

La *nouvelle orthographe* et son usage dans l'internet français

Jean-Luc AZRA

Introduction

En décembre 1990 paraît au Journal Officiel le *Rapport du Conseil supérieur de la langue française sur les rectifications de l'orthographe*, qui propose, sous forme de recommandations approuvées par l'Académie, des modifications relativement mineures de l'orthographe du français. Certaines sont déjà largement en usage (par ex. *asséner* plutôt qu'*assener*). D'autres modifient plus radicalement la physionomie des mots (un *tirebouchon*, une *île*). S'il est bien accueilli par les institutions et par les usagers de la langue en Belgique, en Suisse et au Canada, ce *Rapport* provoque aussitôt en France une violente levée de boucliers.

Quinze ans après cette parution, la langue est véhiculée par un média entièrement neuf, internet. Nous disposons également d'outils extrêmement puissants d'analyse de documents, dont les membres du Conseil supérieur de la langue française de l'époque ne disposaient pas : les moteurs de recherche tels qu'Alta Vista ou Google. Dans ce contexte, il m'a paru intéressant de faire une évaluation de la manière dont « ancienne » orthographe, « nouvelle » orthographe, et orthographes alternatives se partagent le terrain. Malgré les réticences de l'époque, la « nouvelle orthographe » a-t-elle fini par prendre pied en France chez les usagers ordinaires ?

Par ailleurs, une telle recherche peut-elle nous apprendre des choses sur les usages langagiers des Français dans leur ensemble ? Peut-on y trouver des éléments pour comprendre pourquoi les rectifications n'ont pas été acceptées, et comment, éventuellement, guider une future réforme ?

I. La “réforme” de 1990

En 1989, le gouvernement Rocard crée le *Conseil supérieur de la langue française*, qu’il charge de préparer un rapport sur des «aménagements orthographiques» destinés à éliminer un certain nombre d’irrégularités et de graphies exceptionnelles dans l’orthographe du français (*Journal Officiel*, 6/12/1990). Le mot *réforme* est soigneusement évité dans toutes les publications officielles.

Le projet initial de Michel Rocard propose de traiter cinq points seulement :

- le trait d’union [c’est à dire (1) l’usage ou non du trait d’union dans les composés, (2) le choix entre trait d’union et soudure]
- le pluriel des noms composés
- l’accent circonflexe
- le participe passé des verbes pronominaux
- diverses anomalies [telles que le *i* de oignon ou le verbe *interpeller*].

Le conseil est composé de personnalités telles que Tahar Ben Jelloun, Claude Hagège, Luc Montagnier, Jean Daniel, Bernard Pivot, Pierre Perret, Jean-Luc Godard, entre autres. Il fait également appel à des experts extérieurs, linguistes comme Bernard Cerquiglini, Nina Catach, Pierre Encrevé, André Goosse ou André Martinet, mais aussi à des lexicographes (venus du *Larousse* et du *Robert*) et au chef-correcteur du *Monde* (Goosse 1991, 21-23).

En 1990, un premier projet est soumis à l’Académie Française, qui le trouve trop radical. Un second projet est accepté (à l’unanimité, selon le JO) et est publié au *Journal Officiel de la République française* le 6 décembre 1990.

Trois points importants sont à souligner :

- Tout d’abord, il s’agit d’un *Rapport* et non d’un projet de loi. Il est donc non prescriptif. Il n’inclut aucune information sur les conditions dans lesquelles la nouvelle orthographe *devra* être appliquée, ni dans le cadre de la presse, ni dans celui de l’éducation. Au mieux, il constitue une série de recommandations, en particulier dans le cas de la création de néologismes.
- Comme il y est dit, «ces rectifications sont modérées dans leur te-

neur et dans leur étendue». André Goosse soutient qu'elle touchent entre 800 et 1200 mots, soit un mot sur 90 d'un grand dictionnaire, et que seuls 6 mots sont modifiés parmi les 1000 mots les plus fréquents du français (1991, 24-25). Dans un livre, un mot sur deux pages serait touché (1991, 2).

- Enfin, et justement parce que ces propositions sont très réduites et qu'elles ont fait l'objet d'un compromis, elles sont approuvées par l'Académie, qui promet alors de les intégrer dans la prochaine édition de son dictionnaire.

Ces rectifications sont bien accueillies en Belgique, en Suisse et au Canada francophone. En Belgique, «les grands mensuels ont adopté les aménagements sans que les lecteurs protestent», et «l'Éducation Nationale a soutenu la réforme» (*Le Français dans le Monde* N° 313, 64). En ce qui concerne le Canada, «dès 1991, l'Office québécois de la langue française s'est déclaré, de façon générale, favorable à l'application des rectifications de l'orthographe». Cependant, en constatant «les réticences, voire l'opposition, qu'elles soulevaient dans divers milieux en France et ailleurs, il n'a pas voulu faire cavalier seul dans ce domaine» (site de l'*Office Québécois de la langue française*). En effet, en France, dès la publication du *Rapport*, «les réactions sont violentes. Plusieurs personnalités de droite et de gauche lancent une véritable croisade contre les propositions du groupe d'experts et de l'Académie, les jugeant le plus souvent, et au mieux, absurdes» (Site de la *DLF*, Direction de la Langue Française de Suisse).

Rappelons brièvement les arguments de l'époque contre les rectifications et les contre-arguments qu'André Goosse et d'autres membres du Conseil ont tenté d'y opposer :

- **l'ortographe allait devenir fonétique.** *Contre-argument* : les rectifications portaient sur un petit nombre de mots et ne touchaient en rien à mille autres problèmes orthographiques (consonnes doubles, consonnes finales, formes conjuguées...) autrement plus problématiques en ce qui concerne la divergence son-graphie. « Orthographe » allait continuer à s'écrire *orthographe*, et «phonétique», *phonétique*.
- **On allait être obligé d'employer la nouvelle orthographe.** *Contre-*

argument : c'était faux, y compris en ce qui concerne la presse et l'éducation nationale (contrairement, notons-le, à la réforme allemande en préparation au même moment, et qui elle, sera imposée à partir de 2007). Le principe était celui de «l'acceptation de variantes».

- **Le principe des variantes allait mener à la gabegie, à l'anarchie.**

On ne saurait plus comment écrire, on allait donc se mettre à tout écrire n'importe comment. Un commentateur annonce «la dissolution d'une culture toute entière» (Goosse 1991, 5). *Contre-arguments* : D'une part, la langue comporte déjà un nombre incroyable de variantes. Catach relève des *milliers* de cas d'orthographe d'usage présentant des variantes (Catach 1978, 49-50). D'autre part, il paraît peu vraisemblable que notre bonne ou mauvaise orthographe soit déterminée par les doublets que nous sommes amenés à rencontrer depuis l'enfance, que nous les remarquons (par ex. *cuillère* et *cuiller*) ou non (par ex. *couvre-pied* et *couvre-pieds*).

- **Il allait falloir réimprimer toute la littérature française.** De plus, il nous serait impossible de lire les classiques dans leur forme originelle (Philippe Sollers, *Le Monde* du 10 janvier 1991). *Contre-argument* : C'est déjà le cas ! Sans parler de *Tristan et Iseult* ou de la *Chanson de Roland*, Molière n'est pas imprimé aujourd'hui dans son orthographe originelle, cf. ces deux vers du *Misanthrope* (qui s'écrivait d'ailleurs sans *h*) : «Et quoy qu'auec ardeur je ueuille uous haïr / Trouuay-je un Cœur en moy tout prest à m'obeïr ?» (cité par Goosse, 1991, 10 ; j'ai rétabli la non-distinction des *u* et des *v*). Même des œuvres très récentes sont réimprimées avec des modifications (*très-dévoué, poëme, grand'rue, entr'ouvrir...*), modifications dans l'ensemble plus nombreuses que celles que les rectifications de 1990 auraient exigé.

Quoi qu'il en soit, à la suite des campagnes de presse dont elle est la cible, et, semble-t-il, de la réaction du public, la "réforme" est rapidement abandonnée en France.

L'Académie revient en partie sur sa position, en déclarant que «les recommandations ont été acceptées par l'Académie française, qui a néan-

moins voulu qu'elles soient soumises à l'épreuve du temps. Elle maintiendra donc les graphies qui figurent dans son dictionnaire jusqu'au moment où elle aura constaté que ces recommandations sont passées dans l'usage» (site de l'Académie française). On verra qu'elle intégrera environ 30% des variantes au nouveau dictionnaire, regroupant les autres en début de volume sous forme de recommandations.

Bernard Cerquiglini, membre du Conseil supérieur, s'étonne à l'époque du «purisme populaire» qui fait que même les gens «que la réforme devait aider» sont révoltés par l'idée qu'on touche à la langue. Quinze ans plus tard, le même présente sur TV5 une émission intitulée *Merci professeur*. Il y reprend en quelque sorte le rôle d'un Max Favaelli, pour expliquer les subtilités d'un participe passé qu'il avait pourtant bien eu l'intention d'élaguer, les concordances de temps les plus invraisemblables, et les règles orthographiques les plus absconses. Après avoir présenté les explications normatives les plus absurdes, dont en tant que linguiste il ne peut pas ne pas savoir qu'elle n'ont aucun sens, il conclut : «Il ne faut pas avoir peur d'être puriste, si c'est à bon escient». C'est un revirement, pour le moins.

