)oil</i>	793
<i>gasoil</i>	13 600	<i>gazoil</i>	1 490
<i>gasol</i>	1 160	<i>gasole</i>	172
<i>gazol</i>	96	<i>gazole</i>	16 500

Dans ce cas, il est indispensable de limiter le champ de recherche à la langue française. La forme adoptée par les rectifications, *gazole*, arrive en tête, mais *gasoil* correspondant à la prononciation populaire française /gazwal/ reste presque aussi importante ; cette forme est conseillée par « Le millepatte sur un nénufar »³.

Les flexions

La recherche des formes fléchies est souvent indispensable. La proportion de l'emploi au singulier ou au pluriel, au masculin ou au féminin, peut varier du tout au tout pour certains mots. Ce type de recherche statistique permettrait d'affiner l'étude de ces variations.

<i>fantasme</i>		%	<i>phantasme</i>		%	total
<i>fantasme</i>	88 400	76,3	<i>phantasme</i>	27 500	23,7	115 900
<i>fantasmes</i>	94 800	85,1	<i>phantasmes</i>	16 600	14,9	111 400
02-11-2003	183 200	80,6		44 100	19,4	227 300

Les deux formes concurrentes sont considérées toutes deux comme correctes. Qui a dit que supprimer les « lettres grecques », comme les autres langues romanes, dénaturerait le français ?

Si l'on n'a pas besoin de connaître la proportion des différentes formes fléchies, il est possible en « recherche avancée » de les écrire dans la case « au moins un des mots suivants ».

fantasme(s) 150 000 81,9 % *phantasme(s)* 33 100 18,1 % 183 100

Cependant le total est plus faible, les pages sur lesquelles figurent à la fois le singulier et le pluriel ne sont comptabilisées qu'une seule fois. Une simple soustraction permet de les dénombrer, ce qui peut être utile dans certaines recherches. Ici, les proportions restent du même ordre.

Le pluriel de *scénario* présente plusieurs variantes.

- 190 000 *scénarios*, pluriel régulier⁴, considérant que le mot est français.
- 26 300 *scenarii*, avec la terminaison italienne.
- 3 650 *scénari*, méconnaissance de la grammaire italienne.
- 2 650 *scénaris*, forme batarde injustifiable.

³ « Le millepatte sur un nénufar » AIROÉ, 2002. Liste des mots avec l'orthographe recommandée des *Rectifications* de 1990.

⁴ Pluriel de l'orthographe recommandée par les *Rectifications orthographiques* de 1990. (NDA 7-2020)

Cette recherche se limite à la variation de la terminaison. Les variables sont beaucoup plus nombreuses si l'on prend en compte les accents. Ce critère mettrait en évidence que les formes qui se veulent italiennes ignorent l'accentuation dans cette langue de *scenàrio*. *Scénario* est bien devenu un mot français⁵. Alors, que les dictionnaires cessent de proposer *scenarii* ou *scénarii* comme forme savante ! Ce n'est souvent que la manifestation d'un snobisme pédant et ignorant.

La démarche est la même pour le genre. Il peut être nécessaire de chercher jusqu'à quatre formes pour une seule orthographe.

Le dérivé de **Prague**, adjectif et gentilé, se présente sous deux formes, une régulière *pragois*, une irrégulière *praguois* qui conserve, sans raison, comme pour les verbes, le *u* après le *g*. Pour donner la mesure, il est préférable d'avoir le maximum d'occurrences et donc de rechercher les formes fléchies.

	<i>praguois</i>	<i>pragois</i>	total
masculin singulier et pluriel	597	3 610	
féminin singulier	216	1 400	
féminin pluriel	71	399	
total	884	5 409	6 293
	14 %	85 %	100 %

- 5 -

La forme régulière est très largement dominante. La forme irrégulière ne se maintient que par sa présence dans les dictionnaires. Ce résultat statistique de l'usage ne justifie-t-il pas de n'y conserver que *pragois* ?

Certains cas sont difficilement exploitables. Celui de ces différentes graphies en est un exemple.

<i>millepatte</i>	78	<i>millepattes</i>	640
<i>mille(-)patte</i>	286	<i>mille(-)pattes</i>	6 830

D'abord, seul un décompte « manuel » permet de distinguer pour les formes en deux mots, celles avec trait d'union.

