

Registre et moyen de communication: interdépendance ou indépendance?

Aurélia ROBERT-TISSOT

Romanisches Seminar, Universität Zürich

In diesem Beitrag werden wir anhand einer empirischen Untersuchung zeigen, dass die *Computer-Mediated Communication* (CMC) nicht als ein vollkommen eigenständiges Register betrachtet werden kann. Empirische Daten aus dem Korpus *sms4science.ch*, sowie die Analyse von Tagebüchern als zweite informelle schriftliche Quelle deuten vielmehr darauf hin, dass bestimmte sprachliche Phänomene wie z.B. die Subjektauslassung einerseits mit dem Kommunikationskontext, andererseits aber auch mit dem medialen Code zusammenhängen. Für die syntaktischen Merkmale spielt die Technologie jedoch keine unmittelbare Rolle. Daher scheint es angebracht für das Französische von einem informell schriftlichen Register, statt von einem CMC-spezifischen Register zu sprechen.

1. Introduction

Dans cette contribution nous abordons d'abord le concept de *registres* d'un point de vue théorique et nous proposons ensuite un procédé concret qui le rend opérationnel pour des bases de données de tailles considérables. À l'aide d'un phénomène linguistique choisi, à savoir l'omission du sujet, nous illustrons ce procédé qui permet de répondre, sur des bases empiriques, à la question plus générale de savoir ce qu'est un registre.

Comme exemple susceptible de constituer un registre, nous avons choisi la *communication médiée par ordinateur* (CMC, angl. *computer-mediated communication*), qui comprend toutes les situations de communication qui sont caractérisées par l'emploi d'une technologie numérique, comme p.ex. les e-mail, les blogs, Facebook ou les SMS (cf. Dürscheid & Brommer 2009: 4). Dans ce contexte, plusieurs auteurs parlent d'un 'langage internet' (p.ex. *netspeak* chez Crystal 2008: vii), qui est censé être une variété linguistique utilisée en CMC¹. De manière similaire, certains auteurs supposent l'existence d'un 'langage SMS' (p.ex. Fairon & al. 2006), propre à la communication par SMS. Mais nous verrons que d'un point de vue structurel, une variété SMS en tant que (sous-)système linguistique², ne peut pas être mise en évidence empiriquement.

¹ Mais cf. Dürscheid (2004: 147) qui argumente contre une telle variété, cependant sans présenter d'analyses empiriques quantitatives.

² Nous définissons une variété linguistique comme un (sous-)système autonome constitué par un ensemble de traits en co-variation et en corrélation avec certains facteurs extralinguistiques,

Nous argumenterons qu'il n'est pas approprié de regrouper toute la CMC sous une variété linguistique, ni de l'isoler d'autres formes de communication graphique qui ne sont pas à priori contraintes par une norme prescriptive, à savoir une norme qui est défendue et divulguée par des institutions officielles, telles p.ex. l'Académie Française ou les écoles. Nous utiliserons la caractérisation de *potentiellement informelle* (dorénavant dans sa version courte *informelle*) pour cette forme de communication, car il s'agit d'un contexte communicatif dans lequel la norme prescriptive peut, mais ne doit pas être respectée, comme nous verrons plus en détail dans la section 2.

La base empirique de notre analyse se constitue de la comparaison de la distribution de sujets omis dans les journaux intimes étudiés par Haegeman (p.ex. 2013) et dans le corpus *sms4science.ch* (cf. section 3). Il s'agit de deux types de production langagière graphique informelle, dont l'une fait partie de la CMC, tandis que l'autre existe depuis bien avant l'émergence de la CMC. La comparaison de ces deux ensembles nous permet d'identifier, à titre d'exemple, si le moyen de communication a une influence sur le phénomène étudié.

Dans la section 2, nous établissons d'abord le cadre théorique et méthodologique dans lequel nous plaçons notre analyse. Dans la section 3, nous présentons les corpus utilisés, avant de passer, à la section 4, à la description du phénomène linguistique, à savoir l'omission du sujet. Dans la section 5, nous présentons nos résultats issus du corpus *sms4science.ch* et les comparons avec ceux des études sur les journaux intimes. La section 6 est consacrée à la discussion. Nous terminerons par nos conclusions dans la section 7.

2. Conception, code, registre et technologie

2.1 La conception et le code

Du fait que dans les corpus analysés nous avons affaire à des données graphiques informelles, la question qui surgit est de savoir comment concevoir la relation entre l'écrit et l'oral.

D'après Koch & Oesterreicher (2001, ²2011: 3-14, dorénavant K&O), il faut distinguer la réalisation médiale d'un énoncé de sa conception communicative, comme illustré dans la Fig.1, une adaptation de K&O (2001: 585).

comme p.ex. les variétés du *français démotique* et du *français classique tardif* chez Massot (2010). Cf. aussi Duffer & Stark (2002) pour une critique du concept de *variété*.

		conception	
		de proximité	de distance
réalisation médiale	code graphique	faut pas le dire	il ne faut pas le dire
	code phonique	[fopal'dir]	[ilnəfopələ'dir]

Fig.1: Le code et la conception de communication selon Koch & Oesterreicher (2001: 585)

Pour la *réalisation médiale*, les auteurs parlent d'une réalisation en code phonique quand le message est réalisé à l'aide de l'appareil articulatoire, et de réalisation en code graphique quand le message est transmis par des lettres alphabétiques, des idéo- ou des syllabogrammes.

La *conception* d'un message dépend de la situation de communication qui est caractérisée par des facteurs extralinguistiques. K&O (2011: 8) donnent une liste de facteurs pertinents qui comprend, entre autres, le degré d'intimité, la proximité physique, le degré de coopération ou le type interactionnel (monologique ou dialogique) du discours. Ces facteurs ouvrent un continuum entre les deux pôles de la proximité respectivement de la distance communicative sur lequel on peut situer chaque message. Ce dernier sera formel, donc proche de la norme prescriptive d'une langue quand la situation de communication relève d'une distance communicative (illustré par la présence de *il*, de *ne* et de *e* caducs dans la Fig.1), tandis qu'il montrera des traits informels, déviants par rapport à la norme dans des situations de proximité communicative (illustré par l'absence des traits mentionnés dans la Fig.1).