Il est intéressant de noter aussi que le journal *Le Monde*, qui avait participé aux expertises en la présence de son chef-correcteur, n'adopte aujourd'hui aucune des recommandations. Des formes que j'ai testées dans cette étude, pratiquement toutes sont présentes dans la version électronique du *Monde* sous leur graphie «ancienne», même quand la graphie «nouvelle» est largement utilisée par les usagers de la langue. Sans même parler des propositions les plus radicales comme la soudure des mots composés ou l'effacement de l'accent circonflexe, *Le Monde* écrit par exemple *événement* dans 933 cas sur 962, alors que les usagers l'écrivent *èvenement* dans 52% des cas. De même, *Le Monde* donne *complètera* dans 53 cas sur 57, alors que *complètera* est préféré sur l'internet français dans 49% des cas. Encore plus surprenant, il écrit *matches* dans 124 cas sur 140, alors que 94% des *matchs* trouvés sur l'internet français n'ont pas de *e*. Comprenons bien que le point ici n'est pas de critiquer les graphies choisies par *Le Monde*, mais de constater que sa participation à l'aventure de 1990 n'a pas influencé son orthographe dans le sens du modernisme. Il est même possible que ce soit le contraire qui se soit produit. On verra que ceci pose

un problème pour d'éventuelles réformes futures (la question du «cercle vicieux» soulevée par le *Français dans le Monde*).

II. La «nouvelle orthographe» en 2006

Les rectifications dans les dictionnaires

Qu'en est-il aujourd'hui en février 2006, presque exactement 15 ans après la publication du rapport ? Selon le *Français dans le Monde* (interrogeant Cerquiglino, Goosse et d'autres), outre l'intégration des variantes nouvelles à la 9^{ème} édition (en cours) du dictionnaire de l'Académie, «les dictionnaires les plus courants suivent avec plus ou moins d'enthousiasme». «Le *Petit Robert* [suit] à 60%, ce qui représente 5% des mots touchés par la réforme [*sic*. On ne voit pas très bien comment se fait ce calcul], le *Petit Larousse* et le *Dictionnaire Hachette* à 40%». On lit ailleurs le chiffre de 85% des formes pour l'ensemble des dictionnaires (site Francité). On trouvera plus de détails sur les sites DLF (*Délégation à la Langue Française* de Suisse) ou APARO (*Association pour l'Application des Recommandations Orthographiques*).

Cependant, ces chiffres ne nous disent pas grand-chose dans la mesure où on ne sait pas exactement à quoi il font référence (nombre d'entrées, de mots, de signes ou autres ; cf. ci-dessus en ce qui concerne le Robert). De plus, ils ne sont pas forcément fiables puisqu'on lit régulièrement, par exemple, que l'Académie a adopté les formes nouvelles mais qu'on apprend ailleurs qu'elle «a entériné de façon définitive 30% des modifications proposées... les 70% restant étant regroupés en début de volume et présentés sous forme de recommandations». Enfin, signalons que les chiffres sont en évolution, dans les deux sens : la réforme fait son chemin dans certains ouvrages (le *Dictionnaire Hachette* mentionne toutes les variantes proposées dans son édition 2002) ; mais elle peut régresser dans d'autres (dans le Larousse 2001, *événement* a repris la place d'*évènement* en entrée) (site DLF).

Les rectifications chez les usagers de la langue

Dans cette étude, on s'intéressera moins au progrès des rectifications auprès des institutions, des journaux, des enseignants et des professionnels de la langue qu'auprès des usagers ordinaires. Pour cela, j'ai simplement

décidé de comparer sur internet certaines formes «anciennes» et certaines formes «rectifiées». Précisons qu'il s'agit de l'internet *français* et non de l'internet *francophone*, puisqu'il s'agit ici de s'intéresser à la réception des rectifications de 1990 dans la France de 2006.

Dans cette première évaluation, je n'ai pas cherché à établir un bilan pour les quelques 1000 à 2000 mots touchés par le rapport de 1990, mais simplement à en traiter un très petit échantillon. Pour cette raison (et aussi en raison du caractère versatile du matériel, comme on va le voir ci-dessous) ce travail ne peut être considéré que comme une première ébauche. Ses conclusions ne peuvent être que provisoires, destinées à jeter les bases d'un éventuel projet de recherche.

Je me suis intéressé aux quelques 25 cas de figure que les rectifications touchent, et pour chacun de ces cas, j'ai traité deux paires de formes : une paire présentant un changement "radical", proposé dans le Rapport mais jamais proposé dans aucun dictionnaire jusqu'en 1990 (type *mille-pattes/millepatte*), et une forme dont on pourrait dire qu'elle est "en usage", proposée dans le *Rapport* mais aussi, déjà, dans certains dictionnaires (type *mille-feuilles/millefeuille*). Pour ce faire, je me suis basé sur l'index établi par André Goosse (Goosse 1991, 93-107), dans lequel il marque d'une astérisque ces formes déjà "en usage". Il s'agissait de vérifier, comme on peut s'y attendre, que ces formes "en usage" sont plus largement utilisées que les formes "radicales", mais aussi, en séparant les deux, de se faire une image plus précise du degré d'utilisation des formes nouvelles "radicales" (telles que *millepatte*).

L'intérêt d'une approche à travers internet est qu'elle permet de quantifier, mot par mot et cas par cas, la manière dont les rectifications sont entrées dans l'usage (ainsi, on le verra, que certaines variantes alternatives auxquels les rectificateurs n'avaient pas songé). On peut ainsi infirmer ou confirmer, dans la mesure de l'échantillon, les informations fournies par les différents organismes qui s'intéressent au problème. Ainsi, en 1990, l'ANROE (*Association pour l'information et la recherche sur les orthographes et les systèmes d'écriture, citée par le Français dans le Monde*), déclare que «le mouvement autorisant les soudures a été bien suivi». Or, pour les quelques mots que j'observe sur l'internet français, ce n'est pas du tout le cas (*croquemonsieur*, avec soudure, ne représente que 2% des occurrences

de toutes les formes de *croque(-)monsieur*. De même, *millepatte* 2%, *train-train* 3%, *bluejean* 3%, *weekend* 4%, *exvoto* 3%, *un apriori* 1,4%, *portecle* 0,2%, *brisetout* 0%). Les seuls mots qui présentent un nombre conséquent d'occurrences sont ceux qui *préexistaient* au Rapport, et dont le Rapport, en quelque sorte, n'a fait que *présenter à la connaissance de tous* l'usage nouveau (*croquemort* 18%, *millefeuille* 28%, *faitout* 58%). Il est également intéressant de constater la répartition de ces mots en deux groupes : un groupe autour de 2% et un autre à 20, 30, 50% et plus. Or, on verra que 2% est proche du chiffre de variation orthographique de *n'importe quel mot*, c'est à dire de la zone des variantes d'inattention, des erreurs, des formes volontairement rebelles, etc. Autrement dit, il n'est même pas sûr que ces 2% présentent réellement des variantes stabilisées. On comprend qu'il en va tout à fait autrement pour des formes présentes dans 30% à 60% des occurrences d'un mot. On a bien là affaire à des variantes largement représentées et non à des erreurs ponctuelles. En ce qui concerne un mot comme *croquemort* (18%) on peut aussi penser qu'il s'agit d'une variante stable chez certains usagers.

L'ANROE déclare également que «la régularisation des pluriels» a été bien suivie. Ici, on peut constater qu'en effet, la plupart des formes présentent sur internet une variante régulière (avec cependant des écarts qu'il faudrait expliquer dans une étude plus détaillée) : la forme *un essuie-main / des essuie-mains* représente 39% des occurrences de ce mot (alors que la forme «ancienne» est invariable : *un essuie-mains / des essuie-mains*). On trouve aussi *un casse-cou / des casse-cous* (17% des occurrences de ce mot, alors que là aussi, la forme ancienne est invariable), *un lave-vaisselle / des lave-vaisselles* (40%), *un brise-glace / des brise-glaces* (45%), *un après-midi / des après-midis* (5%), *un après-ski / des après-skis* (35%). Cette courte liste semble donc montrer que le pluriel régulier des noms composés est une chose acquise pour de nombreux usagers de la langue. Pourtant, elle est insuffisante pour s'assurer qu'il s'agit bien d'une «régularisation». En effet, il faut constater qu'*après-midi* résiste. On peut se demander si pour l'usager, il fait bien partie de la même catégorie de mots que, par exemple, *après-ski*.

Ensuite, on relève aussi sur internet quantité de pluriels *irréguliers* qui ne sont ni ceux des dictionnaires, ni ceux du Rapport. Par exemple, si

pour le pluriel de *brise-glace*, on analyse l'ensemble des formes, on trouve : 373 *brise-glaces* (34%) et 461 *brise-glace* (43%), mais aussi 240 *brises-glace(s)* avec un *s* à *brises* (soit 22% des formes) — ce qui semble montrer que certains *brise-glaces* ne touchent que le sommet de l'iceberg. Les formes qui émergent n'ont sans aucun doute rien à voir avec le Rapport de 1990 et les recommandations de la Délégation à la Langue Française. On peut penser qu'elle relèvent plutôt de phénomènes concernant directement l'utilisateur et son rapport personnel à la langue. On ne peut donc pas vraiment dire avec l'ANROE que telles ou telles suggestions sont plus ou moins bien suivies.

