

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

L'ÉLISION DU ⟨I⟩ DES PRONOMS DANS LES TEXTOS EN FRANÇAIS QUÉBÉCOIS :
ÉTUDE SOCIOLINGUISTIQUE SUR LE LANGAGE SMS

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN LINGUISTIQUE

PAR

VINCENT ROUSSEAU

SEPTEMBRE 2025

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je souhaite d’abord remercier mon directeur de mémoire, Philip Comeau, pour son soutien constant, ses précieux conseils et commentaires ainsi que son expertise. Ils auront grandement contribué à l’aboutissement de ce travail de recherche. Merci d’avoir maintenu mon enthousiasme pour mon projet, et ce depuis le début dans ton cours de sociolinguistique alors que je formulais encore mes idées pour mon mémoire. Ton aide m’aura permis de mener à terme mon projet le plus important.

Je souhaite aussi remercier les membres de mon jury Patrick Drouin et Mireille Tremblay. Merci d’avoir accepté de me consacrer du temps et de l’énergie. J’espère avoir produit un mémoire à la hauteur de vos attentes.

Je remercie également Daphnée Simard pour son soutien inestimable. Merci d’avoir remarqué mon potentiel et d’avoir pris une chance avec moi. Merci de m’avoir donné le courage de me lancer dans une maîtrise. Sans tes encouragements, je ne me serais probablement jamais lancé dans un projet d’une telle envergure et c’est avec beaucoup de fierté que je vois aujourd’hui que tu avais raison de croire en moi.

Finalement, je souhaite souligner la généreuse contribution de M. Philippe Langlais et M. Patrick Drouin, de l’Université de Montréal, qui m’auront donné accès à leur gigantesque corpus sans lequel ce projet n’aurait pas pu voir le jour.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX	vii
RÉSUMÉ	viii
ABSTRACT	ix
INTRODUCTION	1
CHAPITRE 1 PROBLÉMATIQUE ET CADRE THÉORIQUE.....	4
1.1 Le langage SMS	4
1.1.1 Quelques distinctions	8
1.2 Linguistique et discours populaire	9
1.2.1 Problème de réputation	10
1.3 Typologie des textismes	14
1.4 Variation orthographique	17
1.5 Questions de recherche	18
CHAPITRE 2 ÉTAT DES CONNAISSANCES	19
2.1 L'élision du /l/ à l'oral	19
2.1.1 Français laurentien	21
2.1.2 Français hexagonal	21
2.2 Travaux de corpus sur les textos	22
CHAPITRE 3 MÉTHODOLOGIE.....	25
3.1 Présentation du corpus <i>Texto4science</i>	25
3.2 Participants	27
3.3 Domaine d'application de la variable	28
3.4 Procédures et analyse statistique (Rbrul)	34
3.5 Facteurs à l'étude	35
3.5.1 Facteurs linguistiques et hypothèses	36
3.5.1.1 Contexte phonologique suivant/précédant l'occurrence	37
3.5.1.2 Catégorie grammaticale de l'occurrence	38
3.5.1.3 Catégorie syntaxique du mot suivant/précédant l'occurrence	39
3.5.1.4 Longueur des textos	40
3.5.2 Facteurs sociaux et hypothèses	42

3.5.2.1	Âge	44
3.5.2.2	Genre	46
3.5.2.3	Niveau de scolarité.....	47
3.5.2.4	Niveau de langue requis pour l'emploi	47
3.5.2.5	Depuis combien de temps le participant texte	50
3.5.2.6	Nombre de textos envoyés par jour	50
3.5.2.7	Capacité de décodage du langage SMS	50
3.5.2.8	Utilisation de dictionnaires.....	51
3.5.2.9	Type de clavier utilisé	51
3.5.2.10	Nombre d'heures de lecture et d'écriture par semaine	52
CHAPITRE 4 RÉSULTATS.....		53
4.1	Taux globaux.....	53
4.2	Analyses linguistiques	55
4.2.1	Pronoms sujets masculins.....	55
4.2.2	Pronoms féminins	58
4.2.3	Pronoms objet indirects.....	60
4.3	Analyse sociologique	61
CHAPITRE 5 DISCUSSION		65
5.1	De l'oral à l'écrit	65
5.2	Contexte phonologique versus contexte graphique	66
5.3	Paradoxe de la longueur des textos.....	68
5.4	Valeur sémantique comme obstacle à l'élision.....	71
5.5	Âge et technologie.....	72
5.6	La scolarité et le langage SMS	73
5.7	Les impondérables.....	74
CONCLUSION		78
6.1	Retour sur les questions de recherche	78
6.1.1	Question 1 : Quels facteurs linguistiques influencent l'élision dans les textos ?	78
6.1.2	Question 2 : Quels facteurs sociaux influencent l'élision dans les textos ?	79
6.1.3	Question 3 : De quelle manière les résultats de l'analyse des textos se comparent-ils aux résultats de l'analyse de la langue orale ?	79
6.2	Perspectives de recherche	80
ANNEXE A Questionnaire original pour l'étude <i>Texto4science</i>		83
ANNEXE B Exemple de SMS annoté dans le corpus <i>Texto4science</i>		86
ANNEXE C Exemples des variantes pour chaque catégorie de pronom.....		87
ANNEXE D Feuille de codage		89

ANNEXE E Commentaires des participants92

RÉFÉRENCES94

LISTE DES FIGURES

Figure 1 - Taux d'élision par tranche d'âge	63
Figure 2 - Distribution des participants par tranche d'âge.....	63
Figure 3 Comparaison des groupes d'âge selon les types de claviers et les taux d'élision.....	73

LISTE DES TABLEAUX

Tableau 1.1 Différences entre les moyens de communication.....	9
Tableau 1.2 Typologie des textismes	15
Tableau 2.1 Études sur l'élision du /l/ à l'oral, comparaison entre les communautés	20
Tableau 3.1 Distribution des participants par région.....	27
Tableau 3.2 Distribution des participants par emploi.....	28
Tableau 3.3 Distribution des pronoms omis	32
Tableau 3.4 facteurs à l'étude	36
Tableau 3.5 Facteurs sociaux omis.....	44
Tableau 3.6 Niveau de langue requis par l'emploi	49
Tableau 4.1 Taux globaux, pronoms sujets masculins	53
Tableau 4.2 Taux globaux, pronoms sujets féminins	54
Tableau 4.3 Taux globaux, pronoms objets indirects.....	54
Tableau 4.4 Comparaison des taux d'élision entre les études.....	55
Tableau 4.5 Analyse linguistique des pronoms sujets masculins.....	56
Tableau 4.6 Analyse linguistique des pronoms féminins	59
Tableau 4.7 Analyse linguistique des pronoms objets indirects	60
Tableau 4.8 Analyse sociologique	62
Tableau 5.1 Distribution des pronoms masculins élidés.....	67
Tableau 5.2 Valeur sémantique des pronoms.....	71
Tableau 5.3 Rapport entre l'attitude des participants envers le langage SMS et leur taux d'élision	76

RÉSUMÉ

Cette étude sociolinguistique examine l'élision du ⟨l⟩ dans les pronoms *il(s)*, *elle* et *lui* dans les messages textes en français québécois. Elle a comme objectif d'identifier les facteurs linguistiques et sociologiques qui influencent ce phénomène et de comparer les résultats avec ceux rapportés dans des études sur l'oral pour des communautés linguistiques similaires. Pour ce faire, une analyse approfondie du corpus de messages textes québécois *Texto4science* (Langlais et Drouin, 2012) a été réalisée afin de révéler les *patterns* de textage de la communauté québécoise.

Les messages textes ont exhibé des taux d'élisions largement inférieurs aux taux observés dans les études sur le phénomène à l'oral. Les résultats révèlent que l'élision est favorisée dans des contextes prévocaux (pour les pronoms sujets) et postvocaux (pour les pronoms objets indirects). Elle était aussi plus prévalente chez les participants plus jeunes. En revanche, l'élision était défavorisée dans les messages plus longs. De plus, la batterie de facteurs influençant l'élision dans les textos s'est avérée différente entre les deux médiums. Notamment, les marqueurs sociologiques classiques (le genre, le niveau de scolarité, le domaine d'emploi, etc.), à la seule exception de l'âge, n'avaient pas d'impact statistiquement significatifs sur les taux d'élision dans les textos. Les facteurs propres à l'usage du langage SMS (le nombre de textos envoyés par jour, la capacité de décodage, le type de clavier utilisé ou l'utilisation de dictionnaire ou de fonctions d'autocorrection et d'autocomplétion) n'avaient pas non plus d'effets importants sur l'élision. Ces résultats suggèrent que le médium de l'écrit et la rigidité du code écrit standard limitent dans une certaine mesure la variabilité du langage SMS.

Mots clés : Messages textes, élision de ⟨l⟩, Analyse de corpus, Français québécois, Sociolinguistique.

ABSTRACT

This sociolinguistic study examines ⟨l⟩ deletion in pronouns *il(s)*, *elle* and *lui* in Quebec French text messages. The objective was to identify the linguistic and social factors that influence ⟨l⟩ deletion in text messages and to compare the results with the results of studies on the oral phenomenon in similar linguistic communities. To do so, an analysis of the Quebecois French text message corpus *Texto4science* (Langlais et Drouin, 2012) was conducted in order to reveal the texting patterns of Quebecois French texters.

Text messages showed rates of ⟨l⟩ deletion that were far inferior to those shown in oral studies. The results revealed that deletion was favored in prevocalic (for subject pronouns) and postvocalic (for object pronouns) environments. It was also favored by younger texters. On the other hand, deletion was disfavored in longer messages. Moreover, the battery of factors influencing ⟨l⟩ deletion turned out to be different between the two medium. Notably, the classical sociological markers (gender, level of education, professional domain, etc.), apart from age, turned out to have no statistically significant effect on deletion rates. Factors specific to the use of SMS language (number of texts sent per day, ease of decoding, the type of keyboard used or the use of dictionaries and functions like autocorrect and autocomplete) had no significant effect on deletion rates. These results suggest that the written medium and the rigidity of the standard written code limit the variability of SMS language to a certain extent.

Keywords : SMS messaging, ⟨l⟩ deletion, Corpus analysis, Quebec French, Sociolinguistics.

INTRODUCTION

En 1992, Neil Papworth envoyait le premier message texte, dorénavant texto, à son collègue Richard Jarvis. Il contenait un simple message : *Merry Christmas [Joyeux Noël]*. À cette époque, la technologie des messages textes était très limitée. Rappelons-nous qu'il y avait une limite de 160 caractères (incluant les espaces) par textos (Crystal, 2008) et qu'il fallait appuyer plusieurs fois sur une même touche pour produire un seul caractère. De plus, les textos étaient coûteux à envoyer. Le langage SMS (ou *textese*) est apparu en réponse à ces limitations afin de les contourner en incorporant abréviations et autres conventions d'écriture qu'on appelle des textismes et qui nous rappellent les télégrammes du 19^e siècle. Ce nouveau code avait pour but d'accélérer la communication entre les texteurs¹, de réduire les coûts d'usage et d'augmenter la quantité d'informations qu'on pouvait transmettre par message. L'efficacité, pas la qualité de la langue, était le mot d'ordre (Verheijen, 2013). Sept ans plus tard, les textos pouvaient maintenant être échangés entre plusieurs réseaux. Ainsi, entre les années 2000 et 2010, le nombre de textos envoyés par mois à travers le monde a augmenté de 12 millions à 135 milliards (Battestini et al., 2010). Les textos deviennent dès lors l'un des moyens de communication les plus populaires sur la planète (Squires, 2012 ; Thurlow et Brown, 2003), particulièrement chez les jeunes. Aujourd'hui, la technologie a grandement progressé et ces limitations ont en grande partie disparu. La majorité des appareils ont un clavier QWERTY répliquant les claviers d'ordinateurs, ce qui facilite grandement la rédaction de messages. Les nouvelles fonctions comme l'autocorrection et l'écriture prédictive rendent les messages textes encore plus faciles et rapides à rédiger en corrigeant automatiquement les fautes et en proposant des mots souvent employés par l'utilisateur, de sorte qu'il soit aujourd'hui possible d'écrire des messages complexes et sans fautes extrêmement rapidement : les messages s'écrivent pratiquement d'eux-mêmes (De Jonge et Kemp, 2012).

Si le code écrit standard est strict, les textos, quant à eux, jouissent d'une liberté sans égal. De cette liberté émerge l'hybridité dans les textos, le fait qu'ils emploient à la fois des conventions orales et écrites (Crystal, 2009). Les textos deviennent un outil avec lequel les texteurs jouent avec la langue et expriment leur identité (Blondeau et al., 2014 ; Crystal, 2008, 2009 ; Holtgraves, 2011 ; Tagg, 2009 ; Waldron et al., 2017). Ils permettent de déformaliser la communication écrite et d'imiter en partie le discours oral, facilitant

¹ Le masculin générique est employé à travers ce mémoire pour en faciliter la lecture. Il n'est pas une marque d'exclusion. Pour plus de discussion sur les effets du genre sur les comportements de textage, voir les chapitres 3, 4 et 5.

l'expression d'émotions et introduisant un aspect ludique et transgressif à la communication écrite. L'hybridité des textos en fait une anomalie linguistique qui intéresse autant la communauté de chercheurs que la société « linguistiquement naïve » (Paveau, 2007), qui en craint les effets néfastes sur la langue normée. Depuis leur apparition à la fin des années 90, les formes orthographiques non standards qui caractérisent le contenu typique des textos font l'objet d'un regard minutieux, inspirant un corps grandissant de recherche sur les textos et leur impact sur le niveau de compétence langagière des auteurs, particulièrement à l'écrit. Si de nombreuses recherches ont cherché à étudier cette particularité, il n'en demeure pas moins que la question persiste encore aujourd'hui : quels modules ou aspects de la langue des textos se comportent comme l'oral et comme l'écrit ? Autrement dit, comment se manifeste l'hybridité dans les textos ?

Le projet de recherche présent porte spécifiquement sur l'élision du ⟨l⟩ (le caractère) dans les pronoms personnels *il(s)* et *elle*², le pronom impersonnel *il* et le pronom objet indirect (POI) *lui* dans les messages textes en français québécois. À l'oral, l'élision du /l/ (le phonème) est un phénomène largement couvert dans la littérature depuis les années 70 (Armstrong, 1996 ; Ashby, 1984 ; Howard et al., 2006 ; Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976 ; Sax, 2003). Ces recherches montrent que l'élision du /l/ à l'oral dans les pronoms est un phénomène qui se manifeste très fréquemment dans l'usage (environ 80% et plus selon les études). Je n'ai trouvé aucune recherche sur ce phénomène à l'écrit parce que, en français, dans une perspective synchronique, on ne trouve pas de variation dans l'orthographe des pronoms dans la langue normée et toutes les formes non standards sont simplement interprétées comme des fautes. Le médium de la MI et le langage SMS, quant à eux, sont beaucoup plus libres. Ils s'avèrent donc être des terrains propices à explorer pour voir si l'élision se manifeste de la même manière à l'oral qu'à l'écrit. Autrement dit, il permettra de voir si nous écrivons vraiment comme nous parlons dans les textos et, ce faisant, il contribuera au débat sur les risques que présentent les textos pour la qualité de la langue des auteurs.

Le projet présent se donne donc comme objectif d'étudier la question du statut hybride des messages textes et d'identifier quels facteurs linguistiques et sociologiques influencent l'usage des formes non standards d'orthographe dans les textos en français québécois. Ces résultats seront ensuite comparés aux résultats rapportés dans les études sur l'élision à l'oral afin de voir comment ces deux phénomènes se

² Pour des raisons qui seront expliquées dans la section 3.3, les pronoms féminins pluriels ont dû être exclues de l'étude.

comparent. Dans le cas où le phénomène serait similaire dans les deux médiums, nous aurions des raisons de croire que la langue employée dans les textos serait un reflet du parler et donc qu'il pourrait y avoir un certain degré d'interférence entre les médiums. Dans le cas contraire, nous pourrions croire que le *textese* se comporte comme une langue à part régie par ses propres règles. Ultimement, le projet actuel servira, je l'espère, à défendre les textos (et par la même occasion toutes les variétés non standards) face aux accusations qui leur sont souvent lancées selon lesquelles leur emploi serait responsable du déclin de la qualité de la langue.