Les résultats statistiques ne sont pas exploitables car les formes en *-s* ne permettent pas de distinguer le pluriel de *millepatte*, du singulier *millepattes*,

⁵ D'autant plus qu'en italien *scenàrio* signifie *décor*, le français *scénario* se traduit par *sceneggiatura*. La question du pluriel sera réglée, dans le milieu professionnel on utilise couramment l'apocope *un scénar, des scénars*.(NDA 07-2020)

identique à son pluriel. Autrement dit la première colonne correspond à un seul cas singulier, tandis que la seconde confond le pluriel de la première avec le singulier et le pluriel de la seconde option. La seule possibilité serait une analyse contextuelle laborieuse.

La conjugaison

L'étude portant sur des verbes peut aussi devenir fastidieuse par la recherche des formes conjuguées. Une recherche limitée aux plus fréquentes sera souvent suffisante. Pour les comparaisons, il faut s'assurer que la base est bien la même.

Les accents

Pour aborder la question de l'accentuation évoquée dans le cas de *scénario*, prenons un autre exemple qui permette d'en aborder les différents aspects. L'interrogation simple sur un mot ne tient pas compte des accents.

Sans revenir ici sur les arguments en faveur de l'une ou l'autre solution d'accentuation de *evenement*, rappelons simplement que le cas n'a pas été posé en 1990 par les rectifications. L'Académie avait adopté *Évènement* en 1975, puis l'avait rejeté en 1984, et l'a enfin approuvé dans les rectifications. Essayons de saisir l'usage actuel.

La saisie de l'une ou l'autre forme accentuée donne la somme de toutes les formes. La solution consiste en « recherche avancée » à inscrire dans la case « aucun des mots suivants » les formes avec les autres accentuations.

événement – (évènement evènement evenement)	568 000
événements – (évènements evènements evenements)	1 400 000
total « événement »	1 968 000
évènement – (événement évènement evenement)	79 300
évènements – (événements évènements evenements)	166 000
total « évènement »	235 300

L'inconvénient de cette procédure est d'éliminer toutes les occurrences quand la forme refusée figure sur la même page. Le résultat statistique n'en est pas moins significatif.

Les majuscules

Les moteurs de recherche ne permettent pas de distinguer les majuscules. Une pratique courante est de ne pas accentuer la majuscule initiale. Beaucoup croient même que cette tolérance, justifiée par d'anciennes contraintes d'imprimerie, est une règle. Il faut donc prévoir de rechercher aussi les formes avec l'initiale sans accent, quand cela est nécessaire. Relançons la recherche en inscrivant les deux formes é- E- dans la case « l'un des mots suivants ». On obtient alors ces résultats :

événement Événement – (évènement evènement evenement)	582 000
événements Événements – (évènements evènements evenement)	1 380 000
total « événement »	1 962 000
« évènement Evènement –évènement –évènement –evenement »	87 000
« évènements Evènements –évènements –évènements –evenements »	235 000
total « évènement »	322 000

La comparaison fait apparaître une augmentation attendue des occurrences de *évènement* (de 235 300 à 322 000, soit 37%). La stagnation et même la diminution de 3 % de *événement* est plus surprenante. L'explication réside dans le fait que le moteur de recherche dispose d'un dictionnaire de référence, analogue à celui d'un correcteur, qui détermine la forme « correcte ». Il propose alors directement les variantes « correctes » notamment avec l'initiale non accentuée, ce qu'il ne fait pas pour une forme considérée « incorrecte ». Une preuve de plus que la mise à jour des correcteurs s'impose. La diminution est due au fait que la recherche sur deux formes ne comptabilise qu'une seule fois les pages sur lesquelles figurent les deux, ici avec ou sans majuscule à l'initiale.

- 7 -

L'homonymie.

la recherche des différentes graphies de *nénufar* sur le « web » donne les résultats suivants :

	singulier	%	pluriel	%
<i>nénuphar</i>	12 100	77,5	10 500	98,8
<i>nénufar</i>	3 520	22,5	133	1,2
total	15 620	100,0	10 633	100,0

La disparité dans l'emploi de *nénufar* entre le singulier et le pluriel semble anormale. l'examen du corpus montre une présence massive de textes en

espagnol. Il est donc nécessaire de mener la recherche sur les pages francophones :

	singulier	%	pluriel	%
<i>nénuphar</i>	8710	97,3	9690	
<i>nénufar</i>	242	2,7	103	1,1
total	8952	100	9793	100

L'ancienne graphie rétablie par les rectifications de 1990, reste marginale, même si une bonne part de la forme en *ph* se trouve dans des titres d'œuvres, des noms de bateaux... Il s'agit bien pourtant d'une rectification, le mot n'est pas d'origine grecque mais arabe. La graphie prétendument savante *nénuphar* (Académie 1935) s'est répandue par contamination de *nymphéa*.

