

Recherches dans un corpus de français écrit naturel : le règne des agents de l'orthographe

Introduction

Le but de cet article est de présenter un petit corpus de français écrit naturel ouvrant la voie à l'étude de l'orthographe réelle des scripteurs du français. Ce corpus est constitué d'un livre d'or de musée et de quatre cahiers des lecteurs de bibliothèques. Nous les utiliserons tout d'abord pour obtenir quelques données chiffrées sur l'usage réel des scripteurs. Puis, en nous focalisant sur les interactions hébergées dans les cahiers, nous rechercherons quelle est la place de l'idéologie associée à la norme orthographique dans l'écrit de tous les jours, à travers l'étude d'une pratique orthographique courante: la correction des fautes.

1. Présentation du corpus

Par l'appellation «corpus de français écrit naturel», nous entendons un corpus de français manuscrit et autant spontané que possible (le message est réel et non provoqué). Manuscrit, car nous partons du principe que les formes graphiques que produit un scripteur donné varient en fonction du canal par lequel ils s'expriment. En effet, selon qu'il écrit à l'aide d'un clavier d'ordinateur, d'un clavier de téléphone, d'un crayon ou de caractères d'imprimerie en fonte, un même scripteur diversifie ses pratiques orthographiques: erreurs de saisie informatique avec un clavier, tendance à l'abréviation avec un téléphone, caractères superposés avec une machine à écrire (si le chariot se bloque), recherche de la perfection orthographique pour un imprimé traditionnel, etc. Lorsqu'il écrit à la main, sur du papier, un scripteur est aussi soumis à des contraintes: il peut lui arriver de ne pas revenir en arrière tracer un signe diacritique, ne serait-ce que pour ne pas perdre le fil de son écriture, ou pour ne pas poser la main sur l'encre encore fraîche. Si l'on ne peut pas prétendre que l'écriture manuscrite est le mode le plus instinctif, le plus naturel, de l'expression écrite, on peut toutefois supposer que ce canal est celui qui laisse la plus grande liberté à l'utilisateur. Sans être en mesure de le démontrer, nous partons du principe que l'écriture manuscrite transcrit fidèlement l'image graphique d'un mot telle qu'un scripteur se la représente avant de la produire.

À cette caractéristique primordiale s'en ajoutent quelques autres. Les écrits du corpus doivent être récents (XXI^e siècle), afin de privilégier la recherche sur l'état actuel de la langue. Ils doivent le moins possible appartenir à la sphère du privé; ceux que nous avons collectés se trouvent dans des lieux accessibles à tous. Ils doivent enfin, dans l'idéal, être courts: nous préférons un petit corpus formé de centaines de messages courts laissés par autant d'auteurs différents à un long manuscrit d'un seul auteur.

Tous ces paramètres excluent un bon nombre d'écrits, parmi lesquels tous les imprimés, les copies d'élèves,¹ les manuscrits de professionnels de l'écriture (journalistes, écrivains) qui sont travaillés plutôt que spontanés, les courriers s'adressant à une hiérarchie (car la langue, et donc l'orthographe, y sont influencées par ces rapports hiérarchiques), les enquêtes de terrain où l'on demanderait à des personnes d'écrire un texte, enfin les dictées, dans lesquelles les scripteurs laissent de côté leurs pratiques orthographiques personnelles pour se rapprocher de la norme.

Signalons que les recherches dans ce sens, menées sur des écrits non littéraires du XX^e siècle, ne sont pas nouvelles. Ainsi Frei (1929) a-t-il utilisé un corpus constitué de lettres adressées à l'Agence des Prisonniers de Guerre de la Croix-Rouge pour étudier l'évolution du français sous l'angle de la linguistique fonctionnelle des années 1920; et des chercheurs de l'Université du Québec² ont-ils observé un corpus de lettres envoyées à un missionnaire entre 1930 et 1936 pour étudier la langue des colons autour du lac Témiscouata. Lucci et Millet (1994), encore plus proches de nos préoccupations orthographiques, ont enquêté sur des écrits ordinaires: courriers personnels et professionnels, journaux. Leurs observations permettent de cerner efficacement la notion de variation graphique et le trouble qui lui est attaché. Le présent travail s'inspire de leur étude.

Le corpus que nous sommes en train de constituer (25 000 mots actuellement) correspond aux critères énoncés plus haut. Il est formé de livres d'or de musées et de cahiers des lecteurs de bibliothèques, textes facilement accessibles. Les livres d'or, que l'on trouve souvent à la sortie des musées, sont des cahiers dans lesquels les visiteurs inscrivent leurs impressions à chaud après leur visite, ou déposent plaisamment une marque attestant leur venue. On peut diviser les messages rédigés dans ces livres en deux groupes principaux, d'après le ton des commentaires. D'une part, les messages positifs: remerciements, éloges; d'autre part, les messages négatifs: plaintes, réclamations, etc. Cependant, il est fréquent que les visiteurs n'écrivent rien dans le livre d'or et détournent l'usage habituel pour lequel il est prévu. Ils le parcourent alors pour lire les messages des autres visiteurs, en s'autorisant parfois à les commenter.