CHAPITRE 1

PROBLÉMATIQUE ET CADRE THÉORIQUE

1.1 Le langage SMS

Le langage SMS est le nom donné au langage qu'on retrouve dans les textos (Crystal, 2008). Ce langage est caractérisé par un relâchement des normes d'écriture standards. Contrairement à la langue normée, il priorise la rapidité de communication (Verheijen, 2013). Il existe un débat quant à la nature du langage SMS. En effet, il n'y a pas consensus à savoir si le langage SMS constitue une nouvelle langue à proprement parler avec des normes écrites distinctes d'autres langues naturelles (Crystal, 2008), un dialecte des langues standards (Shafie et al., 2010 ; Sutherland, 2002) ou un registre de langue (Tagliamonte, 2016). Bien que ce débat ne soit pas au cœur du projet de recherche actuel parce que mes données ne me permettent pas de me positionner dans celui-ci, il demeure tout de même important de l'aborder puisqu'il existe un lien clair entre la perception des non-linguistes (médias, enseignants et parents) et le statut du langage SMS. En effet, si l'on considère que le langage SMS est un dialecte de la langue standard, on pourrait croire que les formes non standards qui caractérisent son usage pourraient avoir un impact négatif sur la compétence des auteurs dans la langue standard (Mphahlele et Mashamaite, 2005) parce qu'ils partagent une grammaire. Inversement, si l'on croit que le langage SMS constitue une nouvelle langue distincte, nous pourrions étudier le langage SMS du point de vue du bilinguisme et de ses bénéfices sur la cognition (Bialystok, 2009 ; Bialystok et al., 2004 ; Bialystok et Martin, 2004 ; Crawford, 1998 ; Martin-Rhee et Bialystok, 2008 ; Swanson et al., 2004). Certains auteurs (Craig, 2003 ; Verheijen, 2013) affirment même que les textos pourraient être un outil pédagogique à l'aide duquel les enfants développeraient leurs connaissances métalinguistiques, ce qui favoriserait à son tour le développement de leur littératie. Finalement, si le langage SMS est un registre, on pourrait concevoir les différences entre la langue standard et le langage SMS de la même manière que l'on conçoit le vernaculaire par rapport à la langue standard (Blondeau et Tremblay, 2022). Il existe également un débat quant à l'approche qui explique mieux la variation dans les textos ainsi que leur caractère hybride : l'approche variationniste (Labov, 1972) et l'approche diglossique (Blondeau et Tremblay, 2022 ; Ferguson, 1959). Les deux approches diffèrent dans le nombre de grammaires qu'elles requièrent pour expliquer la variation entre la langue standard et les variétés, que ce soit le vernaculaire, les variantes régionales ou le *textese*. Selon l'approche variationniste (Labov, 1972), le langage SMS serait conçu comme partageant la même grammaire que le français standard. La variation qu'on retrouve dans les textos serait conditionnée par des facteurs sociaux et

linguistiques et par le contexte de locution. À l’opposé, selon l’approche diglossique (Blondeau et Tremblay, 2022 ; Ferguson, 1959 ; Massot, 2008; Palasis, 2013; Rowlett, 2013), on conçoit que les locuteurs emploient différentes grammaires en relation diglossique, c’est-à-dire qu’une des grammaires a un statut privilégié et l’autre un statut dévalorisé. Ultimement, peu importe comment on considère le langage SMS, il y aura toujours une grande proportion de non-linguistes (et certains linguistes) qui verront le langage SMS comme une menace à la langue, que ce soit une menace de l’interne par dégradation du standard ou de l’externe par contact de langues. À noter ici que le langage SMS est surtout étudié en anglais dû à son statut de *lingua franca* (Internet Society Foundation, 2023), mais il serait réducteur de limiter la portée du langage SMS à l’anglais seulement. En effet, même si les textismes ne sont pas tous les mêmes dans toutes les langues, on retrouve le langage SMS avec de nombreuses langues. Le terme *langage* renvoie donc à l’idée selon laquelle le langage SMS n’est pas une langue à proprement parler, mais un terme collectif qui regroupe toutes les « langues SMS » et qui englobe les pratiques de textismes qui caractérisent son usage. Pour l’étude présente, la variété étudiée est le *textese* québécois. Cette variété est conçue comme une variété à part entière qui partage de nombreux traits avec le français standard, mais qui se distinguent suffisamment dans sa grammaire et ses pratiques langagières pour mériter un statut à part.

Malgré son apparence chaotique, (a) le langage SMS (toutes langues confondues) est fonctionnel, (b) il répond à des principes et (c) il est expressif (Tagg et al., 2012). D’abord, le langage SMS est fonctionnel dans le sens qu’il émerge de l’usage et qu’il remplit une fonction de brièveté et de vitesse, ce qui explique pourquoi on retrouve entre autres des abréviations (*cbr* = *crampé bin raide*), une orthographe phonétique (*skool* = *school*) et des homophonies de lettres, de nombres et de symboles (*a+* = *à plus*). Ensuite, on dit que le langage SMS répond à des principes parce qu’il répond tout de même à une certaine forme de régularité, même s’il n’adhère pas strictement aux normes standards de l’écrit. Autrement dit, les variations orthographiques sont flexibles, mais demeurent interprétables en suivant les règles d’écriture standards. L’orthographe phonétique, par exemple écrire le mot *école* comme *ekol* dans les textos, illustre bien ce fait. D’abord, les accents sont souvent omis dans les textos puisqu’ils sont largement superflus quand vient le temps d’interpréter le contenu d’un texto (Stark et al., 2018), en plus d’être laborieux à produire sur certains claviers de téléphone. Il est effectivement facile de lire le *e* dans l’exemple *ekol* parce que le graphème ⟨e⟩ se lit souvent comme un [e] ((effet) [efe]) et que le reste du mot ressemble suffisamment à *école* pour qu’il soit interprété comme tel. Ensuite, le graphème ⟨k⟩ se lit comme le phonème /k/, tout comme le graphème ⟨c⟩ lorsque suivi d’un ⟨a, o, u⟩. On pourrait peut-être même imaginer que ⟨k⟩ serait plus facilement interprétable pour ceux qui présente des problèmes de lecture

puisqu'il se lit toujours comme un /k/ et qu'il a une forme plus facilement lisible sur des petits écrans comme des écrans de téléphones. À ce sujet, les textos offrent un environnement alternatif, neutre et libre des contraintes de la norme écrite (McCulloch, 2019), ce qui pourrait favoriser l'engagement des enfants avec des troubles d'apprentissages en lien avec l'écriture (Crystal, 2008). Finalement, le langage SMS est expressif parce qu'il contribue à la performance de l'identité (Tagg, 2009) en signalant l'appartenance à certains groupes par l'adoption de normes communautaires (Blondeau et Tremblay, 2022 ; McCulloch, 2019) ou en exprimant l'esprit créatif du texteur. Ces normes communautaires s'opposent aux normes institutionnelles qui régissent la langue standard par leur caractère plus permissif. Les textos sont généralement réservés pour la communication entre personnes proches (Essoh et al., 2014 ; Tremblay, 2020), d'où la prévalence du registre familier dans ce médium. Pour cette raison, les textos sont un moyen propice à l'étude du vernaculaire authentique libre du paradoxe de l'observateur (Blondeau et Tremblay, 2022 ; Labov, 1972 ; McCulloch, 2019).

Comme mentionné plus haut, l'hybridité dans les textos émerge du fait que les textos répondent à la fois à des conventions du code oral et écrit. Cependant, si de nombreuses études ont conclu que les textos sont une réalisation hybride du langage (Blondeau et al., 2014 ; Crystal, 2008, 2009 ; Endong et Essoh, 2015 ; Soffer, 2010 ; Tagliamonte, 2016 ; Tagliamonte et Denis, 2008), peu d'études ont identifié les aspects du langage qui se comportent comme l'oral ou l'écrit (ou un mélange des deux). Par exemple, sur le plan morphosyntaxique, l'alternance entre le *on* pluriel et le *nous* à l'oral en français québécois se comporte essentiellement de la même manière dans les textos (Blondeau et al. 2014) qu'à l'oral (Laberge, 1977). Par ailleurs, Blondeau et al. (2014) observe aussi l'alternance entre les formes complexes *nous/vous-autres* et les formes simples *nous/vous* et remarque que les formes complexes en *-autres*, souvent associées au français vernaculaire, se retrouvent relativement rarement dans le corpus *Texto4science* (le même que celui employé dans l'étude actuelle) par rapport aux formes simples, qui elles sont souvent associées au français normé. Ceci pourrait constituer un contrargument au fait que le lexique se manifeste similairement à l'oral et dans les textos, mais les auteures ajoutent que les formes complexes en *-autres* sont en déclin en français parlé, signalant que les tendances dans les textos suivent les changements en cours dans la langue parlée et que les textos sont sensibles aux changements linguistiques au niveau du lexique. Sur le plan grammatical, le langage SMS a une tendance forte à l'élimination ou au relâchement des mots fonctionnels (Androutsopoulos et Schmidt, 2002 ; Dürscheid, 2003 ; Fairon et al., 2006a), ce que Fairon et al. (2006a) appellent le « syndrome SMS ». Cependant, dans leur étude sur l'ellipse des déterminants en français et en allemand dans les discours électroniques écrits, Stark et al. (2018) affirment

que les parties grammaticales ne sont pratiquement pas affectées par le langage SMS, contrairement aux contenus lexicaux et graphiques, qui eux sont la cible de nombreuses innovations orthographiques. Par exemple, elles remarquent que les marques d'accords sujet-verbe et objet-participe passé en français étaient présentes même dans les messages hautement informels. Inversement, un regard de survol sur le corpus québécois révèle une tendance contraire où les marques d'accords étaient souvent absentes ou mal accordées/écrites. Ceci suggère que ces tendances sont propres aux différentes communautés linguistiques ou aux différentes langues.³ Dans des langues à noms nus, les textos ne permettent pas les noms seuls là où le code écrit standard ne le permettrait pas non plus, et ce malgré leur objectif de rapidité et leur relâchement de la norme. Aussi, les régularités grammaticales comme le besoin de matériel fonctionnel à certains endroits dans la structure syntaxique ne semblent pas sujettes à la variation. Ceci suggère que du point de vue de la syntaxe, les textos se comportent largement comme la langue écrite standard. Stark et al. (2018) clarifient tout de même que certains textos étudiés présentaient un style d'écriture télégraphique, ce qui aurait pu augmenter les chances qu'on retrouve des noms seuls. Cependant, même en tenant compte du style, ces occurrences demeurent suffisamment rares dans leur étude (4% des occurrences de *determiner-drops* avec les sujets définis) pour qu'on puisse conclure qu'il s'agit d'erreurs de performance typiques chez des adultes (Stark et al., 2018). Dans le corpus *Texto4science* étudié ici, on retrouve de nombreux cas de sujet absent pour les pronoms explétifs comme dans « [...] Faut vrm que jlamene demain! »^{4, 5} (110250, *Texto4science*), un phénomène très fréquent à l'oral (Djuikui Dounstop et al., 2020). Étonnamment, on retrouve également des sujets absents pour des pronoms personnels singuliers et pluriels dans le corpus *Texto4science*, tout comme dans des corpus suisses (Stark et Robert-Tissot (2017)).

Il semble donc difficile de trancher clairement quels aspects du langage SMS se comportent comme l'oral ou l'écrit. Cela demanderait une partition extrêmement fine de chaque aspect de la langue (syntaxe, lexique, grammaire, phonologie, morphologie, etc.), de chaque comportement de textage, dans chaque

³ Tremblay et al. (2020) montrent, par exemple, que l'emploi du futur dans les textos varient entre les communautés belges et québécoise. Les Belges tendent à employer davantage le futur synthétique (futur simple) alors que les Québécois emploient surtout le futur analytique ou périphrastique (ALLER + verbe).

⁴ Les textos cités dans le texte sont repris de manière identique à la forme qu'ils avaient quand ils ont été donnés par les participants. Certains messages sont tronqués afin de faciliter la lecture. Lorsque je le jugerai nécessaire, je fournirai la normalisation (la « traduction ») faite par Langlais et Drouin avec le corpus.

⁵ *Il faut vraiment que je l'amène demain!*

langue, et encore là, il faudrait prendre en compte le style idiosyncratique de chaque texteur afin de dresser un portrait même flou de l'hybridité des textos.

1.1.1 Quelques distinctions

Avant de continuer, il est important de distinguer le langage SMS du langage Internet (ou *Internet Slang*), des salles virtuelles de clavardage (*chatrooms*) et de la MI (Verheijen, 2013). Ces termes sont souvent utilisés de manière interchangeable parce qu'ils partagent certains traits, notamment une norme d'écriture relâchée et le fait d'être, pour la plupart, des moyens de communication médiée par ordinateur (CMO). Pourtant, ils diffèrent tous légèrement au niveau des contraintes communicationnelles de leur médium respectif. Ces caractéristiques sont présentées dans le tableau 1.1 ci-bas. Crystal (2004, 2008, 2009) différencie le langage Internet (*Netspeak*) du langage SMS (*Textspeak*). Les salles virtuelles de clavardage et la MI sont des moyens de communication synchrones dans lesquels les participants communiquent à tour de rôle à un débit qui simule une conversation en personne, en prenant en compte le temps de rédaction des messages. Ces deux médiums diffèrent au niveau du nombre de participants impliqués dans la conversation. La MI implique typiquement une conversation à deux participants (mais elle ne s'y limite pas) alors que les salles virtuelles de clavardage peuvent soutenir des conversations à plusieurs participants. Les textos, quant à eux, ne sont pas des moyens de CMO à proprement parler puisqu'ils sont rédigés sur nos téléphones. Aussi, ils sont un moyen de communication asynchrone où les délais entre les réponses sont conventionnellement plus longs. En effet, les auteurs communiquent souvent en mouvement ou lors d'activités, divisant leur attention entre les tâches qu'ils accomplissent et leurs conversations en cours (Tagliamonte et Denis, 2008). Ce partage attentionnel est probablement l'une des sources des « erreurs » qu'on retrouve dans les textos. Le langage SMS diffère du langage Internet en ce qui a trait aux contenus typiques des messages et à la modalité de communication (petits claviers de 12 touches sur téléphone versus un clavier complet sur ordinateur). Si aujourd'hui il est possible d'envoyer des messages d'un ordinateur à un téléphone et vice-versa, le travail présent porte spécifiquement sur le langage SMS tel qu'employé dans des conversations entre téléphones mobiles.

Tableau 1.1 Différences entre les moyens de communication

Caractéristiques	Langage SMS	Langage Internet	Salles virtuelles de clavardage	Messagerie Instantanée
Langue normée	Non	Non	Non	Non
CMO	Non	Oui	Oui	Oui
Synchronicité	Asynchrone	Asynchrone	Synchrone	Synchrone
Médium	Claviers à 12 touches ou QWERTY	QWERTY	QWERTY	QWERTY
Participant	1 à 1	Plusieurs	Plusieurs	1 à 1
Exemple	Textos	Twitter	Twitch	Messenger

1.2 Linguistique et discours populaire

Les champs d'études ayant l'appellation folk (*folk linguistics*, *folk economics*, etc.) désignent des domaines où l'objet d'étude porte sur les perceptions populaires ou non instruites par rapport aux domaines étudiés. Elle se traduit en français par l'étiquette populaire. La linguistique populaire (Preston, 1993, 1996) porte sur la manière dont les croyances et les attentes socialement construites informent la perception qu'ont les locuteurs des usages langagiers (Paveau, 2007). Si ce champ d'études est reconnu aux États-Unis depuis les années 50, il demeure peu exploré et surtout peu apprécié dans la francophonie. En effet, la linguistique populaire représente à cette époque une théorie ou une opinion primitive (Devitt et Sterelny, 1989) que plusieurs jugent impertinentes à étudier dans le cadre d'une théorie qui se veut proprement scientifique. Après tout, quelle est l'importance de l'opinion des gens dans le développement d'un cadre théorique? Pourtant, les travaux de Labov (1972) suggèrent au contraire que la perception du public importe beaucoup dans l'élaboration d'une théorie sociolinguistique du langage. En plus d'étudier les communautés linguistiques, il faut aussi étudier les normes (Kauhanen, 2006). Le problème est que les normes sociales sont comparées aux normes écrites et que le vernaculaire ou toutes autres formes de langage non standards sont définis et qualifiés à partir du standard. Autrement dit, parce qu'on tient le standard comme sacrosaint, l'étude des autres normes est dévalorisée. Dans cette perspective, l'étude

des textos serait ignorée parce que les textos sont perçus comme une dégradation de la langue qui diverge de la norme standard du code écrit et qui ne vaut même pas la peine d'être étudiée. L'idée d'étudier les textos dans le cadre d'une étude linguistique relève d'un non-sens pour ceux qui les voient comme des barbarismes.

Niedzielski et Preston (2000) viennent redorer le blason de la linguistique populaire en requalifiant l'usage du mot *folk* comme faisant référence à ceux qui n'ont pas de formation académique sur le sujet d'étude et en clarifiant qu'ils ne font pas référence à des individus ou des groupes rustiques, ignorants, non éduqués, arriérés, primitifs, minoritaires, isolés, marginalisés ou de statut inférieur. La linguistique populaire observe donc la perception des non-linguistes par rapport au langage (Preston, 2011). En ce sens, l'étude de l'attitude populaire par rapport aux textos s'inscrit dans le domaine de linguistique populaire et est au cœur du débat sur le risque que présentent les textos pour la qualité de la langue.

1.2.1 Problème de réputation

Ce n'est pas nouveau que les messages textes soient perçus comme une menace pour la langue. Nombreux sont les changements à la langue qui attirent la foudre des parents, des enseignants et des grammairiens. Toutes les vieilles générations ont tendance à critiquer le langage des jeunes (Cameron, 2012 ; Craig, 2003) et les nouvelles technologies sont souvent victimes de l'opinion publique (Kling, 1996 ; Standage, 1999). La popularité des textos, particulièrement chez les jeunes, et leur caractère transgressif en font des cibles de choix. Aux États-Unis, 75% des jeunes de 12 à 17 ans utilisent la MI comparativement à 42% pour les adultes, et 48% de ces jeunes disent l'utiliser plus d'une fois par jour (Lenhart et Madden, 2005). L'usage des téléphones et de plateforme de MI sont aussi des pratiques courantes à l'école. Une étude canadienne sur l'usage d'applications de MI sur les campus universitaires révèle que 97% des étudiants de 268 universités canadiennes utilisent différentes plateformes de MI tous les jours (Quan-Haase, 2007). De plus, les nouvelles générations débutent de plus en plus tôt l'usage de MI (Verheijen, 2013). L'utilisation de plateformes de MI est donc évidemment une pratique extrêmement courante. Il ne faut donc pas s'étonner que la MI, omniprésente dans la vie de tous les jours, particulièrement chez les jeunes (Lenhart et Madden, 2005 ; Quan-Haase, 2007), soit perçue par certains comme une menace pour la qualité de leur langue.