Le code n'est pas lié de manière univoque à une certaine conception, raison pour laquelle on trouve des productions phoniques proches du pôle de la conception de distance communicative (p.ex. des conférences de presse, illustré par la prononciation [ilnəfopələ'dir] dans la Fig.1), mais aussi des productions graphiques qui révèlent certains traits typiques de la conception de proximité communicative (p.ex. une note pour le colocataire sur le réfrigérateur, illustré par <faut pas le dire> dans la Fig.1).

Cependant, Dürscheid (2003) souligne la difficulté d'établir un profil communicatif pour un mode d'interaction (p.ex. par ordinateur, par téléphone, en face-à-face), selon le modèle de K&O, car contrairement aux genres textuels (p.ex. les contes de fées), un mode d'interaction n'est pas homogène dans sa conception communicative.

La conception communicative des SMS, qui font clairement usage du code graphique, est variée. A titre illustratif, nous donnons un exemple d'un SMS entre des amoureux en (1), qui diffère nettement en termes de conception communicative de celui en (2), probablement plutôt de nature professionnelle.

- (1) Bonjour mon amour [...] Ah j'allais presque oublié j'ai très envie de toi tes lèvres me manque! [...] Ø T'embrasse fort Ø suis dispo ce soir si tes pulsions grimpe aussi (10586)³
- (2) Bonjour. Je viens de vous envoyer les citations par email. Cordialement. Nicolas [LastName] (8540)

Vu que les SMS en tant qu'ensemble ne peuvent pas être caractérisés de manière homogène selon le modèle de K&O, nous privilégions un autre modèle de la variation linguistique, à savoir le concept de *registres* selon Biber (1995).

2.2 Le registre et son opérationnalisation

Les facteurs extralinguistiques qui définissent la conception de communication déterminent également les *registres* selon Biber (1995). Ces facteurs extralinguistiques incitent le locuteur à effectuer un choix quant à la forme linguistique concrète qu'il va adopter pour chaque énoncé. L'ensemble qui se forme à partir de ces énoncés constitue un registre empiriquement observable. Bien que par exemple Zwicky & Zwicky (1981) fassent une différence entre le *style*, qui se choisit en fonction de la relation entre les interlocuteurs, et le *registre*, qui est lié à un certain contexte communicatif, Biber (1995) considère le style comme partie intégrante du registre. Pour des raisons de simplicité, nous suivons Biber (1995), puisque le style au sens de Zwicky & Zwicky (1981) dépend, lui aussi, de facteurs extralinguistiques.

La Fig.2 schématise l'influence des facteurs extralinguistiques sur la production linguistique du locuteur.

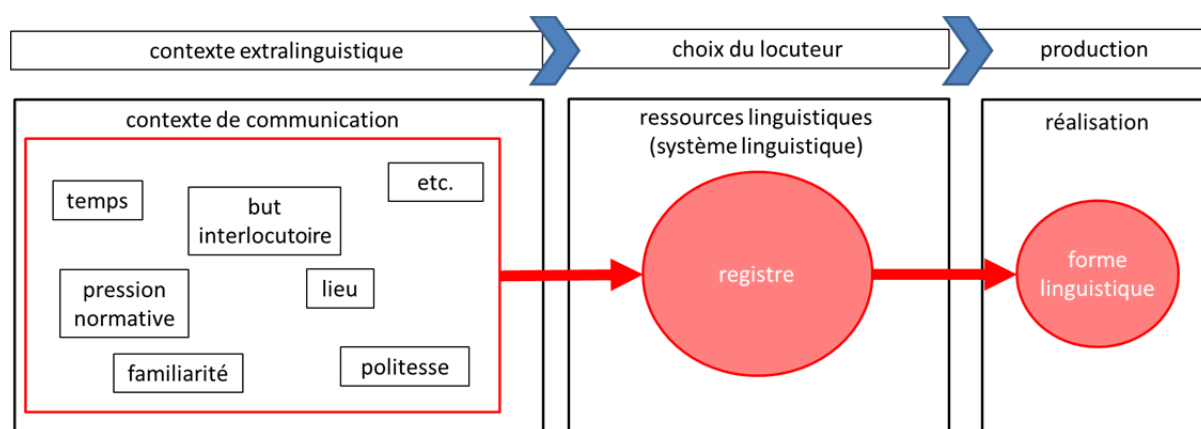


Fig.2: Contexte extralinguistique et registre

Nous avons ajouté la pression normative aux facteurs extralinguistiques cités par Biber (1995) et K&O (2011) dans le premier cadre de la Fig.2, car elle

³ Les numéros entre parenthèses se réfèrent au numéro d'identification du SMS dans le corpus (cf. section 3). L'orthographe originale du message est maintenue. Le symbole Ø note l'omission d'un argument. Les noms et les chiffres sont anonymisés.

joue un rôle clé pour la description de nos données (que nous avons caractérisées comme à priori non contraintes par des normes prescriptives, cf. section 1).

Les ressources linguistiques disponibles pour articuler un concept sont définies par le système linguistique respectif. Dans ce dernier s'inscrivent obligatoirement les traits qui constituent un registre, comme représenté dans le deuxième cadre de la Fig.2. La forme linguistique spécifique finalement choisie par le locuteur doit donc répondre aux exigences du registre approprié à une situation de communication et elle doit être conforme au système linguistique en question. Le locuteur produira une forme linguistique unique qui dépend de son interprétation individuelle des facteurs extralinguistiques, comme illustré par le troisième cadre de la Fig.2⁴.

Sur la question de savoir comment il faut intégrer les registres dans la compétence linguistique des locuteurs, il n'y a pas unanimité. Zwicky & Zwicky (1982) parlent d'un choix parmi les structures à disposition, tandis que Ferguson (1982) conçoit les différents registres plutôt comme des systèmes à part qui peuvent être activés à volonté. Or, si chaque registre est un système à part, les traits récurrents doivent être encodés dans chacun de ces systèmes avec leur fréquence respective, ce qui conduit à un stockage d'informations énorme et souvent répétitif. Nous préférons suivre la conception de Zwicky & Zwicky (1982) d'un registre qui se définit par un choix spécifique parmi les ressources linguistiques disponibles, qui ne sont pas choisies parce qu'elles appartiennent à un système spécifique, mais parce qu'elles sont appropriées dans un certain contexte extralinguistique. Le registre est, dans cette conception, un épiphénomène des choix effectués et n'existe, en tant que système cohérent, qu'en description linguistique.