Les cahiers des lecteurs, dans les bibliothèques, ont une tout autre fonction, celle d'assurer la communication entre le public et les bibliothécaires, à travers la commande d'ouvrages. C'est

¹ Comme celles de l'inspecteur Beuvain, de Chervel et Manesse (1989) et de Manesse et Cogis (2007).

² C. Portelance *et al.*, *Un corpus du Témiscouata (1930-1936), Des lettres des pays d'en bas*, <<http://www.ling.uqam.ca/forum/corpus/messages/3.html>> (consulté le 27 août 2007).

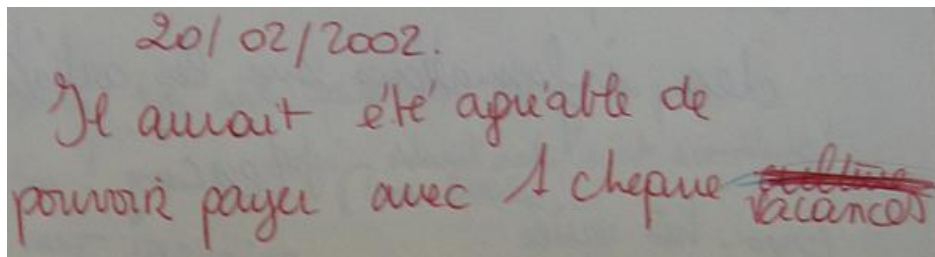
dans ce but que les pages des cahiers des lecteurs sont ordonnées en colonnes: un groupe de colonnes dévouées à la description de l'ouvrage (auteur et titre notamment) et la dernière colonne réservée à la réponse des bibliothécaires. Cependant, cette organisation préétablie des pages n'empêche pas les lecteurs, lorsqu'ils le désirent, de formuler des plaintes, des réclamations, des remerciements, tout comme dans les livres d'or.

Chaque cahier est ajouté au corpus en deux étapes. La première consiste à recopier tels quels les messages du cahier dans un fichier informatique, en suivant leur mise en forme (colonnes, retours à la ligne) et surtout leur orthographe. Les données personnelles concernant les scripteurs sont alors anonymisées. Il est également nécessaire de retranscrire, dans la mesure du possible, toutes les ratures et les séquences qui se chevauchent, comme dans cet extrait d'un livre d'or, dans lequel un mot est barré puis supplanté par un autre:

- 1a) 20/02/2002.
Il aurait été agréable de pouvoir payer avec 1 cheque ~~culture~~ vacances

La deuxième étape consiste à photographier toutes les pages du cahier, afin de pouvoir les consulter ultérieurement, et à anonymiser les photographies:

- 1b)



La collecte du corpus se heurte à deux difficultés techniques. La première difficulté réside dans le fait qu'il est très souvent difficile de lire les séquences barrées, et de voir, quand deux mots sont écrits au même endroit, lequel est «par-dessus» l'autre.³ Cette difficulté ne peut généralement pas être résolue avec les photographies. Il est donc nécessaire de procéder à la copie des cahiers directement sur les pages originales.

La deuxième difficulté rencontrée se situe au niveau de la lecture, lorsqu'il faut interpréter un signe pour le faire correspondre à un graphème. Le cas de figure le plus fréquent concerne les accents, signes qui, en fin de compte, ne sont que de petits traits placés au-dessus de la ligne,

³ Remarquons au passage que cette rature avec correction est une trace d'hésitation du scripteur, qui n'est pas sans rappeler les phénomènes d'hésitation et de reprise, fréquents dans la langue parlée spontanée (cf. Blanche-Benveniste et Jeanjean, 1987).

d'ailleurs pas toujours en face de la voyelle affectée, et qui pointent dans une direction donnée. Dans la pratique, les accents sont souvent illisibles, mais il faut toutefois les prendre en compte. Sauf s'ils étaient absents ou s'ils pointaient résolument dans une mauvaise direction, nous les avons interprétés comme correctement formés.

2. Matériau pour l'établissement de statistiques orthographiques

Dans les messages du corpus, chacun est en droit d'écrire comme il le veut. De fait, la norme communément admise dans les imprimés est balayée dans les cahiers de notre corpus par les usages spontanés de chacun. Cette plongée au cœur de l'usage orthographique des scripteurs offre un cadre de choix pour une étude objective de l'orthographe dans son état naturel. En effet, alors que l'orthographe française en cours dans les imprimés (où des correcteurs, des relecteurs, des éditeurs modifient les graphies de première main) ne cesse de se rigidifier de siècle en siècle depuis le premier *Dictionnaire de l'Académie française* (1694)⁴ jusqu'à devenir norme unique aujourd'hui, l'usage en cours dans les manuscrits est bien plus soumis à la variation.

À terme, le corpus de français écrit naturel devrait comporter plusieurs dizaines de milliers de messages, soit quelques millions de mots dont la graphie, rappelons-le, est copiée telle quelle. À partir de cet ensemble, il sera possible d'observer facilement l'usage orthographique en interrogeant le corpus sur un mot donné, pour trouver toutes les formes graphiques que les scripteurs d'aujourd'hui donnent à ce mot. Les études sur l'orthographe se libèreraient alors du cadre de l'imprimé auquel elles se confinent souvent (dictionnaires, presse, littérature). Nous donnons ici un avant-goût des possibilités qu'offre le corpus dans son état actuel.

