

Typologies de variation graphique dans l'écrit sms

Louise-Amélie Cougnon, Sophie Roekhaut, Richard Beaufort
Cental, Institut Langage et Communication, Université catholique de Louvain

Introduction

L'évolution de l'orthographe s'est vue récemment bouleversée par une nouvelle pratique : la *communication écrite médiée par ordinateur* (ou *CéMO* : Cougnon et François, 2011). En effet, depuis une dizaine d'années, l'utilisation du courrier électronique, de la messagerie instantanée et de l'*sms* – pour ne citer qu'eux – s'est intensifiée de manière spectaculaire au point de devenir, au sein des tranches d'âge les plus jeunes, la pratique principale – et soutenue – de l'écrit. Or, une des propriétés essentielles de cette CéMO est son altération de la graphie. Pour répondre à des besoins d'économie de cout, d'espace et de temps, les usagers de la CéMO montrent une tendance manifeste à l'abréviation.

Le sujet de l'orthographe dans ces nouvelles formes de communication, et en particulier dans les sms, est délicat. Il suscite autant l'excitation, l'engouement (principalement des utilisateurs), que la crainte, le scepticisme (p.ex. chez les professeurs ou auprès des parents), ou encore des excès de prohibition (p.ex. sur les forums de discussion¹). C'est pourquoi il est nécessaire d'obtenir des données empiriques qui viendront confirmer ou infirmer les opinions variées.

Un article portant sur les sms peut paraître inadéquat au moment de célébrer les 20 ans des réformes de l'orthographe, en ce que l'on aurait tendance à confondre *Rectifications orthographiques* (dorénavant *Rectifications*, pensées, justifiées, argumentées) et liberté orthographique autoproclamée par les usagers. Or, il est évidemment incorrect d'assimiler les deux phénomènes. C'est pourquoi notre propos fera précisément la part des choses. L'objectif de cet article est de présenter les variations graphiques présentes dans l'*écrit sms* (eSMS : Cougnon et François 2011, ou *écriture sms*, Panckhurst 2010).

Nous présentons d'abord sommairement (1) les corpus à la base de notre recherche et la méthode globale de collecte. Nous détaillons ensuite une part importante des variations graphiques propres à la CéMO : nous commençons par établir (2.1) une typologie des stratégies utilisées dans les sms comme procédés de simplification de l'encodage ; nous présentons également (2.2) une typologie des erreurs orthographiques dans les sms, typologie dont sont exclues les simplifications présentées en (1) et basée sur la grille typologique dressée par Catach (1980) ; enfin, nous abordons (2.3) la question des Rectifications en examinant de plus près les points semblables – entre les règles régissant l'écrit sms et les principales rectifications orthographiques adoptées en 1990.

1. Un corpus de sms pour la linguistique

1.1 Le corpus francophone de 60.000 sms

Le corpus de sms francophones sur lequel est basée notre étude recouvre cinq zones² géographiques : la Belgique francophone (30.000 sms), La Réunion (12.000 sms), la Suisse (4.000 sms), le Québec (5.000 sms), et les Hautes-Alpes et l'Isère (9.000 sms).

Ces corpus de sms ont été assemblés dans le cadre d'un projet plus vaste de collecte de sms plurilingue nommé *sms4science*³. Ce projet a été fondé en 2007 par le Cental (Centre de traitement automatique du langage), à l'Université catholique de Louvain (Belgique), sous l'impulsion d'un premier projet-pilote belge organisé en 2004. Depuis, des chercheurs de plusieurs pays, isolés ou en équipe, s'associent afin de constituer rigoureusement de vastes corpus de sms plurilingues et de profils

¹ On voit très souvent notée, sur la page d'accueil du forum ou au sein de son règlement interne, une indication de type « Langage sms interdit ». Notons, paradoxalement, par ce type de signalisation au sein des forums, qu'il ne s'agit plus d'un langage propre au médium sms.

² Nous appelons « zone » un espace géographique qui ne coïncide pas toujours avec les frontières territoriales d'un pays.

³ Pour plus d'information, visiter : www.sms4science.org.

d'usagers. Cette constitution de corpus a pour ambition d'étudier à la fois les messages eux-mêmes, les propriétés extralinguistiques de ce type de communication, mais également l'identité et les pratiques de leur auteur.

À l'heure actuelle, le consortium international compte 11 partenaires officiels⁴. Quant aux corpus, ils recouvrent à ce jour 130.000 sms plurilingues et 5.000 profils sociolinguistiques.

1.2 Méthodologie de collecte

Le but du projet *sms4science* peut se résumer en quatre points majeurs : obtenir des données (1) « véritables » (c'est-à-dire des copies de messages échangés entre usagers du sms et non rédigés expressément pour l'enquête) et (2) « objectives » en obtenant des corpus comparables (issus d'une méthodologie identique), (3) « de façon automatique », en utilisant l'« informatique » pour la collecte, le stockage et le recopiage des données. Ces trois premiers points amènent à une méthodologie dont le but principal n'est pas, de façon idéaliste, de supprimer les biais, mais de *tendre* vers des données (4) « non biaisées », autant que possible.

Cette méthodologie et le traitement des données permettent de tendre vers une certaine représentativité de la population et des phénomènes observés. Les études à partir de ces corpus sont rendues possibles grâce à la taille des corpus récoltés, mais également grâce au respect des aspects légaux (p.ex. par l'anonymisation des données). Enfin, la transcription en graphie standard, encouragée dans chaque zone de collecte, permet des recherches lexicales et syntaxiques approfondies.