Dans la perspective de la linguistique populaire, le statut non standard des textos ainsi que le message véhiculé par les médias poussent les gens linguistiquement naïfs à les voir comme une menace pour la

qualité de la langue. De nombreux journalistes ont publié des articles décrivant les dangers des textos (Humphrys, 2007 ; Ross, 2007 ; Sutherland, 2002), les décrivant comme des barbarismes linguistiques. Certains des passages les plus sévères se retrouvent en (1-3).

- (1) "Text and instant messaging are negatively affecting students' writing quality on a daily basis, as they bring their abbreviated language into the classroom" (Ross, 2007)⁶
- (2) "[Textese] is snot-talk, unimaginative, bleak, bald, sad shorthand, drab shrinktalk[...]" (Sutherland, 2002)⁷
- (3) "vandals who are doing to our language what Genghis Khan did to his neighbours eight hundred years ago. They are destroying it: pillaging our punctuation; savaging our sentences; raping our vocabulary." (Humphrys, 2007)⁸

Dans son analyse d'un corpus international de 101 articles portant sur la CMO parue entre 2001 et 2005, Thurlow (2006) fait ressortir les tendances dans le discours populaire et médiatique par rapport aux textismes et aux autres formes d'écriture non standards qui caractérisent la CMO. Il observe que l'une des critiques les plus récurrentes est celle qui reproche aux nouvelles technologies de venir détruire les formes conventionnelles ou standardisées du langage. L'opinion populaire est que la variation chaotique des messages textes nuit au sens au lieu d'en créer (Tagg et al., 2012). Les médias ont aussi tendance à exagérer les textismes qu'on pourrait retrouver dans les textos, suggérant que les textismes sont utilisés à outrance dans les messages textes. Pourtant, Tagg et al. (2012) ont révélé que la plupart des textos contiennent nettement moins de textismes qu'on aurait pu penser selon le message véhiculé par les médias. En effet, les abréviations correspondent à moins de 20% du contenu des textos (Thurlow et Brown, 2003). Crystal (2009) rapporte des taux similaires : 10% des mots dans les messages textes se manifestent avec des textismes. De plus, lorsqu'on retrouve de la variation dans l'orthographe, ce sont souvent les mots plus courts ou d'usage courant qui sont atteints (Grinter et Eldridge, 2003). Ceci pourrait s'expliquer par le fait que les mots courants sont facilement reconnaissables et qu'une modification de leur orthographe ne nuit pas à l'intelligibilité. Crystal (2004) dresse une liste d'abréviations souvent rapportées dans les médias mais qui sont peu fréquentes dans l'usage réel. Toutefois, ce sont souvent ces abréviations qui sont pointées du doigt comme étant la manifestation du déclin des compétences

⁶ *Les textos et la messagerie instantanée affectent quotidiennement la qualité de la langue écrite des jeunes alors qu'ils introduisent leur langage abrégé dans la classe.* (ma traduction)

⁷ *Le [langage SMS] est un jargon de morveux ; il s'agit d'une sténographique sans fantaisie, terne, insipide et triste ; d'un bla-bla morne.* (ma traduction)

⁸ *Des vandales qui infligent à notre langue ce que Gengis Khan a infligé à ses voisins huit cents ans plus tôt. Ils la détruisent : ils pillent notre ponctuation, saccagent nos phrases, violent notre vocabulaire.* (ma traduction)

langagières des texteurs. Aussi, ces abréviations ne se traduisent pas toujours en compression importante dans les textos. La compression est mesurée en pourcentage en comparant le nombre de caractères dans un texto au nombre de caractères dans sa version normalisée, c'est-à-dire « traduite » en langue standard. Langlais et Drouin (2012) rapportent un taux de compression de seulement 10% dans leurs corpus du français montréalais, ce qui fait écho aux résultats mentionnés par Crystal (2009) et contredit le discours des médias. Ainsi, la grande majorité des formes standards dans les textos sont masquées par les quelques formes non standards qui s'y retrouvent parce qu'elles pourraient être plus saillantes et attirer davantage l'attention des lecteurs (Blondeau et al., 2014). Une situation similaire se retrouve dans des études sur le français en milieu minoritaire (Flikeid, 1989 ; Poplack, 1988), où on rapporte que les mots d'origine anglaise ne constituent souvent que 10% des énoncés, mais qu'on dit souvent que cette variété est « moitié anglaise, moitié-française ».

La perception négative des textos est empirée par tout le métadiscours employant des citations numériques superlatives qui contribuent à créer une forme de panique statistique chez la population (Thurlow, 2006). Ces articles, tirés de Thurlow (2006), parlent du nombre de messages textes envoyés tous les jours (4), de la vitesse de croissance de ces nouvelles pratiques (5), des revenus engendrés (6) et du nombre croissant d'utilisateurs (7). Les journalistes emploient des nombres énormes comme appâts à clics :

- (4) "It has an estimated 75 million users sending more than 700 million real-time messages a day [...]" (Ruane, 1999)⁹
- (5) "The fad picked up steam in the U.K. two years ago. In March alone, 864 million messages were sent in the country, up from 372 million in the same month a year ago." (Ellison, 2001)¹⁰
- (6) "U.S. carriers are starting to tap into this market, estimated at \$20 billion worldwide." (Stroud, 2004)¹¹
- (7) "Nearly three-quarters of online teens use instant messaging, compared with less than half of adults, according to a 2000 survey by the Pew Internet & American Life Project." (Dunnewind, 2003)¹²

Ces statistiques donnent une impression de démesure qui déforme la perception de la population à l'endroit du *textese* et donne un semblant d'apparence scientifique aux articles, ce qui les rend plus

⁹ Il (système de MI de America Online) compte environ 75 millions d'utilisateurs qui envoient plus de 700 millions de messages en temps réel par jour [...] (ma traduction)

¹⁰ La mode a pris de l'ampleur au Royaume-Uni il y a 2 ans. Seulement en mars, 864 millions de messages ont été envoyés dans le pays, contre 372 millions durant le même mois l'année dernière. (ma traduction)

¹¹ Les compagnies américaines commencent à exploiter le marché, estimé à 20 milliards de dollars dans le monde. (ma traduction)

¹² Selon une enquête réalisée en 2000 par le Pew Internet & American Life Project, près du trois quarts des adolescents en ligne utilisent la MI, contre moins de la moitié des adultes. (ma traduction)

crédibles aux yeux de la population linguistiquement naïve. Ce discours démesuré alimente la panique de la population, d'autant plus que, comme Thurlow le mentionne dans son article, les figures citées par les journalistes ne sont pas toujours accompagnées de sources.

Le problème de perception ne se limite pas aux médias. Plusieurs chercheurs rapportent avoir interrogé des enseignants sur la qualité des travaux de leurs élèves et sur les causes du faible niveau de littératie des jeunes apprenants et le suspect habituel revient toujours. En effet, si l'on se fiait simplement aux faits rapportés par les enseignants de langues, qui sont en contact direct avec les travaux de leurs élèves, il semblerait que les textismes et la CMO en général soient la cause évidente d'un déclin en littératie. De plus, les enseignants ne sont pas les seuls à sauter aux conclusions. Certains chercheurs dans la communauté linguistique arrivent aux mêmes conclusions et publient des articles propageant cette idée selon laquelle les textismes sont une menace pour la littératie des jeunes. Mphahlele et Mashamaite (2005) sont parmi les premiers à s'être penchés sur l'impact des textismes sur le développement de la compétence à l'écrit des apprenants d'anglais en Afrique du Sud. Ils reconnaissent l'impact positif de la MI dans l'amélioration de la communication, mais croient aussi que la MI, et plus précisément les textismes, ont un effet néfaste sur la compétence langagière des apprenants. Notamment, ils avancent que le langage SMS affecte l'habileté des élèves à s'exprimer avec éloquence à l'oral comme à l'écrit et leur capacité à utiliser des mots qui sont appropriés au contexte. Ils rapportent que les étudiants mélangent des textismes à leurs travaux scolaires, produisant des formes orthographiques non standards et des phrases agrammaticales à cause de leur exposition quotidienne aux textismes. Essentiellement, ils avancent que les élèves ne font pas de distinction entre les contextes appropriés pour les textos et ceux qui ne le sont pas (Mphahlele et Mashamaite, 2005). Cette conclusion va à l'encontre de résultats obtenus par Volckaert-Legrier et al. (2015). Dans une étude qui compare le niveau de compétence linguistique en langue standard d'élèves avant, pendant et après avoir commencé à texter, ils montrent que le niveau relatif de compétence demeure stable, c'est-à-dire que les élèves plus forts sont restés forts et les élèves plus faibles sont restés faibles, peu importe leur degré d'utilisation de textismes. Ces résultats font écho à Bernicot et al., (2014). Si la conclusion de Volckaert-Legrier et al. (2015) s'appuie sur des données empiriques, on ne peut pas en dire autant pour l'article de Mphahlele et Mashamaite (2005). En effet, un regard rapide sur le contenu de l'article révèle qu'ils n'emploient pas de protocole expérimental qui leur permettrait de conclure, grâce à des données probantes, que les textos sont une réelle menace.

Autrement dit, le tout serait une question d'opinion, une chasse aux sorcières. Il faut reconnaître les risques d'avancer des propos de la sorte sans s'appuyer sur des expériences et des données parce qu'on pourrait se priver des bénéfices que pourrait apporter le langage SMS au développement de la compétence langagière. En effet, on trouve un corps grandissant de recherches appuyant l'idée que les textos pourraient avoir un impact positif sur le développement de compétences en littératie (Arthur-Nyarko, 2017 ; Craig, 2003 ; Drouin et Davis, 2009 ; Drouin et Driver, 2012 ; Verheijen, 2013), en acquisition du vocabulaire ainsi qu'en développement de fluence orale (Blake, 2009). De plus, les textismes pourraient même servir d'outils pédagogiques pour faciliter l'enseignement des langues maternelles et secondes (L2) (Derakhshan et Kaivanpanah, 2011 ; Lopéz Rúa, 2007 ; Tabatabaei et Goojani, 2012). Notons finalement que le discours anti-texto est beaucoup moins virulent aujourd'hui, notamment parce que les études n'ont pas réussi à établir un lien de causalité entre l'usage des textos et le déclin de la compétence langagière et parce que l'emploi des autocorrecteurs facilite l'emploi de formes langagières standards (Tremblay, 2020).

1.3 Typologie des textismes

L'étude des textismes est complexifiée par l'absence de typologie largement acceptée par la communauté de chercheurs. Le tableau 1.2, ci-bas, dresse une liste (non exhaustive) des textismes typiquement catalogués dans les travaux sur les textos. La liste est une compilation des typologies proposées par Thurlow et Brown (2003), Lopéz Rúa (2007), Plester et al. (2009), Verheijen (2013), Arthur-Nyarko (2017) et Totanes et Lintao (2019). Les exemples qui s'y trouvent sont inspirés de ceux trouvés dans les typologies citées plus haut. Les textismes sont parfois nommés différemment parce que la catégorisation des différentes formes de textismes est un processus subjectif au cours duquel les perceptions des chercheurs sont imposées sur les données (Tagg et al., 2012). Ce qui constitue un textisme pour un chercheur ne l'est peut-être pas pour un autre. Par exemple, écrire *thru* pour dire *through* serait classifié comme une *abréviation phonologique* par Verheijen (2013), une *réécriture phonétique* par López Rúa (2007), une *orthographe phonologique* par Arthur-Nyarko (2017) ou encore une *transcription phonético-graphique* par Blondeau et Tremblay. (2022). De façon plus concrète pour l'étude présente, l'élision du <l> compte comme une contraction ou une orthographe phonétique simulant l'absence de prononciation du /l/ à l'oral.

Tableau 1.2 Typologie des textismes

Type de textisme	Définition	Exemples
Homophonie de lettres/nombres/symboles (rébus)	Remplacer une syllabe par une lettre, un chiffre ou un symbole qui se prononce de manière similaire.	<i>Cya l8r</i> = « see you later » <i>A+</i> = « à plus »
Raccourcis de mots	Couper la fin de mots communs.	<i>app</i> = « application »
Contraction / abréviation	Enlever des lettres (souvent les voyelles) à l'intérieur de mots communs.	<i>Msg</i> = « message »
Orthographe phonétique	Modifier l'orthographe d'un mot pour suivre plus directement la prononciation du mot.	<i>Skool</i> = « school » (école) <i>Gonna</i> = « going to » (je vais)
Initialismes / acronymes	Utiliser les premières lettres de mots d'une expression commune ou figée.	<i>lol</i> = « laugh out loud » (rire aux éclats)
Stylisation d'accent	Répliquer un accent particulier à travers l'orthographe des mots.	<i>Anuva</i> = « another » (un autre) <i>Ze</i> = « the » (<i>the</i> français).
Symboles	Utiliser des symboles pour remplacer des syllabes ou des mots.	<i>@6</i> = « at 6 » (à 6h) <i>w/</i> = « with » (avec)
Émoticons	Utiliser des caractères pour représenter des émotions.	☺ ;) ☹_☹ :P ☹ :D
Hypercapitalistion / hyperponctution / redoublement de caractères	Représenter des émotions ou l'allongement par l'usage à outrance de lettres majuscules et de ponctuation.	QUOOOOI?!?!?! ????
Omission de capitalisation, d'accent et de ponctuation	Omettre les lettres majuscules, les accents et les marques de ponctuation pour réduire le nombre de caractères et accélérer la communication	<i>je suis alle au cine cet ete peut etre quon devrait</i>
Néologismes / slang	Nouveaux mots dérivés de l'usage des textismes ou emploi de slang dans les messages.	<i>Lolz</i> <i>trudat</i>

Malgré le manque de consensus quant au nom donné aux différents textismes, la grande majorité des chercheurs attestent des mêmes formes de textismes, du moins lorsqu'on compare des données à l'intérieur d'une même langue. En effet, les langues SMS n'emploient pas tous les mêmes textismes (Weimers, 2008), et des textismes fréquents dans une langue ne le sont pas forcément dans une autre langue (Winzker et al., 2009). Comme mentionné plus tôt, la très grande majorité des études de corpus sur les textos ont porté sur les formes attestées en anglais, mais les catégories énumérées dans le tableau 1.2 (ci-haut) se retrouvent aussi dans un large inventaire d'autres langues. Dans une étude sur le langage SMS en afrikaans, Weimers (2008) commente qu'on retrouve peu de rébus en afrikaans puisque peu de nombres sont des homophones de lettres ou de symboles. Par exemple, en anglais, le nombre 4, prononcé [fɔːr], est souvent utilisé pour remplacer la graphie ⟨for⟩ (*pour*), mais en afrikaans, 4 est prononcé comme *vier* [fiə] alors que la préposition *for* se dit *vir* [fɜːr]. Aussi, certains textismes peuvent être regroupés en une catégorie par certains chercheurs et diviser en différentes sous-catégories par d'autres. Dans le tableau 1.2, l'homophonie de lettres/nombres/symboles est considérée comme une seule catégorie, mais d'autres chercheurs les divisent. De plus, certains types de textismes peuvent se retrouver dans plusieurs catégories simultanément parce que les auteurs peuvent utiliser plusieurs textismes à l'intérieur d'un même mot. Certaines catégories sont plus controversées. Par exemple, les omissions de capitalisation, d'accents et de ponctuation ne sont pour certains que des erreurs (López Rúa, 2007 ; Thurlow et Brown, 2003). À ce sujet, il est généralement difficile de distinguer entre des erreurs causées par de l'inattention ou de l'ignorance de la part du auteur et un choix stylistique d'utiliser un textisme. Il s'agit d'une limitation importante pour toutes études sur les textismes et la compétence des auteurs.

Drouin et Driver (2012) suggèrent qu'il existerait une distinction de perception entre des textismes positifs et négatifs. Les textismes positifs seraient ceux qui nécessitent un certain degré de maîtrise de la langue pour être utilisés. Typiquement, ces textismes sont ceux qui ont une dimension phonétique ou phonologique. L'homophonie de lettres et de nombres, par exemple, est seulement utilisable si le auteur détient une certaine conscience au niveau de la phonétique et de la phonologie. Dans un exemple comme *cya l8r*, il faut savoir que le ⟨c⟩ doit être décodé comme un *see*, le ⟨ya⟩ comme un *you*, le ⟨8⟩ comme un */eɪt/* pour former le mot *later*. Les textismes négatifs, quant à eux, sont ceux qui sont typiquement associés à de la paresse, comme l'omission d'accents, de capitalisation et de ponctuation. Cette distinction pourrait s'inscrire dans les idées préconçues véhiculées par la linguistique populaire.