Soit par exemple le mot *ouvrage*, si fréquent dans les cahiers des lecteurs. On en trouve 112 occurrences, singuliers et pluriels confondus, et on ne peut que constater que ce mot ne connaît pas de variation graphique. Quant à *œuvre*, fréquent lui aussi dans les cahiers et les livres d'or, on en trouve 39 occurrences avec *o* et *e* séparés: *oeuvre* ; et 36 avec *o* et *e* entrelacés: *œuvre*. Le taux de variation pour ce mot semble à première vue très élevé, mais il est probablement trompeur. Car en réalité la différence entre *æ* et *oe* est infime; un lecteur reconnaît l'un ou l'autre selon que le signe *e* est collé au signe *o* ou bien séparé de lui par un espace millimétrique. La grande variation dans l'écriture du mot *œuvre* trouve sa source dans un artéfact graphique. Le soin nécessaire au tracé du graphème *æ* n'est pas toujours déployé par les scripteurs.

⁴ Cerquiglini (2004:166).

Mais d'autres mots sont porteurs d'une variation bien plus palpable. Par exemple, *intéressant* apparaît sous quatre formes dans le corpus: *intéressant* 10 fois, *interessant* sans accent 5 fois, *interressant* sans accent et avec deux *r* 3 fois, et *intèrressant* une fois. Pour écrire ce mot, plusieurs scripteurs redoublent le *r* plutôt que d'employer l'accent aigu. Ce n'est pas la norme, mais c'est une structure graphique tout à fait valable en français, que l'on retrouve d'ailleurs dans *erreur*, *territoire*, *perroquet*, etc., quand ces mots sont prononcés avec [e].⁵ Ou peut-être ont-ils cherché à noter le son [ɛ]?⁶ Enfin, il est également possible que les scripteurs aient écrit *interressant* par analogie avec *interroger* ou *interrompre*, trois mots dans lesquels le préfixe latin *inter-*, bien qu'opaque, peut être senti intuitivement. Quoi qu'il en soit, la graphie *interressant* nous dévoile que son scripteur a mis en place une stratégie graphique pour noter le son [e] ou [ɛ], stratégie tout à fait conforme au système, mais non conforme du point de vue de la norme orthographique établie.

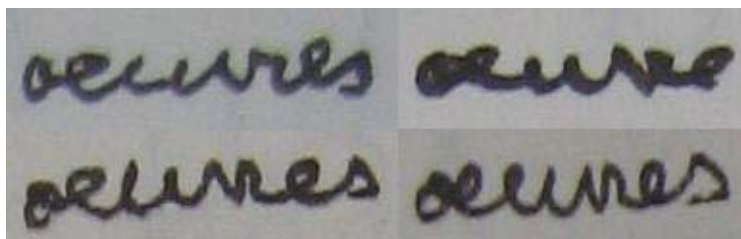
Toujours à propos de ce mot, remarquons que plusieurs se passent de noter l'accent aigu devant *r* simple. Cette hésitation autour de l'accent n'est pas caractéristique du seul mot *intéressant*. On la retrouve également dans les graphies de l'adverbe *très*: nous avons relevé *très* avec accent grave 67 fois, *trés* avec accent aigu 17 fois, et *tres* sans accent 5 fois. Et d'autres mots connaissent le même phénomène (7 occurrences de *édition* sans accent aigu, autant de *étudiant*). En fait, ce ne sont pas tant les mots eux-mêmes qui varient; c'est l'emploi des accents qui n'est pas stable, tout comme celui de *æ*. On peut avancer l'hypothèse que le système accentuel du français n'est pas utilisé dans toute sa latitude parce que sa mise en œuvre diminue la vitesse d'écriture, et parce que son absence n'empêche pas la lecture (sauf quelques cas d'ambiguïté). En un mot, il n'est pas pertinent.

La présence d'une forte variation graphique dans les manuscrits nous révèle des éléments sur la langue, mais aussi sur les scripteurs. Nous avons suggéré plus haut que le recours à la forme non normée *interressant* dévoilait une opération logique chez son scripteur, et non une insuffisance mentale. Un cas de figure tout à fait particulier vient étayer cette idée. Il s'agit du cas où un scripteur, dans un même message, utilise plusieurs fois un même mot, mais sous des graphies différentes. Voici trois collages obtenus à partir de trois messages:

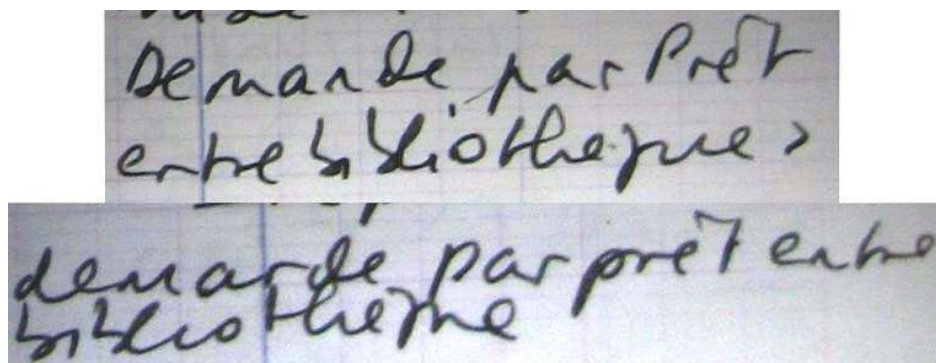
2)

⁵ D'après les données du *Petit Robert 2008* et du *Trésor de la langue française*, ces trois mots connaissent les deux prononciations, avec [ɛ] ou [e].