Enfin, L'ANROE fait un « constat d'échec » sur le déplacement du tréma et la suppression des accents circonflexes sur *i* et *u*. En effet, dans les quelques formes que j'ai explorées, je constate un maintien très net des formes « anciennes » : (*abîme* 1%, *brûler* 9%, *gagéüre* 0.01%, *aigüe* 10%) mais avec des écarts qu'il conviendrait de vérifier et, le cas échéant, d'expliquer. On verra ci-dessous que l'outil est versatile et que les écarts peuvent varier. Cependant, en admettant que les écarts entre *abîme* et *brûler* d'une part, entre *gagéüre* et *aigüe* d'autre part soient significatifs, je peux émettre l'hypothèse selon laquelle les mots *abîme* et *gagéüre* seraient emblématique des particularités orthographiques du français (« il faut savoir écrire *abîme* et prononcer *gagéüre* »). Par leur rareté, leur difficulté, et ce caractère emblématique, celles et ceux qui les utilisent tendraient à s'attacher à leur orthographe particulière. En revanche, *brûler*, *aiguë*, plus fréquents et moins directement désignés comme complexes, seraient plus sensibles aux variantes. Et là encore, il est intéressant de constater que les usagers proposent une solution originale qui n'est pas issue du Rapport : *aigue* (sans aucun tréma) est encore plus présente sur l'internet français que la forme rectifiée (30800 contre 21000 dans mon échantillon). De même, *ambigue* sans tréma est presque aussi représentée que la forme rectifiée *ambigüe* (607 contre 888). Voilà une solution orthographique que les rectificateurs n'avaient pas prévue, et qui présente de multiples avantages (simplicité, régularité de la série *aigu/aigue*).

On voit donc que même l'analyse d'un nombre très limité de formes permet des constats intéressants et dissipe des généralités.

III. Rechercher des formes orthographiques sur internet

Phénomène stupéfiant par sa soudaineté à l'échelle historique, internet, qui existait à peine il y a quinze ans, présente aujourd'hui quelque neuf milliards de pages accessibles. Un moteur de recherche tel que Google peut en faire une analyse virtuelle en quelques centièmes de secondes. L'internet français (les pages "France") présente de l'ordre de deux-cent millions de pages (ce qui paraît peu en comparaison du total). En réduisant la recherche aux pages des domaines *.fr* (ce que j'ai fait), on peut accéder à cent millions de pages. A titre d'exemple, une recherche sur le mot «avec» donne 35 700 000 résultats en 0.28 secondes.

Néanmoins, un certain nombre d'objections peuvent être adressées à une recherche linguistique sur internet, tant sur le contenu que sur la forme :

- **On trouve n'importe quoi sur internet.** On ne peut pas distinguer entre les types de sites, les types de textes. On y trouve aussi bien de la littérature que des pages perso de lycéens, les grands journaux que des sites scientifiques confidentiels ou des petites annonces. On y rencontre toutes les manières d'écrire, les orthographes les plus fantaisistes et les grammaires les plus fautives.

Réponse à cette objection : C'est exact, et c'est aussi ce qui en fait l'intérêt. Google brasse le corpus le plus complet qu'on puisse trouver d'une langue sous sa forme écrite. Sur des millions de pages, il offre précisément un panorama de tous les usages possibles de la langue que l'utilisateur (locuteur, lecteur ou scripteur) va rencontrer dans sa vie quotidienne : sur internet, précisément, mais aussi à travers les autres productions langagières, médiatiques ou autres. En gros, internet nous fournit ce qu'aucun autre corpus ne pouvait nous proposer jusqu'à présent : l'ensemble des productions linguistiques auxquelles les usagers sont confrontés, et à partir desquels ils forment aussi leurs manières d'écrire.

On peut également objecter que l'échantillon est biaisé par la surreprésentation sur internet de certaines formes de la langue: littérature du domaine public en ligne, blogs, sites commerciaux ou autre. Cependant, dans l'état actuel des choses, et considérant la quantité phénoménale de pages auxquelles on peut accéder, je ne pense pas tel

ou tel type de pages puisse être considéré comme dominant. On trouve effectivement de tout, et ce tout représente, tautologiquement, ce à quoi les usagers de la langue ont effectivement accès.

Par ailleurs, il n'est pas tout à fait vrai qu'on ne peut pas distinguer entre les types de site. On peut utiliser les fonctionnalités de la recherche avancée pour sélectionner un échantillon plus étroit. Avec quelques acrobaties, on peut arriver à ne prendre en compte que des sites personnels ou des sites commerciaux. Plus simplement, on peut choisir de ne travailler que dans un site donné. Par exemple, pour chercher les occurrences de ma liste dans le site en ligne du *Monde*, plutôt que d'utiliser le moteur de recherche interne au site, j'ai utilisé la recherche avancée de Google de la façon suivante : « *terme* site:lemonde.fr -blog -forum -forums -chat » (Les quatre derniers termes étant destinés à ne laisser que les articles, les dépêches et les pages de présentation. Les raisons pour lesquelles je n'ai pas utilisé la recherche interne au site sont les suivantes : elle est plus lente ; elle ne sélectionne que les articles ; et surtout, elle ne présente pas les phrases où apparaissent les occurrences mais seulement les premières lignes de l'article).

Un autre des intérêts de ces moteurs de recherche est qu'ils présentent leurs résultats de telle manière qu'on peut, en quelques minutes, balayer du regard plusieurs centaines d'occurrences dans leur contexte. On peut en effet sélectionner une option qui permet de voir 100 résultats par écran. Apparaissent alors le titre de la page et quelques morceaux de phrases comprenant l'occurrence désirée, l'ensemble fournissant une indication presque toujours claire du type d'écrit.

- **Google ou les autres moteurs de recherche ne traitent pas les accents.** Comment alors rechercher des termes dont les différences portent le plus souvent sur des accents ?

Réponse à cette objection : En fait, Google traite les accents, mais la manière dont il le fait n'est pas claire. Il me semble même qu'elle ait changé au cours des années. Bizarrement, je ne suis pas parvenu à en trouver sur internet de description fiable. On lit communément l'assertion selon laquelle « Google ne prend pas en compte les accents » et qu'une entrée comme *cede* renverra aussi bien à *cede*, *cédé*, *cède* ou

autres *cèdè*, *cedé*, *cède*, etc. C'est en effet ce qui semble se passer quand on fait une recherche simple. Dans le cas de la recherche avancée, les choses sont moins évidentes: il semble que dans certains cas (quand on sélectionne un site unique par exemple), Google prenne en compte les accents.

Le "truc" consiste à passer par l'internet japonais, qui traite les lettres accentuées différemment. Cependant, il semble là encore que cette prise en compte soit pondérée par un algorithme de priorité entre lettres accentuées et lettres non accentuées. Autant que je puisse le comprendre, une entrée telle que *cédé* renverra en priorité aux pages qui contiennent *cedé*, mais aussi, dans certains cas, à celles qui contiennent *cede* ou autre. Rien ne garantit alors que le nombre de pages affichées concerne uniquement l'entrée désirée. La solution consiste à interdire systématiquement les formes indésirables. Par exemple, pour connaître le nombre d'occurrences de *crèneler*, on entrera : «crèneler -crèneler -creneler» et on vérifiera, dans les premières et les dernières pages de résultats (c'est à dire sur quelques centaines de sites), qu'il n'y a pas de mélanges. Les choses se compliquent encore lorsqu'on veut connaître le nombre d'occurrences d'un mot *non accentué* par comparaison avec sa version accentuée. Non seulement on veut sélectionner le mot non accentué et seulement celui-là, mais aussi on veut éviter les sites qui, pour des raisons techniques, ne mettent aucun accent. En effet, dans ce dernier cas, l'absence d'accent ne saurait être considérée comme un choix typographique relatif à ce mot précis. Pour ce faire, j'ai choisi de ne m'occuper que des sites qui contiennent la séquence *à la* avec un accent, indicateur certain du fait que le site en question utilise ordinairement les accents. Prenons le cas du mot *abîme*. Pour comparer la fréquence de *abîme* et *abime*, on cherchera d'abord «"un abîme" "à la" -a-la -abîme» puis «"un abime" "à la" -a-la -abime». On peut objecter que, par exemple, *un abîme* ne donnera pas les mêmes fréquences que *l'abîme* ou une autre forme, mais comme on va le voir plus bas, l'important est de s'assurer de ce que les résultats ne contiennent pas de formes parasites, et que les formes qu'on cherche à comparer soient traitées, comme ici, de façon parallèle.

Notons aussi qu'outre celui des accents, Google et les autres moteurs de recherche posent le problème des traits d'unions. Dans ces moteurs, le trait d'union ne sert que de ligateur de séquence, au même titre que les guillemets. Ainsi "*vingt et un*", "*vingt-et-un*" ou *vingt-et-un* correspondent au même objet. Pour cette raison il est impossible de distinguer directement les formes avec trait d'union des formes sans, et cette fois il n'y a pas de truc pour en sortir. La solution consiste à chercher d'abord ensemble les deux formes à comparer (par exemple *vingt et un* et *vingt-et-un*), puisque le moteur de recherche ne peut pas les distinguer. On obtient alors une série d'écrans de 100 résultats chacun, contenant les deux formes. On utilise ensuite la fonction *Recherche dans la page* (Control-F) qui, elle, distingue les formes avec ou sans trait d'union. Il suffit alors de compter manuellement, sur quelques centaines de résultats (c'est à dire sur un échantillon), combien de fois apparaissent chacune des deux formes. C'est l'affaire de quelques minutes.

- **La sélection de la langue n'est pas fiable.** Même en choisissant uniquement les sites en français dans la recherche avancée, on obtient des sites en anglais ou dans d'autres langues.

Réponse à cette objection : C'est exact, et c'est inévitable : il arrive que des pages contiennent des textes en plusieurs langues ; les tags qui fournissent des indications aux moteurs de recherches sont parfois erronés, etc. Ce qui est important est de s'assurer que les résultats que l'on recueille sont bien en français. Une solution consiste à proposer une courte séquence (par exemple : *un blue jean*) plutôt qu'un mot seul (ex. : *blue jean*). On recueille évidemment moins de pages, mais encore une fois, ce n'est pas le nombre absolu de pages qui est important mais la *comparaison* entre le nombre d'occurrences de *deux formes particulières* (par exemple *un blue jean* et *un bluejean*) dans des contextes identiques.