2. De la variation graphique dans les sms

2.1 Typologie des stratégies de simplification linguistique dans l'écrit sms

Une analyse superficielle de l'eSMS laisserait volontiers croire que cette pratique est truffée d'erreurs orthographiques et syntaxiques. Nous proposons de regarder de plus près le message suivant :

- (1) Il voulu se venG psk en cour la class fé tro d brui!é il a trouV un moyen..!mé bon c fé c fé...2m1 en fizik on sassoir tjr ensembl ou koi? [smsFF]⁵

À la vue d'un tel message, qui contient de très nombreuses abréviations en tout genre, il est légitime de se demander de quelle manière distinguer ce qui est méconnaissance des règles et ce qui est stratégie de communication. Afin d'éclaircir ce point, nous avons entrepris de répertorier clairement les phénomènes graphiques employés dans l'eSMS et qui ressortent de stratégies réductionnelles. Une telle typologie permet, *a minima*, d'identifier les productions qui n'appartiennent pas à ces phénomènes, et de les rapprocher d'une maîtrise moins effective des règles linguistiques traditionnelles (cf. 2.2).

La plupart des stratégies utilisées entraînent une abréviation de la forme standard :

1. L'apocope : on y retrouve l'apocope simple (*univ* pour *université*), l'abréviation sémantisée⁶ (la lettre *t* pour *tu* ou *te*), l'abréviation suivie d'un point (*auj.* pour *aujourd'hui*), la neutralisation des

⁴ Belgique, Chypre, Espagne, Grèce, Italie, France, Canada, Roumanie, Royaume-Uni, Allemagne et Suisse.

⁵ Exemple provenant du corpus de sms francophones de l'île de la Réunion. La transcription des exemples de sms se trouve en annexe.

⁶ C'est-à-dire réduite à l'initiale (Panckhurst, 2008).

graphèmes muets en finale absolue (*ils port* pour *ils portent*)⁷, ainsi que les sigles et acronymes (*mdr* pour *mort de rire*, *asv* pour *âge sexe ville*, etc.) ;

2. L'aphérèse (*lut* pour *salut*)⁸ ;

3. La syncope : on y retrouve la syncope simple (*aple* pour *appelle*), l'effacement du schwa à l'intérieur du mot (*ramner* pour *ramener*), la réduction de certains digrammes ou trigrammes (*sel* pour *seul*) ainsi que les squelettes consonantiques⁹ (*prbl* pour *problème*) ;

4. La phonétisation¹⁰ est également un phénomène récurrent : on recense la phonétisation par la lettre (*r* pour *air*), la phonétisation par le chiffre ou par le nombre (*2mande* pour *demande*, *10cute* pour *discute*), la phonétisation par le signe (*pl@* pour *plate*) et la phonétisation par des graphies supposées plus proches du phonétisme réel, plus proche d'un accent régional ou d'un « style sms » (*koi* pour *quoi*) ;

5. La casse fonctionnelle : contrairement aux conventions traditionnelles associées à l'usage de la casse (début de phrase, distinction adjectif/substantif, unicité/pluralité, noms propres, etc.), la majuscule revêt d'autres fonctions dans les sms. Retenons les deux fonctions majeures : (1) le marquage d'une séparation entre des mots agglutinés (fonction typographique) et la prosodie¹¹ (fonction expressive) ;

6. La chute massive de l'accentuation, qu'elle permette de distinguer des homographes ou non. Celle-ci ne participe pas à une abréviation du message dans le sens d'un raccourcissement de ce dernier, mais bien à une simplification d'encodage (les lettres accentuées nécessitant un plus grand nombre de pressions dans la méthode de saisie multi-pressions).

Afin de clore ce résumé des phénomènes graphiques propres aux sms, notons qu'il existe des lieux de résistance à l'abréviation. Ainsi, nous retiendrons toute une série de digrammes et de trigrammes qui sont très majoritairement maintenus. Il semblerait par exemple qu'on remplace rarement un *th* (comme dans *théâtre* ou *math*) par un *t* ou un *ph* par un *f*¹², bien que ces pratiques faciliteraient l'encodage de caractères (Fairon et al., 2006). De la même façon, il y a une tendance manifeste à garder les consonnes doubles, surtout lorsque leur simplification entraînerait une altération de la valeur phonique.

Ce bref panorama des tendances graphiques dans les sms a permis de montrer qu'il n'existait pas réellement de « cacophonie orthographique », mais bien des règles propres à la CéMO, conscientes ou inconscientes, applicables ou non.

(2) Salut {Nom}. Je n'ai toujours pas reçu ta lettre... Ce n'est pas normal ! Es ce que tu m'as envoyer la lettre ?? [smsFB¹³]

Dans un message tel que (2), il est dès lors possible d'interpréter *envoyer* comme une erreur orthographique, étant donné qu'il n'entre dans aucune des catégories énumérées et ne participe pas à une abréviation du message.

2.2 Typologie des erreurs orthographiques de l'écrit sms

Suite au constat de phénomènes tels qu'illustrés par (2), pour lesquels la variation graphique n'abrège pas le message et ne simplifie pas l'encodage de caractères, nous avons entrepris d'examiner la présence d'erreurs orthographiques dans l'écrit sms. Pour étudier ces erreurs, nous avons choisi

⁷ Un sous-type qui entraîne conséquemment la suppression des marques de flexion et qui permet plus difficilement la distinction entre stratégie d'abréviation et faute d'orthographe.

⁸ Un phénomène qui marque plus souvent un effet de style qu'une volonté réelle d'abrégé.

⁹ Il s'agit de l'abréviation d'un mot quasiment charpenté autour de ses consonnes (Prochasson et al., 2007) ; ce type d'abréviation met en avant la plus forte valeur informative des consonnes.

¹⁰ Anis (2002 : 60) la nomme « graphie à connotation oralisante ».

¹¹ L'étude des phénomènes de l'accentuation et de l'intonation (variations de hauteur, de durée et d'intensité).