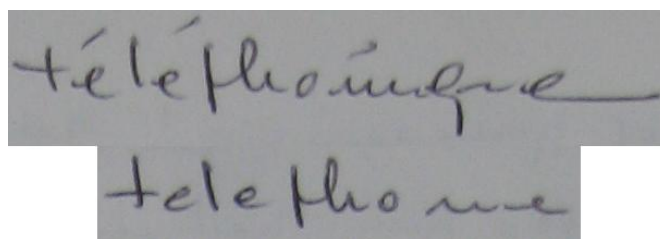
⁶ Cette prononciation de *intéressant* est également attestée par le TLF.



3)



4)



Dans ces messages, le cas de figure évoqué se produit: un scripteur écrit le mot *oeuvre* sans prendre soin de lier l'*e* à l'*o*, puis il écrit *œuvre* avec le soin dont il vient de se passer; et cela, deux fois de suite. De même dans *Prêt entre bibliothèques* avec *s* final, immédiatement suivi de *prêt entre bibliothèque* sans *s* final; et idem entre *téléphonique* pourvu d'accents d'une part, et *telephone* sans accents d'autre part. Nous imaginons que les trois scripteurs observés connaissent la norme du mot qu'ils écrivent, et qu'ils savent l'appliquer. Mais alors, pourquoi l'appliquent-ils irrégulièrement? Parce qu'utiliser la norme graphique de façon absolue n'est pas primordial dans la communication écrite.

Cependant, le scripteur d'un message ne doit pas oublier qu'il partage l'utilisation de celui-ci avec son lecteur. Et malheureusement, au-delà d'être un enjeu pour la compréhension du message, l'application de la norme graphique représente pour le lecteur une référence en vue d'émettre des jugements de valeur sur le scripteur. C'est en considérant les cahiers comme des supports d'interaction linguistique que nous proposons maintenant d'étudier le corpus.

3. Le manuscrit comme lieu d'interaction

Un point commun entre livres d'or et cahiers des lecteurs retient toute notre attention. Il s'agit de la situation d'interaction linguistique dans laquelle se trouvent les scripteurs, car les cahiers, en tant que lieux publics d'écriture, sont un carrefour linguistique entre utilisateurs de l'écrit. Ce cadre interactionnel est pour le moins original, puisqu'il se situe à l'écrit. Et bien sûr, les outils de la linguistique interactionnelle que l'on pourrait dire *traditionnelle* ne sont pas directement transposables à la situation d'interaction écrite qui nous préoccupe. Le cas le plus approchant de ceux que nous allons détailler est celui qu'observe Lopez-Bouchet (2004) sur des inscriptions graffitées. Elle constate que le mur support de graffiti devient un lieu d'interaction lorsque plusieurs tagueurs l'utilisent. L'une des caractéristiques de cette interaction nous intéresse particulièrement: lorsqu'un premier graffiti est recouvert d'un second graffiti dessiné par-dessus, cela s'interprète comme une revendication de la part du second tagueur (généralement une revendication de territoire). D'ailleurs, le graffiti est parfois accompagné d'un texte explicite à ce sujet.

Dans nos cahiers, il y a interaction entre deux scripteurs dès lors qu'un message (message 1, laissé par le scripteur 1) reçoit une réponse, un commentaire ou une modification (message 2, laissé par le scripteur 2). L'interaction écrite présente diverses caractéristiques remarquables. Premièrement, le temps y est détendu, dans le sens où le message 2 survient après un temps indéterminé, variant de quelques secondes à plusieurs années. Deuxièmement, le scripteur 1 n'est pas informé de la survenance d'un message 2 en réponse à son message initial, qui se déroule en son absence; et il ne peut donc pas y répondre à son tour. On peut dire de cette interaction qu'elle n'est pas dialogique, et difficilement constructive. Troisièmement, le personnel des musées n'étant pas censé écrire dans le livre d'or, la plupart des messages 1 qui y sont laissés s'adressent à des personnes qui n'y répondent pas, tout en ne s'adressant pas aux scripteurs qui se permettent d'y répondre.

Les cahiers des lecteurs abritent des interactions plus classiques, dans le sens où une colonne est toujours prévue pour recevoir la réponse des bibliothécaires au message laissé. Dans la plupart des cas, le message 1 formule la demande d'un ouvrage dont le scripteur a besoin et dont il estime qu'il manque à la bibliothèque; et le message 2 contient une réponse, soit positive («commandé le 6.01.04», «OK. à l'étude», etc.), soit négative («Nous n'achetons pas cette collection», «épuisé», etc.), auquel cas elle est souvent argumentée.