Il est indispensable de vérifier que les références données par le moteur correspondent bien au mot recherché. Même dans les pages en français, peuvent sortir des extraits en langues étrangères, particulièrement en anglais, fréquemment dans les adresses. L'homonymie existe aussi à l'intérieur de la langue.

Par exemple, en cherchant les dérivés du nom de la ville d'*Aden* au Yémen, on trouve un grand nombre d'occurrences de *adénite*, formé avec un suffixe classique en français pour les noms arabes. L'examen montre que la quasi totalité des cas se trouvent sur des sites médicaux, l'adénite étant le nom d'une inflammation des ganglions.

Dans l'étude du pluriel des mots en *-ou* ci-dessus, nous avons dû renoncer à présenter le résultat pour *chou*, la recherche de *chous* faisant ressortir de nombreux *bout'chous*, *boud'chous*, *chous-chous*... Dans le cas où la réponse est indispensable, il faut pratiquer un décompte « manuel ». Il est souvent possible de simplifier le contrôle par élimination de résultats parasites en utilisant la case « aucun de ces mots », par exemple *bout* dans le cas présent, sachant que tous les *chous* sur la même page qu'un *bout* seront éliminés avec.

Ce tri, qui oblige à lire le contexte, parfois à ouvrir la page, et souvent utiliser la fonction « rechercher dans la page » pour retrouver les occurrences, n'est possible que pour un nombre restreint d'occurrences. Le travail peut être considérable quand le nombre de pages est très grand. Reste alors à faire une estimation sur un échantillon ; ce peut être suffisant dans certains cas.

La recherche du genre.

La recherche du genre, qui a des implications sur l'orthographe grammaticale, peut se faire par une analyse contextuelle syntaxique automatique en utilisant la case « cette expression exacte » en « recherche avancée ».

un orque 1 430 / ***une orque*** 295

Cet exemple est limité à la simple recherche avec l'article indéfini. Les autres articles sont ambigus du fait de la voyelle initiale. Le résultat est suffisamment significatif ainsi. Il montre que l'usage ignore la norme qui veut que ce nom de l'épaulard soit féminin. Les dictionnaires qui veulent suivre l'usage ne peuvent-ils pas prendre cette évolution en compte ? Que le mot latin *orca* soit féminin ne devrait pas influencer. Les cas de changement de genre sont nombreux. La plus grande partie des noms dont le genre est mal maîtrisé ont une voyelle initiale dont seul l'article indéfini est discriminant mais atténué par la liaison. Prenons un exemple à consonne initiale qui permet une recherche avec différents articles, sans oublier les formes contractées du masculin.

<i>planisphère</i>			
un	1240	une	256
le	1010	la	282
au	156		
du	262		
ce	82	cette	16
	2750		554 3304
	83,2 %		16,8 %

- 9 -

Ce corpus n'est que partiel, il ne prend pas en compte les occurrences sans article, avec un article indéterminé *de* ou remplacé par un nombre, ou dans lesquelles l'article est séparé du nom, par un adjectif par exemple. Pour en prendre la mesure, on peut comparer le total ainsi obtenu avec la recherche du nom seul, sachant que parmi ces occurrences beaucoup viennent de contexte non discursif.

Le nombre ainsi obtenu pour *tique* avoisine les 100 000. L'examen par sondage du corpus fait apparaître de nombreuses troncatures de mots en *-tique*, *inform@tique*, des formes conjuguées du verbe *tiquer*... Dans ce cas, l'analyse automatique n'est pas pertinente. Si l'on veut vraiment ce résultat, il faut revenir l'étude classique fastidieuse du corpus. Pensons quand même à ce que représenterait la réunion d'un tel corpus par les moyens classiques.

Constitution de corpus pour l'analyse contextuelle.

L'étude de l'orthographe grammaticale, accord de l'adjectif, du participe passé..., impose l'examen du contexte. Le moteur de recherche ne peut pas distinguer la nature grammaticale des mots, ni déterminer les relations entre eux. La seule possibilité est de sélectionner des chaînes de mots dans « cette expression exacte », méthode employée pour les articles, ci-dessus.