¹² Le corpus belge montre en moyenne 10% de telles pratiques rédactionnelles.

¹³ Corpus de sms francophones provenant de Belgique (francophone).

comme corpus échantillon le corpus belge de 30.000 messages collecté en 2004, le seul à posséder une transcription au mot près, nécessaire à l'analyse. Nous avons exclu de l'analyse les formes répondant aux stratégies linguistiques présentées au point précédent, ne gardant que les formes dont la version en brut avait un nombre de caractères égal ou supérieur à la forme standard. Lorsqu'il y avait hésitation sur la nature de la variation (stratégie ou erreur orthographique), nous avons fait le choix méthodologique d'exclure la forme¹⁴. Parmi ces formes, nous avons également écarté les variations d'accentuation, étant donné qu'elles peuvent également participer aux stratégies de simplification rédactionnelles (telles qu'énoncées au point 2.1 : *chute massive de l'accentuation*), les formes plus longues qui répondent à une fonction expressive de l'eSMS, ainsi que les erreurs liées à l'utilisation du clavier (confusion entre les lettres correspondant à une même touche) ou du dictionnaire T9¹⁵.

Ces choix méthodologiques engendrent des précautions que nous devons signaler avant la lecture même des résultats : d'abord, il ne s'agira pas ici de connaître le taux d'erreurs par rapport au taux de formes correctement orthographiées mais de présenter la distribution des types d'erreurs au sein de l'ensemble des variations qui ne répondent pas à des stratégies linguistiques. Ensuite, le rejet de certaines formes entraîne un biais que nous pensons important dans nos résultats, tel que la confusion *a/à* (code 420), certainement la plus récurrente dans l'ensemble des homophones grammaticaux ; pour la même raison encore, le code 2000 (omission ou adjonction de voyelle, de digramme vocalique et d'accent) voit son nombre d'occurrences terriblement chuter. Pourtant ce choix de rejet se justifie : dans un très grand nombre de cas, la chute de l'accent a effectivement lieu, qu'elle soit le résultat d'une stratégie d'abréviation ou d'une déformation du dictionnaire d'aide à la rédaction (cf. 2.3)¹⁶.

La distribution des erreurs (Figure 1) montre une nette prédominance des erreurs de nature morphogrammique.

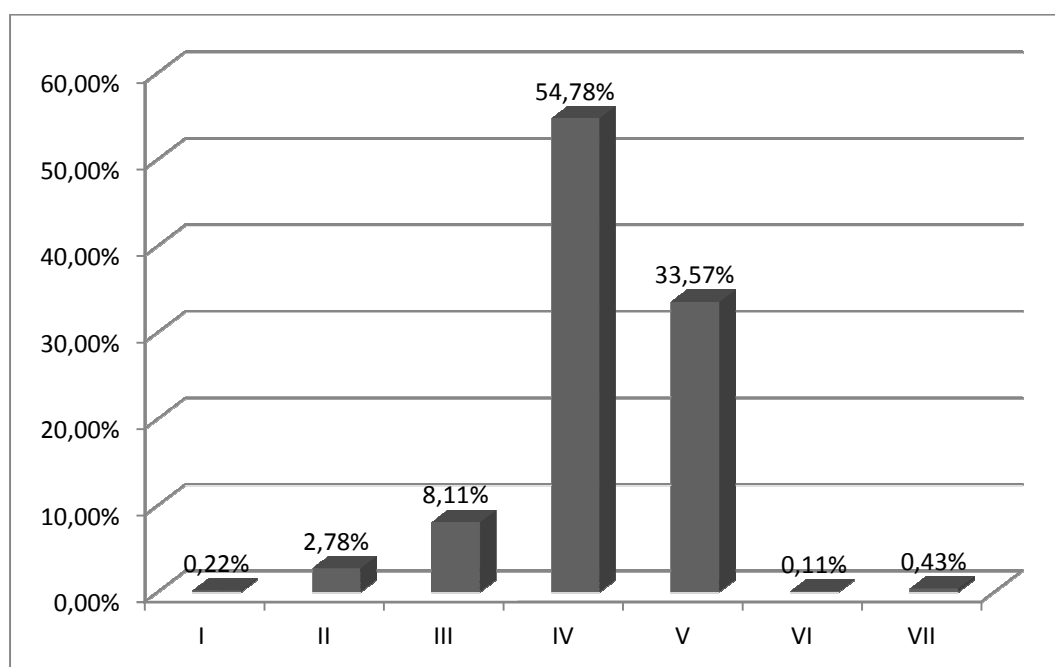


Figure 1. Distribution des erreurs orthographiques de l'écrit sms suivant la classification de Catach (1980)

¹⁴ Il est, dans la majeure partie de ces cas, impossible de trancher quant à la nature de la variation.

¹⁵ Un système mis sur le marché par la société Tegic Communications.

¹⁶ On évite donc d'identifier comme faute des variations de type (3) *T amenés la twingo vers 19h30...* [smsFB] où *amenés* est clairement une proposition du T9 lorsque l'auteur du message voulait encoder *amènes*.

La typologie (Tableau 2) précède quelques commentaires explicatifs. Chaque catégorie est suivie de son nombre d'occurrences, entre parenthèses, et est illustrée d'exemples non limitatifs.