Dans cette étude, comme il s'agissait d'explorer l'internet français et non l'internet francophone, j'ai non seulement limité la recherche à *français* mais aussi au seul domaine *.fr*. Comme dit précédemment, cette limitation divise le nombre de pages explorées par deux environ (il reste néanmoins une centaine de millions de pages).

- **On peut faire des erreurs énormes.** Comme on ne contrôle pas les pages qu'on traite, on peut obtenir des résultats sans rapport avec la forme qu'on cherchait.

Réponse à cette objection : C'est exact, et c'est pourquoi il faut contrôler les occurrences obtenues en jetant un coup d'œil sur au moins quelques centaines de résultats. En effet, il peut se passer plusieurs choses :

- (1) Il se peut que la forme recherchée apparaisse dans le titre d'un roman célèbre, d'un film, dans le nom d'un groupe de rock à la mode, etc., et cette forme est alors citée des milliers ou des dizaines de milliers de fois. Parmi les formes que j'ai explorés, ça a été le cas par exemple pour *Placebo* (groupe de rock, éclipsant la forme *placébo*), pour *île* (à cause du roman de Houellebecq) ou encore pour *dentellière* (à cause du roman et du film *La Dentellière*). Il est alors intéressant d'éliminer ces cas massifs pour observer les formes dans des contextes plus ordinaires. Ainsi, pour éliminer *Placebo* (le groupe, qui représentait au moins 82 000 occurrences), j'ai limité ma recherche à la comparaison entre *un placébo* et *un placebo*. Pour *dentellière / dentellière*, j'ai comparé *une dentellière* et *une dentellière*, ce qui éliminait les références au film (qui apparaissent toujours avec l'article défini).
- (2) Il se peut que la forme recherchée apparaisse dans des noms de chaînes de magasins, de sites ou de revues, se trouvant donc répétée dans de nombreux sites sous une forme particulière. Cela ne signifie pas pour autant nécessairement que cette forme est la plus fréquente dans l'usage. Là encore, on traitera une petite séquence, par exemple en ajoutant l'article indéfini (ex. : *un croquemonsieur*).
- (3) Enfin, il se peut que les résultats renvoyés par le moteur de recherche concernent non seulement la forme désirée mais plusieurs formes homonymes. Par exemple, *dissout* peut renvoyer non seulement à la forme «nouvelle» *dissout* du participe passé de dissoudre (forme «ancienne» : *dissous*), mais aussi à d'autres formes du verbe : *il dissout, il se dissout...* Pour éviter ce problème, il faut traiter une courte séance non ambiguë : comparer *a dissout*

et *a dissous* par exemple.

- **Les résultats sont versatiles.** Une même recherche répétée plusieurs fois peut renvoyer des résultats divergents.

Réponse à cette objection: Il est exact que les résultats changent beaucoup dès qu'on modifie un détail dans le contenu de la recherche. Pour exemple, si on compare *une dentelière* et *une dentellière*, la première représente 7.38% du total des deux formes ; mais si on compare *la dentelière* et *la dentellière*, la première ne représente plus que 1.4% du total. De même, selon qu'on sélectionne ou pas le domaine (*.fr*, *.com*, etc.), qu'on traite les pages françaises ou les pages francophones, qu'on élimine ou pas une séquence parasite (ex. : *-brises-glace*), on obtient des résultats très différents. C'est pourquoi :

- (1) il est important de contrôler les résultats en regardant quelques centaines de résultats (en particulier les premiers écrans renvoyés par Google, où apparaissent les mots à la mode et les formes les plus représentées sur internet).
- (2) Il est imparatif d'utiliser exactement la même formule pour les expressions qu'on cherche à comparer. On ne peut les comparer que deux à deux, dans le même contexte. En particulier, les chiffres absolus de fréquences n'ont aucune valeur en tant que tels : tel mot a été traité avec l'article défini, tel autre avec une préposition, etc. Par exemple, on ne peut pas directement conclure que *croque-monsieur* (817) est moins fréquent que *gageure* (49400). Les contextes dans lesquels ces mots apparaissent sont différents, et de plus, j'ai extrait ces mots de façon différente. En revanche, on peut conclure que *gageure* (49400) est plus fréquent que *gageüre* (3) car ces deux formes ont été traitées de façon parallèle.
- (3) Enfin, l'exemple de *dentelière* et d'autres montrent qu'il ne faut pas prendre les pourcentages obtenus ici avec trop de précision. Si on peut considérer une différence de 20 à 50% comme significative, une différence de 10% ou moins est à prendre avec précaution. Comme on le verra, une différence de 2 à 3% se rapproche, de toute façon, de la marge d'erreur orthographique ordinaire.

IV. Quelques formes et commentaires

Dans ce qui suit, je reprends la présentation d'André Goosse (Goosse 91, 37-83), qui elle-même reprend les points du *Rapport*, mais dans un ordre différent. Comme dit plus haut, pour chaque point qu'abordent les rectifications, je traite deux paires de formes (à quelques exceptions près) : une paire présentant un changement "radical", proposé dans le Rapport mais jamais proposé dans aucun dictionnaire, et une forme dont on pourrait dire qu'elle est déjà "en usage" en 1990, proposée dans le Rapport mais aussi, déjà, dans certains dictionnaires. Pour ce faire, je me base sur l'index dans lequel André Goosse marque d'un astérisque ces formes "en usage" (93-107). La présentation des nouvelles règles suit également les formulations de Goosse. Je ne reproduis cependant pas les listes de mots, ni les exceptions, qu'on trouvera sur internet (par exemple sur www.orthographe-recommandee.info, ou ailleurs).

1.1. *E est remplacé par É : mots dont la prononciation a changé*

Goosse cite le Rapport comme suit : «La lettre *e* est remplacée par *é* dans les mots dont la prononciation a changé ou dont l'accent avait été omis par le dictionnaire de l'Académie : *asséner*, *bélitre*, *bésicles*, *démiurge*... »

"En usage": asséner 25400 ← assener 753 97.1%

"En usage" signifie ici que la forme est déjà présente dans «certains dictionnaires», selon Goosse. Ici, exceptionnellement, je ne peux comparer cette paire avec une paire "radicale" car tous les mots de cette catégorie étaient déjà tous "en usage".

Parmi toutes celles que j'ai testées, cette forme est celle qui semble la plus universellement adoptée. On peut considérer que la forme *assener* n'existe plus (même *Le Monde* utilise la forme moderne).

1.2. *E est remplacé par É : mots latins ou étrangers*

«La lettre *e* est remplacée par *é* dans les mots latins ou étranger : *critérium*, *déléatur*, *délirium trémens*... »

"Radical" : placebo 582 ← placebo 130000 0.45%

“En usage” : facsimilé 73 ← **facsimile 114** 39%

Les accents sur les mots latins et étrangers ne semblent pas très répandus, mais seule une étude de détail permettrait de le prouver. De plus, on voit qu’au moins certains mots, comme *facsimilé*, présentent une variante accentuée courante. Notons qu’elle précède les recommandations du Rapport, donc qu’on ne peut pas dire, là encore, que celles-ci auraient été *suivies*.

2.1. É devient È devant E muet : inversion du sujet (dussè-je, etc.)

«La lettre *é* est remplacée par *è* devant une syllabe contenant un *e* muet à la fin de formes verbales quand il y a inversion du pronom *je*, dans l’interrogation ou dans d’autres cas : *Aimè-je, puissè-je, dussè-je, etc.*»

“Radical” : dussè-je 440 ← **dussé-je 579** 43.2%

Ici il n’y a pas de forme “en usage” car cette forme orthographique n’était pas représentée dans les dictionnaires au moment du Rapport. Elle fait donc partie de la série “radicale”, et sa bonne représentation est surprenante. En fait, on va voir (à quelques exceptions près qu’il faudra expliquer) que les formes qui inversent simplement la direction de l’accent sont parmi les mieux acceptées (*dussè-je* 43%, *abrègement* 80%, *cèdera* 63%, *complètera* 49%, *évènement* 52%...). Si on émet l’hypothèse selon laquelle ces changements ne modifient pas beaucoup la physionomie du mot, contrairement à ceux qui ajoutent, retirent ou déplacent un accent (dont on va voir qu’elles sont nettement moins bien représentées), on peut penser que la physionomie du mot a plus d’importance que ne l’ont préjugé les auteurs du Rapport.

2.2. É devient È devant E muet dans un certain nombre de mots

«La lettre *é* est remplacée par *è* dans les mots suivants : *abrègement, affèteerie, allègement...*»

“Radical” : crèneler 6 ← **créneler 202** 2.9%
“En usage” : abrègement 310 ← **abrègement 78** 79.9%

La forme *crêneler* contredit ce qui a été dit en 2.1. ; néanmoins il s'agit peut-être d'un effet de spécialité. Mot technique, il serait utilisé par une catégorie particulière de la population qui s'attache à sa graphie classique.

2.3. *É devient È devant E muet dans les futurs et conditionnels*

«La lettre *é* est remplacée par *è* dans les futurs et les conditionnels des verbes qui ont un *é* fermé à l'avant-dernière syllabe de l'infinitif : *je cèderai, je délèguerai...*»

“Radical” :	il cèdera 430	← il cédera 253	63%
“En usage” :	il complètera 405	← il complètera 426	48.7%

Ces deux formes ont déjà été commentées en 2.1. On voit que tant la forme “en usage” que la forme “radicale” présentent autant de graphies «nouvelles» que de graphies «anciennes». C'est sans doute un indice fort du fait qu'il y a peut-être là un changement dans les habitudes (*Le Monde*, cependant, reste résolument conservateur).