Du point de vue linguistique, un registre est reconnaissable en termes quantitatifs par les *register features* (cf. Biber 1995: 28). Ces *register features*, des traits linguistiques élémentaires (p.ex. noms, pronoms, phrases subordonnées), caractérisent un certain registre par leur distribution quantitative spécifique correspondant au registre respectif. De plus, il existe les *register markers*, qui sont des propriétés linguistiques distinctives, à savoir qu'on retrouve uniquement dans un registre spécifique.

Par la fréquence des *register features* et la présence de *register markers* un registre devient empiriquement descriptible et quantifiable. Ce concept permet d'identifier empiriquement une fourchette avec un maximum et un minimum d'occurrences d'un certain trait par message, ou même de calculer sa

⁴

Le sujet de conversation influence également le choix de la forme linguistique, comme le fait remarquer un relecteur anonyme. Nous ne pouvons pas thématiser cette dimension ici, car elle supposerait une analyse de contenu des messages détaillée qui dépasserait le cadre de cette contribution. Quand il y a une contradiction résultant de la présence de différents interlocuteurs, nous supposons que l'exigence la plus formelle est celle qui déterminera le choix du locuteur.

moyenne d'occurrence sur l'ensemble d'un corpus (cf. Biber 1995: 314 à propos de la variation à l'intérieur de registres). Cette approche a l'avantage de rendre possible un procédé *bottom-up*, à savoir de partir de données concrètes au lieu de prendre comme point de départ un registre défini par l'un ou l'autre facteur extralinguistique (procédé *top-down*). Plus précisément, il devient possible de vérifier si deux ensembles définis théoriquement par des facteurs extralinguistiques montrent une distribution différente de *register features* et de *register markers*, à savoir si les deux ensembles se différencient également au niveau linguistique proprement dit.

Par exemple, on peut définir l'ensemble *journaux intimes* et l'ensemble *CMC* en termes extralinguistiques en se basant sur les différentes circonstances de production et de transmission comme suggéré dans la Fig.3. Ayant établi des corpus qui correspondent à ces ensembles, ceux-ci peuvent être analysés pour certains traits qu'on suppose être typiques de l'un ou l'autre ensemble, comme illustré schématiquement par les traits A-D à droite dans la Fig.3.

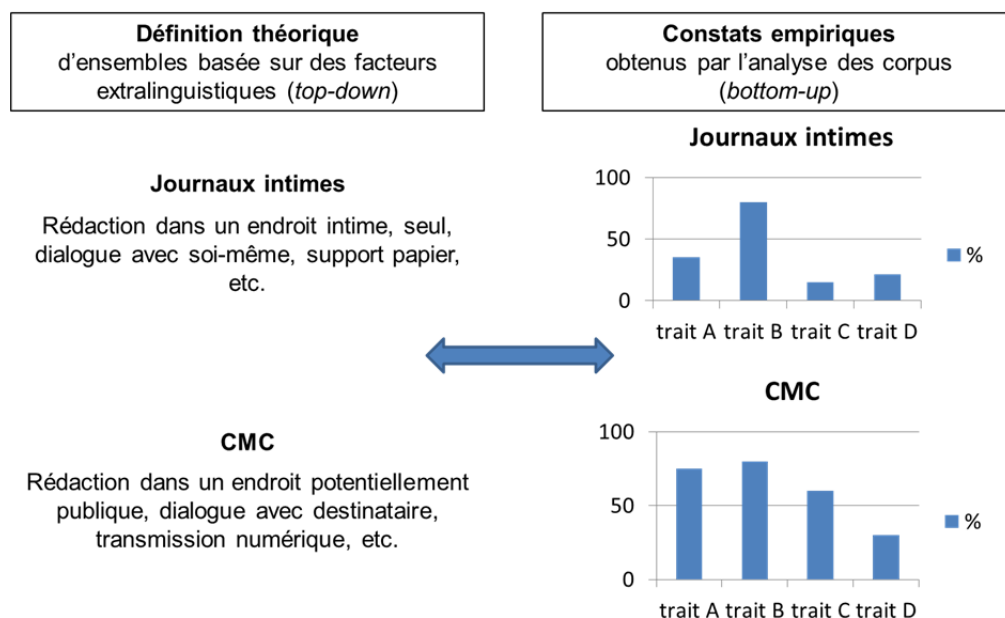


Fig.3: Les approches *top-down* et *bottom-up*

Dans l'exemple fictif de la Fig.3, on pourrait constater que les journaux intimes et la CMC se distinguent par la distribution des traits A et C, mais pas de manière significative par celle des traits B et D.

Puisque certains traits varient en fonction du contexte extralinguistique, tandis que d'autres montrent une distribution similaire (ce qui équivaut à dire que les traits ne sont pas nécessairement tous en co-variation), il est primordial de considérer les traits dits spécifiques d'un registre un par un, de manière à donner une caractérisation nuancée de ce registre et de le délimiter par rapport à d'autres registres.

Si la CMC constitue une variété ou un registre à part, comme suggéré dans la littérature (cf. section 1), il faut se poser la question de savoir en quoi la CMC

forme un ensemble qui s'oppose à d'autres ensembles du point de vue linguistique. Dans la prochaine sous-section, nous thématisons les traits linguistiques dits typiques de la CMC.

2.3 Les traits linguistiques dits typiques de la CMC

Panckhurst (2009: 37) donne une liste de caractéristiques (para)linguistiques des textes numériques de la CMC. Y figurent, entre autres, les erreurs de frappe de toutes sortes, la présence dominante du présent de l'indicatif, la néologie et la néographie ou encore beaucoup d'ellipses (p.ex. de sujets; ce qui est répertorié sous "[l']omission de mots grammaticaux" chez Fairon & al. 2006: 43).

Cet aperçu de Panckhurst (2009), que nous avons choisi à titre d'exemple, réunit toutes sortes de niveaux d'analyse et ne s'interroge pas sur l'origine des phénomènes constatés. Pourtant, ils ne sont pas tous déclenchés par les mêmes facteurs. Notamment, le facteur pertinent n'est que très rarement la technologie de transmission. Par exemple, nous attribuerons les fautes de frappe aux claviers des divers transmetteurs électroniques. Par contre, la présence dominante du présent de l'indicatif résulte probablement de l'interaction de type dialogique des communications CMC, car les dialogues se situent typiquement dans le présent, contrairement aux narrations classiques, souvent au passé. L'invention de nouveaux mots et de nouvelles formes graphiques peut être attribuée à une basse pression normative normalement rencontrée dans les contextes de CMC, ce qui permet plus de créativité et un emploi individuel des signes disponibles. Pour les ellipses, on pourrait invoquer un principe d'économie. Cependant, nous verrons dans la section 4 que les ellipses de sujets suivent des règles syntaxiques précises et qu'une explication en termes d'économie tout court est trop simpliste.

