



HAL
open science

Short Message Service (SMS) : typologie et problématiques futures.

Rachel Panckhurst

► **To cite this version:**

Rachel Panckhurst. Short Message Service (SMS) : typologie et problématiques futures.. Polyphonies, pour Michelle Lanvin, Université Paul-Valéry Montpellier 3, pp.33-52, 2009. hal-00443014

HAL Id: hal-00443014

<https://hal.science/hal-00443014>

Submitted on 26 Dec 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Panckhurst R. (2009), « Short Message Service (SMS) : typologie et problématiques futures », in Arnavielle T. (coord.), *Polyphonies, pour Michelle Lanvin*, Université Paul-Valéry Montpellier 3, p. 33-52.

Rachel Panckhurst, UMR 5267 CNRS-Université Paul-Valéry Montpellier 3
rachel.panckhurst@univ-montp3.fr

À Michelle

qui m'a demandé dès 1992 d'organiser un stage d'initiation aux TICE pour les enseignants du département. Elle avait compris très tôt l'importance des nouvelles technologies dans le cadre de l'enseignement supérieur. Quelques années plus tard, lorsqu'un de ses étudiants, Christophe Pujol, préparait un mémoire de maîtrise sous sa direction, elle m'a invitée à être membre du jury. Dans ce cas, un logiciel spécialisé était utilisé, et elle était pleinement consciente que la pluridisciplinarité pouvait être fédératrice. Enfin, depuis quelques années, l'écriture de type texto-mini-messages-SMS commence à envahir l'espace communicationnel virtuel de l'université, d'où ce nouveau parcours de recherche. J'aurais aimé écrire cet article — qui touche directement la spécialité professionnelle de Michelle — plus tôt pour en recueillir son avis...

Short Message Service (SMS) : typologie et problématiques futures¹

Rachel Panckhurst, UMR 5267 CNRS-Université Paul-Valéry Montpellier 3
rachel.panckhurst@univ-montp3.fr

1. Introduction

Le discours électronique médié constitue notre champ de recherche depuis plus d'une décennie :

Quand l'ordinateur est utilisé pour le courriel, les forums de discussion et les chats, en tant qu'outil permettant la communication entre individus, il devient un véritable *médiateur* ; son utilisation modifie notre discours et ainsi notre façon de communiquer avec autrui. Émerge alors un nouveau « genre de discours »², le « discours électronique médié (DEM) ». Le DEM contient des marques linguistiques et extra-linguistiques qui lui sont propres et il entre dans le cadre plus global de la « communication médiée par ordinateur » (CMO). (Panckhurst, 2006a : p. 345)

Plus précisément, nous avons effectué des analyses linguistiques et informatiques de courriers électroniques, puis de forums de discussion, et plus récemment de messageries instantanées (ou chats), et ce, dans un contexte d'enseignement supérieur et de recherche (Panckhurst 1999a, 1999b, 2001, 2003, 2006a,b, 2007 ; Panckhurst & Bouguerra, 2003).

Tout au long de ces recherches, nous avons été confrontée, à maintes reprises, à une certaine évolution créative. Crystal (2001) évoque ces modifications d'écriture :

« Il est vraiment remarquable qu'autant de personnes se soient si rapidement adaptées aux exigences linguistiques de ces situations nouvelles et aient su exploiter tout le potentiel du nouveau medium de manière suffisamment créative pour favoriser l'émergence de nouvelles formes d'expression. Cela s'est produit en quelques décennies seulement. J'en conclus que la faculté linguistique de l'être humain est en bonne forme. L'apparition du parler Net nous révèle ce que *l'Homo loquens* a de meilleur. ». (Crystal, 2001 : p. 242, notre traduction).

Nous avons proposé une expression néologique pour le français : « communication médiée par ordinateur » (« médiée » par opposition à « médiatisée » : Panckhurst, 1997), ou plus précisément dans un contexte linguistique « discours électronique médié » (Panckhurst, 1999, 2006a, 2007). D'autres chercheurs proposent des variantes terminologiques pour décrire les phénomènes d'analyse linguistico-informatique de ces types de discours. Entre autres : « communication médiatisée par ordinateur » (Marcocchia, 2000), « communication électronique scripturale » (Anis, 2003), « nouvelles formes de communication écrite » (Guimier de Neef & Véronis, 2004), « communication électronique » (Anis, de Fornel, Fraenkel, 2004) et, pour l'anglais : « Netspeak » (Crystal, 2001), « computer-mediated communication » (Herring, 1996). Les termes retenus dans un contexte (journalistique) plus vaste (comprenant également la

¹ Mes remerciements à Michel et Anne Bresson-Lucas, Sophie David, Laurent Fauré, Claudine Moïse pour des discussions fructueuses à propos des SMS.

² Pour une justification de notre utilisation de « genres de discours », cf. Panckhurst 2006a.

publication électronique) incluent notamment : « cyberl@ngue », « cyberlangage » (Dejond, 2002, 2006).

En une décennie, la situation s'est modifiée, car la téléphonie mobile a connu un essor remarquable, en apportant l'utilisation massive de SMS³. Pour ne pas exclure un moyen de communication aussi important, nous proposons d'élargir la terminologie initiale (qui ciblait prioritairement l'ordinateur) à : « communication électronique médiée » (CEM). Cependant, l'appellation plus contextualisée en linguistique : « discours électronique médié » (DEM) convient également. Par ailleurs, la notion de médiation est toujours au centre du débat : si l'on suppose que l'ordinateur agit en tant que véritable médiateur, d'autres médiateurs, humains, sont effectivement parfois présents : l'utilisation du courriel et de la téléphonie mobile n'implique pas des médiateurs humains, alors que ceux-ci peuvent (ou non) être présents en tant que « modérateurs » dans le cadre de listes de diffusion et de chats/messagerie instantanée. Anis (2003) préfère, dans ce cas, éliminer la notion de « médiation » pour se concentrer sur la « communication électronique » écrite (ou *scripturale*). Nous continuons à préférer CEM ou DEM, car même si des acteurs humains « modérateurs » peuvent être absents, nous sommes convaincue que l'ordinateur, ou le téléphone portable, en tant que simple outil ou « support » de médiation, modifie notre communication, notre discours avec autrui.

Dans le cadre de cet article, nous focalisons nos recherches sur l'étude des SMS, pour en dresser une typologie et réfléchir sur cette évolution afin de comprendre quelles peuvent être les problématiques et les éventuelles conséquences dans un domaine comme celui du traitement automatique des langues (TAL).

2. Contexte de la recherche

Dans un contexte d'enseignement supérieur et de recherche, il peut paraître étonnant que l'on trouve des échanges de messages comme celui figurant dans l'exemple 1, ci-dessous, recueilli à partir d'un forum de discussion entre étudiants dans le cadre d'un travail collaboratif.

« je c ke c chian de devoir se connecter mé en fait on é noté ossi sur le nombre d'échange et ce qu'on dira dc si vous avez un ordi à disposition ca seré cool d'envoyer des infos sur d sites par ex pr le dossier ou se genre de truc pr le site je pense ke ca devré allé vite je vou envéré un message ds les jours suivants sur chaque page k on peu faire et chacun poura dire ce ki lui plé le plus de faire moi je m en fou voila. »

(Exemple 1 : Forum entre étudiants, travail en groupe, 7/2/05, cours informatique généraliste en licence, cf. Panckhurst (2006a), p. 361)

Dans ce contexte précis, les étudiants avaient été informés que nous accédions à leurs discussions, voire que leurs discussions entraient en ligne de compte pour l'évaluation du travail. Cet exemple, datant de février 2005, constitue le premier exemple riche d'écriture de type SMS (désormais eSMS) apparaissant dans le cadre de nos recherches

³ Le SMS (« Short Message Service » ou « service de messages succincts »), , est un « service que peut offrir un système de communication sans fil, permettant aux usagers d'envoyer ou de recevoir des messages alphanumériques courts (autour de 160 caractères), qui s'affichent à l'écran de leur terminal mobile, généralement un téléphone portable. », http://www.journaldunet.com/encyclopedie/definition/190/50/20/short_message_service.shtml, consulté en février 2008. Dans la suite de cet article, nous utiliserons le terme SMS, davantage utilisé dans un contexte francophone que ses synonymes « mini-message » et « texto ». Les anglosaxons ont généralement recours au terme « text » et « texting » pour ce même phénomène.

portant sur l'analyse des discours médiés depuis 1996. Cela a été déterminant dans l'orientation des recherches que nous menons : depuis cette date nous avons élargi notre champ afin de prendre en considération les eSMS.