I. Erreurs à dominante phonétique (6)				
1101¹⁷ Confusion t-d (3) <i>descendre/descentre</i> (4) ¹⁸ <i>Tu sais vider la bassine qui est dans ma chambre, la descentre et mettre l'essuie à lavé ?</i>		1102 Confusion f-v (1) <i>ferai/verai</i> (5) <i>haa mechan lou epargne mwa je te verai ce tu voudra</i>		1103 Autres confusions (2) <i>fatigué/fatiqué</i> (6) <i>J ai eu une nuit hyper courte et je sui fatiqué</i>
II. Erreurs à dominante phonogrammique altérant la valeur phonique (77)				
	Voyelles/digrammes voc.	Semi-voyelles	Consonnes/digrammes cons.	Consonnes simples/doubles
Omission ou adjonction	2000 (4) <i>crevée/creuvée</i> (7) <i>y a peu d chances d m voir tonight car j suis creuvée</i>	2001 (6) <i>filé/fillé</i> (8) <i>Salut c {Nom} ben vla {Nom} ma fillé ton num et cest pr voir si cest b1twa!</i>	2002 (1) <i>enregistré/enregistré</i> (9) <i>Oui un trop marrant que g enregistré</i>	2003 (27) <i>appeler/appeller</i> (10) <i>Tu peux m'appeller ou m'envoyer un sms</i>
Confusion	2010 (16) <i>envoies/envoyes</i> (11) <i>tu m'envoyes chier</i>	2011 (4) <i>ennuyer/ennuyer</i> (12) <i>Désolée de t'ennuyer avec ça!</i>	2012 (17) <i>excuse/escuse</i> (13) <i>Bb je m escuse de te tel si longtenp</i>	2013 (2) <i>dilemme/dilemne</i> (14) <i>Ou se situer face à ce dilemne ?</i>
III. Erreurs à dominante phonogrammique n'altérant pas la valeur phonique (225)				
	Voyelles/digrammes voc.	Semi-voyelles	Consonnes/digrammes cons.	Consonnes simples/doubles
Omission ou adjonction	2100 (1) <i>garantis/guarantis</i> (15) <i>j'guarantis pas mon ambiance!</i>		2102 (8) <i>absence/absceance</i> (16) <i>Tout était bien sauf ton absceance</i>	2103 (4) <i>ensemble/enssembl e</i> (17) <i>jme disai kn pourrait se mettre tt enssemble</i> 21030 (67) <i>acompte/accompte</i> (18) <i>Serait-il possible de ravoit un accompte lundi?</i>
Confusion	2110 (85) <i>commence/commance</i> (19) <i>ça commance dans 4 min</i>	2111 (1) <i>pièce/pyece</i> (20) <i>y avai 1 pyece kole</i>	2112 (59)¹⁹ <i>concentrer/consentrer</i> (21) <i>tu veu bien me laisser me consentrer</i>	

¹⁷ Codes introduits par Catach (1980).

¹⁸ Numéro du message. Les messages se trouvent tous transcrits en graphie standard en pièce jointe.

¹⁹ À cette catégorie appartiennent les nombreux cas de phonétisation par des graphies supposées plus proches d'un « style sms » (cf. 2.1) tels que tous sons /k/ orthographiés uniformément k- ou encore tous sons /z/ orthographiés uniformément z-. Nous avons choisi de les inclure dans la catégorie tout en envisageant dans certains cas qu'il s'agisse uniquement d'un jeu sur la langue.

IV. Erreurs à dominante morphogrammique (1519)
Morphèmes grammaticaux (1470)

Relation mal établie entre :

3000 les catégories grammaticales (4) <i>bienvenue/bienvenues</i> (22) <i>Bienvenues à vous 3</i>	3002 adjectif et nom (32) <i>beau/beaux</i> (23) <i>Merci Lol il est trop beaux</i>	
3001 les formes du pluriel (7) <i>bleus/bleux</i> (24) <i>tu as les cheveux bleux</i>	3003 sujet-verbe (865) <i>aimerais/aimerait</i> (25) <i>j'aimerait justement savoir!</i>	30040 avec être (24) <i>né/née</i> (26) <i>Le petit {Nom, masc} est née ce 22.09.04</i>
3010 nom et déterminant (29) <i>classe/classes</i> (27) <i>c toute des bombes dan÷ ta classes</i>	30120 infinitif/participe (204) <i>assumé/assumer</i> (28) <i>Heyyy catiiiiiin! T'as pas assumer toi aujourd'hui!</i> 30121 infinitif/indicatif (40) <i>embêtez/embêter</i> (29) <i>Si vous vous embeter vous pouvez nous rejoindre à la doud</i> 30122 indicatif/conditionnel (112) <i>demanderais/demanderais</i> (30) <i>ok ma mimi que j'aime je lui demanderais</i>	30041 avec avoir sans C.O.D (27) <i>décidé/décidés</i> (31) <i>Mes parents ont encore décidés de m'em...</i>
3011 nom, adj + cplmt (2) <i>bonheur/bonheurs</i> (33) <i>je vs souhaite bcp d bonheurs</i>	30123 autres (77) indicatif/impératif (34) <i>amuses-toi bien ce soir</i> indicatif/subjonctif (35) <i>c ce que je t ai dis hier je n aie po les notes</i> indicatif/participe (36) <i>je suis degoutai</i>	3014 Temps (1) <i>dirai/disais</i> (37) <i>il voudrait que vous soyez là... Je te disais quel week end</i>
Morphèmes lexicaux (49)		
	313 Ignorance du maintien ou non du radical (1) <i>fatigant/ fatiguant</i> (38) <i>tu crois que je vais monter pour te dire bonjour? Trop fatiguant ;o)</i>	314 Ignorance des lettres finales justifiables (48) <i>fois/foix</i> (39) <i>on les regarderai une foix a 2 d ak ?</i>