3.0. *Les verbes en ELER et ETER font ÈLE, ÈTE au présent*

«Dans les verbes terminés à l'infinitif par *-eler* et *-eter*, le *e* du radical se change en *è* quand la syllabe suivante contient un *e* muet : *il amoncèle, il étiquète...*»

“Radical” :	il étiquète 5	← il étiquette 78	6.02%
“En usage” :	il amoncèle 0	← il amoncelle 76	0%

Le rejet très net des formes «nouvelles» pour ces deux types de formes pourrait confirmer l'impression dégagée plus haut. (1) Ce n'est pas le fait qu'une forme ait été proposée par le Rapport (ni même, dans ce cas, par les dictionnaires) qui fait qu'elle est acceptée ou non. Les changements semblent se dérouler sur d'autres critères, qui restent à déterminer. (2) Ce n'est pas le fait de s'accorder avec la logique ou avec la prononciation qui rend une forme plus acceptable pour le public. Par exemple, s'il est facile de passer de *cédera* à *cèdera*, il l'est moins de passer d'*amoncelle* à *amoncèle*. Ainsi, un des critères serait peut-être l'étendue du changement apporté à

la physionomie du mot.

4.0. *Suppression de l'accent circonflexe sur Î et Û.*

«L'accent circonflexe ne se met plus sur les lettres *i* et *u* : *voute*, *cout*, *abime...*»

“Radical” :	abime 191	←	abîme 17500	1.08%
“Radical” :	bruler 20400	←	brûler 208000	8.93%

Ici aussi, pas de forme “en usage” car aucun dictionnaire ne proposait de supprimer ces accents au moment du Rapport. Ce point a été l'un des plus critiqués par les journaux et par les auteurs, et il a entraîné Bernard Pivot à se désolidariser de la Commission.

Ces formes confirment ce qui est dit en 3.0. On voit que les formes «nouvelles» ne sont pas, ou peu, présentes, et ce pourrait être dû au fait que la perte de l'accent circonflexe change trop la physionomie du mot.

J'ai proposé aussi plus haut une hypothèse pour expliquer l'écart éventuel entre les deux paires : les mots les plus courants perdraient un petit peu plus facilement leur accent circonflexe ; les mots savants, techniques ou emblématiques le conserveraient à tout prix. Si on observe deux autres paires, on en a une première confirmation :

“Radical” :	voute 288	←	voûte 22100	1.28%
“Radical” :	il connait 18000	←	il connaît 119000	13.14%

On peut supposer que l'accent circonflexe, comme on l'a vu à travers les réactions contre son effacement, est porteur d'une valeur iconographique forte, d'autant plus marquée dans les mots rares.

5.0. *Déplacement ou ajout du tréma sur Ü*

«Le tréma est placé sur la lettre *u* dans les adjectifs féminins *aigüe*, *contigüe*, *ambigüe...* et dans les noms *cigüe*, *ambigüité...*»

«Le tréma est ajouté sur le *u* dans les diverses formes du verbe *argüer* et dans les noms *gageüre*, *mangeüre...*»

“Radical” :	gageüre 3	←	gageure 49400	0.01%
“En usage” :	aigüe 21000	←	aigüè 180000	10.4%

On trouve également ici la dichotomie *mot rare / mot courant* avec un écart similaire entre les résultats. On constate également, comme il a été dit plus haut, que les formes qui changent la physionomie du mot, comme *gageüre*, sont massivement rejetées. Cependant, le cas de *aigüe* est intéressant car les usagers de la langue proposent également une forme, *aigue* sans tréma, qui n’a pas été prise en considération par les rectificateurs. Or cette forme est encore plus représentée que la forme «nouvelle» *aigüe*. Il en va de même pour *ambigue*, pour laquelle la forme sans accent est presque autant représentée que la forme «nouvelle» *ambigüe*. Ceci montre à nouveau qu’on ne peut affirmer que les formes alternatives sont *inspirées* par le Rapport, que le rapport est bien ou mal *suivi*, etc. : les usagers de la langue se débattent avec l’orthographe et ses variantes (parmi lesquelles celles issues du Rapport, s’il en ont eu connaissance, constituent une source parmi d’autres).

6.1. Mots composés : soudure de mots dont le premier élément est un verbe

«Sont soudés les mots dont le premier élément est un verbe à l’origine : *d’arrachepied, boutentrain, chaussetrappe...*»

“Radical” :	croquemonsieur 17	←	croque(-)monsieur 817	2.04%
“En usage” :	croquemort 179	←	croque(-)mort 829	17.8%

Les données concernant la soudure des mots composés ont été commentées plus haut : les formes nouvelles “radicales” avec soudure se situent entre 0 et 4% de leurs occurrences respectives, alors que celles qui sont déjà “en usage” au moment du Rapport vont de 18% pour *croquemort* à 58% pour *faitout* (cependant, la soudure ne se fait pas du tout dans le cas des mots latins et étrangers, cf. 6.4.). Par ailleurs, je n’ai pas traité pour ces mots la différence entre les mots non soudés avec trait d’union (*croque-monsieur*) et les mots non soudés sans trait d’union (*croque monsieur*) mais les seconds sont nombreux. Dans le cas de *porte(-)clef*, moins de la moitié des occurrences comportent un trait d’union ! Ces faits appellent deux

commentaires :

- (1) Encore une fois, on voit que ce ne sont pas les recommandations du Rapport qui sont en jeu mais la relation complexe entre les usagers et les mots, pris un à un. Les mots soudés sont rares, et ceux qui le sont ont une histoire qui précède les rectifications de 1990.
- (2) Ce n'est pas non plus la norme conservatrice qui est à l'œuvre. Pour exemple, les mots non soudés se présentent, dans une grande proportion, sans trait d'union alors que la norme en impose un.

J'é mets à nouveau l'hypothèse selon laquelle les usagers acceptent des variantes à partir du moment où elle ne modifie pas trop la physionomie du mot. Ainsi, *croque monsieur* est accepté parce qu'il se distingue visuellement et fonctionnellement très peu de *croque-monsieur*. Ce n'est pas le cas de *croquemonsieur*, qui semble un mot nouveau, qu'on ne sait où couper à la ligne, etc. En termes de «théorie de l'optimalité» (Prince & Smolensky 2004), on pourrait dire que les usagers de la langue génèrent toutes sortes de variantes mais ne conservent que celles qui répondent le mieux à un ordonnement de contraintes. L'une de ces contraintes de haut niveau (qu'on peut mettre en parallèle avec la contrainte phonologique PARSE) serait que la variante comporte le même nombre d'unités (ici de lettres, d'espaces ou d'accent) que sa "forme profonde" (ici constituée par la représentation de base que l'usager se fait du mot).

6.2. Mots composés : soudures de mots dont les éléments sont des noms ou des adjectifs

«... les mots dont les éléments sont nominaux ou adjectivaux : *arcboutant, autostop, bassecour...*»

"Radical" :	millepatte 10	← mille(-)patte(s) 743	1.3%
"En usage" :	millefeuille 350	← mille(-)feuille(s) 857	28.9%

Alors que ces deux paires de formes semblent rigoureusement parallèles, elles présentent une répartition radicalement différente. Même commentaire qu'en 6.1.

Par ailleurs, on trouve parmi ces formes 179 *mille(-)patte* (sans *s* à *patte*) et 424 *mille(-)feuille* (sans *s* à *feuille*). On verra plus loin dans le cadre du

traitement des pluriels que les usages sont très variés et, là encore, sans rapport avec les rectifications de 1990.

6.3. Mots composés : mots composés d'éléments expressifs

«... les mots d'origine onomatopéique ou expressive : *blabla, bouiboui, coincoin...*»

“Radical” :	coincoin 500	←	coin-coin 864	36.7%
“En usage” :	traintrain 533	←	train-train 18400	2.82%

Ces formes semblent contredire ce qui a été dit plus haut puisque c'est la forme “radicale” qui semble la mieux représentée. Cependant, le cas de *coincoin* appelle un commentaire : sa présence actuelle semble correspondre à un néologisme récent (*un coincoin* = *un coin*. Peut-être popularisé par une émission de télévision?). Dans la plupart des cas, il ne s'agit pas de la soudure du *coincoin* des canards.

Quand à *traintrain*, tout ce que je peux en dire est que si la forme a été représentée dans certains dictionnaires antérieurement à 1990 comme l'atteste la liste de Goosse, elle n'a pas trouvé son chemin dans le public.

6.4. Mots composés : soudure de composés issus d'une autre langue

«... les mots d'origine latine : *apriori* (nom), *exlibris, exvoto...basketball, blackout, bluejean...*»

“Radical” :	exvoto 12	←	ex voto 356	3.26%
“En usage” :	un apriori 352	←	un a priori 25000	1.39%

«... ou étrangère : *basketball, blackout, bluejean...*»

“Radical” :	bluejean 14	←	blue jean 531	2.57%
“En usage” :	weekend 103000	←	week-end 2570000	3.85%

Toutes ces formes montrent à quel point la soudure est impopulaire (même dans le cas de mots anglais, comme *basketball* ou *weekend*, que les anglophones eux-mêmes écrivent aujourd'hui soudés).

Une hypothèse sur cette résistance est proposée en 6.1. : la soudure, contrairement à l'usage ou non du trait d'union, modifierait trop l'aspect général du mot.