Ces pratiques d'écriture en contexte universitaire sont cependant loin d'être évidentes. Lorsqu'on pose explicitement la question aux étudiants⁴, on constate qu'ils sont réfractaires à l'idée d'utiliser un style qui leur paraît trop « relâché » pour un contexte institutionnel, tout au moins dans une situation d'échange vertical hiérarchique (étudiants/enseignants) :

« Dans le cadre de l'enseignement du Master, je pense qu'il serait mal vu de ne pas écrire en toutes lettres. » Étudiant A, M2 Gestion des connaissances, apprentissages, Formation ouverte et à distance (GAF), octobre 2007.

Dans un fil de discussion initié par les étudiants de M2 GAF en janvier 2008, au sein d'un réseau d'échanges pédagogiques pour étudiants en eLearning (REEL : <http://reelgaf.ning.com/>) que nous animons à partir d'un espace de réseau social sur Internet (Ning : <http://www.ning.com>), les étudiants précisent leur réticence à utiliser les SMS ou l'eSMS (à l'aide de téléphones mobiles ou d'ordinateurs permettant la communication (a)synchrone) dans un cadre d'enseignement exclusivement à distance, en évoquant un problème de niveau de langue et d'introduction d'erreurs :

« [Le langage SMS] est trop familier pour qu'il soit légitime dans un tel cadre comme l'enseignement à distance. » Étudiant B, M2 GAF, 10/1/08.

« [Les SMS] peuvent amener à prendre l'habitude de commettre des fautes d'orthographe. », Étudiant C, M2 GAF, 8/1/08.

Quant aux situations de communication horizontale (entre pairs) ou croisée (horizontale et verticale : essentiellement entre pairs, mais avec la présence participative de l'enseignante), la situation peut différer, selon le mode de communication, et parfois en fonction du niveau d'études. Dans un de nos cours de 3^e année de licence, afin de familiariser les étudiants avec un système de parrainage étudiants en présentiel/étudiants à distance, nous leur avons proposé d'effectuer un chat à partir de la plateforme d'enseignement à distance, utilisée au sein de notre université, WebCT : <http://karolus.univ-montp3.fr>. Malgré les précisions apportées, concernant à la fois notre participation et la possibilité d'enregistrer les échanges écrits synchrones afin de les visualiser *a posteriori*, les étudiants se sentaient visiblement en situation quasi-privée et ainsi libres de s'exprimer comme ils l'entendaient. L'échange suivant se passe entre les deux premiers étudiants connectés (É1 et É2) qui ne se connaissent pas encore en début de semestre ; l'eSMS est de mise :

⁴ Nous avons posé la question suivante aux étudiants en enseignement à distance dans le cadre du M2 Sciences du langage, spécialité Gestion des connaissances, apprentissages, Formation ouverte et à distance (MGAF), Université Paul-Valéry, 2007-2008 : « Pensez-vous, si ce n'est déjà le cas, que l'utilisation d'une écriture de type SMS (pour vos communications asynchrones et synchrones) sera de plus en plus répandue dans le cadre de la formation ouverte et à distance ou pas ? »

É1 : salu ;
É2 : bonjour ;
É2 : mais qui es tu? ;
É1 : ben c moi ;
É1 : *et toi t ki* ;
É2 : *c'est cool ce tchat on pourra tricher au partiels*

(Exemple 2 : Échange synchrone entre étudiants de L3, janvier 2006⁵)

Cet échange est intéressant à double titre : 1) les étudiants passent outre les consignes et s'approprient l'outil selon leurs propres besoins ; 2) la situation de communication semble leur « appartenir » ; est-elle quasi-privée, peut-être, car instantanée ?

En revanche, le réseau REEL (pour les échanges pédagogiques entre étudiants en M2 GAF), et notamment l'espace de discussion asynchrone, ne semble pas induire une utilisation d'eSMS.

Les commentaires des étudiants de M2 GAF ainsi que les exemples ci-dessus montrent que l'usage vécu de l'eSMS dépasse le strict cadre technologique. L'eSMS n'est plus réservée uniquement à la téléphonie mobile, mais s'étend effectivement à d'autres situations de communication et à d'autres outils (par ex. chats, forums, etc. sur ordinateur). De ce fait, pourrait-on caractériser l'eSMS en termes de pratiques sociales et culturelles, impliquant ainsi un investissement dans un nouveau genre (ou forme) de discours ? La question est à creuser.

3. Classification du DEM

Le discours électronique médié, au sens large (incluant donc courriels, forums, chats, blogs, dans lesquels apparaissent ou non des eSMS), se caractérise par un ensemble de phénomènes qui nous rappelons ci-dessous (cf. Panckhurst, 2006a, 2007, Véronis & Guimier de Neef, in Sabah, 2006) :

1. les *didascalies électroniques* (cf. Mourlhon-Dallies & Colin, in Anis, 1999), par exemple, les *binettes* (ou *smileys*) permettant d'introduire des aspects sémiologiques non-verbaux ; les signes typographiques spécifiques : majuscules, allongement, répétition de caractères (permettant de simuler, dans certains cas, la communication para-verbale), chevrons ou barres verticales (permettant une répétition du discours entre l'énonciateur et son destinataire) ;
2. les erreurs ou « ratages » orthographiques, typographiques et grammaticaux et l'absence (ou la diminution) de ponctuation (cf. Panckhurst, 1998b, Véronis & Guimier de Neef, in Sabah, 2006, pour une synthèse) ; Véronis (1988) distingue les erreurs de *compétence* et de *performance* ; dans Panckhurst (1988b) nous distinguons les erreurs *machinales*, *discriminantes* (probablement engendrées uniquement par l'utilisation de l'ordinateur) des erreurs *floues*, *non-discriminantes* (plus difficiles à déterminer, provenant soit d'une méconnaissance de règle, soit d'une erreur due au moyen utilisé) ;
3. la *néologie* et la *néographie* (cf. Véronis et Guimier de Neef, in Sabah, 2006), entre autres, par exemple, les emprunts linguistiques et les abréviations, troncations, notations *sémio-phonologiques* (cf. Leinard, 2007) ou graphies phonétisées typiques des SMS.

⁵ Les erreurs éventuelles ne sont pas modifiées.

Des marques plus spécifiquement syntaxiques sont apparentes également :

- une utilisation prédominante du présent de l'indicatif ;
- une utilisation importante de déictiques, essentiellement des pronoms à la 1^{re} et à la 2^e personne ;
- un pourcentage moins important de verbes par rapport à d'autres formes de l'écrit plus normé ;
- une utilisation importante d'ellipses (par exemple : *Vous remerciant ; Impossible de trouver le document à enregistrer sur disque.*)

Sur un plan extra-linguistique, bien que les caractéristiques soient diversifiées en fonction des supports, on peut, entre autres, citer :

- le relationnel : divers sentiments et attitudes (rapidité, angoisse, impulsivité, agressivité, réajustements ultérieurs, accoutumance, incapacité à assumer des rencontres en face-à-face) ;
- la situation de communication : réduction, voire absence de formules d'ouverture et de clôture, bouleversement des tours de parole, de l'ordre, de la séquentialité, etc.