V. Erreurs concernant les homophones (931)	
40 Homophones de discours (6) <i>tes/t'es</i> (40) <i>T'es messages ne sont pas tjs les plus clairs au monde</i> 41 Homophones lexicaux (38) <i>court/cours</i> (41) <i>Elle l'est, mais un cours instant!</i>	Homophones grammaticaux (912) 422 ce/se (45) (42) <i>je te phone ce soir et d2m1 pour qu on ce voit</i> 423 ces/ses (6) (43) <i>T as d ses idée choux jte jure...</i> 425 on/ont (8) (44) <i>ont est tj au resto</i> 428 autres (828) <i>sa/ça</i> (45) <i>Bien dormi? Moi comme un prince sans ça princesse</i>

VI. Erreurs concernant les idéogrammes (3)		
501 Ajout d’apostrophe (2) <i>dorénavant/d’orénavant</i> (46) <i>D’orénavant tu sais quoi</i>	502 Ajout de trait d’union (1) télécommande/télé-commande (47) <i>J’ai une nouvelle télé-commande</i>	
VII. Erreurs concernant les lettres non fonctionnelles (12)		
		614 Ajout de finales particulières (12) <i>cauchemar/cauchemard</i> (48) <i>je nage de nouveau en plein cauchemard</i>

Tableau 1. Typologie des erreurs orthographiques (inspirée par Catach : 1980)

Remarques générales

- (1) Le tableau originel proposé par Catach (1980) n’a pas pu être conservé tel quel, notamment parce qu’il nous paraît y avoir confusion entre l’intention du locuteur (ex. code 313 : *Ignorance du maintien ou non du radical*) et le résultat graphique (ex. code 210 : *Omission de voyelle*). Ce genre de confusion a rendu l’attribution des catégories difficile. Ensuite, même si le tableau est uniquement basé sur les intentions des locuteurs, il semble très difficile d’interpréter ces intentions. Prenons un de nos exemples : le substantif *chien* écrit *chient*. Le locuteur a-t-il souffert d’un code 314 : *Ignorance des lettres finales justifiables* ou a-t-il simplement confondu ponctuellement (code 2110) les digrammes vocaliques ? Bien sur, Catach (1980) propose de croiser les critères et d’apposer plusieurs étiquettes à une même erreur. Mais cela n’aurait-il pas été plus simple de créer une grille d’analyse uniquement basée sur le résultat graphique, de sorte que chaque erreur n’appartienne qu’à une seule catégorie et ne soit pas sujette à une interprétation subjective ? Dans le cas de l’analyse des sms, cette approche nous paraît plus appropriée.
- (2) Certaines catégories proposées par Catach (1980) ne sont pas représentées dans notre corpus d’écrit sms. Plusieurs explications se profilent ; notamment :
 - a. Certaines catégories semblent redondantes, par ex. les codes 3004 et 3002 pour *l’accord du participe passé et du nom sans auxiliaire*. Pourquoi le code 3004 ne représente-t-il pas la catégorie des participes passés en général ? Face à ce questionnement, nous n’avons rien étiqueté à l’aide du code 3004.
 - b. Certaines catégories ont été évitées pour contourner la subjectivité. Ainsi le code 312 : *Ignorance des pré- et suf-fixes* qui pouvait à chaque fois être interprété comme un code 2110 : *Confusion digramme vocalique*.
 - c. Certaines catégories concernent clairement des variations abrégeant/simplifiant la forme standard (ex. code 610 : *Omission des consonnes grecques*) et ont donc été éliminées de l’étude avant même d’entamer l’analyse selon la grille de Catach (1980).
 - d. Certaines catégories participent du mouvement plus général de chute massive de l’accentuation et de mutation fonctionnelle de la casse (ex. code 424 : *ou/où*) et ont donc été, comme c., éloignées de l’analyse dès ses prémices.
 - e. Nous ne traitons pas des erreurs de ponctuation (503 et variantes) ni des erreurs de casse (500 et variantes) car l’eSMS ne répond pas à l’usage standard de la ponctuation et de la casse.
- (3) L’analyse des erreurs a fait également apparaître des catégories absentes de la grille telle que la variation dans les expressions figées, que nous n’avons pas pu traiter car ce cas n’était pas repris dans la grille d’analyse. Par ex. *de nouveaux* pour *de nouveau*. Ces formes ont donc été écartées de l’analyse.
- (4) Les catégories d’erreurs orthographiques les plus présentes dans le corpus sont les erreurs à dominante morphogrammique et les erreurs d’homophones. Les erreurs rencontrées touchent donc moins les variations entraînant une altération phonique.

2.3 Rectifications orthographiques et sms

Si l'écrit sms est pertinent à comparer aux Rectifications, c'est parce qu'il se présente comme un lieu de communication écrite spontanée (Cougnon, 2008) où se trouvent des tendances naturelles inhérentes aux auteurs des messages. Les résultats présentés dans cette partie valent pour tous les messages de notre corpus, quelle que soit la méthode d'encodage des caractères, c'est-à-dire qu'il s'agisse de la technique à pressions multiples ou de la technique par saisie prédictive (qui passe par un dictionnaire du téléphone)²⁰. Or, la saisie prédictive, telle que proposée par le système répandu T9, impose un certain nombre de modifications graphiques telles que la syncope de l'accent circonflexe. Les résultats ainsi obtenus ne doivent pas être interprétés comme une volonté personnelle de l'utilisateur de modifier le mot, mais comme une modification imposée par le système.

Afin de regarder de plus près les convergences entre phénomènes du sms et Rectifications (cf 2.3), nous avons suivi une méthodologie pluridimensionnelle, schématisée par la Figure 2.