Les formes de textismes les plus fréquents sont celles qui dépendent du décodage phonologique comme l'orthographe phonétique et l'homophonie de lettres, nombres et symboles (Waldron et al., 2017). Ces types de textismes demandent un bon niveau de compréhension de la phonologie et de principes orthographiques. Sans cette compétence, les enfants auraient beaucoup de difficultés à produire et déchiffrer des textos. Ainsi, les textos peuvent être vus comme des outils pédagogiques utiles (Andujar, 2016 ; Blake, 2009 ; Derakhshan et Kaivanpanah, 2011 ; Drouin et Davis, 2009 ; Plester et al., 2008 ; Plester et al., 2009 ; Tabatabaei et Goojani, 2012 ; Wood et al., 2011).

1.4 Variation orthographique

Selon Thurlow et Brown (2003), la variation orthographique dans les messages textes répond à trois maximes ; la brièveté et la rapidité de communication, la restitution paralinguistique et l'approximation phonologique (Thurlow et Brown, 2003). Premièrement, la maxime de brièveté et de rapidité de communication dicte qu'un texto doit être le plus court possible (en termes de nombre de caractères). C'est pour ça qu'on retrouve beaucoup d'abréviations, d'orthographe phonétique et autres textismes qui permettent aux auteurs de réduire le nombre de caractères sans nuire à l'intelligibilité du message. Deuxièmement, la restitution paralinguistique dicte qu'on écrit comme on parle. Par exemple, dans un texto comme « [...] HAHahaha TROP PAS MANE ! »¹³ (Texto4science, 112337), l'hyperc capitalisation du message prend plus de touches à produire que si la auteuse avait simplement écrit son message en lettres minuscules. Cependant, le texto transmet l'émotion de la auteuse et suggère au lecteur la manière dont elle aurait prononcé son énoncé. Il en va de même pour des cas de répétition de caractère comme « niiniiice je checkerai ak Tremblay [...] »¹⁴ (Texto4science, 112950), où la répétition du (i) rend le texto plus long à écrire mais exprime l'intonation du auteur. Troisièmement, l'approximation phonologique est la maxime qui dicte que les auteurs cherchent à écrire phonétiquement. La rectitude de l'orthographe est moins importante que l'intelligibilité du message lui-même. Cette approximation libère le auteur des conventions d'écriture et lui permet de s'exprimer librement et de jouer avec une langue et un registre plus informel et plus approprié au médium des textos. Les trois maximes ne sont pas aussi importantes les unes que les autres. En effet, si autrefois on pouvait imaginer que la maxime de brièveté et de rapidité de

¹³ HAHahaha trop pas man !

¹⁴ Niiiiice je regarderai avec Tremblay [...]

communication primait à cause des limitations des technologies, il semble que les deux autres sont aujourd'hui plus déterminantes dans l'émergence des variantes orthographiques (Tagg et al. 2012).

1.5 Questions de recherche

Blondeau et al. (2014) ont démontré que, du côté de la morphosyntaxe, les textos reflètent les comportements oraux (dans leurs cas, l'alternance *on* pluriel/*nous* et formes fortes *nous-autres/nous*) sans toutefois les imiter complètement. Autrement dit, le lexique semble se manifester similairement à l'oral et à l'écrit, avec un impact marqué de la norme écrite qui influence le choix de mots des auteurs dans certains contextes. La question maintenant est de voir si d'autres aspects de la langue comme ceux ayant trait à la phonologie exhibent des comportements similaires à l'oral et dans les textos. L'étude de l'élision du ⟨l⟩ dans les textos est une occasion unique de voir si un processus phonologique comme l'élision du /l/, qui se retrouve catégoriquement à l'oral, se manifeste à des taux semblables dans la langue des textos. Ainsi, le projet présent se donne comme objectif de répondre aux questions de recherches suivantes :

Question 1 : Quels facteurs linguistiques influencent l'élision du ⟨l⟩ dans les textos?

Question 2 : Quels facteurs sociologiques influencent l'élision du ⟨l⟩ dans les textos?

Question 3 : De quelle manière les résultats de l'analyse des textos se comparent-ils aux résultats de l'analyse de la langue orale?

Si l'élision du ⟨l⟩ est influencée par les mêmes facteurs que l'élision du /l/ à l'oral, on pourrait conclure que la phonologie opère aussi à l'écrit dans les textos, influençant la manière dont les auteurs rédigent leurs messages pour refléter leur parler. Si, au contraire, on remarque une différence dans les facteurs qui opèrent sur la variable, il faudrait rejeter l'idée que les textos sont un reflet direct de la langue parlée, du moins en ce qui a trait de l'élision. Nous aurions alors davantage de raisons d'appuyer la position selon laquelle le *textese* est une langue à part.

CHAPITRE 2

ÉTAT DES CONNAISSANCES

2.1 L'élision du /l/ à l'oral

L'élision du /l/ à l'oral est un phénomène phonologique qu'on retrouve dans les pronoms personnels et impersonnels, les pronoms objets (*le, la, les, lui, leur, leurs*), les articles définis (*le, la, les*) et dans certains mots communs (*table, /tab(l)/*). L'élision serait le résultat d'une réduction à la voyelle thématique (Bougaïeff et Cardinal, 1980), un processus diachronique qu'on retrouve dans les langues romanes. C'est par ce processus, par exemple, que les articles /lo/ *le* et /los/ *les* en latin vulgaire sont devenus /o/ et /os/ en portugais moderne. Selon ces auteurs, la réduction de l'article et du pronom à la voyelle thématique dans le discours familier, sous la forme de l'alternance vocalique /ɪ/ [i] et /ɛ/ [æ], n'est que la manifestation, en français québécois, d'un processus latent dans les langues romanes. Le projet de recherche présent se penche exclusivement sur les pronoms personnels *il(s)* et *elle(s)*, le *il* impersonnel et le POI *lui*. Les raisons pour cette sélection sont présentées en 3.3.

L'élision est un phénomène observé en français depuis le 12^e siècle (Pope, 1952). Durant cette période, le /l/ en position préconsonantique commençait à s'effacer et le /l/ à l'intérieur du mot se neutralisait, comme dans *alba/aube, cultello/couteau, pulmone/poumon* (Bourciez, 1889). De manière plus pertinente pour l'étude présente, le /l/ en position préconsonantique dans les pronoms clitiques *il(s)* et *elle(s)* aussi commençait à s'effacer à cette période, de sorte que *qu'il* et *qui* ainsi que *s'il* et *si* sont devenus des homophones (Pope, 1952). On remarque que certains contextes ou entourages phonologiques encouragent l'élision. La position intervocalique (/sy a kloʃ/ *sur la cloche*) favorise l'amuïssement des consonnes (Bougaïeff et Cardinal, 1980, page 92), mais ne saurait être valable avec les formes lexicales. Selon Bougaïeff et Cardinal (1980), l'élision du /l/ survient principalement pour les formes grammaticales, c'est-à-dire les articles *la, les, l'* et les pronoms *la, les, lui, il(s)* et *elle(s)*. Les formes lexicales, par exemple *village, allumer, lumière*, ne sont pratiquement jamais éliminées dans l'usage. On retrouve cependant de l'élision dans les mots terminant en -ble, -cle et -ple (*coupable, oncle, couple*) résultant d'une règle générale d'effacement des liquides dans les codas complexes en fin de mot.

En français moderne, la présence du /l/ est fortement associée au niveau cultivé, et il disparaît presque entièrement au niveau populaire, un fait aussi relevé par Sankoff et Cedergren (1976). François (1974) observe aussi de l'élision en position prévocalique. Trois sources pourraient expliquer l'émergence de

l'élision du /l/ à l'oral : la phonologie (Pupier et Légaré, 1973), la sociolinguistique (Sankoff et Cedergren, 1976) et la morphosyntaxe profonde (Bougaïeff et Cardinal, 1980). Ces derniers précisent que selon eux, « l'élision du /l/ est une variante sociolinguistique dont la manifestation phonologique est déterminée par la fonction des morphèmes dans l'énoncé » (Bougaïeff et Cardinal, 1980). La présence du /l/ est non marquée dans le français standard (international). Elle est cependant marquée dans le dialecte québécois (ainsi que d'autres variétés du français) et dénote un statut social élevé (Sankoff et Cedergren, 1976). On s'attendrait donc, si l'on écrit comme on parle dans les textos, à en retrouver peu dans un corpus de textos puisque le contenu des textos relève souvent d'un rapport familier entre les auteurs (Essou et al., 2014 ; Tremblay, 2020).

Bien qu'en général l'élision du /l/ soit un phénomène fort en français, elle ne se manifeste pas de la même manière dans toutes les communautés francophones. En effet, on remarque des taux d'élision légèrement plus prononcés en français laurentien par rapport aux communautés européennes. Pour Bougaïeff et Cardinal (1980), ceci montre qu'il s'agirait d'une innovation québécoise, mais dans la mesure où nous observons de l'élision depuis le 12^e siècle, ceci semble peu probable. Les résultats rapportés dans différentes études sont présentés dans le tableau 2.1 ci-bas. Les sections qui suivent discutent du phénomène en distinguant entre le français laurentien et le français européen.

Tableau 2.1 Études sur l'élision du /l/ à l'oral, comparaison entre les communautés

Pronoms	Canada		France	
	Sankoff et Cedergren, 1976 (Montréal)	Poplack et Walker, 1986 (Ottawa-Hull)	Ashby, 1984 (Tours)	Armstrong, 1996 (Dieuze)
<i>Il personnel</i>	89%	100%	63%	98.4%
<i>Il impersonnel</i>	98%	100%	88%	Omis (catégorique)
<i>Ils</i>	92%	99%	88%	Omis (catégorique)
<i>Elle</i>	63%	84%	26%	34.1%
<i>Elles</i>	Trop rares	33%	42%	Trop rares
<i>Lui</i>	Omis (rare)	99%	15%	Omis

2.1.1 Français laurentien

Dans leur étude sur 16 sujets montréalais, Sankoff et Cedergren (1976) rapportent des taux d'élision déjà très prononcés à cette époque. Ceci montre que l'élision est bel et bien un phénomène fort en français montréalais. Elles ne montrent pas de résultats pour *elles* à cause de la rareté d'occurrence de ce pronom. En effet, le genre est typiquement neutralisé au pluriel de sorte que les formes plurielles soient synchrétiques. Ainsi, une phrase comme *les femmes, elles aiment ça* se prononcerait [lɛfæmjɛmsɑ], avec le [j] masculin pluriel. Ces résultats sont répliqués par les études rapportées, ce qui suggère que le synchrétisme se retrouve dans plusieurs communautés linguistiques. Dans leur étude sur la communauté francophone de Ottawa-Hull, Poplack et Walker (1986) rapportent que des taux d'élision très similaires (100% pour *il* personnel et impersonnel et 84% pour *elle*, 99% pour *ils* et seulement 33% pour *elles*). Le problème du nombre d'occurrences pour *elles* se répète ici. Ces résultats sont essentiellement identiques dans l'étude d'Ostiguy (1979) et montrent qu'à l'oral en français montréalais, l'élision du /l/ est pratiquement catégorique, c'est-à-dire qu'on retrouve presque exclusivement la forme non standard élidée. La forme standard, quant à elle, fait office de marque d'hyperstyle et signale un niveau de langue relevé.

2.1.2 Français hexagonal

Le portrait est similaire en France. Dans son étude sur le français tourangeau, Ashby (1984) rapporte que l'élision se produit dans 71% des occurrences ($n=4033$), ce qui inclut les pronoms clitiques sujets (*il, ils, elle, elles*) et objets (*lui, la, les, leur*). L'élision est nettement plus prononcée pour les clitiques sujets (76%) que pour les clitiques objets (16%). Les pronoms féminins sont plus rarement élidés que les masculins (88% pour *ils*, 88% pour *il* impersonnel, 63% pour *il* personnel versus 42% pour *elles* et 26% pour *elle*). Il remarque aussi que tous les participants élidait variablement le /l/, c'est-à-dire que dans un même contexte morphosyntaxique, un participant n'élidait pas toujours le /l/, suggérant que d'autres facteurs entrent en jeu pour déterminer si un locuteur élide ou non le /l/. En se basant sur ces résultats, on pourrait s'attendre à ce que l'élision dans les textos soit aussi variable. Ashby (1984) exclue les pronoms non clitiques de son analyse puisqu'il ne trouve aucune occurrence d'élision pour eux, contrairement aux formes clitiques. De plus, il remarque que l'élision est plutôt rare dans les mots autres que les clitiques. Par exemple, dans la phrase *Il me gêne*, le /l/ du pronom pourrait être élidé, alors que dans la phrase *Cet imbécile me gêne*, le /l/ n'est pas élidé même s'il se retrouve dans un contexte phonologique identique. Autrement dit, la source du phénomène semble se retrouver au niveau morphosyntaxique et cibler principalement les pronoms (Ashby, 1984 ; Bougaïeff et Cardinal, 1980). Armstrong (1996), quant à lui,

étudie le français de Lorraine et rapporte des taux d'élision plus proches des résultats des études québécoises (98.4% pour *il*, 34.1% pour *elle*), mais n'incluent pas de données pour les formes plurielles et ne précise pas si le *il* comptabilisé est le pronom personnel ou impersonnel (ou une combinaison des deux).

2.2 Travaux de corpus sur les textos

Comme tous autres corpus, les corpus de textos sont des outils indispensables pour voir comment les locuteurs (ou les texteurs) emploient leur langue et de révéler ce qui se trouve véritablement dans leur répertoire. Plus important pour nous dans le cadre du projet actuel, ils permettent aussi de déceler des tendances dans les données et de mesurer l'impact de différents facteurs sur les comportements de textage.

Les corpus de textos, surtout ceux qui portent sur d'autres langues que l'anglais, sont relativement rares. Quelques corpus généralement plus petits existent déjà. Notamment, Tagg (2009) a construit le corpus *Cortxt* contenant plus de 11 000 textos anglais. Treurniet et al. (2012) ont construit *SoNaR*¹⁵, un corpus de 53 000 messages en néerlandais provenant de la Hollande et de la Belgique tirés de *chats*, de *tweets* et de textos. Le *National University of Singapore (NUS) SMS corpus*¹⁶ (Chen et Kan, 2013) comporte, quant à lui, environ 50 000 textos en anglais et en mandarin. On retrouve aussi le *Spanish/English Bilingual Youth Texts corpus* (McSweeney, 2016)¹⁷ qui compte environ 44 000 textos. D'autres projets indépendants existent aussi en France et en Italie (Fairon et Paumier, 2006), mais les corpus ne semblent plus accessibles.

Le projet de recherche international *Sms4science* (Fairon et al., 2006b) est le plus grand projet de recherche sur les textos à ce jour. Il regroupe des équipes de chercheurs d'à travers le monde et avait comme objectif de construire une grande base de données de textos authentiques dans différentes langues. Ce projet s'est décliné en plusieurs sous-projets internationaux comme le projet (du même nom) *Sms4science*¹⁸ produit en Suisse (Dürscheid et Stark, 2011), le projet *Sud4science* (Panckhurst et al., 2013) (et son corpus français *88milSMS* comptant près de 88 000 textos) et plusieurs autres. Le plus important

¹⁵ <https://taalmaterialen.ivdnt.org/download/tstc-sonar-corpus/>

¹⁶ <https://metatext.io/datasets/nus-sms-corpus>

¹⁷ <https://byts.common.gc.cuny.edu/>

¹⁸ <https://sms.linguistik.uzh.ch/>

pour le projet actuel est le corpus *Texto4science* (Langlais et Drouin, 2012), le seul corpus de textos portant sur le français québécois à ce jour. Plus de détails sur le corpus sont donnés à la section 3.1.