Enfin, nos recherches plus récentes (2005-2008) ont montré également une variation importante d'utilisation de catégories syntaxiques⁶ selon l'outil de communication employé (Panckhurst, 2006b, p. 635) :

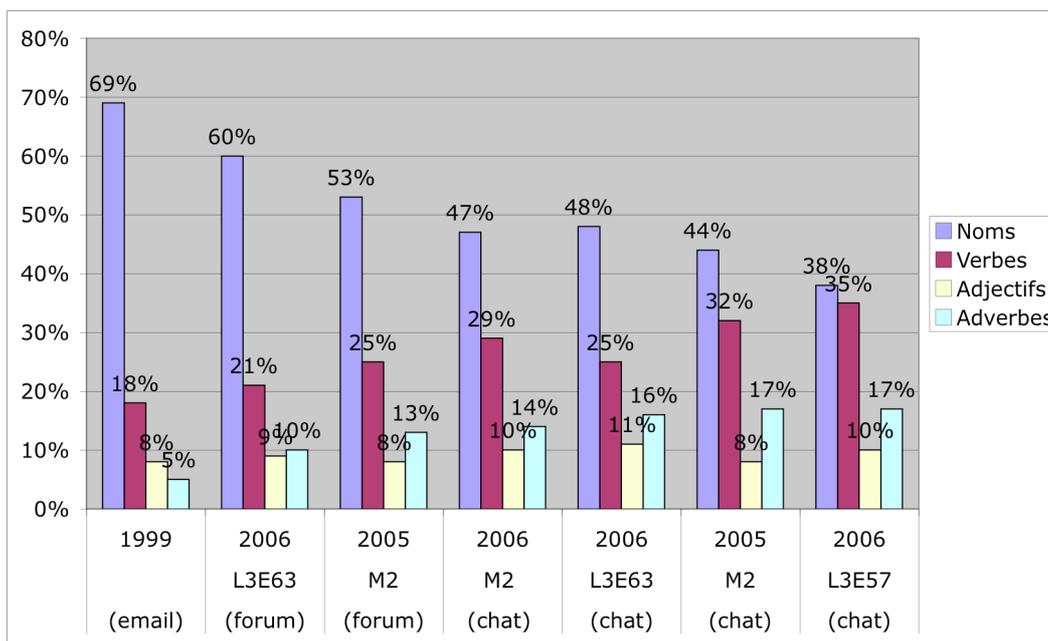


Tableau 1 : Catégories syntaxiques (occurrences) utilisées (courriel, forums, chats) entre 1999 et 2006

⁶ Nous préférons utiliser le terme de « catégories syntaxiques » ici plutôt que celui de « classes grammaticales », car ces catégories sont précisément déterminées à partir de l'utilisation d'un analyseur syntaxique automatisé, qui fournit des étiquetages suite à l'analyse du groupement structurel phrastique.

Jusqu'en 2005, nos corpus montraient une similitude entre les catégories syntaxiques utilisées dans le DEM et celles apparaissant au sein d'autres formes écrites plus conventionnelles (c'est-à-dire une utilisation importante de noms et une quantité relativement faible de verbes). À partir de nos corpus de 2005 (et le corpus de 2006 le confirme), un changement notoire s'est effectué : figurent une réduction de noms, une stabilité adjectivale et une augmentation de verbes ainsi que d'adverbes modalisateurs. Cela démontre une évolution linguistique radicale ; les résultats de l'analyse des catégories syntaxiques indiquent que le DEM peut se rapprocher de l'oral ou de l'écrit, et cela varie selon l'outil de communication adopté. Au sein du tableau 1, le chat apparaît comme correspondant à un moyen de communication plus « oral », tandis que les forums et le courriel semblent être davantage rattachés à un moyen plus « écrit ».

4. eSMS : présentation

À partir de cette classification du DEM, l'analyse de l'eSMS nous permet d'approfondir notre champ de recherche et de poursuivre notre étude sur l'évolution du langage. Par ailleurs, apparaissent au sein des SMS des phénomènes incluant l'oralité, l'écrit, et le lien entre les deux, d'une manière très créative, que ce soit dans une dimension néologique ou néographique. La *néologie* (créativité lexicale : essentiellement l'utilisation de langues étrangères — souvent des anglicismes dans nos corpus — ou de verlan) et la *néographie* (nouvelle orthographe) nous intéressent au premier chef, car elles sont fréquemment utilisées dans le cadre de l'eSMS.

Si l'étude des SMS permet aux différents domaines relevant de la linguistique théorique et au-delà (phonétique/phonologie, morphologie, lexique, syntaxe, morphosyntaxe, sémantique, stylistique, pragmatique...) de se côtoyer, Marty (2005), citant Anis (2004), rappelle les contraintes imposées et le dépassement du cadre purement linguistique :

« Le modèle proposé [par Anis] part des contraintes imposées : linguistiques (concision ; rapport avec l'oral ; dialogisme ; rapidité ; jeu avec les normes) mais aussi techniques (taille de l'écran ; nombre de caractères), communicationnelles (communication rapide et discrète), psychosociales (intimité ; langage du groupe) et économiques (coût : moins cher qu'un appel vocal). », Marty (2005 : 65).

Faire une typologie de l'eSMS est complexe ; la stabilisation véritable est absente et la créativité lexicale est très importante. Par ailleurs, certains chercheurs linguistes et informaticiens ont déjà proposé des typologies (*cf.* en annexe, un tableau récapitulatif de celles proposées par Anis, 2004, Fairon *et al.*, 2006b, Liénard, 2007, Véronis et Guimier de Neef, 2006). Dans ce cas, est-il réellement utile d'en proposer une autre ? Dans quel but ? De notre point de vue, ces différentes typologies, qui ont le mérite d'exister, contiennent l'essentiel (de manière parfois très détaillée) des phénomènes rencontrés dans l'eSMS. Cependant, nous avons souhaité clarifier certains aspects qui ne nous semblent pas toujours distingués de manière limpide dans les typologies catégorielles existantes. Les points majeurs sont évoqués ci-dessous :

- La description effective (graphique, phonétique, etc.) du phénomène en question semble parfois être mise au même niveau que la portée significative de l'interprétation ultérieure du message.
- La même terminologie est utilisée, selon les chercheurs, pour recouvrir des phénomènes différents.
- Une catégorie unique contient des exemples qui éclairent le phénomène exposé, mais ceux-ci peuvent également appartenir à d'autres catégories.
- Une catégorie est présentée comme étant une combinaison d'autres catégories existantes.

Les typologies existantes permettent des analyses linguistiques, sociopragmatiques, etc. fines et nous interpellent positivement. Celle que nous proposons ici ne prétend ni être exhaustive ni être exempte d'erreurs. L'étude de l'eSMS est précisément passionnante car elle fait intervenir des aspects linguistiques et extra-linguistiques très divers et complexes. Notre seul souhait est que cette typologie supplémentaire permette d'approfondir le débat. Seront utilisés des exemples incluant des eSMS émanant de nos chats et forums pédagogiques et également à partir des consultations de la base logicielle *SMS pour la science* (Fairon et al. 2006a) et des 4 typologies figurant en annexe.

Dans la mesure du possible, nous nous efforcerons de proposer une typologie qui décrit clairement, sans mélange de niveaux catégoriels ou descriptifs, des phénomènes *simples*. Ensuite intervient la comparaison avec des aspects *complexes* qui correspondent soit à des concaténations de phénomènes simples, soit à des phénomènes *de facto* complexes, correspondant, par exemple à un écrasement de signes (cf. Liénard, 2007 : *kestufé* « Qu'est-ce que tu fais ? »).