En résumé, la majorité des caractéristiques de la liste de Panckhurst (2009) n'est donc pas conditionnée par la technologie et se retrouve également dans d'autres contextes communicatifs (cf. aussi Thurlow & Poff 2013: 164). Ceci laisse déjà surgir des doutes quant à l'existence d'une variété ou d'un registre spécifique de la CMC ou des SMS.

Avant de passer au phénomène que nous avons choisi d'analyser à titre d'exemple pour un trait dit typique de la CMC, à savoir l'ellipse du sujet dans les SMS français (cf. p.ex. Panckhurst 2009; Fairon & al. 2006), nous présentons brièvement les corpus qui ont servi de base à cette étude.

3. Le Corpus SMS et les journaux intimes

Le corpus SMS étudié se compose de 4628 messages dont la langue principale est le français. Il s'agit d'une sous-partie d'un corpus plus vaste de

SMS de Suisse qui a été établi entre octobre 2009 et février 2010 (cf. Dürscheid & Stark 2011: 309 et www.sms4science.ch).

Les auteurs des messages sont presque exclusivement de langue maternelle française (95%), dont une bonne moitié a une formation de haute école.

Pour l'analyse syntaxique, le corpus a été annoté manuellement. Nous avons relevé un total de plus de 14'000 sujets que nous avons subdivisés en sujets nominaux et pronominaux. Ces derniers ont ultérieurement été classifiés comme toniques, clitiques ou zéro et comme référentiels ou explétifs.

Nous comparerons les résultats issus de ce corpus avec ceux rapportés par Haegeman (p.ex. 1999, 2013), qui a analysé des journaux intimes anglais et français. Pour l'anglais, l'échantillon de Haegeman contient 10 auteurs de journaux intimes qui ont été édités et dont le texte original le plus ancien date de 1936 (Woolf 1985) et le plus récent de 1997 (Isaacs 1997). Pour le français, Haegeman se base sur les journaux de Léautaud (1956 et 1986) dont l'original contient des entrées de 1917 à 1933.

Comme mentionné, il est possible de définir en termes extralinguistiques le registre SMS (représentatif de la CMC) d'une part, et les journaux intimes d'autre part (cf. section 2.2). À côté de caractéristiques communes comme le fait de ne pas devoir se conformer à priori à une norme prescriptive, les deux ensembles se différencient par exemple en ce qui concerne l'interlocuteur (une deuxième personne ou l'auteur-même), le lieu de rédaction (potentiellement public ou privé) ou le mode d'écriture et de transmission (manuscrit sur papier ou tapé sur un clavier et numérique).

Par la suite, il sera question de déterminer si ces différentes conditions de production et de transmission ont un impact sur la possibilité d'omettre le sujet, un trait dit typique de la CMC (cf. section 2.3).

4. Le phénomène linguistique observé: l'omission du sujet

Nous avons choisi d'analyser l'omission du sujet dans le corpus SMS, car les ellipses (de sujets) sont dites caractéristiques de la CMC (cf. section 2.3), bien qu'en français standard, le sujet soit obligatoire avec les verbes conjugués (sauf au mode impératif). C'est un phénomène qui a déjà été étudié dans les journaux intimes anglais et français (cf. ci-dessous).

L'anglais et le français standard ne font pas partie des langues à sujet zéro⁵ (*pro-drop*, dans lesquelles le sujet non-marqué n'est pas exprimé), ni des langues à omission du topique (*topic-drop*, qui ont la possibilité d'omettre le topique continu). Par contre, dans des registres qui ne sont pas ou peu soumis

⁵ Il est important de noter que l'anglais standard se comporte comme le français standard en ce qui concerne l'expression obligatoire du sujet. Ceci est une condition indispensable si on veut comparer les résultats issus de journaux intimes majoritairement anglais de Haegeman et les données des SMS français.

à la pression normative, comme p.ex. les journaux intimes, le sujet est parfois omis tant en français qu'en anglais. Selon Haegeman (2013), il ne s'agit cependant pas d'omissions du type *pro*- ni *topic*-drop.

Un argument pour le fait qu'il ne s'agisse pas de *pro*-drop est que l'omission du sujet dans les journaux intimes n'est pas possible après des interrogatifs, comme illustré en (3), ou des compléments préposés, comme en (4), avec l'italien *pro*-drop dans les exemples a, qui s'oppose au français respectivement à l'anglais dans les exemples b.

- (3) a. Quando *pro* tornerà?
b. Quand reviendra-t-*(elle) à Paris? (Haegeman 2000: 140)
- (4) a. Questo libro, *pro* non lo voglio.
b. This book, *(I) didn't like. (Haegeman 2000: 94, citant Wilder 1997)

En phrases subordonnées, un contexte typiquement sans sujet dans les langues *pro*-drop comme illustré en (5), Haegeman ne constate pas non plus de sujets omis dans les journaux intimes et elle suppose qu'une phrase du type (6) n'y est pas possible⁶.

- (5) I ragazzi cantano quando *pro* lavorano.
- (6) *Before Ø went, Clive came to tea. (Haegeman 2013: 94)

Haegeman (2013) exclut également la possibilité que l'omission du sujet soit un reflet de *topic*-drop. L'omission du topique dans les langues *topic*-drop n'est pas limitée au sujet, mais elle touche également les objets, comme dans l'exemple allemand en (7).

- (7) (Den) habe ich schon gesehen. (Haegeman 2013: 96, adapté)

Par contre, selon Haegeman (2013: 97), "[t]here is no systematic object drop in the diary register". Et finalement, les sujets explétifs, à savoir des expressions pronominales non référentielles, ne peuvent pas être des topiques et, par conséquent, ils ne peuvent pas être omis en langues *topic*-drop comme illustré en (8).

- (8) *(Es) wurde viel getanzt. (Haegeman 2013: 96)

Cependant, les sujets explétifs sont régulièrement omis dans les journaux intimes comme en (9).