Plusieurs raisons peuvent expliquer la rareté des corpus de textos. D’abord, comme pour tous autres corpus, la construction d’un corpus de textos soulève des considérations à propos de la taille du corpus, de son niveau de représentabilité, de l’échantillonnage ainsi que des considérations éthiques (Tagg, 2009). De plus, de par la nature des textos, d’autres difficultés surviennent comme le choix de la bonne méthode de collecte de données, les considérations éthiques qui viennent avec la collecte de données électroniques, les problèmes de sélection de messages par les participants (le fait que les messages soumis au corpus soient filtrés par les participants) et les erreurs de transcriptions (le fait que les textos n’apparaissent pas exactement comme ils sont dans l’historique de messages des participants) (Fairen et Paumier, 2006 ; Tagg, 2009). D’abord, selon les méthodes de collectes de textos et la publicité faite, il est possible que l’échantillon soit biaisé. Par exemple, si la publicité est faite uniquement dans des universités, les textos obtenus seront moins diversifiés que si la publicité est faite à la grandeur du pays par l’entremise de la télévision, des journaux, de la radio et autres (Fairen et Paumier, 2006). Aussi, comme les données qu’on cherche à collecter sont des textos sauvegardés dans les téléphones des participants, il faut avoir la coopération des nombreux participants ou des compagnies de service téléphonique et les logiciels de soutien nécessaires afin d’y avoir accès (Chen et Kan, 2013). Ceci peut mener à des problèmes de distribution et de représentabilité. Par exemple, dans l’élaboration de *SoNaR*, Treurniet et al. (2012) remarque que les femmes semblent plus portées à envoyer leurs textos que les hommes, un problème qui s’est manifesté dans les données de l’étude actuelle (voir la section 3.2 pour une discussion sur les participants). Pour le projet *Sms4science* (Fairen et Paumier, 2006) et pour le corpus *Texto4science* (Langlais et Drouin, 2012), les chercheurs ont créé un numéro gratuit auquel les volontaires pouvaient envoyer leurs textos et un questionnaire en ligne pour collecter les métadonnées des participants. Ensuite, sur le plan éthique, étant donné la nature confidentielle des textos et des métadonnées recueillies (réponses de questionnaires), et parce que l’un des objectifs de ces corpus est qu’ils soient accessibles librement par la communauté de chercheurs (Chen et Kan, 2013 ; Treurniet et al., 2012), il faut s’assurer d’obtenir le consentement de tous les participants. Les participants fournissaient uniquement leurs messages et non les réponses de leurs interlocuteurs. Il n’était donc pas nécessaire d’obtenir le consentement de tous les interlocuteurs dans la conversation. Aussi, il arrive souvent que les participants acceptent de fournir leurs textos, mais refusent de répondre aux questionnaires, ou encore y répondent de façon incomplète. Les corpus doivent ensuite passer par un processus d’anonymisation afin d’assurer

la confidentialité des textos. Ce processus masque les noms des participants en les changeant en code alphanumérique, en les annotant comme des prénoms (par exemple changé *Stéphanie* par <PRENOM>) ou en les changeant pour d'autres prénoms fictifs. Ceci peut avoir une incidence sur le nombre de caractères contenu dans le message, un facteur retenu pour l'analyse linguistique des textos proposée pour le travail présent, mais comme nous le verrons dans la section 3.5.1.4, cet impact est relativement mineur.

Une fois les données collectées, de nouveaux problèmes surviennent. Le fait que les textos soient aussi libres au niveau de l'orthographe fait en sorte que les fonctions de recherche (*Ctrl+F*) sont incapables de collecter toutes les occurrences d'une variable de manière automatique parce qu'ils reposent sur la recherche par chaîne de caractères précis. Pour contourner ce problème, selon la nature de la variable étudiée, les corpus de textos doivent typiquement passer par un processus de normalisation, c'est-à-dire une traduction plus ou moins automatique en langue standard (ou du moins en orthographe standard), avant d'être utilisables par des chercheurs. Nombreux sont ceux qui ont cherché à produire des outils de normalisation automatique afin d'agrandir la taille des corpus, ce qui est devenu en soit son propre champ d'études (Aw et al., 2005, 2006 ; Koehn et al., 2003 ; Papineni et al., 2002). Une fois un texto normalisé, il est plus facile de chercher des occurrences par leur forme standard, mais comme il faut observer les textos dans leur forme originale, la normalisation ne suffit pas à elle seule à rendre le processus de dépouillement des données automatique et rapide. Par exemple, si l'on cherche à étudier le phénomène de l'absence du pronom dans les textos, il serait impossible de trouver automatiquement les occurrences parce qu'elles ne seraient évidemment pas détectées si elles sont absentes. Il faudrait procéder en deux étapes en observant toutes les occurrences de pronoms normalisés puis en les comparant avec les textos originaux pour confirmer leur présence ou absence. Certains chercheurs (Dürscheid et Stark, 2011; Langlais et Drouin, 2012) ont effectué un travail herculéen d'annotations de corpus en cataloguant, entre autres, les formes d'orthographe non standards, les abréviations, l'usage d'émoticons et l'absence de mots comme les cas d'absence du sujet, ce qui rend leur corpus beaucoup plus puissant que d'autres corpus qui ne sont qu'une collection de données.

CHAPITRE 3

MÉTHODOLOGIE

3.1 Présentation du corpus *Texto4science*

Le corpus *Texto4science*¹⁹ (Langlais et Drouin, 2012) est la partie franco-qubécoise d'un projet de recherche international nommé *Sms4science*. Le projet a été lancé en octobre de 2009 et s'est terminé en septembre 2010. Afin de recueillir des participants, les chercheurs ont d'abord collaboré avec les services de communication de l'Université de Montréal afin de faire la promotion de leur projet de recherche en posant des affiches. Ils ont aussi animé des discussions sur les réseaux sociaux (Facebook et Twitter). Ils ont ensuite distribué l'information dans des endroits stratégiques à Montréal (des cafés, des bars, d'autres universités et d'autres écoles). Par la suite, les chercheurs ont fait la promotion de leur projet à travers des émissions de télévision et de radio québécoises et des articles de journaux. À cause des méthodes de recrutement des chercheurs, le corpus regroupe principalement des messages textes issus de la région de Montréal. Cependant, il compte aussi des textos venant de l'ensemble du Canada. La très grande majorité des textos sont en français, avec une plus petite partie en anglais et très petite partie en espagnol. À cause du choix de la variable, l'étude présente se concentre sur les occurrences en français, mais n'exclut pas les occurrences issues de textos bilingues ou qui contiennent du *code-switching*. Le français montréalais étant marqué par son contact avec l'anglais et d'autres langues, ces occurrences sont probantes pour l'étude de la variété montréalaise.

Le corpus comporte 7274 messages textes provenant d'environ 360 participants. Langlais et Drouin (2012) précisent qu'il s'agit en réalité de 360 téléphones différents puisqu'il était possible pour un participant de soumettre des textos pour le corpus à partir de plusieurs téléphones. Ce faisant, il était possible qu'un participant fournisse le même texto à partir de différents appareils, créant ainsi des doublons dans le corpus. Cependant, parce que les auteurs ont identifié chaque participant par un code numérique unique de six chiffres, il était possible d'identifier les textos qui étaient dédoublés dans le corpus et de les retirer. Ceci sera tenu en compte lors de l'analyse des données. Des 7274 messages textes contenus dans le corpus, l'étude présente retient seulement 822 occurrences de la variable étudiée divisées en quatre analyses distinctes : une première analyse (515 occurrences) des facteurs linguistiques influençant l'élision dans les pronoms sujets masculins (incluant les sujets explétifs), une deuxième analyse (103 occurrences) portant

¹⁹ <http://rali.iro.umontreal.ca/rali/?q=en/texto4science>

sur les facteurs linguistiques influençant l'élision dans les POIs, une troisième analyse (204 occurrences) pour les facteurs linguistiques influençant l'élision dans les pronoms féminins singuliers et une quatrième (706 occurrences) portant sur les facteurs sociologiques influençant l'élision. Les raisons justifiant le besoin de faire 3 analyses distinctes pour les facteurs linguistiques sont présentées dans la section 3.5.1. L'analyse des facteurs sociologiques englobe toutes les variables (pronoms masculins, féminins et POIs) puisque les conditions linguistiques qui me forcent à faire différentes analyses pour les facteurs linguistiques ne sont plus problématiques pour l'analyse sociologique. Aussi, cette analyse couvre moins d'occurrences que les autres (combinées) parce que certaines données sociologiques sont absentes du corpus. Les participants avaient reçu un questionnaire à remplir (annexe A), mais certains participants n'y ont pas répondu du tout ou pas répondu entièrement. L'outil statistique utilisé dans la présente étude, Rbrul (Johnson, 2009) (voir section 3.4 pour plus de détails), est incapable de traiter les trous dans les données et donc, des 822 occurrences retenues, 119 (14.48%) ont dû être omises de l'analyse sociologique parce que leurs réponses étaient incomplètes. Ceci a un impact relativement mineur sur la représentabilité des données pour les facteurs sociaux. Ces occurrences sont cependant conservées pour l'analyse des facteurs linguistiques. Il est important de noter que puisque l'étude présente se concentre sur le phénomène de l'élision du ⟨l⟩ dans les textos, le nombre d'occurrences considérées pour l'étude représente un sous-ensemble du corpus *Texto4science*. En effet, les pronoms les plus souvent utilisés dans les textos sont les *je* et les *tu* et les femmes en particulier emploient plus de pronoms de la première personne (Baron, 2004 ; Holtgraves, 2011), des résultats qui s'alignent avec des résultats obtenus dans le cadre d'étude sur l'oral (Mehl et Pennebaker, 2003 ; Pennebaker et King, 1999). La proportion relativement plus petite de pronoms de 3^e personne, en plus de la problématique des participants ayant mal répondu aux questionnaires, de la grande proportion de femmes dans l'étude et des nombreux cas de sujets absents relevés dans le corpus, expliquent pourquoi le nombre total d'occurrences considérées pour l'étude sociologique est comparativement bas par rapport au nombre total de textos dans le corpus.

L'annotation des textos s'est faite à partir d'un schéma XML (annexe B) et de l'éditeur XML <oxygen/>. Les chercheurs ont créé un protocole d'anonymisation des noms propres pour transformer les noms masculins en ⟨Florent⟩ ou ⟨Alain⟩, les noms féminins en ⟨Amélie⟩ ou ⟨Virginie⟩, les cas où les noms étaient ambigus en ⟨Aloïs⟩ et ⟨Claude⟩ et tous les noms de familles par ⟨Tremblay⟩. Les chercheurs ont procédé à la normalisation des textos dans le corpus, sur laquelle l'analyse présente se base presque entièrement. Dans la section 3.3, je décris la seule différence dans la normalisation que je considère pour mon domaine d'application, nommément la façon de normaliser les phrases de types *il y a*.

3.2 Participants

Pris dans son ensemble, le corpus comporte 298 personnes (63% de femmes). Cependant, comme mentionné plus haut, parce que l'étude présente se concentre sur l'élimination du <I> et qu'une partie des participants n'ont pas répondu correctement au questionnaire sociologique, le nombre de participants pris en compte dans mon analyse sociologique est nettement plus petit. Des 298 personnes ayant répondu au questionnaire (Langlais et Drouin, 2012), seulement 85 participants (74.12% de femmes) ont produit des occurrences de la variable étudiée. Pour les analyses linguistiques, 94 participants ont produit des occurrences de pronoms sujets masculins, 53 participants des pronoms féminins et 38 participants des POIs. Les femmes sont aussi responsables d'un plus grand nombre d'occurrences dans le corpus (85.01%). Cette disparité dans la distribution des participants a mené à de nombreux problèmes dans les analyses des facteurs sociaux qui seront discutés dans les sections 3.5.2.1 à 10. L'âge moyen des participants est 24.05, s'étalant de 15 à 60. L'échantillon penche vers une majorité de jeunes, ce qui aura des conséquences dans le codage des facteurs et dans l'interprétation des résultats. De plus, grâce aux codes postaux que les participants ont donnés, la provenance des participants retenus pour mon étude a pu être identifiée précisément. Le tableau 3.1 montre la distribution des participants par région ou municipalité.

Tableau 3.1 Distribution des participants par région

Région	N	%	Région	N	%	Région	N	%
Beloeil	1	1.15	Mascouche	1	1.15	Saint-Jérôme	2	2.30
Boucherville	1	1.15	Montréal	43	49.43	Saint-Léonard	1	1.15
Chambly	1	1.15	Sainte-Anne-de-Bellevue	1	1.15	Salaberry-de-Valleyfield	1	1.15
Châteauguay	1	1.15	Nepean	1	1.15	Sherbrooke	1	1.15
Delson	1	1.15	Louiseville	1	1.15	Saint-Jean-sur-Richelieu	1	1.15
Gatineau	3	3.45	Sainte-Thérèse	3	3.45	Terrebonne	1	1.15
Hull	3	3.45	Québec	5	5.75	Toronto	1	1.15
Lanaudière	1	1.15	Rimouski	1	1.15	Val-d'Or	1	1.15
Laval	2	2.30	Saint-Constant	1	1.15	Saint-Hubert	1	1.15
Longueuil	1	1.15	Saint-Hyacinthe	1	1.15	Saint-Jacques	1	1.15
Saint-Luc	1	1.15	Senneville	4	4.60			

Bien que le corpus *Texto4science* soit décrit comme un corpus de français québécois, près de la moitié des participants de mon étude viennent de Montréal. Le deuxième groupe le plus nombreux vient de Gatineau et de ses environs, puis le troisième vient de la ville de Québec. Les autres participants viennent d'un peu partout au Québec et hors du Québec (Nepean, Toronto), mais se concentrent dans des régions près de Montréal. Malgré la large distribution géographique des participants, tous déclarent avoir le français comme langue maternelle. Une large proportion des participants rapporte être des étudiants (32.94%), de loin le groupe le plus nombreux. Les autres emplois sont tous relativement peu représentés dans l'échantillon. Le tableau 3.2 ci-dessous présente la distribution des participants par domaine d'emploi tel que codé initialement par Langlais et Drouin (2012). Cette distribution n'a cependant pas été retenue pour les besoins de l'analyse sociologique. Les détails sur la nouvelle forme que prend ce facteur sont présentés dans le tableau 3.6 à la section 3.5.2.4.

Tableau 3.2 Distribution des participants par emploi

Emploi	Nombre	Pourcentage	Emploi	Nombre	Pourcentage
Étudiant	28	32.94	Foresterie	2	2.35
Technologie	6	7.06	Service à la clientèle	2	2.35
Vente	6	7.06	Administration	1	1.18
Communication	5	5.88	Assurance	1	1.18
Bureau	5	5.88	Bar	1	1.18
Économie	5	5.88	Évènementiel	1	1.18
Santé	5	5.88	Courrier	1	1.18
Éducation	4	4.71	Garderie	1	1.18
Arts	4	4.71	Génie	1	1.18
Service public	2	2.35	Sciences	1	1.18
Droit	2	2.35	Sports	1	1.18

3.3 Domaine d'application de la variable

La variable à l'étude est l'élision du <l> dans les pronoms personnels *il(s)* et *elle*, le pronom impersonnel *il* et le POI *lui* dans les messages textes. Le domaine d'application de la variable s'inspire des études précédentes sur le même phénomène à l'oral (Armstrong, 1996 ; Ashby, 1984 ; Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976). Il diffère partiellement du domaine d'application de ces études.

D'abord, la plupart de ces études incluait les articles *le*, *la*, *les*. Cependant, dans le corpus *Texto4science*, on ne retrouve que très peu de variation au niveau de l'orthographe de ces articles. En réalité, pour ces articles, ce sont davantage les voyelles qui sont éliminées plutôt que les ⟨l⟩, comme dans « [...] a voulait lfaire chier [...] »²⁰ (111030, *Texto4science*). Il s'agit d'un autre phénomène morphophonologique, l'élision du schwa, attesté en français oral (Morin, 1979 ; Walker, 1984 ; Côté, 2022), mais hors de la portée du projet actuel. Dans son analyse de l'élision du /l/ à l'oral, Sax (2003) se concentre sur les pronoms sujets et exclut les clitiques objets directs (*le*, *la*, *les*) et indirects (*lui*, *leur*) de son analyse. Dans le corpus *Texto4science*, on ne retrouve aucune élision avec les formes *leur*. Ces formes sont donc exclues. Le POI *lui*, cependant, présente de la variation et est maintenu dans l'analyse.

En ce qui concerne les formes féminines, les pronoms pluriels *elles* sont exclus de l'analyse des pronoms féminins parce que le corpus ne contenait que très peu d'occurrences ($n=12$, 5.88% des données féminines) et presque aucune occurrence n'était éliminée ($n=3$). Cette quasi-absence d'occurrences ne permettrait pas de faire des analyses à régressions multiples avec l'outil statistique employé ici. Ceci se reflète aussi dans les études sur l'oral (Armstrong, 1996 ; Sankoff et Cedergren, 1976) qui omettaient le pluriel féminin parce qu'il était trop rare. Aussi, rappelons que les pronoms pluriels sont syncrétiques à l'oral en français québécois, ce qui explique en partie leur absence dans le corpus. En effet, les données du corpus révèlent que le syncrétisme se retrouve aussi dans les textos. Il n'existe aucune forme plurielle non standard propre au féminin dans les textos. On retrouve plutôt les mêmes ⟨i⟩ ou ⟨y⟩ qu'au masculin, et ce même lorsque le texteur fait référence à un groupe composé uniquement de femmes, comme dans « [...] nan mais laa les filles y capotai a chk foi [...] »²¹ (110713, *Texto4science*). Dans les cas où le genre des référents était clairement féminin, le pronom était comptabilisé comme une occurrence du pronom féminin pluriel. Cependant, dans plusieurs cas, comme nous n'avons pas accès au contexte complet des textos, il est possible que des occurrences de féminin aient été traitées comme masculin à cause de ce syncrétisme et du manque d'information. Pour la forme singulière, on retrouve trois formes distinctes : *elle*, *a*, et *al*. Autrement dit, le féminin se distingue du masculin parce qu'il y a une combinaison d'effets qui détermine les formes employées. On retrouve à la fois de l'élision du ⟨l⟩ et de l'alternance vocalique (*elle* /ɛl/ → [æ]) ou [æ] au singulier), alors qu'avec le masculin, on ne retrouve que de l'élision. Cette alternance vocalique à l'oral pourrait s'expliquer par le besoin de distinguer plus facilement entre les pronoms *il* [i] et *elle* [ɛ]

²⁰ [...] *elle* voulait le faire chier [...]

²¹ Non mais là les filles elles capotaient à chaque fois [...]