4.1. eSMS : typologie

La typologie porte essentiellement sur la *néographie*, d'un point de vue variationnel : substitutions, réductions, suppressions, absence ou raréfaction, augmentation, de caractères (lettres, chiffres, signes de ponctuation, symboles spéciaux et mathématiques, etc.). Cela permet d'apercevoir, dans un premier temps (cf. § 4.1.1), les dominantes : dimensions morphologiques, morpho-lexicales, lexicales, morpho-syntaxiques, syntaxiques, sémantiques, discursives, etc. Dans un second temps (cf. § 4.1.2), nous identifierons les processus combinatoires qui sont omniprésents dans l'eSMS.

Dans le cadre de ce travail, nous ne mentionnons pas les *erreurs/ratages* (cf. § 3) qui figurent dans les SMS (éventuellement dues à différents facteurs : erreurs machinales de saisie, orthographiques, grammaticales, etc.). La *néologie*, ou *l'emprunt néologique*, faisant intervenir l'usage de langues étrangères (par exemple, un mot comme *now*, en anglais, peut être privilégié par rapport à son équivalent français *maintenant*, à cause de l'économie des caractères) et le verlan n'est pas non plus étudié dans le contexte présent.

4.1.1 Néographie : phénomènes simples

I. Substitutions	
I.1. phonétisées :	
	I.1.1. entiers : remplacer un son par des caractères uniques (lettres ou chiffres). L'orthographe du lexème est totalement modifiée : o (eau), 7 (cet).
	I.1.2. partiels : remplacement de digrammes et trigrammes, qui transcrivent un phonème. L'orthographe du lexème est ainsi partiellement modifiée : <i>ossi</i> (aussi), <i>allé</i> (aller), <i>bo</i> (beau) ; « s » intervocalique : <i>bizes</i> (bises)
	I.1.3. avec variation : <i>bisoo</i> (bisou)
I.2. graphiques :	
	I.2.1. élision, typographie, majuscules : remplacement de l'apostrophe d'élision ou d'un trait d'union, etc. par l'espace, « m en » (m'en), « est ce que » (est-ce que) ; mise en majuscules de l'ensemble d'un message ou, au contraire substitution majuscules/minuscules.
	I.2.2. icônes, symboles mathématiques, caractères spéciaux, rébus : (*, + = => @) ; à + (à plus), <i>de grandes @</i> (de grandes oreilles)
	I.2.3. avec variation : <i>bisoux</i> (bisous) <i>mwa</i> (moi)
II. Réductions	
II.1. phonétisées :	
II.1.1. abrégements morpho-lexicaux :	
	II.1.1.1. troncations : <i>ordi</i> (ordinateur, apocope), <i>'lut</i> , <i>Net</i> (salut, Internet, aphérèse).
	II.1.1.2. sigles/acronymes : <i>ASV</i> (âge, sexe, ville), <i>mdr</i> (mort de rire), <i>tvb</i> (tout va bien), <i>t/m</i> (tout le monde), <i>lol</i> (laughing out loud, § 4.1.2.).
	II.1.2. variation : ui (oui), i (il)
II.2. graphiques :	
	II.2.1. suppression de fins de mots muettes : <i>échange</i> (échanges), <i>vou</i> (vous), <i>peu</i> (peut), <i>chian</i> (chiant), <i>fou</i> (« m'en fous »), chute de e instables : <i>douch</i> (douche)
	II.2.2. squelettes consonantiques & abréviations : <i>dc</i> (donc), <i>pr</i> (pour), <i>ds</i> (dans) ; consonnes doubles : <i>ele</i> (elle), <i>poura</i> (pourra) ; abréviations sémantisées (abréviations réduites à l'initiale) : <i>t</i> (te/tu) <i>p</i> (peux).
	II.2.3. agglutinations : <i>jattends</i> (j'attends).
III. Suppression/absence ou raréfaction :	
III.1. graphiques	
	III.1.1. typographie & ponctuation : cf. exemple 1 dans l'article
	III.1.2. signes diacritiques : <i>ca</i> (ça), <i>voila</i> (voilà)
IV. Augmentations et ajouts :	
IV.1. graphiques	
	IV.1.1. Répétition de caractères et/ou de signes de ponctuation : <i>suuuuppppeerrr !!!!!</i>
	IV.1.2. Représentations sémiologiques (Smileys/binettes) :-)
	IV.1.3. Ajout de caractères : <i>oki</i> (ok), <i>les zamours</i> (les amours)
	IV.1.4. Onomatopées : <i>mouarf</i> , <i>arfff</i> , <i>bof</i> .

Tableau 2 : Typologie de phénomènes simples de l'eSMS

Commentaires sur le tableau 2.

La nouvelle orthographe inclut à la fois des variations graphiques et phonétisées⁷ de l'écriture ou plus précisément de la saisie. D'un point de vue linguistique, celle-ci se situe sur un plan à dominante morphe-lexicale, dans un premier temps, mais pas uniquement, comme nous le verrons dans l'analyse ci-dessous.

La *substitution* correspond à un remplacement de la graphie ou une partie de la graphie par une autre. On obtient généralement un nombre inférieur de caractères (mais pas toujours, cf. *bisoo/bisou*). La *réduction* correspond à un enlèvement de certains caractères et résulte nécessairement en un nombre inférieur de caractères. La *suppression/absence* ou *raréfaction* correspond à un enlèvement total (suppression de ponctuation, signes diacritiques, etc.). Ces trois premières distinctions correspondent souvent (mais pas nécessairement) à un souhait de *simplification* du message. Le cas de *l'augmentation/ajout* montre que le SMS ne vise pas toujours l'économie. La dimension expressive, du point de vue para et non-verbal est essentielle ici. Fayada (2007), dans sa recherche portant sur l'utilisation de l'eSMS par des jeunes (15-18 ans), montre que l'aspect stylistique paraît plus important que l'économie de caractères, ce qui résulte parfois en des augmentations de longueurs de messages.

Nous utilisons le terme *rébus* pour une « devinette graphique mêlant lettres, chiffres, dessins, dont la solution est une phrase, plus rarement un mot, produit par la dénomination, directe ou homonymique, de ces éléments. » (définition TLF-i, <http://atilf.atilf.fr/tlf.htm>). Des exemples comme ceux proposés par Fairon *et al.* (2006b) entrent dans ce cadre : *mobi** (mobistar), *de grandes @* (de grandes oreilles), etc. En revanche, ceux qui incluent une graphie phonétisée complexe, incluant parfois des lettres et des chiffres (*a2m1*, pour « À demain ») ne nous semblent pas correspondre à de véritables rébus (cf. Anis, 2004, Véronis et Guimier de Neef, 2006), mais plutôt à des substitutions phonétisées multiples (cf. § 4.1.2.)⁸.

La distinction *abréviation* (uniquement graphique) et *abrégement* (résultant d'un aboutissement phonétique) est adoptée ici. Par ailleurs, la *troncation* sera exclusivement réservée à l'apocope et à l'aphérèse, et la réduction à l'initiale (*p*, pour *peux*, par exemple) est incluse dans les *abréviations sémantisées*. Nous adoptons cette terminologie afin de préciser que l'abréviation ne peut être comprise que dans un contexte sémantique, lorsque le message est interprété dans sa globalité.

4.1.2. Néographie : phénomènes complexes

Une des particularités principales de l'eSMS est que de multiples phénomènes apparaissent simultanément, ce qui rend leur analyse complexe. Le phénomène évoqué peut être *de facto* complexe, car il s'agit d'un écrasement ou d'un compactage d'énoncé en un seul signe linguistique, ou bien il y a concaténation de phénomènes simples juxtaposés. Nous présentons quelques exemples de manière non-exhaustive ci-dessous :

⁷ Nous préférons utiliser le participe passé passif (graphie *phonétisée*), plutôt que l'actif (graphie *phonétisante*), afin de cibler plus précisément le résultat de la phonétisation.

⁸ On pourrait éventuellement souhaiter étendre la définition du TLF-i pour *rébus* afin d'inclure « le jeu sur la valeur phonétique du code numérique » (discussion avec L. Fauré). Ce point reste à débattre. Néanmoins, même élargie, cette définition ne permettrait pas de classer des exemples comme *a2m1* dans le tableau 2 ci-dessus, qui ne concerne que les « phénomènes simples » de la néographie.