6.5. Mots composés : soudure de composés avec **-TOUT, -PARTOUT**

«... les noms composés dont le second élément est tout ou *partout* : *brise-tout, faitout, fourretout...*»

“Radical” :	brisetout 0	←	brise-tout 30	0%
“En usage” :	faitout 956	←	fait-tout 689	58.1%

Même commentaire qu'en 6.1. On peut cependant se demander comment *faitout*, qui non seulement réduit la taille du mot, change son aspect mais aussi lui fait perdre une lettre, a pu s'imposer dans plus de la moitié des occurrences. Cependant, si l'étymologie est bien fait-tout, on a affaire ici à un produit commercial. La forme soudée peut avoir été véhiculée par des affiches ou par des réclames.

6.6. Mots composés : soudure de composés commençant par **PORTE-**

«... les noms composés dont le premier élément est *porte-* : *porteclé, portecrayon...*»

“Radical” :	porteclé, etc. 46	←	porte(-)clefs, etc. 24000	0.19%
“En usage” :	portemine 17	←	porte(-)mine 324	4.99%

Là encore, même commentaire qu'en 6.1. et 6.4 : les formes soudées sont peu représentées, même quand elles étaient déjà présentes dans certains dictionnaires avant 1990.

Dans le cas de *porteclé*, j'ai additionné toutes les occurrences possibles de formes soudées : *porteclé* 5, *porteclés* 1, *porteclef* 28, *porteclefs* 12 (total 46). De même, pour les formes non soudées, avec ou sans trait d'union, j'ai additionné toutes les occurrences : *porte(-)clé* 697, *porte(-)clés* 737, *porte(-)clef* 21500, *porte(-)clefs* 848 (total 24000).

Or, le cas de *porte(-)cle(f)(s)* appelle un commentaire important. On constate que les formes les plus représentées (et ceci même dans les cas de for-

mes soudées) comportent un *f*. Ceci soulève une question intrigante : pourquoi, alors qu'il existe une variante plus simple, plus régulière, plus courte (*clé*), les usagers de la langue préfèrent-ils, et dans une proportion écrasante, une variante archaïque, irrégulière, et plus longue (*clef*) ? La réponse à cette question devrait être importante pour la préparation d'une éventuelle future réforme destinée à opérer des simplifications.

Serait-ce, comme il a été proposé plus haut, que les usagers sont attachés à une norme traditionnelle ? On peut rejeter cette hypothèse quand on sait que plus de la moitié des formes de *porte(-)clef(s)* font l'économie du trait d'union (*porte clefs*, etc., en deux mots). On verra également comment les pluriels des mots composés sont traités.

Est-ce parce que les gens préfèrent les orthographes compliquées ? Il y a peut-être de ça, mais encore faut-il que ces "complexités" soient facilement repérables, comme l'accent circonflexe ou le *f* de *clef*. Cette hypothèse va dans le sens développé plus haut de la question de la physionomie du mot. Finalement, même si une variante non fautive plus simple *clé* est disponible, *clef* a une physionomie plus facilement repérable. Ceci expliquerait pourquoi les deux variantes restent en concurrence depuis 1835 (Catach 1995). En revanche, quand les "complexités" en jeu impliquent un ensemble de règles ou d'exceptions (accord du participe passé, pluriel des mots composés, pluriel des mots étrangers), elles glissent plus facilement vers une régularisation. On le voit aussi dans le cas de l'accent circonflexe: indestructible sur *abîme*, il faiblit dans *paraît* car il s'agit d'une désinence grammaticale.

Il en va de même, comme on l'a vu, quand les graphies impliquées sont moins visibles (cas de l'orientation de l'accent sur *e*). À ce propos, notons un cas intéressant, celui de la ligature du *æ* de *cœur*. Celle-ci fait rarement l'objet de discussions lorsque sont débattues des questions orthographiques. Or, sur l'internet français, *plus de la moitié* des mots concernés sont écrits sans ligature (par ex. *coeur* 4 100 000, *cœur* 3 220 000, soit 56% sans ligature). On peut réduire la question en disant que les Français ne savent pas taper la ligature et que, faute de mieux, ils la remplacent par *oe*. Mais cette explication me paraît insuffisante. Je gage que si une Commission proposait de supprimer le *æ* il réapparaîtrait dans tous les sites. Dans le cadre de l'hypothèse émise jusqu'à présent, je suppose que le

æ est négligé parce que son absence ne défigure pas suffisamment le mot pour que cela soit repéré par les usagers de la langue. De façon encore plus intéressante, *Le Monde*, qui se montre radicalement conservateur en ce qui concerne toutes les autres graphies, accepte dans les parties rédactionnelles de son site 61% de formes sans ligatures (*coeur* 33500, *cœur* 20800). Ce n'est donc pas une question de savoir ou ne pas savoir taper la lettre en question (ce ne devrait pas être un problème au *Monde*), mais bien de la repérer comme importante ou non.

6.7 Mots composés : TIRE-BOUCHON et COUVRE-PIED(S)

«sont soudés ... *tirebouchon* (ainsi que *tirebouchonner*) et *couvrepiéd*. »

“Radical” :	tirebouchon 5	←	tire-bouchon 856	0.58%
“Radical” :	couvrepiéd 1	←	couvre-piéd 159	0.63%

Même commentaire que précédemment à propos des composés soudés.

7.0.a Pluriel des mots composés

«Les noms composés formés, avec trait d'union, d'un verbe suivi d'un nom ou d'une préposition suivie d'un nom prennent la marque du pluriel au second élément quand et seulement quand le nom composé lui-même est au pluriel : ... *un* (ou *une*) *après-midi*, *des après-midis* ; *un essuie-main*, *des essuie-mains*... »

“En usage” :	un essuie-main 143	←	un essuie-mains 224	39%
“En usage” :	des casse-cous 41	←	des casse-cou 199	17.1%

Dans toutes les formes précédentes, afin de ne traiter que les singuliers, j'ai fait les recherches avec l'article indéfini (*un tirebouchon*, etc.). Ici, comme il s'agit de repérer la présence des formes «nouvelles», j'ai cherché la forme qui entre en contradiction avec la forme «ancienne». Ainsi, pour *essuie-main*, c'est la forme du singulier (sans s), qui est réellement «nouvelle». Inversement, en ce qui concerne *casse-cou*, c'est la forme du pluriel, avec s, qui est à comparer avec la forme traditionnelle sans s.

Dans un premier temps, j'ai émis l'hypothèse selon laquelle les formes

comprenant des parties du corps seraient moins bien acceptées que les autres formes, si elles entraînaient des contradictions entre le nombre de parties du corps (par exemple *deux mains*, *un cou*) et les formes nouvelles (ex. : *un essuie-main*, *des casse-cous*). Cette hypothèse s'est avérée sans fondement. En effet, on voit ci-dessus qu'un *essuie-main* est aussi bien représenté que *des lave-vaisselles*, l'un et l'autre étant pareillement en contradiction avec les règles «anciennes» du pluriel des mots composés.

“Radical” : des lave-vaisselles 227 ← des lave-vaisselle 347 39.5%
“En usage” : des brise-glaces 373 ← des brise-glace 461 44.7%

Ce qui apparait de façon nette, c'est une large diffusion de la forme de pluriel que préconisait le Rapport : pas de *s* au singulier, un *s* au second élément au pluriel. Néanmoins, deux bémols sont à apporter :

- (1) Non seulement les pluriels “réguliers” tels que ceux proposés par le Rapport sont présents, mais également toutes sortes de pluriels, y compris ceux qui mettent un *s* au premier élément même si c'est un verbe. Dans le cas de *brise(s)-glace(s)*, les formes avec un *s* à *brises* représentent 22% de l'ensemble des formes ! C'est une proportion très importante qui montre encore une fois que l'usage réel stabilise des formes auxquelles les rectificateurs n'avaient pas pensé.
- (2) Certaines formes “résistent” à la régularisation. C'est le cas d'*après-midi* :

“Radical” : des après-midis 833 ← des après-midi 14500 5.43%
“En usage” : des après-skis 91 ← des après-ski 166 35.4%

Peut-être s'agit-il d'une de ces formes emblématiques que les usagers de la langue ont facilement repérées et qu'ils ne veulent pas modifier? En tout cas, elle montre qu'il n'y a pas de régularisation générale.

8.1. Les numéraux s'écrivent tous avec des traits d'union

«Les numéros composés sont unis par des traits d'unions : *trois-mille-deux-cent-vingt-et-un*»

“Radical” : mille-deux-cent-vingt 3 ← mille deux cent vingt 176 1.67%
“En usage” : quarante-et-un 33 ← quarante et un 267 11%

C'est encore là un cas où on peut parler de modification de la physionomie des mots. De façon intéressante, les usagers de la langue semblent tout simplement généraliser la règle selon laquelle «on met des traits d'union en dessous de cent». Ainsi, *quatre-vingt* est écrit avec trait d'union dans 77% des cas, *quarante et un* (qui selon la règle «classique» ne prend pas de trait d'union) est écrit avec trait d'union dans 11% des cas, et les mots au dessus de cent pratiquement jamais.

9.0. «*Laissé*» suivi d'un infinitif ne s'accorde pas

«Le participe passé *laissé* suivi d'un infinitif reste invariable: ...*ils se sont laissés mourir de faim.*»

“Radical” : les a laissé partir 58 ← (laissés+laissées) 44 56.9%
“En usage” : se sont laissé mourir 18 ← (laissés+laissées) 75 19.4%

Là encore, on observe une régularisation qui va dans le sens de ce que préconisait le Rapport. Néanmoins, on voit que toute “règle”, si on voulait en écrire une, serait diversement suivie selon les formes. Et encore, il ne s'agit là que d'un aspect très limité de la question des participes passés, qui laisse ouvert nombre de situations épineuses. Comme dans le cas des pluriels des mots composés, la meilleure réponse à apporter en matière de réforme serait sans doute, plus que de nouvelles règles, une tolérance envers les usages réels. On voit bien ici qu'il serait absurde de considérer ces formes comme fautives.