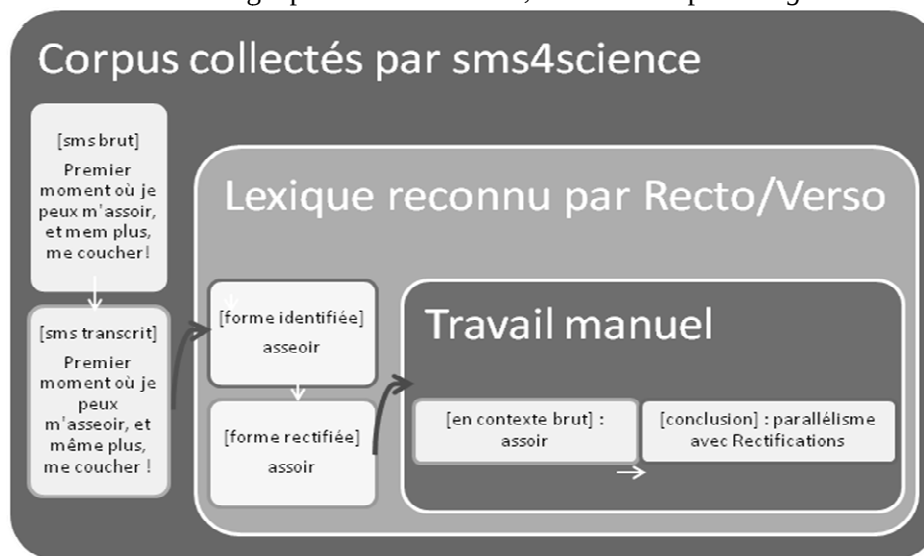


Figure 2. Méthodologie de comparaison eSMS/Rectifications

Tout d'abord, nous avons entrepris d'appliquer l'outil de rectification automatique Recto/Verso, développé par le Cental, aux corpus de sms. Recto/Verso a pour but d'introduire automatiquement les Rectifications de 1990 dans un texte (Beaufort et al., 2009). La tâche de rectification automatique nécessite donc d'analyser le texte et d'identifier des mots concernés par la réforme. Or, dans un message de type sms, les formes concernées sont difficilement identifiables, puisqu'elles sont tronquées par les usagers (phonétisation, agglutination, apocope, etc.). Heureusement, le projet *sms4science* comprend, comme nous l'avons dit précédemment, une étape de transcription (manuelle ou automatique²¹) des messages en graphie standard. L'outil Recto/Verso est donc appliqué à un ensemble de messages dont la graphie a été normalisée. Finalement, l'outil permet d'identifier les formes lexicales concernées par les Rectifications, qu'il faut ensuite manuellement analyser afin d'identifier l'appartenance (ancienne/nouvelle orthographe) de la forme originelle²².

²⁰ Il semblerait toutefois que la majorité des usagers du sms n'emploie pas le dictionnaire d'aide à la rédaction de leur téléphone portable ; en effet, les données sociolinguistiques concernant les personnes enquêtées montrent que 66,59 % des personnes enquêtées n'utilisent pas le dictionnaire d'aide à la rédaction ; si on limite l'analyse à l'Ile de la Réunion, ce chiffre atteint même les 83,45 %.

²¹ La « transcription manuelle » est réalisée par les chercheurs eux-mêmes ou par des étudiants, tandis que la « transcription automatique » est exécutée grâce à une méthode hybride de normalisation des sms détaillée par Beaufort et al. (2010). Dans les deux cas, la transcription ne suit pas les Rectifications.

²² Notons que ce travail manuel n'est pas un travail fastidieux puisque, comme l'explique Villers (1998), seuls un peu plus de 2.000 mots sont concernés par ces *Rectifications*, c'est-à-dire environ 5% de la nomenclature des dictionnaires de grande diffusion. Nous constatons également qu'il ne s'agit pas forcément d'un lexique d'usage courant.

L'analyse réalisée sur nos cinq corpus montre deux phénomènes récurrents cités dans les Rectifications : la syncope de l'accent circonflexe sur *i* et *u* et la soudure des mots composés avec trait d'union. Ces deux phénomènes répondent parfaitement à la logique de simplification de l'eSMS, telle qu'elle a été présentée précédemment : faire court et encoder facilement, tout en restant compréhensible.

La syncope de l'accent circonflexe, phénomène le plus largement représenté dans nos corpus, répond à une facilité d'encodage (les lettres accentuées²³, comme nous l'avons dit, demandent plus de pressions dans la méthode multi-pressions que les lettres non accentuées) ainsi qu'à un gain d'espace, étant donné qu'une lettre accentuée par un accent circonflexe coûte de l'espace à l'utilisateur du sms. D'un point de vue technique, l'utilisation de caractères spéciaux (accents circonflexes, cédilles, etc.) force un encodage sur 16 bits, ce qui, sur de nombreux téléphones, réduit à 70 caractères la taille d'un sms simple au lieu de 160 (encodage sur 7 bits). La syncope de l'accent circonflexe serait donc, ici encore, motivée pour des raisons techniques (économie d'espace) et ne serait donc pas le miroir de pratiques linguistiques réelles. La syncope de l'accent apparaît dans 99 % des cas pour *û* et 97 % pour *î*. Les unités lexicales les plus concernées sont *sûrement*, *boîte*, *plaît*, *paraître* et *dîner*. Retenons toutefois que la syncope de l'accent circonflexe ne se cantonne pas aux seules lettres *i* et *u* et qu'elle se manifeste quelle que soit la lettre qu'elle accentue.

La soudure des mots composés avec trait d'union permet aussi un gain en termes de caractères : l'agglutination à la place du tiret abrège le message d'un caractère. Dans le corpus, les unités lexicales les plus concernées par ce phénomène sont *week-end*, *pique-nique*, *porte-monnaie* et *strip-tease*. Si l'on prend le mot *week-end* seul, le phénomène est plus qu'intéressant. En effet, bien que l'on s'attende à voir apparaître la forme abrégée *w-e* qui permet un gain d'espace considérable, 62 %²⁴ des sms analysés comportent la forme non abrégée, dont non moins de 40 % connaissent la soudure en *weekend*.