(Armstrong, 1996). L'importance de distinguer les formes n'est pas aussi présente à l'écrit dans les textos. Pourtant, la forme ⟨a⟩, par sa forme inhabituelle, pourrait être plus saillante que les formes en ⟨i⟩ ou ⟨y⟩. Les occurrences de ⟨al⟩ sont traitées comme des occurrences « standards » malgré l'alternance vocalique parce que le ⟨l⟩ demeure présent. Dans le corpus étudié, la majorité des cas de ⟨al⟩ ont la forme ⟨a_l⟩, où le ⟨l⟩ est connecté à la première lettre du mot suivant comme dans « [...] pis a la juste 8 ans [...] »²² (110250, *Texto4science*). On le distingue des formes en ⟨a⟩ parce que le mot suivant, dans ce cas-ci le verbe *avoir*, ne commence pas par un ⟨l⟩. Il est donc évident qu'il s'agit du ⟨l⟩ du pronom en attaque de la syllabe suivant. Notons au passage qu'il semble y avoir aussi de l'apparence d'alternance vocalique pour le masculin, mais ce n'est en fait qu'une alternance graphique entre le ⟨i⟩ et le ⟨y⟩ présumément conditionnée par le contexte graphique entourant le pronom. La forme ⟨i⟩ est relativement rare ($n=35$, 6.78%) et se retrouve seulement 5 fois comme caractère seul (isolé des mots adjacents par un espace). Presque toutes les autres occurrences de pronoms ⟨i⟩ ont la forme ⟨ki⟩ (*qu'il*) ($n=30$). En contrepartie, la forme sujet ⟨y⟩, comme dans « *Non lol kan y repond pa c ke y la lus mai y vx pa repondre pcq sa y tente pa* »²³ (112339, *Texto4Science*), se retrouvent beaucoup plus fréquemment dans le corpus ($n=164$, 31.84%), surtout en caractère seul lorsque suivi d'une consonne ou collé lorsque suivi d'une voyelle (voir le tableau 5.1 de la section 5.2 pour une plus ample exploration de la distribution des formes masculines élidées en fonction de leur contexte graphique).

Armstrong (1996) exclut les pronoms impersonnels parce que ses résultats montraient qu'il n'y avait pas de variation. Le /l/ était toujours élidé dans ses données. Cependant, l'étude présente remarque de la variation pour les pronoms impersonnels à l'écrit dans les textos, ce qui suggère que le médium exerce une certaine influence sur la forme écrite du pronom et encourage la rétention du ⟨l⟩. Pour cette raison, les pronoms impersonnels sont conservés dans l'analyse. Ils font partie de la même analyse que les pronoms personnels masculins, mais sont codés différemment afin de les distinguer et de voir si la valeur sémantique, c'est-à-dire la quantité d'information transmise par le pronom (Sankoff et Cedergren, 1971, 1976), joue un rôle dans l'élision du ⟨l⟩.

Les pronoms emphatiques ⟨lui, elle⟩, comme dans « *lui, il écoute la game* » (110250, *Texto4Science*) ne varient pas dans le corpus et sont exclues de l'analyse présente. Les cas où le pronom est complètement élidé sont aussi omis car il ne s'agit pas du même phénomène. Les pronoms ⟨lui-même, elle-même⟩ ne

²² [...] et elle a juste 8 ans [...]

²³ *Non lol quand il ne répond pas c'est qu'il l'a lu mais il ne veut pas répondre parce que ça ne lui tente pas.*

varient pas non plus au niveau du ⟨l⟩ et sont omis, tout comme les pronoms inversés dans les phrases interrogatives comme « *Alain es t il chez toi ?* » (112310, *Texto4science*) pour les mêmes raisons.

Les différences qui suivent sont attribuées au changement de médium entre l'oral et l'écrit. En effet, le passage de l'oral à l'écrit change la façon dont on perçoit les données. Par exemple, Sax (2003) excluait les pronoms sujets lorsqu'ils étaient suivis de mots commençant par /l/ (par exemple l(l) lit le livre/l(l) l'aime bien) puisqu'il était impossible de distinguer le /l/ final du pronom et le /l/ initial du verbe et de l'objet à l'oral. Ces occurrences sont conservées dans l'étude actuelle puisque, malgré le fait que les frontières de mots soient plus fluides dans les textos, cette distinction est généralement marquée à l'écrit par un espace entre les mots (Il/Y/i l'aime). Il est donc possible de distinguer entre les formes élidées et les formes standards. Par exemple, dans « *Mais y la tu aek fido ton cell ?* »²⁴ (112339, *Texto4science*), le ⟨l⟩ qui suit le pronom ⟨y⟩ est séparé du sujet par un espace, ce qui indique qu'il s'agit du POI ⟨l'⟩ et non pas le ⟨l⟩ final du pronom élide ⟨y⟩.

En revanche, les occurrences de pronoms explétifs ayant la forme *il y a* ont dû être exclues parce qu'elles étaient trop ambiguës. En effet, il est impossible de différencier entre de l'élision ou des occurrences d'absence du sujet. Ce problème se manifeste lorsqu'on observe comment les chercheurs ont normalisé ces occurrences dans différents corpus. Dans le corpus *Texto4science* (Langlais et Drouin, 2012), les formes explétives non standards ⟨y⟩, comme dans *yavait*, sont normalisées comme *il y avait*. Dans le corpus suisse *Sms4science* (Dürscheid et Stark, 2011), un corpus produit dans le cadre du même projet international, ces occurrences sont normalisées comme *y avait* (sans pronom sujet). En accord avec la normalisation dans le corpus *Texto4science*, il est possible que les auteurs élident d'abord le ⟨l⟩ du *il* pour former ⟨i⟩, puis transforment le ⟨i⟩ en ⟨y⟩ (plus fréquent en forme seule) pour finalement combiner les deux ⟨y⟩ en un seul caractère pour éviter la redondance de caractères. Lorsque le ⟨y⟩ est un pronom sujet comme dans *yavait* (*il avait*), nous n'avons pas le même problème d'ambiguïté parce qu'on peut clairement identifier l'absence du ⟨l⟩ du pronom sujet et qu'il n'y a pas de ⟨y⟩ additionnel pour rendre l'interprétation ambiguë. À l'inverse, en accord avec la normalisation de Dürscheid et Stark (2011), il est également possible que les pronoms soient entièrement omis comme c'est le cas dans les langues *pro-drop* (Chomsky, 1993). Les langues *pro-drop* sont celles où le sujet peut être omis lorsqu'il est pragmatiquement ou grammaticalement identifiable. En espagnol, par exemple, le sujet est généralement identifiable par la

²⁴ *Mais il l'a-tu avec Fido ton cell ?*

flexion sur le verbe. Il est donc moins important pour bien comprendre la phrase que le pronom sujet soit réalisé. Similairement avec les pronoms explétifs, on peut facilement les omettre de la phrase en français à l'oral parce qu'ils sont essentiellement vides de sens et présents uniquement pour répondre aux contraintes grammaticales du français. Comme les occurrences de verbes sans pronom impersonnel ne sont pas rares à l'oral (Leeman, 2006 ; Zimmermann et Kaiser, 2014) et dans le corpus, voir par exemple « *Damn. C lundi. Fo ty metr .sinn koi2 9 ?* »²⁵ (111590, *Texto4Science*), l'analyse des chercheurs suisses semble plus probable. Les cas où les pronoms sont absents sont souvent des pronoms explétifs. De plus, il semble plus probable que le ⟨y⟩ complément soit retenu que le ⟨y⟩ pronom. On ne retrouve effectivement aucun cas où le ⟨y⟩ complément est éliminé comme dans **il a un chien dans la rue*. Ceci suggère que le ⟨y⟩ qu'on retrouve dans des constructions comme *yavait (il y avait)* est bel et bien le complément et qu'il s'agit de cas de pronom absent. Ultimement, le propos du travail présent n'étant pas de trancher entre ces deux analyses, il faudrait exclure ces occurrences.

Tableau 3.3 Distribution des pronoms omis

FORMES	N	Totaux	Proportion (%)
Pronoms masculins			
Singuliers personnels	3	301	0.99
Singuliers impersonnels	99	231	42.86
Météorologiques	0	5	0
Pluriels	12	92	13.04
Pronoms féminins			
Singuliers	12	216	5.56
Pluriels	1	13	7.69

Les cas de pronoms absents doivent aussi être omis de l'analyse puisqu'on ne peut pas parler d'élimination du ⟨l⟩ quand il n'y a simplement pas de pronom. Cependant, il est tout de même intéressant de voir comment ces pronoms se manifestent dans le corpus. La distribution des occurrences se retrouve dans le tableau 3.3 ci-haut. Sans surprise, on voit que les pronoms impersonnels sont de loin les cas les plus fréquents de pronoms absents (42.86%). Les pronoms météorologiques semblent cependant interdire l'omission du sujet malgré le fait qu'ils soient explétifs comme pour les pronoms impersonnels. On retrouve, par

²⁵ *Damn. C'est lundi. Il faut t'y mettre. Sinon quoi de neuf ?*

comparaison, très peu de pronoms singuliers personnels omis (0.99%). Il semble que le pluriel favorise l'omission (13.04%) pour les pronoms personnels masculins. Pour les pronoms féminins, la tendance est renversée, mais c'est probablement attribuable au fait qu'il y a si peu d'occurrences de pronoms pluriels féminins.

Un dernier point doit être considéré pour déterminer le domaine d'application de la variable. Considérant la nature des textos et des libertés orthographiques qui caractérisent leur usage, il arrive parfois que les occurrences comportent des fautes d'accord ou d'orthographe ou qu'il ait la forme de rébus. Ces occurrences sont comptabilisées en fonction de leurs formes normalisées. Par exemple, dans « [...] *Les Habs ils faut continuer a jouer [...]* » (110292, *Texto4science*), le pronom est écrit comme la forme plurielle alors qu'il s'agit du *il* explétif. Cette occurrence serait donc traitée comme un *il* explétif. Les pronoms comportant des fautes d'orthographe (coquilles) sont comptabilisés de la même manière, en fonction de la présence ou de l'absence du ⟨l⟩ et non en fonction de la justesse de la forme orthographique employée.

En résumé, les variantes considérées pour l'étude sont présentées en (8-15). Par soucis de brièveté, ces occurrences ne montrent qu'un exemple pour chaque variante. Pour une liste exhaustive d'exemples pour chaque catégorie de pronoms, consultez l'annexe C. Les formes avec ⟨l⟩ correspondent à la forme standard. Elles incluent *il* sujet (8), *il* explétif (incluant les pronoms météorologiques) (9), *ils* (10), *elle* (11) et le POI *lui* (12). Les formes sans ⟨l⟩ observées dans le corpus sont *i* (13) et *y* (14) pour les formes de masculin (incluant les explétif, les pronoms météorologiques et les pluriels ainsi que le clitique objet *lui*) et *a* (15) pour les formes de féminin singulier.

- (8) « Kk mais attends jvais voir si **il** peut avant..²⁶ » (110250, *Texto4science*)
- (9) « **Il** en reste » (110492, *Texto4science*)
- (10) « Ca fait longtemps qu'**ils** y sont ? » (111032, *Texto4science*)
- (11) « Ben **elle** mamene :P donc elle retourne chez elle.. » (110250, *Texto4science*)
- (12) « [...] On **lui** a prescrit du repos. » (110637, *Texto4science*)
- (13) « Daccord :) kes ki faut que jamene ? :)²⁷ » (110250, *Texto4science*)

²⁶ *Okay, okay, mais attends : je vais voir s'il peut avant...*

²⁷ *D'accord, :) qu'est-ce qu'il faut que j'amène? :)*

(14) « Y pleut ?²⁸ » (113010, *Texto4science*)

(15) « Pis la webcam a s'ouvre, a se ferme [...] » (110250, *Texto4Science*)

Pour des raisons qui seront explorées davantage dans la section 3.5.1, je suis forcé de considérer les pronoms masculins, féminins et les POIs comme trois domaines d'application distincts, et donc d'en faire l'analyse séparément. L'analyse des facteurs sociolinguistiques porte cependant sur l'ensemble des variables.

3.4 Procédures et analyse statistique (Rbrul)

Afin de relever toutes les occurrences de la variable dans les corpus, le corpus, en format XML, est importé dans *Excel*. Le fait que les textos aient déjà été normalisés en facilite énormément l'usage. En effet, comme l'orthographe est flexible dans les textos, travailler à partir des textos bruts peut être extrêmement difficile parce que les outils de recherche sont incapables de les traiter correctement. Par exemple, les pronoms sont souvent joints à d'autres mots sans marque de ponctuation, ce qui les rends difficiles à distinguer d'autres homographes. Il arrive souvent que *qu'il* soit écrit comme <ki>, comme dans « *Daccord :) kes ki faut que jamene ? :)* » (110250, *Texto4science*). Or, <ki> peut faire référence à *qui*, *qu'ils* ou *qu'elles*. Aussi, effectuer des recherches à partir des formes élidées <i> ou <a> serait encore plus inefficace parce que le logiciel est incapable de distinguer les <i> ou <a> pronoms des lettres dans n'importe quels autres mots. Les textos normalisés permettent d'éviter ces problèmes. En utilisant la fonction de recherche d'*Excel*, on peut utiliser les formes normalisées et chercher spécifiquement toutes les occurrences des caractères <il>, <elle> et <lui>. Ça permet aussi de faire ressortir les formes plurielles en même temps. Cette fonction n'est toutefois pas parfaite puisqu'elle détecte quand même les caractères <il>, <elle> et <lui> dans d'autres mots comme *Vaudreuil* (111818, *Texto4science*), *appelle* (111930, *Texto4science*) ou *celui* (110250, *Texto4science*). Il faut aussi vérifier si les pronoms dans les textos normalisés sont présents ou non dans le vrai texto afin d'identifier les cas de pronoms absents. Il faut donc malgré tout passer ligne par ligne pour déceler toutes les occurrences, et seulement les occurrences concernées par le domaine d'application énoncé en 3.3. Une fois les occurrences extraites du corpus, on peut consulter les réponses au questionnaire sociologique, aussi sous la forme d'un fichier XML, afin d'associer chaque occurrence à ses données sociologiques correspondantes.

²⁸ *Il pleut ?*

²⁹ *Pis la webcam elle s'ouvre, elle se ferme [...]*

Le script Rbrul (Johnson, 2009) dans la plateforme R (version 4.3.2) (R Core Team, 2023) est utilisé pour procéder aux différentes analyses quantitatives proposées. Afin d'assurer la validité des analyses et d'éviter des problèmes de colinéarité dans les facteurs, j'ai procédé à une série de vérifications. Ces différentes vérifications m'ont permis de vérifier l'indépendance des facteurs et ont motivé des changements au domaine d'application de ma variable et à la façon dont mes données ont dû être codées.

3.5 Facteurs à l'étude

Le choix des facteurs linguistiques et sociologiques sélectionnés pour l'étude actuelle est inspiré par les études faites sur le phénomène de l'élision à l'oral (Armstrong, 1996 ; Ashby, 1984 ; Bougaïeff et Cardinal, 1980 ; François, 1974 ; Howard et al., 2006 ; Ostiguy, 1979 ; Poplack et Walker, 1986 ; Pope, 1952 ; Pupier et Légaré, 1973 ; Sankoff et Cedergren, 1976 ; Sax, 2003) puisque aucune étude n'a encore été faite sur le sujet de l'élision à l'écrit. Ils ne sont cependant pas identiques. Certains facteurs, comme les traits phonologiques spécifiques des segments entourant l'occurrence, ne se prêtent pas bien à une étude sur l'écrit et d'autres, comme la méthode de rédaction, doivent être pris en compte spécialement parce qu'on étudie les textos. Ceci est dû en partie au fait que le langage SMS ne remplit pas la même fonction que la langue écrite en général. Bien que la modalité soit l'écrit dans les deux cas, les textos jouent généralement un rôle plus conversationnel et ludique que la langue standard. La rectitude grammaticale et orthographique du contenu des textos y est moins importante. Les conditions de production et de perception des textos comme la méthode de rédaction (claviers à 12 touches, claviers QWERTY, autocomplétion) et les limites de caractères, ainsi que des paramètres situationnels comme la relation entre les texteurs et le but de l'interaction influencent la composition des messages (Hård af Segerstad, 2002).

Le tableau 3.4 ci-dessous dresse la liste des facteurs retenus pour l'étude actuelle. Pour voir la feuille de codage, consultez l'annexe D. Les données sociales sont tirées du travail de Langlais et Drouin (2012)³⁰, mais dans certains cas, le codage utilisé à l'origine a dû être modifié pour accommoder les données utilisées ici et régler des problèmes de distribution. Le cas échéant, les changements sont documentés dans la section du facteur impliqué.

³⁰ Les réponses aux questionnaires sociologiques m'ont gracieusement été partagées par Patrick Drouin et Philippe Langlais.