10.0. Le pluriel des mots étrangers

«Les noms empruntés à d'autres langues font leur pluriel à la française : *des matchs, des solos, des maximums...*»

“Radical” : des supermans 105 ← des supermen 706 12.9%
“En usage” : des matchs 3850000 ← des matches 2325000 94.3%

Le pluriel des mots étrangers constitue un terrain passionnant qu'il conviendrait de creuser plus en détail. D'abord, constatons que le pluriel régulier *matches* est totalement adopté par les usagers de la langue (même si *Le Monde* continue à n'utiliser que le pluriel anglais). *Sandwitchs* est également bien adopté (67%).

Plus intéressant est le cas du pluriel de *superman*. Si on ne compare que le pluriel régulier et le pluriel anglais, c'est le pluriel anglais qui semble largement l'emporter ; mais on rencontre aussi 514 *des superman*, sans s, qui, s'ils sont pris en compte, représentent 39% des formes ! Ceux-ci ne peuvent constituer un simple incident orthographique. Pour des mots comme *frite* ou *autoroute* par exemple, l'oubli du s au pluriel n'excède pas 1%. Dans le cas de *match*, les pluriels sans s ne représentent que 0.3%. Pour justifier cette forme, on peut émettre l'hypothèse suivante : d'une part les usagers, dans l'ensemble, savent que la forme *supermans* entre en contradiction flagrante avec le pluriel anglais attendu ; d'une part, ces mêmes usagers prononcent bien, en français, *des superman* et non *des supermen*. Ce dilemme les amène à considérer *superman* comme une sorte de nom propre et à le laisser sans marque du pluriel.

Là encore, on voit que les usagers de la langue proposent collectivement des solutions auxquels le rectificateur n'avait pas pensé.

11.1. Les finales en **-ILLIER, -ILLIÈRE** deviennent **-ILLER, -ILLÈRE**

« Régularisations de finales : *-er* au lieu de *-ier* dans *joailler, marguiller, quinquaiiller...* »

“Radical” :	serpillère 684	← serpillière 566	54.7%
“En usage” :	quincailler 244	← quincaillier 272	47.3%

Ces formes confirment-elles ou infirment-elles l'hypothèse émise plus haut selon laquelle une variante est mieux acceptée si elle ne change pas la physionomie du mot ? Il me semble ici que le *i* des formes « anciennes » est en effet peu perceptible, pour des raisons optiques et phonétiques. Comparez par exemple la paire *ognon / oignon* et la paire *quincailler / quincaillier*. Dans le premier cas, le *i* est en tête de mot ; son point altère la forme du mot ; enfin, s'il est prononcé, il modifie sensiblement la pronon-

ciation du mot. Tous ces éléments font qu'il est *facilement repérable*, pour reprendre le terme qu'on a déjà utilisé à propos du *f* de *clef*. Dans la deuxième paire, en revanche, le *i* est en fin de mot ; il est caché dans l'ombre du *ill*, qu'on repère comme une seule graphie ; la présence ou l'absence de son point n'altère pratiquement pas la forme du mot ; enfin, qu'il soit prononcé ou non, il ne change pas la prononciation du mot. Toutes ces raisons expliqueraient pourquoi les variantes simplifiées de ces mots sont bien reçues.

11.2. Les finales en *-OLLE* deviennent *-OLE*

«... *-ole* au lieu de *-olle* : *barcarole*, *bouterole*, *corole*...»

“Radical” :	corole 155	← corolle 14900	1.03%
“En usage” :	guibole 287	← guibolle 168	63.1%

Comme en 3.0., en l'état actuel des choses, la simplification des doubles consonnes modifie trop la physionomie des mots pour être acceptable. Le mot *guibole* constitue sans doute une exception parce que c'est un mot d'argot et que ceux-ci, lorsqu'ils apparaissent sous forme écrite, sont essentiellement des transcriptions phonétiques. Il ne s'agit pas dans ce cas de *simplification* d'une forme en une autre mais plutôt de deux transcriptions alternatives.

12.0. Simplification de consonnes doubles irrégulières

«Simplification de consonnes doubles dans les mots *dentelière*, *interpeller*, *prunelier* et *lunetier*.»

“Radical” :	pour interpeler 132	← pour interpeller 13400	0.98%
“En usage” :	une dentelière 27	← une dentellière 339	7.38%

Même commentaire qu'en 11.2. Malgré l'avantage que représenterait la simplification sur le plan de la correspondance son-graphie (le mot *interpeller*, par exemple, ne se prononce pas comme il s'écrit), la réduction de deux consonnes à une induit un changement trop repérable de la physionomie du mot. Par ailleurs, le mot *interpeller* est un de ces mots em-

blématiques sur lesquels se sont focalisés les médias au moment du Rapport. Il apparaît surtout dans des sites d'information dont on peut penser que les auteurs sont particulièrement attachés à son orthographe.

13.0. Certaines graphies particulières sont modifiées.

«Rectifications de mots particuliers. Ce sont les participes *absout* et *dissout*; les infinitifs *assoïr*, *rassoïr*, *sursoïr*... les mots suivants... : *Appâts* (pour *appas*), *bizut*, *bonhomme*...»

“Radical” :	avec <i>bonhomie</i> 112	← avec <i>bonhomie</i> 482	18.9%
“En usage” :	<i>boursouffler</i> 58	← <i>boursoufler</i> 266	17.9%
“En usage” :	<i>nénufar</i> 24	← <i>nénuphar</i> 959	2.44%
“Radical” :	a <i>dissout</i> 335	← a <i>dissous</i> 663	33.6%

Là encore, la forme la plus représentée est celle qui comporte une désinence grammaticale, et qui de surcroît modifie peu l'aspect du mot (le participe passé *dissout*). Le cas de *boursouffler* et surtout celui de *bonhomme* m'interpellent, puisque les variantes «nouvelles» sont visuellement très différentes des graphies «anciennes». Cependant, on peut noter que ces variantes n'atteignent pas les quelque 50% de *serpillère*, *quincailler*, *complètera* ou *évènement* ; par ailleurs, elles viennent s'inscrire chacune dans une série (*homme*, *prudhomme*, *bonhomme* ; *souffler*, *siffler*...) ce qui peut être un facteur.

L'orthographe ordinaire sur internet

Ainsi peut-on conclure que les formes recommandées par le Rapport de 1990 ne sont pas *suivies* ou rejetées en tant que telles. Quinze ans après, les formes alternatives les plus employées sont sans doute celles qui *précèdent* le Rapport et non celles qu'il instigue. Tout se passe comme si les usagers avaient leur vie orthographique propre, et que c'étaient eux qui créaient, acceptaient ou rejetaient les variantes.

Il est vrai qu'internet donne parfois une impression de chaos orthographique. Certaines pages, écrites en français SMS, présentent des formes qu'on peut vraiment qualifier de radicales ou de révolutionnaires (la forme SMS *c ki*, par exemple, représente 12% de toutes les formes de type

c'est qui, qui c'est, qui est-ce, ce qui est considérable).

Mais qu'en est-il de l'orthographe ordinaire ?

- (1) Pour ne prendre qu'un exemple d'orthographe grammaticale, l'oubli du *s* pluriel pour un mot donné est de l'ordre de 1%. C'est peu, mais ça signifie en gros qu'en moyenne, sur cent mots au pluriel, on trouvera un oubli.
- (2) D'un autre côté, l'orthographe lexicale semble être une spécialité des usagers d'internet. Un mot raisonnablement compliqué comme *raisonnablement* n'est mal orthographié que dans 0.5% des cas (*raisonnablement, résonnablement*, etc.). Un mot comme *essentiel*, dont je m'attendais à ce qu'il présente des variantes nombreuses, n'est mal orthographié que dans 0.06% des cas !
- (3) En revanche, le mot *ascenseur* est orthographié de façon non conventionnelle dans 9.5% des cas. On approche du chiffre de 10% qui indique la présence d'une variante stable. Il s'avère que celle-ci n'est pas la plus simple (*acenseur*, par exemple), mais *ascenceur* avec *sc* puis un *c* (encore une fois, peu de changement dans la physionomie du mot). Si on accepte les deux variantes *ascenseur* et *ascenceur*, on s'aperçoit que ce mot n'est mal orthographié que dans 0.15% des cas, ce qui nous ramène aux taux précédents.

Conclusion et perspectives

L'exemple d'*ascenceur* ci-dessus est particulièrement instructif en ce sens qu'il nous montre que l'usage réel recèle des variantes stables qui sont demeurées ignorées des auteurs des différentes "réformes" et autres aménagements. Au cours de cette étude, on a pu constater également que certains usagers de la langue avaient parfois choisi des solutions différentes de celles qui avaient été préconisées (par exemple, dans le cas du pluriel des mots composés ou dans celui du pluriel des mots étrangers, dans le cas d'*aigue* sans tréma, etc.). En ce qui concerne les mots composés, on a vu que les usagers se souciaient peu d'employer ou de ne pas employer le trait d'union, à partir du moment où les éléments du mot restaient séparés au moins par un espace.