Mais les interactions qui nous intéressent ne se situent pas à l'endroit où on les attend. Voici la copie d'un message particulièrement parlant. Nous transcrivons le message 1 en caractères maigres et le message 2, très réduit, en caractères gras :

- 5) Il y a de très belles choses à visiter en France et dans chaque region Malheureusement pour beaucoup d'entre nous familles nombreuses les «chèques vacances» ne sont pas accept~~er~~^{és}. Alors un peu d'effort un peu plus pour notre patrimoine et une ouverture pour nos enfants. A Vulcania _ Les chèques sont acceptés. Merci

Le message 1 est une requête formulée par un père ou une mère de famille nombreuse souhaitant que les chèques-vacances soient acceptés à l'entrée du musée. Le message 2, laissé par une autre personne, dont on ne sait rien, consiste en cinq *s* de pluriel ajoutés là où ils sont manquants (*belles choses, familles nombreuses* et *enfants*) ainsi qu'en la correction du participe passé du verbe *accepter*, dont la finale est barrée puis réécrite.⁷ C'est la nature de l'encre et la taille des lettres qui nous ont permis de faire le départ entre les deux messages: le premier est écrit à l'encre bleue, et le second au stylo à bille bleu. En outre, les *s* de pluriel ajoutés par le scripteur 2 sont identiques entre eux et différents des *s* formés par le scripteur 1.

Le message 2 est non verbal et ne constitue pas vraiment un énoncé, mais il est porteur de sens. Il vient se greffer au message 1 dans le but d'en modifier la forme. À travers son intervention, le scripteur 2 a ramené l'usage graphique du scripteur 1 à la norme. On dirait couramment qu'il lui a corrigé les fautes d'orthographe.

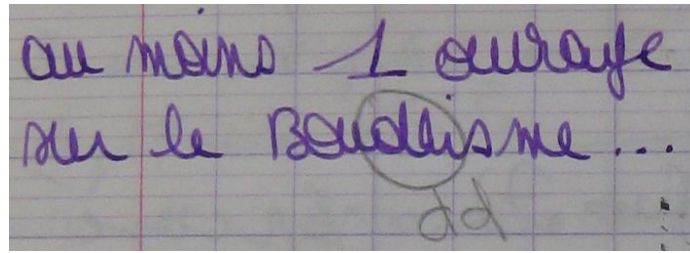
On trouve beaucoup d'autres exemples de ce type d'interaction basée sur la forme des messages dans les cahiers. En voici un florilège. Les ajouts du scripteur 2 apparaissent en caractères gras dans les séquences 6), 7) et 11):

- 6) Les adjectifs qui qualifient le musée (exceptionnel, unique...) me semblent un peu excessifs **(un lecteur indiscret..)**
- 7) Ça serait bien si c'était touts⁸ les jours «gratuit» comme les Musées du Département
- 8) Il serait bienvenue⁹ que dans une bibliothèque universitaire [...]
- 9)

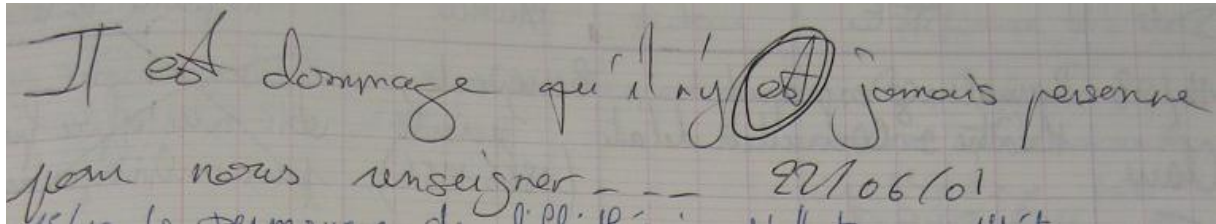
⁷ Le scripteur 1 ayant d'abord écrit «ne sont pas accepter», le scripteur 2 a ensuite barré le *r*, puis a ajouté un accent aigu et un *s* final.

⁸ Le *t* final de *tout*, au stylo noir, est surchargé par un *s* au stylo bleu. Le double soulignement est au stylo noir.

⁹ Le *e* final de *bienvenue*, au stylo bleu, est barré au crayon et au stylo noir.



10)



- 11) [...] je trouve que l'espace qui nous est réservé est incomplet. Nos professeurs nous confirment eux aussi, d'un ton scandalisé le manque d'ouvrages importants pour notre section. Le déplacement à la BNF est de plus en plus dur^e [...]. comment puis-je réussir ma 1^o année si la BU ne s'enrichi^et pas ? [...] Je vous consei^le de consulter les professeurs et responsables [...] pour remédi^er à cette situation désagréable pour vous comme pour tous les étudiants notam^ent de 1^o année un peu perdus dans le monde universitaire. Nos professeurs nous ont assurés^s qu'ils seraient enthousiastes de travaill^esr avec vous ou d'être écoutés^s. Je vous demande enfin, de nous mettre à disposition les [...] qui semblent manqués^s¹⁰. [...]