- agglutinations + compactage + écrasements : *chui* (je suis), *chais* (je sais), *yora*⁹ (il y aura), *kestufé ?* (qu'est-ce que tu fais ?)
- réduction graphique en agglutination & suppression de fin de mot muette & substitution phonétisée entière (*7éta* « cet état »)
- substitutions phonétisées multiples : *2manD* (demander), *6T* (citer), *A2M1* (à demain), *kikoo* (coucou)
- ajout de caractères & substitution graphique avec variation : *mwouah*, *moua* (moi)
- néologie (anglicisme) et néographie (abrégement en acronyme) : *lol* (laughing out loud)
- réduction graphique avec suppression de fin de mot et substitution phonétisée avec variation : *pô* (pas)
- substitution phonétisée partielle (c->k) et avec variation (ou->oo) et réduction phonétique avec variation (ou->i) : *kikoo* (coucou)

4.1.3 Au-delà du cadre morpho-lexical

D'autres aspects dépassant nettement le cadre morpho-lexical figurent à maintes reprises au sein de l'eSMS. Nous les empruntons essentiellement à Fairon *et al.* (2006b) :

- morpho-syntaxe (changement de catégorie grammaticale) : **sms-moi** *qd tu arrives.*
- syntaxe (ellipse) : *t'appelle demain ; pas eu mon exam*
- sémantique lexicale (polysémie) : v (vais, veux/veut), c¹⁰ (c'est, ces, ses...)
- discours : *je regarde la TV. je veux bien ton massage...les magasins de Bruxelles. Ça a l'air d'aller toi ?* (réponses en rafale à des questions multiples) ; *ça va ? moi oui. tu fais quoi dimanche ? moi je suis censé dormir.* (questions/réponses fournies immédiatement).

5. Problèmes liés au traitement automatique des langues (TAL)

L'eSMS pose d'innombrables problèmes pour le TAL : comment segmenter correctement les messages, par exemple pour repérer les agglutinations ou pour déterminer les frontières de mots dans des messages saisis sans espaces ? Comment corriger les mots, alors que différents phénomènes entrent en ligne de compte (graphies phonétisées, abréviations et abrégements, rébus, etc.), et constituent des altérations (substitutions, réductions, suppressions, augmentations) qui doivent être rectifiées ? Comment gérer l'absence ou la diminution de ponctuation et de signes typographiques, de même que l'éllision ? Si la néologie concerne essentiellement le lexique, et ainsi une possible

⁹ L'exemple de *yora* semble inclure à la fois une réduction graphique avec agglutination et une réduction avec graphie phonétisée, ou plutôt le "y" ne correspond pas à une graphie phonétisée en soi, mais une influence de l'écrit car le "il y" est abrégé en [i] + [j] puis en [j]. Nous remercions Laurent Fauré pour une discussion à ce propos.

¹⁰ Le cas de « c » ou de « v » est intéressant car outre la dimension sémantique, les exemples où « c » renvoie à *ses, sais/sait* et « v » à « *vais* », nous pouvons l'inclure dans la catégorie 1.1 du tableau 2, à savoir : substitutions phonétisées entières ; en revanche, « c » et « v » qui renvoient à (*c'est* ou *ces, veux/veut...*) correspondent à II.2.2, à savoir réductions graphiques avec abréviations sémantisées (abréviations réduites à l'initiale).

augmentation d'entrées dictionnaires, la néographie est beaucoup plus complexe à traiter.

Guimier de Neef et Véronis (2004) évoquent ces problèmes qui touchent différentes parties de la chaîne de traitement automatisé, au niveau de l'édition et de la segmentation des messages, du lexique, de la morpho-syntaxique et au-delà. Guimier de Neef *et al.* (2007) et Guimier de Neef & Fessard (2007) proposent un système de transcription et de correction de SMS intitulé TiLT. Ce logiciel novateur, que nous avons pu tester dans sa version de démonstration, permet plusieurs niveaux d'analyse et d'intervention : corrections de mots, mise en forme du message, reconnaissance et traitement spécifique d'abréviations (essentiellement des sigles et des acronymes), identification de smileys (binettes) et une éventuelle détection des langues utilisées (cf. Tableau 3).

corrections de mots
phonétique : <i>g ht du kfé</i> accentuation : <i>moi ca va, soiree mexicaine epicee!</i> typographie : <i>J ai simplemdnt besoin d amour</i> troncation : <i>chuis alé o taf trankil ce mat1, ormi la gueul de bois !</i> séparation : <i>jatend son cou 2fil biz jane</i> modification expressive : <i>g u 16 en fisic suupppeer geeeeniaal</i>
mise en forme du message
élision : <i>j ramene l argentin au centre et j arrive</i> mise en majuscules de noms propres : <i>Bjr jean claude ca va ?</i> gestion des tirets : <i>est ce que tu vi1 ?</i>
abréviations, smileys, langues étrangères
abréviations : <i>dsl pr staprem, MDR le 6né, gsper ke ta pac 1 bon wk</i> identification des binettes : <i>pq tapel pa :(</i> détection de langues : <i>pense à faire le plein ; how do you do ?</i>

Tableau 3 : TiLT : phénomènes traités

Après la phase de transcription/correction, est proposée une vocalisation des messages, faisant intervenir également des problèmes phonologiques poussés. Cette dimension de synthèse vocale peut être particulièrement utile pour une utilisation par des personnes ayant des déficiences visuelles.

Si le système paraît prometteur, un certain nombre de problèmes demeurent, surtout lorsqu'un message contient des phénomènes multiples ou distincts à traiter (par exemple, une liaison qui évolue en agglutination, accompagnée d'une abréviation de type squelette consonantique : *mn nange* ; des séparateurs de mots multiples manquants) ou des problèmes concernant la désambiguïsation catégorielle pour des lettres isolées (*c* qui peut renvoyer, selon le contexte à *ces, c'est, sais/sait...*), etc. Par ailleurs, même si un recensement des formes pour des mots particulièrement fréquents est effectué (*aujourd'hui, demain, etc.*) le système est parfois démuné face à une multitude de

possibilités¹¹. En revanche, le système est doté de « méthodes dynamiques qui permettent de réagir face à de nouvelles orthographes phonétiques [...] telles que *Btise*, *6Dré* etc. »¹².

Nous l'avons stipulé plus haut ; les problèmes posés par les SMS pour le TAL sont particulièrement complexes. Les logiciels comme TiLT, mis en œuvre pour la transcription, la correction et la vocalisation de SMS, ne peuvent qu'approfondir nos connaissances et aboutir à une meilleure compréhension des phénomènes traités.

6. Conclusion

Nous savons qu'une des caractéristiques fondamentales de l'eSMS est une variété impressionnante de formes, une absence quasi totale de normes et une créativité lexicale extrêmement riche¹³. Dans cette perspective, proposer des dictionnaires du *langage SMS* (comme ceux qui fleurissent sur Internet), qui fixeraient l'usage, semble illusoire, car l'usage est précisément libre, et le lexique est particulièrement mouvant. En revanche, établir une typologie claire est un exercice utile, car cela permet de cerner l'étendue des phénomènes rencontrés, comprendre les types de difficultés, notamment pour le TAL, et de fournir des pistes de réflexion ultérieure.