Une autre constatation importante a été de remarquer que les usagers de la langue ne préféraient pas nécessairement les formes les plus *simples*,

mais sans doute les plus repérables (*f* de *clef*, *i* de *oignon*, accent circonflexe de *abîme*, doubles consonnes...). Une hypothèse qui nécessiterait plus de développement, mais qui s'est bien confirmé d'exemple en exemple, est que les usagers tendent, de manière générale, à rejeter les variantes qui induisent trop de changement dans la physionomie du mot. C'est aussi la raison pour laquelle les formes soudées des mots composés ne constituent pas, à quelques exceptions près, des variantes stables.

Retour sur les motivations du rejet des rectifications de 1990

Les motivations des Français à refuser les rectifications ont été beaucoup débattues : « purisme populaire » selon Cerquiglini, refus de remettre en cause ce qu'ils ont eu tant de mal à apprendre, refus de toucher à l'esthétique de la langue, etc. J'é mets maintenant une hypothèse nouvelle : ce ne serait pas tant, chez le public ordinaire, des raisons « puristes » ou motivées esthétiquement qui auraient été les principales causes du rejet, mais des raisons de *nécessité*, ou plutôt d'*absence de nécessité*. En effet :

- (1) L'usager de la langue n'a sans doute pas ressenti la nécessité de modifier ce qu'il maîtrise sans trop d'effort (*oignon*, *tire-bouchon*, *île*).
- (2) De plus, il modifiait déjà certains pans de la langue, que ce soit pour des formes que les rectifications que la Commission touchaient (*asséner*, *matches...*) ou non (*aigue*, *des superman...*).

En effet, quelles étaient les motivations de ces "rectifications" ?

(1) Le *Français dans le Monde* parle de rectifications « dont le propos était de faciliter l'apprentissage de la langue ». Mais pour qui ? Trois cas différents pourraient faire penser à des réformes différentes :

- Les enfants,
- Les adultes pauvres en éducation, qui subissent une discrimination du fait de leur orthographe déficiente,
- Les étrangers (étudiants ou résidents).

Cependant, pour chacun de ces groupes, la "réforme" proposée aurait été une plaisanterie. Ce n'est pas la modification de *corole* ou du pluriel d'*essuie-main* qui aurait pu changer le rapport d'aucune de ces catégories de personnes avec l'orthographe du français. Si ç'avait été ce qu'elle visait, la "réforme" proposée aurait à la fois manqué d'ampleur, de rigueur,

et aurait souffert d'un grave déficit de communication.

(2) Ses motivations étaient-elles donc ailleurs ? Les experts interrogés en 2000 par le *Français dans le Monde* déclarent aussi que Michel Rocard était «désireux de montrer le français comme une langue moderne» (N° 313, 62) «Plus que les transformations elles-mêmes, l'important était de montrer la capacité de la langue à se réformer». Mais pourquoi le français aurait-il eu besoin de se montrer comme une langue moderne, capable de se réformer ? Émettons deux hypothèses.

- Parce que la langue est une métaphore du pays lui-même : donner du français l'image d'une vieille langue sclérosée, c'est donner l'image d'un vieux pays sclérosé. Si c'était là le propos, il y a eu là aussi un déficit de communication du gouvernement vers les acteurs publics, en particulier les entreprises en contact avec l'étranger. Il est possible que cette manière de voir la réforme, si elle a été ignorée par une France qui ne se soucie pas beaucoup de son image à l'étranger, ait compté dans l'acceptation immédiate des recommandations par les institutions belges, suisses et canadiennes.
- Parce que l'action sur la langue est une métaphore de l'action sur le pays, marquer la langue, c'était marquer symboliquement le pays. Comme me l'a déclaré l'un des membres du comité d'expert, «Mitterrand et Rocard voulaient laisser une marque durable».

Toutes ces motivations, qui appelaient à des rectifications différentes et à une communication différente, étaient intriquées et n'ont pas été comprises. C'est peut-être là une des raisons de l'échec. Pour que des modifications de la manière d'écrire soient acceptées par le public, il faut des raisons fortes qui dépassent le simple rapport du linguiste avec son objet. On le voit dans le changement survenu ces dernières années dans la féminisation des noms de métier (*auteure*, *professeure*, etc.). Venu du Canada, relayé par le Conseil supérieur de la langue française (dont Bernard Cerquiglini), accepté par les médias (dont *Le Monde*), ce mouvement est également accepté par les usagers de la langue (au niveau de la lecture en tout cas). La raison de cette acceptation est claire. Ces formes dépassent le simple cadre de la *langue*, de la *règle*, de la *simplicité*, etc., pour toucher à

des problèmes sociétaux qui les concernent directement : la place des femmes et leur représentation dans la société.

Que pourrait être une future réforme de l'orthographe du français ?

Un autre erreur commise a sans doute été celle qui a consisté à s'intéresser aux *règles* plutôt qu'aux *formes*. En ce sens, les rectificateurs se sont tout autant comportés en puristes que ceux qui les ont combattus.

Prenons l'exemple du pluriel des mots composés : plutôt que de proposer, en quelque sorte, une *norme alternative* sous forme de règles (avec exceptions, de surcroît), il aurait été plus simple, plus productif et plus juste d'entériner la tolérance sur les pluriels. Autrement dit, de ne plus considérer comme *fautives* des formes déjà en usage (comme on le voit par exemple dans le cas de *brise-glace*, pour lequel les pluriels *brise-glace*, *brises-glace*, *brise-glaces*, *brises-glaces* se disputent le terrain de façon presque équitable). Plus simple, parce qu'ouvert sur l'avenir et sur l'observation des usages ; plus productif, parce que plus facile à mettre en place qu'une nouvelle norme ; et enfin plus juste, parce que constituant un réel soulagement pour ceux qui ont des difficultés avec l'orthographe. Le même raisonnement s'applique à l'accord du participe passé.

Ne nous méprenons pas : il ne s'agit pas de déclarer comme *correctes* n'importe quelles variantes dans n'importe quelles circonstances, mais seulement celles qui apparaissent très représentées dans les usages. Nous avons vu dans cette étude que nous disposions avec internet d'un outil nouveau, qui nous permet de regarder assez précisément comment les gens écrivent et quelles sont les productions écrites auxquelles ils sont confrontés. C'est une occasion nouvelle de faire évoluer la langue et de sortir du «cercle vicieux» dont parle le *Français dans le Monde* : dans l'état actuel des choses les auteurs de dictionnaire cherchent «l'usage» dans les journaux et la littérature, alors que journalistes et auteurs, de leur côté, utilisent les graphies proposées par les dictionnaires.

En conclusion, ce que pourrait faire une future "réforme" de l'orthographe du français, se serait :

- Plutôt que d'en inventer (*tirebouchon*), porter attention aux graphies nouvelles qui apparaissent chez les usagers (*aigüe*).
- sélectionner et entériner les variantes en fonction de leur

fréquence. Ce qui était un travail impossible en 1990 est maintenant un jeu d'enfant : on peut repérer les variantes nouvelles (comme *ascenceur*), et celle qui disparaissent (comme *assener*). Les variantes qui atteignent une certaine représentation (10%, 20% ?) pourraient être automatiquement considérée comme non fautives.

- Plutôt que de proposer de nouvelles *règles*, entériner les *zones* où se manifestent des groupes de variantes (par exemple le pluriel des noms composés, l'usage du trait d'union). On ne parlerait alors plus de variantes d'un même mot mais plutôt de *manières d'écrire*. Un mot comme *brise-glace* n'aurait pas quatre pluriels, mais une seule forme qui répondrait simplement à une règle générale de formation des pluriels des mots composés (mettre facultativement un s à l'un des deux éléments du mot ou aux deux).
- Enfin, plutôt que de mener tous les vingt ans des tentatives de "réforme" ou de "rectifications", il pourrait y avoir processus continu, sous forme de publications régulière (au JO ou *via* l'Académie) de brèves listes de nouvelles graphies acceptées ou recommandées. On pourrait par exemple annoncer aujourd'hui que la variante *gageüre* de *gageure*, insuffisamment représentée, est abandonnée ; que la forme sans tréma de *aigue* sera acceptable ; ou que le pluriel sans s des mots étrangers (*des superman*, *des adagio*) ne sera plus considéré comme fautif.

Références

- Burnet, Pierre, 1955, 1967, *L'orthographe*, Que sais-je 685, PUF.
- Catach, Nina, 1978, 1997, *L'orthographe*, Que sais-je, PUF (remplace l'ouvrage de Pierre Burnet, avec un contenu différent).
- Catach, Nina, 1995, *Dictionnaire historique de l'orthographe française*, Larousse.
- Collectif, 1990, «Rapport du Conseil supérieur de la langue française sur les rectifications de l'orthographe», *Journal officiel de la République française* du 6 décembre 1990.
- Collectif, 2001, Dossier «Orthographe, la faute à qui ?» *Le Français dans le Monde*, N°313, Janv.-Fév. 2001.
- Goosse, André, 1991, *La "nouvelle" orthographe : Exposé et commentaires*,

Duculot.

Prince, Alan & Smolensky, Paul, 2004, *Optimality Theory: Constraint Interaction in Generative Grammar*, Blackwell.

Sites internet

Académie française : huitième édition du dictionnaire en ligne, premier volume de la neuvième édition, liste des graphies nouvelles.

APARO (Association pour l'Application des Recommandations Orthographiques) : vadémécum abrégé, liste de huit-cent mots modifiés.

Banque de dépannage linguistique (Office Québécois de la langue française) : FAQ sur les rectifications de 1990.

DLF: Direction à la Langue Française de Suisse : historique, quelques chiffres.

Francité : site de la maison de la francité (Belgique) : état des lieux, informations, faits, références.

Le «Site d'information sur la nouvelle orthographe française officiellement recommandée» : résumé des règles, guide au format pdf.