Considérons dans un premier temps les séquences 5) à 10). Chacune d'elles est le cadre d'une interaction linguistique lors de laquelle un scripteur 2 a modifié une partie du message d'origine. Pour cela, ce scripteur a eu recours à trois procédés. Premier procédé, illustré par les séquences 5) et 6), il corrige lui-même l'écart à la norme en écrivant la forme juste ou en ajoutant quelques graphèmes manquants. Deuxième procédé, il barre la forme erronée, comme dans les séquences 7) et 8). Enfin, troisième procédé, il entoure l'erreur; c'est ce que l'on observe dans les séquences 9) et 10). Remarquons que ces différentes façons d'intervenir sur l'orthographe du message 1 sont combinables: dans les séquences 7) et 9), le scripteur 2 a barré et corrigé, ou entouré et corrigé le message original. Dans tous les cas, on peut s'interroger sur l'intention du scripteur 2: puisque le scripteur 1 est absent, à qui s'adresse-t-il? Peut-être à ce scripteur 1, pour lui signifier que son message est mal formulé; ou plutôt aux

¹⁰ La partie erronée des mots *dure*, *enrichie*, *remédié*, *notamment*, *assurés*, *écouté* et *manqués* est entourée au stylo bleu. La fin des mots *conseil* et *travaillés* est entourée au stylo vert. De plus, le *e* final de *dure*, le *e* final de

futurs lecteurs, parmi lesquels l'éventuel destinataire, pour leur montrer qu'il a décelé un écart à la norme?

Ce comportement du scripteur 2 amène d'ailleurs une autre interrogation: pourquoi s'écarte-t-il de l'utilisation normale du cahier (l'écriture d'un message porteur de sens) pour corriger la forme d'un message qu'un inconnu a laissé? Certainement pas pour apporter une réponse à ce message. En effet, dans la séquence 7), le scripteur 2 n'a pas le pouvoir de rendre l'entrée du musée gratuite; dans le message 9), il n'est pas qualifié pour commander un livre sur le bouddhisme; dans la séquence 10), il n'est pas en mesure de renseigner le scripteur 1, qui est d'ailleurs absent... En résumé, le message 2 n'apporte jamais la réponse que le message 1 requiert. Le but du scripteur 2 semble plutôt d'appliquer et de faire appliquer autour de lui la norme linguistique, comme un maître d'école. On lit dans le comportement du scripteur 2, toujours anonyme même quand il signe (voir la séquence 6), un besoin primordial de faire vivre la norme en elle-même et pour elle-même. Peu importe pour le scripteur 2 qu'il ne signe pas, qu'il ne laisse pas son propre message de compliment ou de réclamation: tant que la norme est respectée, il est satisfait. Le scripteur 2 s'est donné la légitimité d'intervenir au nom de la norme: il s'est autoproclamé agent de l'orthographe.¹¹

Considérons à présent le message 11), qui exprime une plainte envers la bibliothèque. Ce message recouvre une page entière d'un cahier, et il est immédiatement suivi d'une page de réponse de la part d'une bibliothécaire. On dénombre dans ce message neuf écarts à la norme. Chacun de ces écarts a reçu une réponse, et même plusieurs réponses, car on constate sur cette page que quatre stylos au moins ont été utilisés. Tout d'abord, le stylo rouge du scripteur 1, à l'origine du long message 1. Puis, un autre stylo rouge a servi au scripteur 2 à corriger quatre des neuf erreurs et à en barrer trois autres. En parallèle, un stylo bleu a été utilisé pour entourer sept erreurs, et un stylo vert pour entourer les deux restantes. Les scripteurs 2, 3 et 4 se sont véritablement acharnés sur le message d'origine, jusqu'à rendre apparentes toutes ses défaillances de forme.¹² Dans le même temps, on observe un autre comportement sur la page suivante, qui héberge la réponse de la bibliothécaire. Une phrase pose problème:

enrichie et le *s* final de *assurés* sont barrés au stylo rouge. Les mots *enrichie*, *conseil*, *remédié*, *travaillés* et *écouté* ont été corrigés; ces corrections figurent ici en caractères gras.

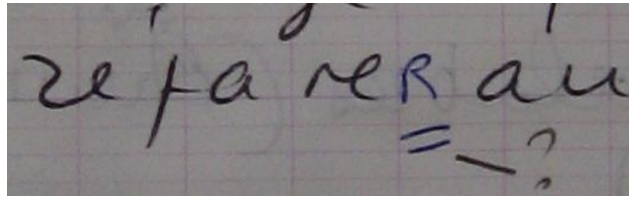
¹¹ Remarquons que personne ne peut avoir cette légitimité, puisque «la perfection en la matière serait un objectif [...] illusoire» (Chervel et Manesse, 1989: 262), c'est-à-dire que personne ne peut l'atteindre. Il est donc possible qu'un message 2 soit corrigé à son tour; c'est ce qui se produit en 12b).

¹² Le message 8) présente un cas de figure similaire: deux personnes ont barré la même erreur.