Dans une prochaine étape de cette recherche, nous espérons pouvoir intégrer une analyse de la dimension interdialogique de l'utilisation des SMS. Pour ce faire, il nous faut recueillir un corpus conséquent de SMS afin de mettre l'accent sur leur contexte dialogique précis. Un projet de recherche (*SMS4science*)¹⁴ mené en collaboration entre la Belgique et des partenaires d'autres pays, dont le démarrage est prévu en 2008, devrait permettre de mener à bien cette avancée. Notamment, nous aimerions avoir accès à des données permettant d'identifier les échanges entre scripteurs et récepteurs. Ces données pourront être encodées avant la phase d'anonymisation afin de permettre des analyses interdialogiques *a posteriori*. Cependant, cette constitution de corpus est difficile à réaliser, dans la mesure où les co-scripteurs ne font pas obligatoirement don de leurs SMS de manière simultanée. Il serait également intéressant d'effectuer des analyses de messages entre un scripteur unique et des destinataires différents, ainsi qu'au sein de tranches d'âge distinctes, afin d'étudier les éventuelles variations. La nature du public lui-même est un facteur important ; dans quelle mesure, les personnes ayant des

¹¹ Nous avons soumis à TiLT les 40 formes graphiques distinctes recueillies pour le mot *aujourd'hui*, (à partir de 654 occurrences), dans le cadre du projet belge (Fairon *et al.* 2006a) : 12 formes ont été correctement reconnues ; pour 7 formes aucune vocalisation n'est proposée et pour les 21 formes restantes, les graphies sont prononcées telles quelles en français (*ajdui*), ou en langue étrangère (par exemple, *ojordwi* est détecté en anglais). Parfois un problème supplémentaire se pose : *jourd'hui* est prononcé « jours huit ».

¹² Communication personnelle avec Émilie Guimier de Neef en janvier 2008.

¹³ Les abréviations les plus fréquentes figurant dans Fairon *et al.* (2006b) sont les suivantes : *pr* (867), *g* (767), *ke* (698), *tt* (561), *ds* (529), *G* (503), *OK* (502), *biz* (433), *qd* (421), *o* (418), *ok* (398), *ms* (386), *ki* (379), *sms* (372), *bcp* (353), *koi* (352), *tjs* (343). Ces chiffres correspondent à un nombre d'utilisateurs (sur 2436) et non à un nombre total d'occurrences.

¹⁴ Le projet *SMS4science* fait suite au projet belge « Faites don de vos SMS à la science », mené par le CENTAL et le CELEXROM à l'Université Catholique de Louvain, qui a permis de recueillir plus de 75.000 SMS entre octobre et décembre 2004.

déficiences visuelles ou auditives, par exemple, se servent-elles de ces technologies ?¹⁵ Comparer l'eSMS utilisée au sein de chats pédagogiques et en téléphonie mobile ne permettait pas, jusqu'à présent, une véritable mise en parallèle des situations de communication, étant donné le clavier téléphonique fruste. Cela change désormais, avec les nouveaux téléphones permettant une saisie de SMS de 4 façons : 1) via le clavier téléphonique standard, 2) en utilisant de manière tactile un clavier d'ordinateur, 3) par reconnaissance de caractères isolés écrits à l'aide d'un stylet, 4) par reconnaissance de mots entiers.

Plus globalement, en linguistique et en sociolinguistique, les points suivants retiendront notre attention dans les recherches ultérieures¹⁶ sur les SMS et l'eSMS : la notion de texte et de textualisation, les marqueurs de dialogisme (interrogation, négation, emphase...), les genres du discours et les interactions verbales, la construction de la sphère intersubjective (entre scripteur et récepteur), les éventuels échanges, échos, reprises, etc. d'un SMS à un autre, l'examen du rapport phonologique/sémiographique, le lien entre l'oral, l'écrit, la temporalité, et également leur inscription dans la réalité sociale en changement. En psychologie cognitive, on se questionnera sur le lien éventuel entre l'utilisation des SMS et la lutte contre le rétrécissement social et l'insécurité inhérente au vieillissement.

Dans cette perspective, plusieurs orientations, approfondissements et questions se dessinent actuellement. Faire participer des chercheurs travaillant dans d'autres disciplines permettra une réflexion pluridisciplinaire élargie et ainsi une meilleure compréhension du discours électronique médié, dans ce cas, les SMS, et la façon dont il est appréhendé dans nos sociétés actuelles.

¹⁵ Une de nos étudiantes en Master GAF (2007-2008), Audrey Gaussuron, effectue actuellement une recherche sur les *SMS et les déficiences visuelles*.

¹⁶ Nous nous associons à d'autres chercheurs afin d'élargir les problématiques de la recherche initiale.

Références bibliographiques :

- Anis, J. (1999), *Internet, communication et langue française*, Paris, Hermès.
- Anis J. (2001), *Parlez-vous texto ? Guide des nouveaux langages du réseau*, Paris : Le Cherche Midi.
- Anis J. (2003), « Communication électronique scripturale et formes langagières », *Actes des Quatrièmes Rencontres Réseaux Humains / Réseaux Technologiques*. Poitiers, 31 mai et 1er juin 2002. « Documents, Actes et Rapports pour l'Education », *CNDP*, p. 57-70. <http://edel.univ-poitiers.fr/rhrt/document547.php> (consulté le 27/11/2007)
- Anis, J., de Fornel, M., Fraenkel, B. (2004, organisateurs) « La communication électronique : Approches linguistiques et anthropologiques », Colloque international, EHESS, Paris, 5-6 février 2004.
- Anis J. (2005, organisateur) « Le SMS : enjeux linguistiques, sociaux et culturels », Paris, Cité des Sciences et de l'Industrie, 20 mai.
- Bove R. (2006), « Étude de quelques problèmes de l'écriture « SMS » pour la synthèse de la parole », in A. Piolat (Éd.). *Lire, écrire, communiquer et apprendre avec Internet*. Marseille : Éditions Solal, p. 367-386.
- Caumont Olivier (2007), *Discours électronique médié et SMS : étude phonologique de quelques morphèmes verbaux en /E / : /ø /, et / ə /*, Mémoire de Master 1, Gestion des apprentissages et formation ouverte et à distance, Département Sciences du Langage, Université Paul-Valéry Montpellier 3.
- Crystal, D. (2001), *Language and the Internet*, Cambridge, Cambridge University Press.
- Dejond A., Mercier J. (2002), *La cyberlangue française*, Bruxelles : La Renaissance du livre.
- Dejond A. (2006), *Cyberlangage*, Bruxelles : Racine.
- Fairon C., Klein J.-R., Paumier S., (2006a), *SMS pour la science. Corpus de 30.000 SMS et logiciel de consultation*, Presses universitaires de Louvain, Louvain-la-Neuve, Manuel+CD-Rom, <http://www.smspouirlascience.be/>
- Fairon C., Klein J.-R., Paumier S., (2006b), *Le langage SMS. Étude d'un corpus informatisé à partir de l'enquête « Faites don de vos SMS à la science »*, Presses universitaire de Louvain, Louvain-la-Neuve, <http://www.smspouirlascience.be/>
- Fairon C., Klein J.-R., Paumier S., (2007), Un corpus transcrit de 30 000 SMS français, in Gerbault, J. (éd.) *La langue du cyberspace : de la diversité aux normes*, Paris : l'Harmattan, p. 173-182.
- Fayada Mélanie (2007), *Le langage SMS des jeunes. Approche lexicale et morpho-syntaxique*, Mémoire de Master 1, Gestion des apprentissages et formation ouverte et à distance, Département Sciences du Langage, Université Paul-Valéry Montpellier 3.
- Gaussuron A. (2008, à paraître), *SMS et déficiences visuelles*, Mémoire de Master 1, Gestion des apprentissages et formation ouverte et à distance, Département Sciences du Langage, Université Paul-Valéry Montpellier 3.
- Gerbault J. (2008, éd.), *La langue du cyberspace : de la diversité aux normes*, Paris : L'Harmattan.