Concernant les anomalies diverses énumérées dans les Rectifications, peu de mots sont réellement concernés dans les sms, puisque non utilisés. L'infinitif *asseoir* fait exception à la règle. La nouvelle orthographe du verbe permet le gain d'un caractère. On note que la moitié des occurrences observées dans le corpus sont effectivement écrites *asseoir*, tel que dans l'exemple (49).

(49) elle doit être chez le veto là. quand {Nom} l'a posée chez {Nom} elle s'est mise à couiner et à ne plus vouloir lacher {Seb} et là elle n'arrivait plus à s'**asseoir** [smsFF]

Un nombre important de formes concernées par les Rectifications montrent, dans le contexte de l'eSMS, une adaptation contextuelle qui ne permet pas de les rapprocher de ces Rectifications. C'est le cas de la règle du tréma, dans les formes finales en *-gue(s)*, se mettant dorénavant sur le *-u*. En contexte sms, on aura, comme pour l'accent circonflexe, une simple syncope du tréma. De la même façon, la règle du trait d'union imposé entre tous les numéraux composés exprimant un nombre entier n'apparaît pas dans le corpus car les nombres sont majoritairement représentés en chiffres arabes, ce qui permet une économie de caractères plus importante.

²³ On notera qu'une exception dans ce domaine est le recours au « é universel » ; dans ce cas les auteurs introduisent une difficulté d'encodage (le *e* accentué) pour abréger les *ai*, *ez*, *er*, etc.

²⁴ Calcul basé sur le corpus belge uniquement.

Conclusions

Notre étude, basée sur des données empiriques, confirme l'apparition d'une nouvelle forme de l'écrit, la CéMO, qui regorge de jeux linguistiques visant principalement à abréger le message. Ces jeux suivent des règles précises que les usagers décident de suivre ou non, consciemment ou non. Certes, cette volonté d'abréger l'objet de communication montre des points communs avec les grands principes des Rectifications, mais ce n'est pas seulement – sinon pas du tout – en vue de suivre la loi du moindre effort, sinon d'optimiser l'acte de communication. Les données résultant de notre étude confirment également le maintien de certaines conventions orthographiques surtout lorsque celles-ci entraînent une altération de la valeur phonique : les catégories I et II représentent moins de 10% des erreurs totales.

En outre, nos données attestent qu'il y a effectivement un panel diversifié de fautes d'orthographe dans l'eSMS, là où certains auraient pensé que les confusions portaient principalement sur les modes et les temps. Il serait intéressant de mettre en perspective ces résultats en les confrontant à la répartition des erreurs orthographiques dans d'autres corpus d'écrits spontanés ou d'écrits plus normatifs. L'eSMS entraîne-t-il l'apparition de nouvelles erreurs ? L'utilisation de stratégies abrégées se retrouve-t-elle davantage dans la rédaction de textes formels qu'auparavant ?

Selon l'étude de Lunsford, citée dans Tatossian (2011), basée sur un corpus de 15 000 documents de 189 étudiants, « les nouvelles technologies n'ont pas d'impact négatif sur les habiletés à rédiger [...] la rédaction de textos ne nuit pas à la rédaction de textes formels. En effet, l'auteure n'a pas trouvé un seul procédé graphique associé au texto dans les travaux écrits des étudiants, ce qui montre que les jeunes adaptent l'écriture à la situation de communication ».

Il y aurait donc, comme l'ont avancé David et Goncalves (2007), un véritable phénomène de *digraphie*, c'est-à-dire une différenciation des registres orthographiques selon les contextes et les interlocuteurs. En conséquence, nous ne serions pas en présence d'une perte de compétence, mais bien d'une double compétence.

Références

- Anis, Jacques (2002). « Communication électronique scripturale et formes langagières : chats et sms ». In *Actes des Quatrièmes Rencontres Réseaux Humains/ Réseaux technologiques*, Université de Poitiers, Poitiers.
- Beaufort, Richard, Roekhaut, Sophie, Cougnon, Louise-Amélie et Fairon, Cédric (2010). « Une approche hybride traduction/correction pour la normalisation des SMS ». In *Proceedings of TALN 2010*.
- Beaufort, Richard, Dister, Anne, Naets, Hubert, Mace, Kévin et Fairon, Cédric (2009). « Recto/Verso. Un système de conversion automatique ancienne/nouvelle orthographe à visée linguistique et didactique ». In *Proceedings of TALN 2009*.
- Louise-Amélie Cougnon and Thomas François (2011). « Étudier l'écrit SMS. Un objectif du projet sms4science ». In Adrian Stähli / Christa Dürscheid / Marie-José Béguelin (eds.) (2011), *La communication par SMS en Suisse. Usages et variétés linguistiques*. Linguistik Online (Themenheft). David, Jacques et Goncalves, Harmony (2007). « L'écriture électronique, une menace pour la maîtrise de la langue ? », *Le français aujourd'hui*, n° 156, pp. 39-48.
- Nina Catach (1980). *L'Orthographe française* (en collaboration avec Claude Gruaz et Daniel Duprez), Coll. fac, Nathan Université.
- Cougnon, Louise-Amélie (2008). "Le français de Belgique dans l'écrit spontané'. Approche d'un corpus de 30.000 SMS". In *Travaux du Cercle Belge de Linguistique*.
- Cougnon, Louise-Amélie et François, Thomas (2010). "Quelques contributions des statistiques à l'analyse sociolinguistique d'un corpus de SMS". In Sergio Bolasco, Isabella Chiari and Luca Giuliano (éds). *Statistical Analysis of Textual Data. Proceedings of 10th International Conference JADT*, 9-11 juin 2010, Sapienza University of Rome, volume 1, pp. 619-630.
- Fairon, Cédric, Klein, Jean et Paumier, Sébastien (2006). *Le langage SMS. Étude d'un corpus informatisé à partir de l'enquête « Faites don de vos SMS à la Science »*. Presses universitaires de Louvain, Louvain-la-Neuve, Cahiers du Cental, 3.1.
- Panckhurst, Rachel (2008). « Short Message Service (SMS) : typologie et problématiques futures ». In Teddy Arnavielle (coord.). *Polyphonies*, pour Michelle Lanvin, Montpellier : Éditions LU.
- Tatossian, Anaïs. 2011. Clavardage et orthographe. Correspondance. Centre collégial de développement de matériel didactique (CCDMD), pp. 26-29.
- Panckhurst R., (2010). « SMS et eSMS : une autre forme de communication électronique médiée ? », Conférence au Laboratoire *Psychologie du Développement et Processus de Socialisation* (EA 1687), Université de Toulouse 2I, 17 novembre, 2010.
- Prochasson, Emmanuel, Viard-Gaudin, Christian et Morin, Emmanuel (2007). « Language Models for Handwritten Short Message Services ». *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR'07)*, Curitiba, Brésil, septembre 2007.
- Tatossian, Anaïs (2011). Clavardage et orthographe. Correspondance. Centre collégial de développement de matériel didactique (CCDMD). pp. 26-29.
- Villers (de), Marie Éva (1998). « La réforme de l'orthographe est-elle restée lettre morte ? », *Correspondance*, vol. 4, n° 1, pp. 5-7.