- 12a) Je serais donc curieuse de connaître les noms des professeurs sous-entendus dans votre message afin que je les contacte et répare au plus vite mon méfait.

Cette phrase est écrite au stylo noir, et un scripteur 2, armé d'un stylo bleu, est venu ajouter un *r* final souligné de deux traits au mot *répare*, pour le changer en *réparer*. Puis, un scripteur 3, au crayon de papier, a commenté cette «correction» par un point d'interrogation:

- 12b)



Il semble ici que le scripteur 2 soit venu corriger une forme qui ne devait pas l'être, et qu'il ait été rappelé à l'ordre par le scripteur 3. Cela confirme qu'une personne qui ne connaît pas bien la norme linguistique peut s'arroger le droit de corriger les autres.¹³

Une question reste en suspens: quel est le but des agents de l'orthographe, rendre service ou punir? À l'école, lorsqu'un professeur corrige une faute d'orthographe à un élève, c'est afin qu'il progresse, donc pour lui rendre service, même si cela passe par une punition symbolique (comme le retrait de points sur une note). Mais dans un cahier, alors que selon toute vraisemblance le scripteur 1 ne viendra pas relire son message après l'avoir écrit, les corrections apportées par le scripteur 2 passent pour un acte de punition, ou plus précisément de décrédibilisation du message. Nous avons vu plus haut que les scripteurs 2 n'étaient pas en mesure de répondre aux attentes exprimées dans les messages 1. Mais, par leur action, ils sont capables de désapprouver ou de décrédibiliser leur contenu. Par exemple, dans la séquence 9) où le message 1 demande «au moins 1 ouvrage sur le bouddhisme», et où le message 2 consiste seulement à entourer le *d* et à écrire la forme normée *dd* dessous, ce message 2 peut s'interpréter comme: «Ta demande d'un ouvrage sur le bouddhisme n'est pas fondée car tu ne sais même pas écrire bouddhisme». Idem dans la séquence 10) où les cercles formés autour du mot non normé signifient «Tu ne peux pas demander quelqu'un pour te renseigner car tu ne sais pas formuler ta demande correctement». Et les requêtes, en recevant une première réponse qui conteste leur formulation, s'en trouvent ébranlées.

¹³ Et cela est facilité par le fait que les scripteurs qui agissent sur le même message sont distants (Koch et Esterreicher, 2001).

Allons plus loin en supposant que corriger une forme fautive, c'est en même temps prendre parti contre son auteur. Cette éventualité s'esquisse dans la séquence 11), dans laquelle un étudiant critique fortement la bibliothèque et s'expose à la colère de son personnel. Dans cet échange, le message laissé par l'étudiant est corrigé par trois personnes. Si ces personnes avaient voulu soutenir la démarche de l'étudiant, elles auraient formulé leur soutien: «je suis d'accord», ou bien «d'autres étudiants pensent comme toi», et elles n'auraient pas cherché à fragiliser cette démarche. Mais rien de tel; la tendance générale sur cette page est à la dénonciation de fautes d'orthographe. Dans la réponse de la bibliothécaire, où un scripteur 2 s'est trompé en corrigeant un mot, le scripteur 3 vient soumettre cette correction au doute, comme pour signifier «Tu es sûr de vouloir toi aussi t'attaquer au personnel?». Les deux messages verbaux, celui de l'étudiant face à celui de la bibliothécaire, inscrivent leurs auteurs dans un conflit. L'orthographe est une arme avec laquelle des intervenants extérieurs prennent parti dans ce conflit, en minant la production écrite de l'étudiant de l'intérieur ou en défendant le message de la bibliothécaire. Et si l'on utilise l'orthographe comme grille de lecture de ce conflit, on constate que celui-ci oppose des scripteurs nantis à des scripteurs qui, non seulement ne sont pas nantis de l'orthographe, mais en plus se mettent en position de demandeurs. C'est ici que notre analyse rejoint les faits attestés par notre corpus : lorsqu'un message de requête est formulé avec des fautes d'orthographe, ce message a plus de chance qu'un autre d'être corrigé par un scripteur 2.¹⁴ La connaissance de l'orthographe est instrumentalisée par ceux qui la détiennent, si bien que celui qui ne possède pas le savoir orthographique ne peut pas prétendre au pouvoir.

Conclusion

En ce début de III^e millénaire, la linguistique voit l'un de ses objets d'étude, la langue écrite, se muter au contact de nouveaux canaux de codage, tels que les claviers d'ordinateur ou de téléphone portable. L'idée de départ dans la constitution d'un corpus de français écrit naturel est que les formes graphiques qui reflètent le mieux l'usage réel du français sont directement issues des mouvements de poignet des scripteurs, et que c'est à cette source qu'il convient de recueillir l'usage. De fait, la description de la langue manuscrite d'aujourd'hui est loin d'avoir

¹⁴ Lors de la présentation orale de ce corpus, à Bristol, en janvier 2007, Stefan Pfänder a émis l'hypothèse que l'opposition entre les messages de réclamations et les messages de louanges était équivalente à une opposition *marqué / non marqué*. D'après cette hypothèse, en tant que non marqués, les messages de louanges n'appellent pas de réponse de la part de ceux qui les reçoivent. Il en découle que les messages de réclamations engendrent plus d'interactions que la moyenne. Nous adhérons à cette hypothèse qui se place en amont de notre explication fondée sur la présence de fautes d'orthographe.

abouti. Mais construire un corpus d'écrit naturel est problématique, comme nous l'avons déjà expliqué.