- Guimier de Neef, E., Debeurme A., Park J. (2007), « TiLT correcteur de SMS : évaluation et bilan qualitatif », Actes de la 14^e conférence sur le traitement automatique des langues naturelles (TALN), p.123-132. Disponible à l'adresse : <http://www.irit.fr/~Dominique.Longin/TALN2007/volume1.pdf> (consulté le 14 février 2008).
- Guimier de Neef E., Fessard S. (2007) « Evaluation d'un système de transcription de SMS », Actes de la 26^e conférence Lexique et Grammaire 2007, Bonifaccio, 2-6 octobre, p. 217-224.
- Guimier de Neef, E. & Véronis, J. (2004, organisateurs), *Journée d'Etude. Association pour le Traitement Automatique des Langues (ATALA), Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.)*, 5 juin 2004, ENST, Paris, Publication en ligne. Disponible à l'adresse : <http://www.up.univ-mrs.fr/~veronis/je-nfce/theme.html> (consulté le 14 février 2008).
- Guimier de Neef, E. & Véronis, J. (2004). « 1 pw1 sr la kestion ;-) » Communication, Journée d'Etude de l'ATALA « Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.) », Paris.
- Herring S. C. (ed), (1996), *Computer-Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives*, Amsterdam/Philadelphia: John Benjamins.
- Liénard F. (2008), « Analyse linguistique et sociopragmatique de l'écriture électronique. Le cas du SMS tchaté », in Gerbault, J. (éd.) *La langue du cyberspace : de la diversité aux normes*, Paris : l'Harmattan, p. 265-278.
- Marcoccia, M. (2000), « La représentation du non-verbal dans la communication écrite médiatisée par ordinateur », *Communication et Organisation*, 18, p. 265-274.
- Marcoccia M. (2004), « La communicaion écrite médiatisée par ordinateur : faire du face à face avec de l'écrit .», *Journée d'étude de l'ATALA « Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.)* », 5 juin 2004, ENST Paris. Publication en ligne. Disponible à l'adresse : <http://www.up.univ-mrs.fr/~veronis/je-nfce/Marcoccia.pdf> (consulté le 14 février 2008).
- Manesse, D. (2007). *Orthographe, à qui la faute ?*, Issy-les-Moulineaux: ESF.
- Marty N. (2005), *Informatique et nouvelles pratiques d'écriture*, Paris : Nathan.
- Melançon B., (1996), *Séguin@Internet. Remarques sur le courrier électronique et la lettre*, Montréal : Fides.
- Mourlhon-Dallies F., Colin J.-Y. (1999), « Des didascalies sur l'internet ? », in *Internet, communication et langue française*, Anis, J. (Ed.), Paris : Hermès, p. 13-30.
- Panckhurst, R. (1997), « La communication médiatisée par ordinateur ou la communication médiée par ordinateur ? », *Terminologies nouvelles*, 17, 56-58.
- Panckhurst, R. (1998a), « Analyse linguistique du courrier électronique », Actes du colloque Les relations entre individus médiatisées par les réseaux informatiques. In *GRESICO*, (p. 47-60) Paris : L'Harmattan.
- Panckhurst, R. (1998b), « Marques typiques et ratages en communication médiée par ordinateur », Actes du colloque *CIDE 98*, (p. 31-43), Paris : Europa Productions.

- Panckhurst, R. (1999a) , « Analyse linguistique assistée par ordinateur du courriel », in *Internet, communication et langue française*, Anis, J. (Ed.), Paris : Hermès, p. 55-70.
- Panckhurst, R. (1999b), « La Communication médiée par ordinateur : un discours autre ? », in *L'autre en discours*, Bres, J., Delamotte-Legrand, R., Madray, F., Siblot, P (éd.), Dyalang-Praxiling, Montpellier : Service des publications de l'Université Paul-Valéry Montpellier 3, p. 307-331.
- Panckhurst, R. (2001), « Distance, open and virtual lifelong learning : shaping the transition within a French University », in *20th World conference on open learning and distance education*, Proceedings ICDE. Oslo, Norway & FernUniversität Hagen, Allemagne.
- Panckhurst, R. (2003), « Computer-mediated communication and linguistic issues in French University online courses » in *Proceedings of the 11th International Conference in Technology supported Learning & training*, Online Educa, p. 454-457.
- Panckhurst R. (2006a), « Le discours électronique médié : bilan et perspectives », in A. Piolat (Éd.). *Lire, écrire, communiquer et apprendre avec Internet*. Marseille : Éditions Solal, p. 345-366.
- Panckhurst R. (2006b), "Mediated electronic discourse and computational linguistic analysis : improving learning through choice of effective communication methods." Proceedings, ascilite conference, p. 633-637. , Sydney, Australie, 3-6 décembre.
- Panckhurst R. (2007), « Discours électronique médié : quelle évolution depuis une décennie ? », in Gerbault, J. (éd.) *La langue du cyberspace : de la diversité aux normes*, Paris : l'Harmattan, p. 121-136.
- Panckhurst, R., Bouguerra, T. (2003), « Communicational and methodological/linguistic strategies using electronic mail in a French University », *Proceedings, 8th International Symposium on Social Communication*, Santiago de Cuba, p. 548-554.
- Piérozak I. (2003), « Le « français tchaté » : un objet à géométrie variable ? », *Langage & Société* n°104 (M. Marcochia et B. Fraenkel, eds, *Ecrits électroniques : échanges, usages et valeurs*), p. 123-144.
- Véronis, J. (1988), *Contribution à l'étude de l'erreur dans le dialogue homme-machine*, Thèse de Doctorat, Université Aix-Marseille III, Marseille.
- Véronis, J., & Guimier de Neef, E. (2006), « Le traitement des nouvelles formes de communication écrite », in Sabah, G. (Éd.), *Compréhension automatique des langues et interaction* (pp. 227-248). Paris : Hermès Science.

Annexe : Tableaux synthétiques – typologies de l'écriture SMS (eSMS)

I. Anis (2004) : variation graphique et aspects morfo-lexicaux	
1.	Graphies phonétisantes
	1. Réductions graphiques
	1. qu : <i>melissa tu as kel age ?</i>
	2. c->k, s->z : <i>kom = comme ; biz = bises</i>
	3. chute des e instables : <i>douch froid</i>
	4. chute des mutogrammes en finale : <i>cour (cours), é (et), pa (pas), salu (salut)</i>
	5. simplification des digrammes et trigrammes : <i>ossi (aussi), vréman (vraiment), bo (beau)</i>
	6. combinaison des deux phénomènes : <i>forfè (forfait), jamé (jamais)</i>
	7. simplifications touchant à la morphologie verbale : <i>é (es), répondé (répondez), pe (peux)</i>
	8. déconstruction de « oi » : <i>moua (moi)</i>
	9. emprunt du digramme « oo » de l'anglais : <i>bizoo (bisou)</i>
	10. réductions avec compactage, lequel dissout les frontières de mots et évoque le mot phonique : <i>cé (c'est), keske (qu'est-ce que), mapelé (m'appeler), vérépa (verrai pas), jsui (je suis)</i>
	2. Réductions avec variantes phonétiques
	1. Variantes vocaliques ou semi-vocaliques : <i>kikoo (coucou), po (pas), ui (oui), moa (moi)</i>
	2. écrasements phonétiques : <i>chais (je sais)</i>
2.	squelettes consonantiques : <i>tt (tout), ds (dans), tjs (toujours), lgtps (longtemps), pb, prb (problème), vs (vous)</i>
3.	syllabogrammes et rébus à transfert : <i>l (elle), c (c'est, sais/sait, s'est, c'é[tait]), d (des), g (j'ai), m (aime), t (t'es), v (vais), oqp (occupé), nrv (énervé), 1 (in, un...), 2 (de), 2m1 (demain), a2line (Adeline), abi1to (à bientôt)</i>
4.	logogrammes : <i>1 (un), 2 (deux), + (plus) et paralogogrammes</i> : <i>asv (âge, sexe, ville), mdr (mort de rire), lol (laughing out loud), brb (be right back)</i>
5.	étirements graphiques : <i>je t'aiiiiiiiime</i>
6.	hétérogénéité (combinaison de procédés) : <i>kel1 (quelqu'un = phonétisme + logographie), ptdr (pété de rire = syllabogramme et siglaison), pkoï (pourquoi = squelette consonantique + phonétisme), k (que = logographie après réduction phonétique)</i>
7.	polyvalence et même polysémie : <i>v (vais, veux) : je v o sport dans une demi heure. [...] tu v une douch froid ?</i>
8.	variation : unités lexicales transcrites différemment par un même (ou plusieurs) scripteur(s). <i>C, c'est, etc.</i>
9.	Particularités morfo-lexicales : Graphies fonctionnant comme allomorphes (variantes). Sigles qui se lexicalisent : <i>Ton ASV stp merci</i> . Étirements graphiques (<i>lol -> looll</i>) -> lexicalisation
10.	Troncations : <i>ordi (ordinateur), lut (salut)</i>
11.	Anglicismes : <i>chatter, kiss</i>
12.	Verlan : <i>sa me fait déprimer a donf</i>
13.	Onomatopées : tenant lieu de segments phrastiques, svt renforcées par des étirements graphiques : <i>mouarf, arfff</i>