Nous cherchons à observer l'usage de l'accord du participe dans la séquence « *passé(e) la barrière* ». Nous ne pouvons pas copier ici le résultat de la recherche, mais il est possible de reproduire la procédure pour suivre le commentaire.

- « *passée la barrière* » donne un corpus de 65 cas. On relève « ...passée la barrière linguistique, ... », « *Passée la barrière de l'Atlas, ...* », « ... une fois passée la barrière des videurs,... », « Dès que sera passée la barrière de l'an 2000... »... Un document sous *Adobe* se révèle inconsultable. Enfin un cas est à rejeter : « ... l'enquête de référence de l'année *passée, la barrière* critique... » ; la ponctuation n'est pas prise en compte. Nous avons donc 63 occurrences dont certaines apparaissent en double. Le moteur fait directement le tri au départ et ne présente que 44 exemples dont 42 sont pertinents.

- « *passé la barrière* » donne plus de 1000 pages. Un premier contrôle révèle de nombreux cas avec *passé*, les accents ne sont pas pris en compte.

En éliminant « *passé* » il ne reste que 138 cas parmi lesquels de nombreuses formes conjuguées avec *avoir*, « ... ils ont passé la b... », « il a p... ». Plutôt que d'éliminer chacune des nombreuses variantes, avec un corpus de cette taille, il est certainement plus rapide de le faire en lecture directe.

En dehors de la conjugaison avec *avoir*, et après avoir éliminé deux cas d'emploi du participe passé au lieu de l'infinitif, « ...la route doit ensuite passé la barrière... », « ... laisse le public passé la b... », il reste 19 cas, dont 9 « Une fois passé la barrière ». Pour établir une comparaison, la recherche sur « Une fois passée la barrière » en donne 18.

On relève 10 mentions du titre de l'ouvrage de Damon Knight « *Passé la barrière du temps* » qui apparaît aussi 5 fois sous la forme « *Passée la b...* »...

Autrement dit, on peut mener une analyse classique en s'aidant pour les comptages des fonctions du moteur. Nous ne tirerons pas ici les conclusions sur ce cas, le but est de montrer que l'internet permet de constituer facilement un

corpus d'emplois réels, ce qui est nettement supérieur à une étude sur des exemples fabriqués.⁶

L'étude de l'évolution de l'orthographe.

Le dérivé de *McDonald's* connaît des formes multiples combinant les variantes, en un ou deux mots, avec ou sans majuscule à l'initiale *M* et au *D*, absence du second *d*, avec ou sans trait d'union, voire apostrophe. L'observation des différentes formes à deux dates différentes montre une évolution dans les graphies de ce néologisme.

juillet 2002

<i>McDonaldisation</i>	83	17,4 %	<i>Macdonaldisation</i>	190	39,7 %
<i>Mc Donaldisation</i>	8		<i>Mac Donaldisation</i>	37	7,7 %
<i>McDonalisation</i>	55	11,5 %	<i>Macdonalisation</i>	100	20,9 %
<i>Mc'Donalisation</i>	1		<i>Mac Donalisation</i>	4	
<i>Mc ...</i>	147	30,7 %	<i>Mac ...</i>	331	69,2 %
total				478	

- 11 -

novembre 2003

<i>McDonaldisation</i>	114	10,9 %	<i>Macdonaldisation</i>	697	66,6 %
<i>Mc Donaldisation</i>	7		<i>Mac Donaldisation</i>	38	3,6 %
<i>McDonalisation</i>	51	4,9 %	<i>Macdonalisation</i>	132	12,6 %
<i>Mc'Donalisation</i>	3		<i>Mac Donalisation</i>	5	
<i>Mc ...</i>	175	16,7 %	<i>Mac ...</i>	872	83,3 %
total				1047	

On se prend à rêver de la possibilité d'une consultation rétrospective. La comparaison fait apparaître, à 15 mois d'intervalle, une évolution significative dans deux directions presque contradictoires. La première touche l'orthographe de la particule *Mac/Mc* introduisant, à l'origine, un nom clanique écossais ou irlandais. En américain, la forme britannique *Mac* est abrégée en *Mc*. La chaîne de respide⁷ *McDonald's* est américaine, elle symbolise, avec *Coca Cola* et *Disney*,

⁶ D'autres exemples d'analyse contextuelle dans « Timor, l'orientale », in *Mots, les langages du politique* n° 71, mars 2003, ENS Éditions. On y trouvera une étude de la dérivation des composés, en dehors des questions orthographiques.