Toutefois, les enseignements auxquels l'analyse d'un tel corpus nous conduit sont riches, notamment en ce qui concerne les pratiques orthographiques réelles des scripteurs. Nous constatons ainsi que loin de la norme orthographique imposée par la sphère de l'imprimé, les manuscrits constituent, comme aux siècles précédents, un vivier de variations graphiques. En pied de nez à cette norme qui passe pour homogène¹⁵ (chaque mot possède une et une seule graphie), notre corpus nous apprend qu'aujourd'hui encore un même scripteur peut utiliser plusieurs graphies pour coder un même mot, et que cela passe inaperçu aux yeux des lecteurs, tellement c'est normal. Nous défendons même l'idée que la survenance de ce qui est communément nommé une «faute d'orthographe» ne prouve pas que son auteur ignore la norme graphique. En effet, ce n'est pas l'ignorance des scripteurs que la plurigraphie généralisée reflète, mais bien plutôt les limites de l'attention du lecteur lambda. Car si personne n'a rajouté d'accents sur *telephone* dans le livre d'or, c'est peut-être parce que personne n'a remarqué leur absence.¹⁶

Ce que l'on pourrait qualifier de normal d'un point de vue linguistique dans les productions écrites recueillies, c'est la variation. Malheureusement, cette norme linguistique va à l'encontre de la norme sociale partagée, en vertu de laquelle un certain usage valorisé doit être appliqué par tous. Pour veiller à cela, les agents de l'orthographe rôdent; nous avons surpris et décrit leurs agissements. Si leur méthode d'action est simple, barrer, entourer, corriger, le but qu'ils poursuivent n'en est pas moins complexe.¹⁷ Au-delà d'un respect zélé de la norme orthographique, puissance souveraine, nous avons lu dans leur action une manière de prendre parti en défaveur du message sur lequel ils agissent. Corriger une faute d'orthographe à quelqu'un revient presque à lui signifier de se taire. Les agents de l'orthographe utilisent leur connaissance de l'orthographe pour régir les rapports sociaux; mais la violence symbolique avec laquelle ils font régner la terreur, dans les cahiers du corpus et partout en dehors, doit amener les linguistes à s'interroger et à planifier une paix orthographique. D'autant plus que la situation dans laquelle nous nous trouvons tous est absurde : car nous sommes tous, par une dualité remarquable, à la fois des agents de l'orthographe et leurs victimes...

¹⁵ Gadet (2007: 27-28)

¹⁶ Cependant, les résultats de Lucci, Millet et Billiez (1994) atténuent cette constatation. Selon eux, le degré du jugement de valeur porté sur une «faute d'orthographe» varie avec la nature de cette faute. Par exemple, les erreurs dans l'emploi des signes diacritiques sont moins lourdement connotées que les fautes dites «grammaticales».

Références

- Blanche-Benveniste C. et Jeanjean, C. (1987) *Le français parlé: transcription et édition*. Paris: Didier Érudition.
- Cerquiglini, B. (2004) *La genèse de l'orthographe française (XII^e-XVII^e siècles)*. Paris: Honoré Champion.
- Chervel, A. et Manesse, D. (1989) *La Dictée, les Français et l'orthographe 1873-1987*. Paris: INRP / Calmann-Lévy.
- Frei, H. (1929) *La Grammaire des fautes*, Paris-Genève: Sadag/Geuthner/Kundig.
- Gadet F. (2^e éd.) (2007) *La variation sociale en français*. Paris: Ophrys.
- Koch P. et Østerreicher, W. (2001) Langage oral et langage écrit . In: *Lexikon des Romanistischen Linguistik*. Tübingen: Max Niemayer Verlag, 584-627.
- Lopez-Bouchet, F. (2004) Inscriptions graffitées: approches (socio)linguistiques interactionnelles. Thèse de doctorat non publiée: Université Stendhal (Grenoble III).
- Lucci, V., Millet, A. et Billiez J. (1994) *L'orthographe de tous les jours, Enquête sur les pratiques orthographiques de tous les jours*. Paris: Honoré Champion.
- Manesse D. et Cogis, D. (2007) *Orthographe: à qui la faute?* Paris: ESF Éditeur.
- Portelance C. et al. (consulté le 27 août 2007) *Un corpus du Témiscouata (1930-1936), Des lettres des pays d'en bas*, <<http://www.ling.uqam.ca/forum/corpus/messages/3.html>>.

Camille Martinez

Lexiques, Dictionnaires, Informatique

UMR 7187 du CNRS, Université de Cergy-Pontoise

¹⁷ D'autant qu'ils n'interviennent pas systématiquement : nous ne sommes pas en mesure de comprendre pourquoi des messages non normés restent vierges de toute correction.