II. Fairon et al. (2006b) : classification générale	
1.	Phonétisation des caractères
1.	Lettres : <i>je t'm</i> (je t'aime) ; <i>KC</i> -> casser ; <i>g pa c 1 super w-e</i>
2.	Chiffres et autres caractères : <i>un peti cl1 d'œil</i> ; <i>Kl1</i> (câlin), <i>7 swaré</i> (cette soirée), <i>2m1</i> (demain) ; mots étrangers : <i>c4trist?</i> (c'est fort triste), <i>2bib</i> (toubib)
2.	Rébus : Séquences mêlant chiffres, lettres et signes divers -> valeur dénomminative : <i>mobi*</i> (mobistar), <i>//</i> (parallèle), <i>tu te x malin</i> (crois), <i>de grande @</i> (oreilles)
3.	Orthographe phonétique
1.	Suppression de fins de mots muettes : <i>Je l'ai pa vue</i> ; <i>par hazar</i>
2.	Simplification digrammes (au, ai, er, ez) <i>G acheté</i> ; <i>j'voulè allé ché toi</i>
3.	trigrammes (eau -> o) <i>baucoup</i> ; (ain c'est en général 1, donc plutôt phonétisation que simplification) ; aucune simplification de : <i>ent, end, ends en an.</i>
4.	consonnes doubles (<i>ele</i> -> <i>elle</i> ; <i>efé</i> -> <i>effet</i>)
5.	lettres <i>k</i> et <i>z</i> (<i>kand</i> , <i>bizou</i>) ; ergonomie : <i>k</i> préféré au <i>c</i> car plus rapide à saisir (2 appuis au lieu de 3) — mais pas systématique.
6.	<i>oua</i> et <i>oa</i> pour <i>oi</i> (<i>mwouah</i> , <i>mwa</i>)
7.	<i>oo</i> pour <i>ou</i> (<i>bizoo</i> , <i>kikoo</i>)
8.	simplification des 3 semi-voyelles du français : <i>paille [J]</i> , <i>huile, lui [V]</i> , <i>chouette, oui [w]</i> : <i>lui [V]</i> et <i>chouette [w]</i> deviennent souvent <i>W</i>
9.	écrasement assimilation consonantique (<i>chui, ché, ztadorme, ztèm</i>)
4.	Phénomènes graphiques
1.	Graphies à fonction expressives (<i>bisouxxx, kissss, t'aimmme, !!!!</i>)
2.	Liaisons évoluant en agglutinations (<i>mn nange, les zamour</i>)
3.	Graphies liées à des réminiscences (<i>po, oux</i> (bisoux) <i>pô</i> (Titeuf) ; <i>oux</i> (exception scolaire qui devient la règle !)
4.	Abréviations (phénomène strictement graphique, à différencier des abrégements) : <i>t</i> pour <i>te/tu</i> <i>p</i> pour <i>peux</i> ; mots non grammaticaux : <i>pr, j</i> (pour, jour)
5.	Icônes et symboles divers
1.	Smileys (émoticones)
2.	symboles mathématiques et logiques (+ remplace et; = remplace est, => utilisé pour une implication, explication)
6.	Phénomènes lexicaux
1.	Troncations (aphérèses - chutes de segments initiaux de mots: <i>tain</i> (putain), <i>man</i> (maman) ; apocopes (chutes de segments finaux : <i>ordi, mat</i>)
2.	Abrégements : Sigles et acronymes (ASV, JAM: just a minute) LOL, MDR, BRB (be rght back), IRL (in real life)
7.	Morphosyntaxe : (dodo, sms, lol, cœur -> verbe : <i>sms-moi qd tu arrives, smser, je pe dodo ds tes bras ? je te cœur tout plein</i>) ; genre comme adverbe : <i>jpeu vnir ms ke jusk genr12h</i>
8.	Syntaxe : (omission de mots grammaticaux : ellipse : <i>j entend rien, pas eu mon exam, je terminerai français demain</i>)
9.	Discours : (réponses en rafale à des questions multiples : <i>je regarde la TV. je veux bien ton massage...les magasins de Bruxelles. Ça a l'air d'aller toi ?</i> ; questions/réponses données immédiatement : <i>ça va ? moi oui. tu fais quoi dimanche ? moi je suis censé dormir.</i>)
10.	Variété des formes : très grande variété des types/formes. Richesse lexicale (création, néologismes ou écritures nouvelles).

III. Liénard (2007) : typologie des formes scripturales	
1.	Simplification
	Abréviations : <i>slt</i> (salut)
	Troncations
	Par apocope : <i>bibli</i>
	Par aphérèse : <i>Net</i>
	Par aphérèse interne : <i>poura</i>
	Élision d'éléments sémiologiques : signes de ponctuation, signes diacritiques, majuscules, articles, conjonctions
	Siglaion : <i>mdr, ASV</i>
2.	Spécialisation
	anglicismes : <i>now</i>
	notation sémiophonologiques
	mono-syllabiques : <i>manG</i>
	bisyllabiques : <i>2manD</i>
	totales : <i>6T, oT</i>
	écrasement de signes : <i>kestufé</i>
3.	Expressivité
	Emoticons
	Répétitions de graphèmes

IV. Véronis et Guimier de Neef (2006) : classification de la néographie	
1.	Graphies phonétisantes : généralement remplacement d'un des graphèmes complexes ou ambigu du français par une version unilette (<i>qu</i> → <i>k</i> , <i>eau</i> → <i>o</i> , <i>s</i> intervocalique → <i>z</i> , etc. : <i>koi, pas bo, bizes</i>)
2.	Squelettes consonantiques : <i>bcp</i> (beaucoup), <i>slt</i> (salut), etc.
3.	Rébus : lettres et chiffres pris dans leur valeur phonétique (<i>j'ai</i> → <i>G</i> , <i>cette</i> → <i>7</i> , <i>à demain</i> → <i>A2M1</i> , etc.)
4.	Troncations : une ou plusieurs lettres finales supprimées pouvant même aboutir à une réduction du mot à son initiale (<i>esp</i> → <i>espérer</i> , <i>p tu</i> → <i>peux-tu</i> , etc.)
5.	Sigles : (<i>tvb</i> → <i>tout va bien</i> , <i>atd</i> → <i>à ta disposition</i> , etc.)
6.	Logogrammes : (<i>en plus</i> → <i>en +</i> , <i>à plus [tard]</i> → <i>A+</i> , <i>en italien per</i> → <i>x</i> , etc.)
7.	Étirements graphiques (marquant généralement l'expressivité) : <i>saaalut ! biiiiizzzz !</i>
8.	Agglutination de mots : certaines séquences morpho-syntaxiques subissent le phénomène de façon privilégiée (<i>jattends/clitique+verbe</i> , <i>7éta</i> → <i>cet état/déterminant+nom</i> , etc.)