Tableau 3.4 facteurs à l'étude

Facteurs linguistiques	Facteurs sociologiques
1. Contexte phonologique précédant (pour POI) ou suivant (pour les pronoms sujets) l'occurrence 2. Catégorie grammaticale de l'occurrence (pour les pronoms masculins) 3. La catégorie syntaxique du mot suivant l'occurrence 4. Longueur du texto (en nombre de caractères)	1. Âge 2. Genre 3. Niveau de scolarité 4. Niveau de langue requis pour l'emploi 5. Depuis combien de temps le participant texte. 6. Nombre de textos envoyés par jour 7. Capacité de décodage 8. Utilisation de dictionnaires 9. Type de clavier utilisé 10. Nombre d'heures de lecture par semaine 11. Nombre d'heures d'écriture par semaine

3.5.1 Facteurs linguistiques et hypothèses

Les facteurs linguistiques considérés pour l'étude présente vont varier en fonction du rôle grammatical de la variable observée. Par exemple, les pronoms sujets et explétifs ont un comportement qui diffère largement des POIs. Notamment, la position du ⟨l⟩ dans la structure syllabique diffère entre les pronoms sujets et explétifs (en coda) et les POIs (en attaque). Lors d'analyses phonologiques de la langue parlée, ce sont les segments adjacents qui importent (Sankoff et Cedergren, 1976). Conséquemment, dans l'analyse des pronoms sujets et explétifs, c'est le contexte suivant l'occurrence qui est observé alors que pour les pronoms objets, c'est le contexte précédant l'occurrence qui est observé. Aussi, l'étude présente diffère dans sa capacité à prendre en compte plusieurs variables qui sont communes dans les études sur l'élision du /l/ à l'oral. Ces études observent l'élision dans toutes sortes de contextes syntaxiques et phonologiques parce qu'elles analysent l'élision dans les noms, les articles, les verbes et les pronoms (Armstrong, 1996 ; Ashby, 1984 ; Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976). Cependant, parce que l'étude présente se concentre spécifiquement sur les ⟨l⟩ des pronoms, les mêmes facteurs ne peuvent pas tous être utilisés de la même façon. Les règles grammaticales du français restreignent fortement les facteurs et causent de nombreuses interactions dans les données. Par exemple, si l'on considère le genre et le nombre grammatical des pronoms explétifs, on remarque rapidement qu'il ne peut y avoir que des pronoms

singuliers et masculins, ce qui cause des effets *knockouts*³¹ dans les analyses (Tagliamonte, 2006). Pour éviter ces effets, on pourrait considérer traiter les pronoms sujets et explétifs séparément. Une telle analyse ne permettrait cependant pas de voir si la valeur sémantique (Sankoff et Cedergren, 1971, 1976) du pronom influence l'élision. En effet, un pronom qui fait référence à une vraie entité, comme le texteur ou son destinataire, défavorise peut-être l'élision par rapport à un pronom explétif qui n'apparaît que pour répondre aux besoins des règles syntaxiques du français. Une meilleure solution serait de traiter le féminin et le masculin séparément. En effet, le féminin se présente différemment dans les textos. Tel que décrit dans la section 3.3, le féminin diffère du masculin par la combinaison d'effet qui produit les différentes formes de *elle*. L'analyse des pronoms féminins devrait aussi être faite à part puisque la distinction personnel/explétif n'existe pas avec les pronoms féminins. Nous créerions à nouveau un problème de distribution en les conservant dans le même modèle statistique.

Ultimement, comme mentionné précédemment, ces problèmes me poussent à devoir faire trois analyses linguistiques distinctes : une première pour les pronoms sujets masculins et explétifs, une deuxième pour les pronoms sujets féminins et une troisième pour les POIs. Les résultats seront présentés séparément dans le prochain chapitre.

3.5.1.1 Contexte phonologique suivant/précédant l'occurrence

Dans son étude sur l'élision du /l/ dans le corpus de Dieuze, une commune en France, Armstrong (1996) avance que les voyelles basses sont plus facilement perceptibles par rapport aux voyelles hautes et favoriserait l'élision du /l/ dans les articles. De même, les voyelles postérieures favoriseraient l'élision. Puisque mon étude porte sur un corpus écrit, il est impossible d'avoir une représentation phonologique exacte de la façon dont un texteur prononcerait le texto qu'il a rédigé. Pour cette raison, il est difficile de préciser l'analyse en observant les phonèmes spécifiques qui suivent ou précèdent les occurrences (selon si on observe des pronoms sujets ou objets respectivement) comme Poplack et Walker (1986) l'ont fait. Dans le cadre du travail présent, je me limite au contraste binaire entre les contextes préconsonantiques et prévoicaliques basé sur les graphèmes, contrairement aux études sur l'oral qui observent aussi le contexte prétyod (comme dans *il y a*) (Ashby, 1984 ; Sankoff et Cedergren, 1976) et parfois d'autres segments plus spécifiques (Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976). Comme les occurrences de type *il y a* sont exclues de l'analyse présente (tel que précisé dans la section 3.3), le contexte prétyod

³¹ Un effet *knockout* se produit lorsque qu'une valeur de 0% ou 100% se retrouve dans une des cellules de l'analyse, signalant qu'il n'y a pas de variation dans les données pour un contexte précis.

n'est pas retenu comme niveau de facteur. Sankoff et Cedergren (1976) et Ashby (1984) rapportent que le contexte préconsonantique entraîne davantage d'élision que le contexte prévocalique. On pourrait s'attendre à des résultats similaires dans les textos si le langage SMS est un reflet de la langue orale, de sorte que les contraintes phonologiques qui influencent l'élision à l'oral exercent une influence similaire à l'écrit. Si, au contraire, les taux d'élisions dans les textos sont largement différents des résultats à l'oral, nous pourrions croire que le médium de l'écrit exerce une influence marquée sur les formes employées dans les textos malgré l'affaiblissement des normes d'écriture des textos. Par exemple, un texteur pourrait choisir d'employer la forme non élidée <kil> (*qu'il*) pour éviter l'ambiguïté avec l'homographe <ki> (*qui*).

3.5.1.2 Catégorie grammaticale de l'occurrence

Un simple regard sur le contexte phonologique ne suffit pas pour rendre compte de l'élision dans les pronoms. En effet, dans des contextes phonologiques identiques, le fait que /l/ se trouve dans un pronom ou dans un nom influence radicalement la possibilité d'élider le /l/. Par exemple³², pour un segment comme /ilva/, l'élision est possible en (16), mais pas en (17). Or, en (16), le /l/ se trouve dans le pronom *il*, alors qu'en (17), le /l/ se trouve dans le nom *fil*. Il apparaît donc évident que la catégorie grammaticale de l'occurrence³³ est un facteur important pour l'élision à l'oral.

(16) *il va*

(17) *le fil va*

Dans le cadre de l'étude présente, comme les seules occurrences observées sont celles des pronoms, cette distinction n'est pas tenue en compte. Cependant, un regard plus fin sur la catégorie grammaticale du pronom va permettre de faire la différence entre les occurrences de pronoms sujets et explétifs. Ceci implique que ce facteur ne sera tenu en compte que pour les pronoms masculins. Dans les études sur l'élision du /l/ à l'oral (Ashby, 1984 ; Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976), les pronoms

³² Exemples tirés de Sankoff et Cedergren (1976, p. 1103).

³³ Il est également possible que l'élision soit encodée dans le lexique et ne s'applique pas en général à une catégorie grammaticale. Le cas échéant, ce seraient les propriétés propres au mot, plutôt que sa catégorie grammaticale, qui déterminerait s'il est susceptible à l'élision. Compte tenu des variables observées, qui sont des catégories grammaticales dont tous les éléments subissent de l'élision (pronom sujet/objet de 3^e personne), l'importance de la catégorie grammaticale comme facteur influençant l'élision est peut-être illusoire. Cette question demeure non résolue dans le cadre de l'étude actuelle. Crucialement, le facteur est retenu pour comparer les taux d'élision avec des pronoms ayant différentes propriétés grammaticales (sujet vs objet, personnel vs impersonnel).

impersonnels entraînent souvent plus d'éllision que les pronoms personnels. Sankoff et Cedergren (1976) avancent que cette différence de taux d'éllision serait attribuable à la valeur sémantique du pronom. Le pronom personnel transmet plus d'information que le pronom impersonnel et serait donc plus fort ou plus résistant à l'éllision. Ces études rapportent aussi que les pronoms sujets favorisent l'éllision par rapport aux pronoms objet. S'il est vrai qu'on écrit comme on parle dans les textos, le même phénomène devrait se manifester dans les données du corpus.

3.5.1.3 Catégorie syntaxique du mot suivant/précédant l'occurrence

Poplack et Walker (1986) rapportent que l'environnement syntaxique suivant l'occurrence exerçait une influence sur les taux d'éllision avec les pronoms objet et les articles, mais qu'il était non significatif pour les pronoms sujets. Les pronoms objet suivis de verbes conjugués, de semi-auxiliaires et d'auxiliaires subissaient davantage d'éllision que les pronoms suivis par des verbes à l'infinitif. En raison de ces résultats, dans le cadre de l'étude présente, j'observe la catégorie syntaxique du mot suivant l'occurrence pour les pronoms sujets et les POIs.

Pour les pronoms sujets, les niveaux de facteurs se limitent aux occurrences précédées par des verbes et celles précédées par des mots « autres ». La catégorie « autres » inclut des adverbes (de négation), des objets directs et indirects ainsi que des pronoms compléments et réflexifs. La mauvaise distribution des données me force à produire un regroupement hétérogène de catégories syntaxiques pour former les « autres ». Bien que ce facteur se soit avéré non significatif pour Poplack et Walker (1986), la saillance des verbes pour les pronoms objet suggère qu'ils devraient favoriser l'éllision par rapport aux « autres » pour les pronoms sujets.

En ce qui concerne les POIs, tout comme Poplack et Walker (1986), j'ai tenté d'observer le contexte syntaxique suivant l'occurrence. Les niveaux de facteurs sont les verbes conjugués (incluant les semi-auxiliaires et les auxiliaires), les verbes à l'infinitif et une catégorie « autres » pour traiter de l'entourage syntaxique des constructions impératives. Ce dernier niveau de facteur n'est pas explicitement traité par Poplack et Walker (1986), mais la configuration syntaxique des impératifs suggère qu'ils forment bien une catégorie à part. En effet, il arrive qu'on retrouve des POIs en fin de proposition comme dans le texto « Tu c kwa? Demande-lui. [...] »³⁴ (111590, *Texto4science*). Dans d'autres cas, comme « Dis-lui qu'elle m'appelle

³⁴ *Tu sais quoi? Demande-lui. [...]*

[...] » (110250, *Texto4science*), le pronom est suivi d'un autre mot dans la même proposition. Évidemment, les cas où les POIs se trouvent en fin de proposition causent un problème au niveau du codage parce qu'il n'y a rien qui suit l'occurrence. Une option pour contourner ce problème aurait pu être d'omettre ces occurrences car elles sont peu nombreuses ($n=10$), mais à l'oral, on observe de la variation au niveau de l'élision du /l/ dans cette construction. Il est donc difficile de justifier l'omission, d'autant plus que le nombre d'occurrences considéré pour l'analyse des POIs est déjà très bas ($n=103$). J'ai plutôt opté de coder les occurrences de constructions impératives comme « autres » afin d'accommoder leur contexte particulier. En effet, dans mes données, les occurrences dans des constructions impératives peuvent être suivies par rien, par des compléments, par des pronoms emphatiques, par des déterminants, par des adverbes ou par des noms. Évidemment, ils ne peuvent pas être suivis d'un verbe puisqu'ils suivent déjà le verbe eux-mêmes. Tout comme pour Poplack et Walker (1986), je m'attends à ce que les verbes conjugués favorisent l'élision par rapport aux verbes infinitifs pour les phrases déclaratives. L'impact de la catégorie « autres » pour les impératives demeure difficile à prédire parce qu'elle représente un vaste ensemble de catégories syntaxiques qui pourraient tous avoir un impact différent, mais indétectable dans un échantillon aussi petit qui me force à les regrouper.

3.5.1.4 Longueur des textos

Spécialement dans le cadre d'une étude portant sur les textos, la longueur des messages pourrait influencer le choix du texteur d'élider ou non le ⟨l⟩ de ses pronoms. La longueur du message est considérée en fonction du nombre de caractères (incluant les espaces) et est traitée comme une variable continue. Pour les pronoms masculins, le message le plus court contient 6 caractères et le plus long, 606. Pour les pronoms féminins, le message le plus court contient 13 caractères et le plus long, 352. Finalement pour les POIs, le plus court contient 16 caractères et le plus long, 352. À noter que le type de clavier pourrait avoir un impact sur ce facteur. Pour les messages rédigés avec des claviers à 12 touches (avec ou sans la technologie T9), avec lesquels on doit appuyer plusieurs fois sur une touche pour produire certains caractères, le nombre de caractères ne pourrait pas être comparé avec les messages rédigés avec des claviers QWERTY. Par exemple, un simple message comme *salut* avec un clavier QWERTY prend 5 touches à écrire, alors qu'il en prendrait 11 avec un clavier à 12 touches ($4 \times 7 = s$, $1 \times 2 = a$, $3 \times 5 = l$, $2 \times 8 = u$, $1 \times 8 = t$). Le problème est pire si l'on utilise le nombre de touches puisque tous les téléphones à clavier QWERTY ne font pas les accents, les symboles ou la ponctuation de la même façon. Aussi, les fonctions d'autocomplétion et d'écriture prédictive comme T9 pourraient aussi avoir un impact sur le nombre réel de caractères que le texteur a dû saisir sur son téléphone pour générer un mot. Pour ces raisons, le nombre

absolu de caractères est retenu comme facteur pour l'analyse linguistique. Les données seront séparées en fonction du type de clavier utilisé pour rédiger le message lors de l'analyse des facteurs sociaux (voir section 3.2.3.9). Parce que le langage SMS sert entre autres à accélérer la communication, je m'attends à ce que plus un message est long, plus il y a de chance que les <I> dans les pronoms soient élidés.

Il est cependant probable que le type de clavier ait un impact, de sorte que la longueur des messages pour les messages rédigés avec des claviers à 12 touches favorise davantage l'élosion que pour les messages rédigés sur claviers QWERTY. Aussi, il est important de noter que le nombre de caractères dans un texte pourrait ne pas être entièrement précis pour deux raisons. D'abord, le protocole d'anonymisation décrit en 3.1 fait en sorte qu'il peut y avoir une certaine disparité entre le nombre de caractères du nom original et le nom anonymisé. Ensuite, certains messages sont tronqués ou incomplets dans le corpus *Texto4science*, mais leur nombre est suffisamment petit ($n=7$ dans toutes les analyses, 0.84% des occurrences) pour avoir un impact négligeable sur la validité des analyses. Ces messages se trouvent en (18-24) ci-bas. Pour (24), l'occurrence de <ils> à la fin du message ne permet pas de voir le mot suivant donc ne peut pas faire partie de l'analyse présente.

- (18) « I decided that demain, mon PP va etre fini et quil va juste me rester a print/plastifier ma creation et a faire les modifications de mon enonce kand Tremblay va m'envoyer la version corrigeed! Jte dit. Je genre sort pas du bureau tant ke c »³⁵ (110250, *Texto4science*)
- (19) « est un techno junkie à bougies faque il va aller se coucher. Snif. »³⁶ (110030, *Texto4science*)
- (20) « Mon modele de lunettes existe plus. Donc il a fallu commander les 2 branches... ils vont les recevoir d'ici mercredi.. Pour les verres de contacts, pour la premiere fois, il faut prendre un rdv.. Et cette semaine il restait juste pendant » (110250, *Texto4science*)
- (21) « Ffuuuuccckkkk! Jpense qu'on doit laisser faire pr a soir/demain! Pcq j'avais oublier que j'avais dit a Virginie et Virginie que je ferais dequoi ds la journee.. Lol j'avais oublier ca :P oops mais ouais, a moins quelles changent la date, je p »³⁷ (110250, *Texto4science*)

³⁵ *I decided that demain, mon PowerPoint va être fini et qu'il va juste me rester à print/plastifier ma création et à faire les modifications de mon énoncé quand Tremblay va m'envoyer la version corrigée! Je te le dis. Je genre sors pas du bureau tant que c'est*

³⁶ *C'est un techno junkie à bougies, ça fait qu'il va aller se coucher. Snif.*

³⁷ *Ffuuuuccckkkk! Je pense qu'on doit laisser faire pour à soir/demain! Parce que j'avais oublié que j'avais dit à Virginie et Virginie que je ferais de quoi dans la journée... Lol J'avais oublié ça. :P Oups mais ouais, à moins qu'elles changent la date, je p*

- (22) « Salut! Jme demandais juste a titre informatif sil restait dla place ds les autos de ceux qui repartent le vendredi de tremblant. Pcq jvx rester jusquau samedi mais mere vient dme dire que messemble que jai un rdv de dentiste le samedi. Et »³⁸ (110250, *Texto4science*)
- (23) « Ouais Jsais quelle fait tjrs sa mais jpensais pas qua le ferait la, alors que Chus sous le choc genre lol si j'avais eu mettons 60 sa »³⁹ (111816, *Texto4science*)
- (24) « Allo. As TU pris tes mails? On s en vient la malade et moi. Stp peux tu défaire son lit je veux le laver et qu elle ne voit pas qu ils » (110660, *Texto4science*)

3.5.2 Facteurs sociaux et hypothèses

Les facteurs sociaux à l'étude sont essentiellement les mêmes qui se retrouvent dans le questionnaire de Langlais et Drouin (2012) (voir annexe A). Cependant, tout comme pour les facteurs linguistiques, ce ne sont pas tous les facteurs sociaux qui sont utilisables dans le cadre de l'étude présente. Certains facteurs, présentés dans le tableau 3.5 ci-bas, n'ont pas été repris pour l'étude présente parce qu'ils n'avaient vraisemblablement pas de portée sur la variable étudiée. Les facteurs qui ont été retenus sont ceux qui témoignent du niveau de compétence en langage SMS et en langue standard. Les facteurs relevant de la langue parlée par les participants sont évidents considérant la nature du corpus et des participants qui le constituent. En effet, compte tenu du sujet du travail présent, il n'est pas étonnant de voir que les participants visés par le corpus rapportent tous le français comme langue maternelle. Il n'est donc pas utile de retenir ce facteur pour mon analyse. Certains des facteurs relevant de l'utilisation des textos ne sont pas utiles à l'étude présente soit parce qu'ils relèvent de facteurs externes aux texteurs (par exemple combien de textos il reçoit par jour) ou qu'ils sont trop généraux et détachés de l'usage concret et immédiat du texteur. Par exemple, le classement des contextes de rédaction est omis parce que savoir dans quels contextes le texteur écrit généralement ses textos est moins pertinent que de connaître le contexte exact d'un envoi, une information qui n'est pas disponible dans le corpus. En revanche, le classement des destinataires (avec qui le texteur communique le plus par jour, question 14 dans le questionnaire en Annexe A) pourrait être un facteur intéressant à considérer dans le cadre d'une étude

³⁸ *Salut! Je me demandais juste, à titre informatif, s'il restait de la place dans les autos de ceux qui repartent le vendredi de Tremblant. Parce que je veux rester jusqu'au samedi, mais mère vient de me dire qu'il me semble que j'ai un rendez-vous de dentiste le samedi. Et*

³⁹ *Ouais, je sais qu'elle fait toujours ça, mais je ne pensais pas qu'elle le ferait là, alors que je suis sous le choc, genre. lol Si j'avais eu mettons 60 sa*

sociologique sur les habitudes de textage et la manière dont les communautés linguistiques influencent ces habitudes.