- (9) Ø Won't be too difficult to reconstruct his argument. (Haegeman 2007: 103, citant Thrasher 1977: 44)

Haegeman fait aussi remarquer l'asymétrie entre compléments et ajouts préposés. Comme nous l'avons vu en (4)b, l'omission du sujet après un complément topicalisé n'est pas possible. Par contre, après des ajouts préposés, le sujet n'est pas obligatoire dans les journaux intimes, comme illustré par l'exemple (10).

⁶ Mais cf. Haegeman & Ihsane (2001) pour les exceptions dans les journaux intimes *fictifs*.

- (10) Jamais Ø ne se serait attendue à cela. (Léautaud, 1^e mars 1924, cité dans Haegeman 2013: 91)

Le Tab.1 résume les conditions d'omission d'arguments de différents types de langues et de registres discutés dans la littérature.

conditions	<i>pro</i> -drop	<i>topic</i> -drop	journaux intimes
omission en début de phrase	✓	✓	✓
omission après interrogatifs	✓	✗	✗
omission après compléments préposés	✓	✗	✗
omission après ajouts	✓	(✗) ⁷	✓
omission en subordonnées	✓	(✗) ⁸	✗
omission de sujets explétifs	- ⁹	✗ ¹⁰	✓
omission d'autres arguments que le sujet	✗ ¹¹	✓	✗

Tab.1: Les conditions d'omission d'arguments de différents types de langues et de registres

5. Résultats

Après avoir décrit les conditions d'omission du sujet dans les journaux intimes, nous présentons les résultats concernant les sujets omis issus du corpus SMS et nous les comparons avec ceux rapportés pour les journaux intimes.

Une remarque préliminaire concerne le taux de sujets omis. Ce dernier est calculé en proportion de pronoms personnels omis par rapport aux pronoms personnels clitiques explicites, puisque les autres sujets (à savoir lexicaux et pronominaux toniques) véhiculent des informations sémantiques et pragmatiques qui ne peuvent pas être remplacées par un élément nul (cf. Sigurðsson 2011: 289). Donc, ces sujets-là ne se prêtent pas à être omis indépendamment du type d'omission dont il s'agit (cf. section 4). De plus, les sujets pronominaux démonstratifs (*ça* et *ce/c'*) ne seront pas inclus dans les considérations qui suivent puisqu'ils ne sont jamais omis sauf dans la locution *comment va* (cf. Robert-Tissot 2015).

Nous pouvons constater qu'à grande échelle, la distribution du sujet omis décrite par Haegeman (2013) pour les journaux intimes français et anglais

⁷ Les langues *topic*-drop avec le verbe en deuxième position (V2) comme l'allemand ne permettent pas l'omission du topique ailleurs qu'en position initial. Les langues *topic*-drop qui ne sont pas soumises à cette restriction, comme p.ex. le portugais, ont la possibilité d'omettre un topique en d'autres positions linéaires.

⁸ Cf. la note 7.

⁹ Les langues *pro*-drop n'ont pas de sujet explétif explicite.

¹⁰ Ceci vaut pour les langues qui ne sont pas en même temps du type *pro*-drop.

¹¹ Ceci vaut pour les langues qui ne sont pas en même temps du type *topic*-drop.

(cf. section 4) se confirme aussi pour le corpus *sms4science.ch*. La Fig.4 offre un aperçu de l'occurrence de pronoms personnels clitiques sujets et de leur omission dans ce corpus. Pour assurer une comparabilité avec les études de Haegeman, nous présentons les données selon le contexte précédant le sujet¹².

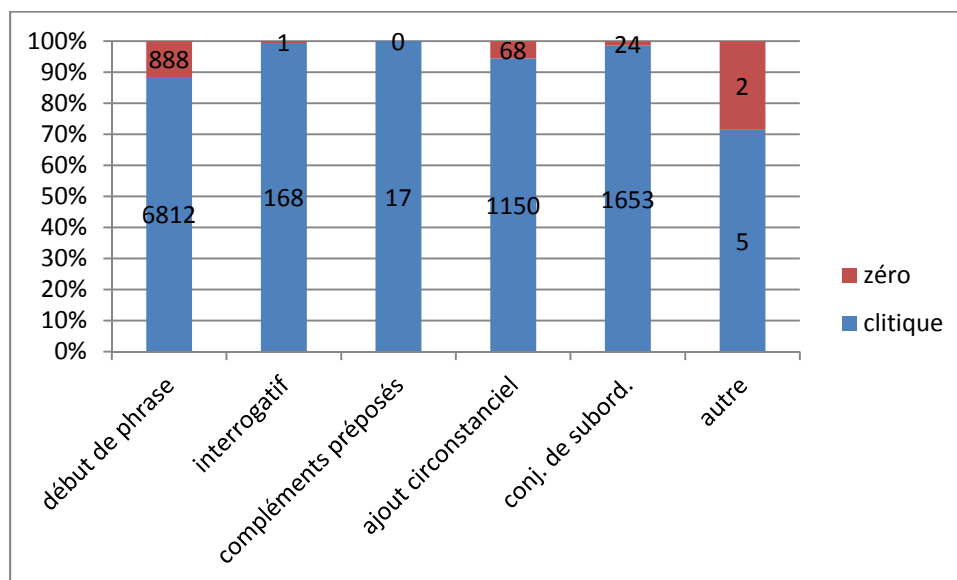


Fig.4: L'omission du sujet selon le contexte syntaxique précédant dans le corpus SMS

La majorité des omissions de sujets, qui ne sont notamment pas limitées à un certain type de verbes ou à une certaine personne grammaticale, se trouvent en début de phrases principales, comme dans l'exemple (11).

- (11) Oups, j'avais pas vu ton message. Ø Arrive ds 10min. A+ (14438)

Comme le montre toujours la Fig.4 ci-dessus, après un mot interrogatif un seul sujet est omis, à savoir un *il* explétif du verbe *y avoir* après la particule interrogative *est-ce que*¹³ dans l'exemple (12).

- (12) C est comment le nom de famille de sergio et est ce que Ø y a un code d entrée?
(9433)

Après des compléments préposés, aucun sujet n'est omis. Par contre, après des ajouts circonstanciels, l'omission du sujet est possible, comme illustré en (13).