Annexe

Transcription des exemples de sms en graphie standard (nouvelle orthographe)

- (1) Il voulait se venger parce qu'en cours la classe fait trop de bruit! Et il a trouvé un moyen...! Mais bon c'est fait c'est fait... demain en physique on s'assoit toujours ensemble ou quoi ? [smsFF]
- (2) Salut {Nom}. Je n'ai toujours pas reçu ta lettre... Ce n'est pas normal ! Es ce que tu m'as envoyé la lettre ? [smsFB]
- (3) Tu amènes la twingo vers 19h30... [smsFB]
- (4) Tu sais vider la bassine qui est dans ma chambre, la descendre et mettre l'essuie à laver ? [smsFB]
- (5) ha méchant loup épargne-moi, je te ferai ce {que} tu voudras [smsFB]
- (6) J ai eu une nuit hyper courte et je suis fatiguée [smsFB]
- (7) y a peu de chances de me voir tonight car je suis crevée [smsFB]
- (8) Salut c'est {Nom} ben voilà {Nom} ma filé ton numéro et c'est pour voir si c'est bien toi ! [smsFB]
- (9) Oui un trop marrant que j'ai enregistré [smsFB]
- (10) Tu peux m'appeler ou m'envoyer un sms [smsFB]
- (11) tu m'envoies chier [smsFB]
- (12) Désolée de t'ennuyer avec ça! [smsFB]
- (13) Bébé je m excuse de te téléphoner si longtemps [smsFB]
- (14) Où se situer face à ce dilemme ? [smsFB]
- (15) Je {ne} garantis pas mon ambiance! [smsFB]
- (16) Tout était bien sauf ton absence [smsFB]
- (17) Je me disais qu'on pourrait se mettre tous ensemble [smsFB]
- (18) Serait-il possible de ravoir un acompte lundi? [smsFB]
- (19) ça commence dans 4 minutes [smsFB]
- (20) {il} y avait 1 pièce collée [smsFB]
- (21) tu veux bien me laisser me concentrer [smsFB]
- (22) Bienvenue à vous 3 [smsFB]
- (23) Merci lol il est trop beau [smsFB]
- (24) tu as les cheveux bleus [smsFB]
- (25) j'aimerais justement savoir ! [smsFB]
- (26) Le petit {Nom, masc} est né ce 22.09.04 [smsFB]
- (27) C'est toutes des bombes dans ta classes [smsFB]
- (28) Hey catin! T'as pas assumé toi aujourd'hui ! [smsFB]
- (29) Si vous vous embêtez vous pouvez nous rejoindre à la Doudingue [smsFB]
- (30) ok ma mimi que j'aime je lui demanderai [smsFB]
- (31) Mes parents ont encore décidé de m'em... [smsFB]
- (32) Je peux agir légalement [smsFB]
- (33) je vous souhaite beaucoup de bonheur [smsFB]
- (34) amuse-toi bien ce soir [smsFB]
- (35) c'est ce que je t'ai dit hier je n'ai pas les notes [smsFB]
- (36) je suis dégouté [smsFB]
- (37) il voudrait que vous soyez là... Je te disais quel weekend [smsFB]
- (38) tu crois que je vais monter pour te dire bonjour ? Trop fatigant ;o) [smsFB]
- (39) on les regarderait une fois à 2 d'accord ? [smsFB]
- (40) Tes messages ne sont pas toujours les plus clairs au monde [smsFB]
- (41) Elle l'est, mais un court instant! [smsFB]
- (42) je te phone ce soir et demain pour qu'on ce voie [smsFB]
- (43) Tu as de ces idées chou je te jure... [smsFB]
- (44) on est toujours au resto [smsFB]
- (45) Bien dormi ? Moi comme un prince sans sa princesse [smsFB]
- (46) Dorénavant tu sais quoi [smsFB]
- (47) J'ai une nouvelle télécommande [smsFB]
- (48) je nage de nouveau en plein cauchemar [smsFB]
- (49) elle doit être chez le véto là. Quand {Nom} l'a posée chez {Nom} elle s'est mise à couiner et à ne plus vouloir lâcher {Nom} et là elle n'arrivait plus à s'asseoir [smsFF]