⁷ *Respide* : néologisme proposé par l'auteur pour remplacer l'anglais *fast food*. Mot centaure composé par troncature et contraction de *res(tauration)+(ra)pide*, ou de *res(restaurant)+(ra)pide* ; *des respides*. Que la fin du mot sonne à l'oral comme l'anglais *speed*, n'a rien de gênant, il n'y a pas incompatibilité sémantique. Pour exprimer

l'impérialisme que les Américains préfèrent appeler maintenant « *globalization* », mondialisation économique et culturelle sur le modèle US. Les formes en *Mc* représentaient encore presque un tiers des occurrences en 2002 ; leur part est quasiment réduite de moitié, au profit des formes en *Mac* qui dépassent 83 % en 2003. *Mc*, et plus encore la séquence *McD*, sont incompatibles avec le système français, le modèle britannique originel permet une discrète et efficace francisation.

La deuxième évolution a un mouvement contraire. Néologisme sauvage, certainement d'abord oral, dans un registre non officiel du discours de contestation politique et économique, l'expression a été transcrite selon sa prononciation française qui élide le *d* après un *l*. Les formes sans *d*, qui représentaient plus d'un tiers des cas, n'atteignent plus les 20 %. *Macdonalisation* reste en deuxième position, mais sa part est passée de plus de 20 % à moins de 13, au profit de *Macdonaldisation* qui représente maintenant deux tiers des occurrences.⁸ L'élision du *d* est pourtant conforme au génie de la langue, mais cette tendance à l'orthographe, qui se veut « savante », pourrait amener à *McDonaldisation* si l'on n'y prend pas garde, introduisant une scorie de plus dans la langue.

Les limites du procédé, qui ne permet pas de distinguer les variantes, avec ou sans majuscules ou trait d'union, ne sont pas gênantes en l'occurrence.

Les différents moteurs de recherche.

Nous avons pris tous nos exemples sur *Google* qui se révèle le plus performant. Les autres moteurs ont leurs qualités propres, il ne faut pas oublier qu'ils n'ont pas été conçus pour ce type de recherches. Il existe des dizaines et même des centaines de moteurs, il n'est pas question de tous les présenter, voyons simplement quelques uns des principaux. *Lycos* est efficace pour la recherche de références, mais il n'a pas de « recherche avancée » ce qui le limite pour la lexicologie. *Voilà* a bien cette fonction, mais n'indique pas le nombre de pages trouvé, il référence les sites, mais ne donne pas directement l'extrait où figure l'occurrence, ce qui oblige à ouvrir chaque page pour en connaître le contenu. *Alta*

l'idée de rapidité, les mots anglais, *fast* (*fast food*, *Fast huîtres*), *speed* (*speedy*, *spizza*), *quick* (*Quick*, chaîne de respide et petit déjeuner), sont depuis longtemps employés en France dans le public. *Respide* est formé de mots français.

⁸ L'étude (polycopiée) sur « les mots de la mondialisation » portait également sur les corrélations de formes et d'occurrences avec *mondialisation*, *globalisation*, *disneylandisation*, mais cela dépasse le cadre de l'orthographe.

vista donne des résultats beaucoup moins nombreux, et ne permet pas de sélectionner une langue, ce qui, comme on l'a vu, est souvent indispensable. Quand aux métamoteurs du type *Copernic*, ils permettent une large exploration dans les différents moteurs, mais plafonnent les résultats, rendant impossible une étude statistique. *Google* est mixte, il complète sa base en allant chercher dans d'autres, sans limiter les résultats. Il est le mieux adapté à nos besoins, il a aussi, comme nous l'avons vu, ses limites. Le fait de ne pas prendre en compte les accents et les majuscules est aussi souvent un avantage. De pouvoir faire une recherche par une chaîne de caractères, comme dans les traitements de texte, serait très utile, par exemple pour les différentes flexions d'un mot, les formes conjuguées, les mots d'une même famille...

Notre objectif est de montrer ici la pertinence d'un outil de recherche lexicologique, d'en montrer les avantages et les limites, sans prétendre faire le tour de la question. Comme pour tout outil, on ne peut en obtenir que ce qu'il est capable de fournir. L'analyse critique des résultats est indispensable comme pour toute autre méthode.

- 13 -

Les lecteurs sont invités à nous communiquer leurs trucs et astuces, pour un fructueux échange autour de cette technique en évolution permanente.