- (13) Parfait lundi Ø suis libre merci, remet toi bien (20425)

En phrases subordonnées, nous avons une seule omission d'un sujet référentiel dans l'exemple (14).

¹² La particule interrogative *est-ce que* a été classifiée sous interrogatif; les phrases relatives sont répertoriées sous conjonction de subordination.

¹³ Cf. la note 12.

- (14) Coucou! Alors ton prog? Moi j te cache pas qu Ø suis pas mott à faire grd chose!
:-) (19036)

Ensuite, on y trouve deux occurrences de sujets explétifs omis avec *falloir* comme en (15).

- (15) [...] 1 truc ki est sur c'est ke Ø fo ke j'apprenne à fermé ma gueule=P !![...] (19459)

De plus, nous avons deux fautes de frappes évidentes, (des <le> pour des <je>) et deux exemples en phrases subordonnées qui sont ambigus quant à leur interprétation¹⁴. Finalement, nous avons 17 occurrences du sujet explétif *il* du verbe impersonnel *y avoir* omis en phrases subordonnées comme en (16).

- (16) Coucou. Bah écoute j'crois qu'il a lieu le cours, j'ai rien reçu qui disait que Ø y avait pas.. Il sèche d'nouveau! :) (15873)

Les exemples à sujet nul sous *autres* sont du type illustré en (17). Il s'agit d'une proposition sémantiquement subordonnée mais pas introduite par une conjonction de subordination.

- (17) [...] Espère Ø éviterons pluie. (14463)

En résumé, nous pouvons affirmer que l'omission du sujet *référentiel* suit les règles syntaxiques établies par Haegeman pour les journaux intimes: le sujet est omis uniquement en début de phrases principales ou après des ajouts et une seule fois dans une phrase subordonnée.

L'omission du sujet référentiel montre donc des parallèles frappants dans les SMS et les journaux intimes. Les déviations par rapport aux règles établies se présentent presque toutes soit avec le verbe *y avoir*, soit avec *falloir*, tous deux requérant un sujet explétif.

6. Discussion

Nous avons constaté que les deux ensembles définis en termes extralinguistiques par leurs conditions de productions différentes, à savoir les SMS et les journaux intimes, suivent les mêmes règles syntaxiques quant à la possibilité d'omettre un sujet référentiel et les proportions d'omissions sont comparables entre les deux ensembles, à savoir d'environ 10% en moyenne (cf. Haegeman 1999 et Robert-Tissot 2015). Ce trait n'est donc pas distinctif pour ces deux ensembles, ou, autrement dit, il ne reflète pas, au niveau linguistique, la distinction théorique se basant sur des facteurs extralinguistiques entre les deux ensembles.

Puisque le comportement de *y a* diffère d'un point de vue qualitatif et quantitatif des autres exemples montrant un sujet omis (le taux d'omission avec *y avoir* dépasse les 50%), nous supposons qu'il s'agit d'une fusion graphique de <il> avec <y>, imitant la fusion phonétique de [i(l)ja] à [ja] très

¹⁴ Pour des raisons d'espace, nous ne pouvons pas discuter ces exemples ici (cf. Robert-Tissot 2015 pour plus de détails).

fréquente dans le code phonique (cf. p.ex. Blanche-Benveniste 2010: 48). Cela n'implique pas nécessairement l'absence de l'explétif au niveau syntaxique. Ces exemples doivent probablement être traités à part. De plus, l'omission des explétifs est aussi largement attestée dans des corpus phoniques (cf. p.ex. Zimmermann & Kaiser 2014) et nous émettrons l'hypothèse qu'il s'agit d'un phénomène d'une nature différente de l'omission du sujet référentiel (cf. Robert-Tissot 2015 et Zimmermann 2014)¹⁵.

Par contre, à notre connaissance, aucune omission systématique du sujet référentiel n'a été signalée pour un corpus phonique de français européen. Par systématique, nous entendons l'omission d'après des règles syntaxiques précises et d'une ampleur dépassant ce qu'on pourrait classer de fautes de performance éparses. K&O (²2011: 258s.) remarquent, certes, l'ellipse du sujet dans un corpus phonique, mais de français québécois. Vu que ce phénomène n'est apparemment pas documenté dans d'autres variétés du français, les auteurs le qualifient de trait diatopique.

De plus, nous avons mené une étude exploratrice sur le corpus OFROM (Avanzi & Béguelin 2012) lors de laquelle nous avons cherché des sujets omis avec la forme verbale *suis* (cf. Robert-Tissot, soumis). Nous y avons relevé un seul passage avec deux omissions consécutives du même sujet dans l'exemple (18).

- (18) ouais d'ailleurs c'est ce qu'il me semblait bizarre au début quand euh j'ai entendu qu'elle faisait la colloc avec toi Ø me suis dit et ben | _ | _ | Ø me suis dit et ben c'est des filles qui ont rien à voir et puis qui se | _ | (OFROM, unifr12-jha, 215-287; _ signale une pause)

L'omission du sujet référentiel ne peut donc pas être considérée comme systématique en français phonique européen.

Nous constatons donc que le français qu'on peut observer dans les SMS montre une plus grande liberté quant aux ellipses de sujets référentiels, non seulement par rapport au français standard, mais aussi par rapport à d'autres registres (potentiellement) informels, surtout en code phonique, où l'omission du sujet référentiel est très rare. Par contre, ce type d'omission est bien attesté dans d'autres types de production informelle *graphique* comme en CMC, p.ex. sur Facebook, mais aussi dans les manchettes de journaux (cf. p.ex. Chow & al. 2008), les notes (cf. p.ex. Janda 1985), ou, comme nous l'avons vu, dans les journaux intimes. Comme exemple illustratif nous donnons en (19) un post sur Facebook.

¹⁵ Selon Zimmermann & Kaiser (2014), l'omission de l'explétif avec *falloir* par exemple (cf. (15)), peut être considérée comme un résidu de stades plus anciens du français. Pour y avoir, cf. ci-dessus.

- (19) Bientôt à Moscou pour signer un accord de coopération scientifique et rencontrer des acteurs culturels. Ø Me réjouis de revoir la Place Rouge (Facebook, 14.12.2012, conseiller fédéral Alain Berset)

Selon le procédé *bottom-up* que nous avons proposé dans la section 2.2, l'analyse empirique concernant l'omission du sujet référentiel suggère donc un regroupement des ensembles journaux intimes et SMS (subsumé sous la notion de 'discours informel graphique') qui se différencie d'un autre grand ensemble qu'on peut réunir sous le 'discours informel phonique'. Ce dernier est, certes, très hétérogène lui aussi, mais pour l'omission du sujet, il semble se comporter de manière uniforme (cf. p.ex. Fonseca-Greber 2004; K&O²2011; Blanche-Benveniste 2010).

La réalisation médiale est une différence frappante entre les deux grands ensembles 'discours informel phonique' et 'discours informel graphique'¹⁶. Nous constatons de manière empirique que le code graphique corrèle avec une répartition spécifique de sujets omis, qui diffère de celle du code phonique. Nous pensons que le choix du code dépend de facteurs extralinguistiques, comme p.ex. de la distance physique ou de la possibilité d'une réception simultanée et doit, par conséquent, être considéré comme un facteur avec une influence du moins intermédiaire sur le choix de la forme linguistique. Rien ne nous empêche donc d'inclure le code comme facteur pertinent pour définir un registre, bien qu'il ne soit pas explicitement mentionné chez Biber (1995) et que K&O le considèrent comme une dimension à part dans leur modèle (cf. section 2.1).

Si, comme dans la Fig.5 ci-dessous, nous représentons les registres définis en termes extralinguistiques par des cercles, ces derniers montrent des zones contenant des traits uniques ou distinctifs (ce qui correspond aux *register markers*, cf. section 2.2; p.ex. les erreurs de frappe dans les SMS, cf. section 2.3, ou les autocorrections dans le code phonique, cf. Blanche-Benveniste 2010: 60), mais aussi des intersections avec des traits en commun (p.ex. l'omission du sujet explétif). Ces derniers peuvent se manifester comme des *register features*, avec des taux d'occurrences plus ou moins élevés selon le registre. Le carré discontinu dans la Fig.5 représente l'espace du 'discours graphique', défini par le code utilisé.

¹⁶

On pourrait naturellement imaginer davantage de facteurs qui distinguent les deux ensembles, p.ex. la nature quasi-synchrone de la communication graphique vs. la synchronie dans la communication par code phonique (cf. Dürscheid 2003 et aussi Robert-Tissot soumis).

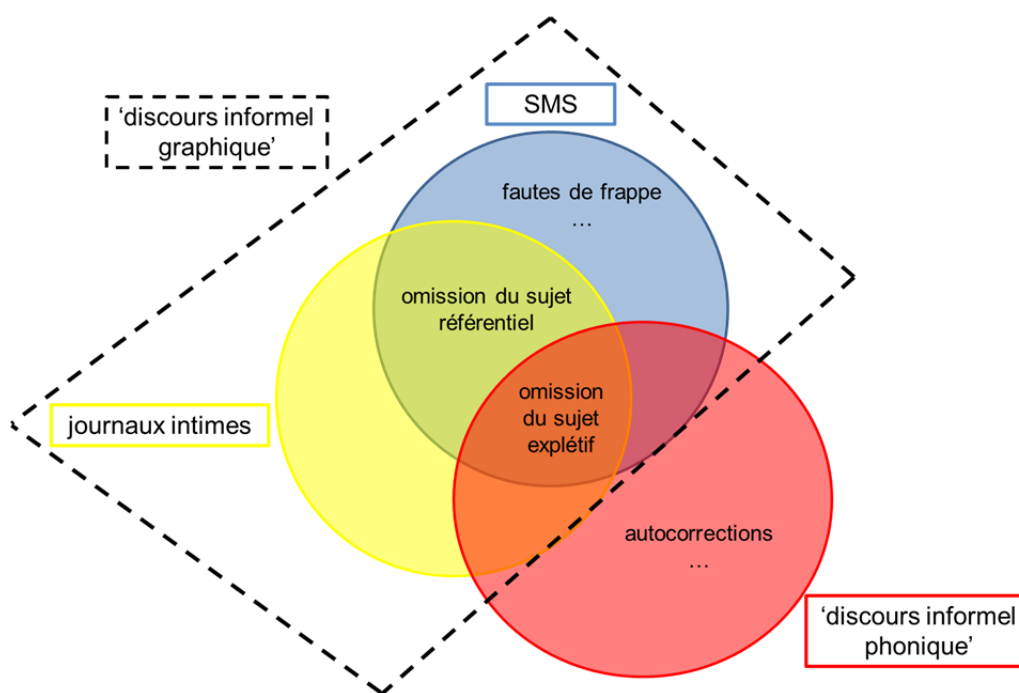


Fig.5: Les traits de différents registres

7. Conclusions

Nous avons suivi Biber (1995) dans l'idée que le contexte extralinguistique définit le registre. Selon notre conception, le registre n'existe pas en tant que système cognitif, mais il est un reflet d'un ensemble de choix qui s'opère selon les facteurs extralinguistiques.

La forme linguistique concrète doit s'inscrire dans le cadre des possibilités du système d'une langue. Ce système impose les restrictions structurelles que nous avons constatées pour la répartition des sujets omis.

Avec l'omission du sujet *référentiel*, nous avons identifié un *register marker* au sens de Biber (1995) pour les registres informels *graphiques*, qui oppose ces derniers aux registres *phoniques*. Nos résultats suggèrent effectivement que le code joue un rôle décisif pour la possibilité d'omettre un sujet référentiel puisque cette dernière reste restreinte au code graphique en français européen. Par contre, la distribution du sujet omis ne reflète aucune différence entre les SMS (faisant partie de la CMC) et d'autres textes graphiques informels, parfois bien plus anciens comme les journaux intimes¹⁷.

Pour répondre à la question posée par notre titre, nous pouvons donc affirmer que, du moins pour le trait que nous avons analysé, le registre est majoritairement indépendant du moyen de communication ou de la technologie de transmission. Ces derniers ne conduisent pas à la constitution

¹⁷ Selon un relecteur anonyme, certains écrivains pratiquent également l'ellipse du sujet, ce qui est parfaitement dans la ligne de notre argumentation.

d'un nouveau registre, bien que certains phénomènes (p.ex. les fautes de frappe) puissent être liés à des conditions techniques de la production. Le moyen de communication n'est alors qu'une petite composante des conditions extralinguistiques qui, dans leur ensemble, déterminent la forme linguistique concrète.

Bibliographie

- Avanzi, M. & Béguelin, M.-J. (2012): Présentation du corpus OFROM – corpus oral de français de Suisse romande. Université de Neuchâtel. Disponible: <http://www.unine.ch/ofrom>. (10.12.2013)
- Biber, D. (1995): *Dimensions of register variation*. Cambridge (Cambridge University Press).
- Blanche-Benveniste, C. (2010): *Approches de la langue parlée en français*, nouv. éd. Paris (Ophrys).
- Chow, I. & al. (2008): La syntaxe, la morphologie et la phonologie de la réduction dans les titres. In: *French Studies Publications*, 89. Disponible: <http://ir.lib.uwo.ca/frenchpub/89>. (7.11.2014)
- Crystal, D. (2008): *Txtng: The Gr8 Db8*. Oxford (Oxford University Press).
- Dufter, A. & Stark, E. (2002): La variété des variétés-combien de dimensions pour la description? Quelques réflexions à partir du français. In: *Romanistisches Jahrbuch* 53, 81–108.
- Dürscheid, C. (2003): Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit. Theoretische und empirische Probleme. In: *Zeitschrift für Angewandte Linguistik*, 38, 37–56.
- (2004): Netzsprache - ein neuer Mythos. In: Beisswenger, M., Hoffmann, L. & Storrer, A. (éds.): *Internetbasierte Kommunikation*. Duisburg (Universitätsverlag Rhein-Ruhr), 141-157.
- Dürscheid, C. & Brommer, S. (2009): Getippte Dialoge in neuen Medien. Sprachkritische Aspekte und linguistische Analysen. In: *Linguistik Online*, 37. Disponible: http://www.linguistik-online.com/37_09/. (7.11.2014)
- Dürscheid, C. & Stark, E. (2011): SMS4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland. In: Thurlow, C. & Mroczek, K. (éds.): *Digital Discourse. Language in the New Media*. Oxford (Oxford University Press), 299-320.
- Fairon, C., Klein, J. R. & Paumier, S. (2006): *Le Langage SMS. Étude d'un corpus informatisé à partir de l'enquête "Faites don de vos SMS à la science"*. Louvain-la-Neuve (Presses Universitaires de Louvain).
- Ferguson, C.A. (1982): Simplified Registers and Linguistic Theory. In: Obler, L. K. & Menn, L. (éds.): *Exceptional Language and Linguistics*. New York et al. (Academic Press), 49-66.
- Fonseca-Greber, B. B. (2004): Zero Marking in French Impersonal Verbs. A Counter Trend to clitic Morphologization ? In: *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 30, 81-92.
- Haegeman, L. (2000): Adult null subjects in non pro-drop languages. In: Friedemann, M.-A. & Rizzi, L. (éds.): *The Acquisition of Syntax. Studies in Comparative Developmental Linguistics*. Harlow et al. (Longman), 129-169.
- (2007): Subject omission in present-day written English: on the theoretical relevance of peripheral data. In: *Rivista di Grammatica Generativa* 32, 91-124.
- (2013): The syntax of registers: Diary subject omission and the privilege of the root. In: *Lingua* 130, 88-110.
- Haegeman, L. & Ihsane, T. (2001): Adult Null Subjects in the non-pro-drop Languages: Two Diary Dialects. In: *Language Acquisition* 9, 329-346.

- Janda, R. D. (1985): Note-Taking English as a Simplified Register. In: *Discourse Processes* 8, 437-454.
- Koch, P. & Oesterreicher, W. (²2011): *Gesprochene Sprache in der Romania. Französisch, Italienisch, Spanisch*. Berlin (de Gruyter).
- (2001): *Gesprochene Sprache und geschriebene Sprache. Langage parlé et langage écrit*. In: Holtus, G., Metzeltin M. & Schmitt, C. (éds.): *Lexikon der romanischen Linguistik*, vol. 1,2. Tübingen (Niemeyer), 584-627.
- Massot, B. (2010): Le patron diglossique de variation grammaticale en français. In: *Langue française* 168, 87-106.
- Panckhurst, R. (2009): Short Message Service (SMS). Typologie et problématiques futures. In: Arnavielle, T. (éd.): *Polyphonies, pour Michelle Lanvin*. Montpellier (Université Paul Valéry Montpellier 3), 33-52.
- Robert-Tissot, A. (2015): Le sujet et son absence dans les SMS français. Une analyse basée sur le corpus de sms4science suisse. Thèse doctorale.
- Robert-Tissot, A. (soumis): Sur quelques propriétés des sujets entre l'oral et l'écrit dans les SMS français.
- Sigurðsson, H. A. (2011): Conditions on Argument Drop. In: *Linguistic Inquiry* 42, 267-304.
- Thurlow, C. & Poff, M. (2013): Text messaging. In: Herring, S., Stein, D. & Virtanen, T. (éds.): *Pragmatics of Computer-Mediated Communication*. Berlin/Boston (de Gruyter), 163-188.
- Zimmermann, M. & Kaiser, G. A. (2014): On expletive subject pronoun drop in Colloquial French. In: *Journal of French Language Studies* 24, 107-126.
- Zimmermann, M. (2014): *Expletive and Referential Subject Pronouns in Medieval French*. Berlin/Boston (de Gruyter).
- Zwicky, A. M. & Zwicky, A. D. (1981): Telegraphic Registers in written English. In: Sankoff, D. & Cedergren, H. (éds.): *Variation omnibus*. Carbondale/Edmonton (Linguistic Research Inc.), 535-544.
- (1982): Register as a Dimension of Linguistic Variation. In: Kittredge, R. & Lehrberger, J. (éds.): *Sublanguage. Studies of Language in Restricted Semantic Domains*. Berlin/New York (de Gruyter), 213-218.

Corpus sms4science.ch:

Demande d'accès via <http://www.sms4science.ch/index.html>

Corpus Haegeman (œuvres citées):

- Isaacs, S. (1997): *Lily White*, London (Penguin Books).
- Léautaud, P., (1956): In: Silve, E. (éd.): *Le Fléau*, Journal particulier. 1917-1930. Mercure de France, 69-70.
- (1986): In: Silve, E. (éd.): *Journal Particulier*. 1933. Mercure de France.
- Woolf, V. (1985): *The Diary of Virginia Woolf. Volume 5 : 1936-1941*, édité par Olivier Bel, A. & McNeillie, A. Middlesex (Penguin Books).