Sur le plan linguistique, il est possible qu'un texteur qui texte plus régulièrement avec ses amis qu'avec sa famille ou des collègues développe un style de rédaction plus propice à l'emploi de textismes. Similairement, le rapport entre le texteur et le receveur pourrait influencer le registre employé dans un échange texto et les choix de formes écrites du texteur. Malheureusement, les données disponibles ne permettent pas de tenir le classement des destinataires en compte dans l'analyse présente. D'abord, quand on considère les destinataires de premier rang (à qui les participants rapportent texter le plus), on remarque rapidement qu'une grande majorité des participants textent davantage avec leurs amis (64.41%), puis avec leur partenaire (22.94%), leur famille (10.74%) et trop peu avec leurs collègues (1.91%). Ceci causerait des problèmes de distribution dans les données. Ensuite, certains répondants ($n=26$) ont identifié plusieurs destinataires avec le même rang (par exemple parce qu'ils textent aussi souvent avec leurs amis et leur famille), ce qui causerait encore plus de problèmes dans le codage des données. De plus, les catégories de destinataires sont problématiques parce qu'elles manquent de clarté. La catégorie « famille » ne précise pas si on fait référence aux parents, grands-parents, frères et sœurs ou autres et on ne texte pas de la même façon à tous les membres de notre famille. Similairement, « collègue » ne précise pas s'il s'agit d'un supérieur ou d'un collègue de même rang, un collègue proche ou une connaissance de bureau. Cette catégorie ne prend également pas en considération le fait que certains collègues pourraient être des amis et il n'est pas clair comment un participant coderait ces collègues avec lesquels il est plus proche. Qui plus est, même au sein de notre groupe d'ami, on ne texte pas de la même manière à tout le monde. Finalement, il n'est pas possible dans tous les cas d'identifier le rapport entre les auteurs dans un message texte et tous les participants n'ont pas fournis d'exemples de textos adressés à chaque catégorie de destinataires. Il est donc difficile de prendre en compte ce facteur dans le cadre de l'étude présente même s'il est envisageable qu'il ait un impact sur la variable.

Enfin, les facteurs relevant des habitudes de vie du texteur (comme le type de lecture et d'écriture) sont omis au profit du nombre d'heures de lecture et d'écriture parce qu'il est difficile de quantifier les différents types de textes en termes d'apport à la compétence linguistique, à savoir si lire un livre est « meilleur » que lire une bande dessinée ou un article de journal. Se limiter au temps consacré à la lecture et l'écriture permet d'éviter de faire des jugements de valeur sur les habitudes des participants.

Tableau 3.5 Facteurs sociaux omis

Facteurs omis	
• Langue maternelle • Langue parlée couramment • Langue parlée le plus fréquemment • Le nombre de textos reçus par jour • Classement des destinataires • Styles de rédaction des textos • Mélange de langues dans les textos	• Utilisation d'autres langues que le français dans les textos • Classements des buts d'utilisation des textos • Classement des contextes de rédaction • Autres modes de communication • Autres types de lecture • Autres types d'écriture

Aussi, les niveaux de certains facteurs ont dû être modifiés afin d'accommoder les données de mon analyse. Par exemple, le niveau d'éducation des participants a été condensé en un facteur à deux niveaux pour éviter des problèmes de distribution dans les données. Ce faisant, on perd en résolution mais on décelle tout de même les tendances des effets des facteurs.

3.5.2.1 Âge

Dans le cadre du projet présent, l'âge est traité comme un facteur à valeur continue (de 15 à 60 ans) et sert à détecter s'il y a un effet générationnel sur l'élimination dans les textos. Cependant, il est difficile de prédire si l'effet en sera un de gradation d'âge ou de changement générationnel (Labov, 1972). En effet, il est possible que les participants qui ont grandi sans téléphone mobile (plus âgés) et ceux qui ont grandi avec des téléphones dès l'enfance (plus jeunes) éliminent moins les <I> que ceux qui ont obtenu leur premier téléphone mobile à l'adolescence, à l'époque où le langage SMS et ses pratiques langagières transgressives se développaient. Comme ce sont généralement les adolescents qui sont à l'avant-garde des pratiques de textage (Eckert, 2017 ; Grinter et Eldridge, 2003 ; McCulloch, 2019), on s'attendrait à ce que la cohorte « adolescente » élide davantage ses pronoms dans leurs textos. Considérant que le langage SMS est devenu populaire au début des années 90 et que le corpus *Texto4science* a été conçu en 2009, cette cohorte « adolescente » correspondrait aux participants qui avaient entre 30 et 40 ans lors de l'étude de Langlais et Drouin (2012). Cependant, cette mesure est approximative parce que les jeunes commencent à texter et à employer la MI de plus en plus tôt (Verheijen, 2013). Les taux d'élimination observés dans les données pourraient montrer une courbe en cloche pour montrer un effet de gradation d'âge (Labov, 1972). Les plus jeunes emploieraient davantage la langue standard parce qu'ils sont à l'école et subissent

l'influence, quoique potentiellement faible (Poplack, 2015), de la norme et du milieu scolaire. Ensuite, les adolescents commenceraient à transgresser les normes (autrement dit à développer le langage SMS). Finalement, les plus vieux abandonneraient le langage SMS en partie peut-être à cause de son association avec la langue des jeunes. Il est aussi possible que l'influence de l'école soit assez faible pour que les jeunes élident autant, voire plus, que les adultes. On obtiendrait alors une courbe en S descendante. Cette hypothèse présume que le comportement de textage des participants évolue à travers le temps. Si, au contraire, le comportement de textage des participants demeurerait stable, nous aurions plutôt affaire à un changement générationnel, où les plus vieux emploient moins de textismes parce qu'ils étaient adultes au moment où le langage SMS est apparu et qu'ils ont moins participé à son développement et n'ont pas acquis les mêmes comportements de textage. Leurs textos reflèteraient donc davantage la norme écrite. La cohorte des « adolescents » (environ 30 à 40 ans) aurait quant à elle appris à texter à l'adolescence et aurait conservé le même style à l'âge adulte, ce qui se traduirait par plus d'élosion pour ce groupe d'âge. Les plus jeunes qui étaient trop jeunes pour avoir des téléphones quand le langage SMS est apparu n'auraient pas acquis les textismes dans la même mesure que la cohorte « adolescente » et donc auraient conservé un taux d'élosion plus bas au moment de la conception du corpus *Texto4science*. L'hypothèse du changement générationnel semble cependant peu probable puisque le langage SMS ne montre aucun signe de disparaître chez les plus jeunes, ce qui pointe plutôt vers un autre facteur pour expliquer un taux d'élosion différent pour eux par rapport aux « adolescents » et aux plus vieux.

En effet, le développement de la technologie pourrait être responsable de la différence de taux d'élosion entre les cohortes. En 2012, dans une étude sur 799 jeunes de 12 à 17 ans, le centre de recherche PEW (Lenhart, 2012) rapportait que seulement le quart des adolescents avaient des téléphones intelligents à clavier QWERTY comme les *BlackBerry* et que la moitié rapportait avoir des téléphones réguliers à 12 touches. À l'inverse, presque la moitié des adultes rapportaient avoir des téléphones intelligents à clavier QWERTY. Compte tenu du fait que ce type de téléphone vient avec des fonctions qui sont indispensables dans le monde professionnel comme l'accès sans fil aux courriels et à internet, il y a plus de chance que les adultes possèdent des téléphones intelligents et qu'ils s'en servent davantage dans le cadre de leur emploi (en employant un registre de langue plus professionnel). Les adultes (les plus vieux en 2012) auraient ainsi moins de chances d'élider les <l> de leurs pronoms parce que le contexte professionnel décourage les textismes, et donc l'élosion. Ceux qui étaient adolescents quand ils ont eu leurs premiers téléphones (les adultes en 2012) ont quant à eux plus de chances d'avoir eu des téléphones réguliers à 12 touches (moins dispendieux et avec moins de fonctions) et de communiquer de façon plus

informelle avec leurs amis et leur famille (Grinter et Eldridge, 2003). Ils auraient donc plus de chance d'employer des textismes et d'élider les ⟨l⟩ de leurs pronoms que les adultes. Quant à eux, les plus jeunes qui ont eu leur téléphone plus tard en termes de développement de la technologie, c'est-à-dire les participants qui étaient trop jeunes dans les années 90 pour avoir des téléphones et qui n'ont pas baigné dans le langage SMS à son apparition, ont plus de chances d'avoir eu des téléphones plus modernes à clavier QWERTY avec des dictionnaires et des outils d'autocomplétion et d'autocorrection plus performants. Les chances que leurs textos contiennent des ⟨l⟩ élidés sont donc plus basses, comme pour les plus vieux. On retrouverait alors la même courbe en cloche, mais pour des raisons différentes.

3.5.2.2 Genre

L'impact du genre sera potentiellement étonnant dans la mesure où le parler des femmes tend à être plus standard que celui des hommes lorsque les normes sont clairement prescrites (Labov, 1990 ; Chambers, 1992 ; Holmes, 1993 ; James, 1996 ; Baron, 2004). Par exemple, à l'oral, Ashby (1984) montre une légère différence de taux d'élision entre les hommes et les femmes, où les hommes sont plus enclins à élider le /l/ que les femmes (80% contre 70%), des résultats similaires à Sankoff et Cedergren (1976) et Poplack et Walker (1986). Cependant, dans le contexte des textos, où la norme est relâchée (l'orthographe standard est moins valorisée) et où le langage SMS côtoie la langue standard, il est difficile de prédire la direction de l'influence du genre sur les comportements d'élision. En effet, les femmes montrent un comportement qui peut paraître contradictoire par rapport à la norme linguistique. Elles tendent à moins employer des formes linguistiques explicitement non standards, mais elles tendent aussi à être la source d'innovation, ce que Labov (1990) appelle le paradoxe du genre. Il est donc possible qu'elles emploient plus de textismes et qu'elles élident davantage leurs ⟨l⟩. Il est aussi possible qu'elles résistent davantage à l'élision et emploient d'autres formes de textismes plus fréquemment que les hommes. Aussi, selon Ostiguy et Tousignant (2008), certains traits phonétiques du français (comme l'élision du /l/ dans les pronoms) sont non marqués et n'auraient pas de valeur sociale. Il n'y aurait donc peut-être pas de différenciation de genre en ce qui a trait à l'élision du /l/ puisque ni son absence ou sa présence ne sont critiquées (Carolanne Paquin-Drouin, 2017 [Manuscrit non publié]). Finalement, il serait important de rappeler la grande disparité entre le nombre d'occurrences produites entre les femmes et les hommes. Ceci pourrait influencer la représentabilité des données.

3.5.2.3 Niveau de scolarité

La langue standard est largement apprise et imposée à l'école. Il semble donc évident que le niveau de scolarité d'un participant soit corrélé avec son taux d'utilisation de textismes, et donc d'élosion des <I>. De plus, des études de niveaux supérieurs (baccalauréat et plus) requièrent généralement un niveau de langue plus soutenu et exposent les étudiants à davantage de textes complexes et demandant linguistiquement. Il en découle qu'on s'attendrait à ce que les participants qui ont fait des études avancées soient plus habiles avec la langue standard et accordent plus d'importance à la qualité de leur langue. Dans le cadre du questionnaire (Langlais et Drouin, 2012), les participants étaient appelés à laisser des commentaires sur leur usage des textos et du langage SMS. Parmi les participants ayant fait des études supérieures qui ont émis des commentaires (et qui font partie de mon échantillon), environ la moitié exprime une attitude négative à l'égard du langage SMS, alors que l'autre moitié exprime une attitude neutre, c'est-à-dire qu'ils emploient le langage SMS parce qu'ils y sont contraints et non parce qu'ils cherchent à l'utiliser. Je n'ai trouvé aucun commentaire positif sur les textos. On s'attendrait alors à ce que cette attitude se traduise par des taux d'élosion inférieurs pour les participants ayant fait des études supérieures. Malheureusement, trop peu de participants ($n=25$) ont laissé des commentaires portant sur leur attitude envers les textos et le langage SMS alors je ne peux pas retenir l'attitude comme facteur dans l'analyse présente⁴⁰. Le niveau de scolarité permettra malgré tout de voir si l'éducation peut influencer l'usage de textismes et l'élosion dans les textos. Dans le questionnaire de Langlais et Drouin (2012), les participants pouvaient répondre primaire, secondaire, Cégep⁴¹, université premier cycle, université cycle supérieur, ou autre. Ici encore, pour des raisons de problèmes de distribution, les niveaux sont réduits à préuniversitaire (primaire, secondaire, cégep et écoles professionnelles) et universitaire (niveau bac et plus), et on prédit que les participants ayant fait des études universitaires élident moins leurs <I>.

3.5.2.4 Niveau de langue requis pour l'emploi

Le manque d'équilibre dans la distribution des emplois fait en sorte que ce facteur doit être recodé différemment de la façon dont Langlais et Drouin (2012) avaient procédé. Or il est difficile de trouver une autre catégorisation qui permettrait de voir comment l'emploi pourrait influencer la variable. Les niveaux de facteur proposés par Langlais et Drouin (2012) regroupent déjà les emplois par domaine. L'approche

⁴⁰ Voir section 5.7 pour une discussion plus approfondie sur les commentaires des participants.

⁴¹ La catégorie Cégep regroupait le niveau cégep pour les participants du Québec, les collèges pour les autres provinces et les écoles professionnelles.

de Sankoff et Laberge (1978) offre une piste de solution. Ils rapportent qu'une simple corrélation entre la variable linguistique étudiée et la classe sociale d'un participant ne rend pas compte du rapport à la langue standard du participant. Ils notent, par exemple, que les professeurs, les acteurs et les réceptionnistes parlent généralement de façon plus standard que d'autres personnes dans des positions sociales ou économiques similaires. De la même façon, des participants rapportant travailler dans un même domaine professionnel n'occupent pas forcément des postes qui requièrent le même niveau de langue. Pour contourner ce problème, Sankoff et Laberge (1978) ont mis au point une échelle d'intégration dans le marché linguistique (Bourdieu et Boltanski, 1975) qui représentait l'importance relative de la langue légitimisée (standard, de l'élite, d'enseignement, etc.) dans la vie socioéconomique du locuteur. Pour ce faire, les chercheurs ont fait appel à 8 sociolinguistes (professionnels et étudiants) avec plusieurs années d'expérience avec la communauté francophone de Montréal et leur ont demandé d'évaluer le degré de participation au marché linguistique, entre 0 (participation minimale) et 1 (participation maximale). Pour ce faire, chaque juge a reçu 120 fiches (1 par participant) contenant chacune quelques détails de la vie socioéconomique d'un participant.

Malheureusement, je ne peux pas faire appel à cette procédure puisque je n'ai pas le même accès aux informations sociales de mes participants. Je suis donc contraint de considérer uniquement le domaine d'emploi des participants et de présumer le niveau de langue requis par le domaine d'emploi. Tout comme on présupposait en 3.5.2.3 que les études supérieures requièrent un niveau de langue typiquement plus soutenu que le niveau collégial, secondaire et primaire, on peut concevoir que différents domaines d'emplois requièrent différents niveaux de langue. Le tableau 3.6 ci-bas montre le découpage fait pour réduire le nombre de niveaux de facteurs à 2 pour Rbrul. Notons tout de suite que ces regroupements sont subjectifs et, en soit, problématiques. Comme le mentionne Sankoff et Laberge (1978), cette approche est problématique puisqu'elle classifie des emplois et non des individus. Aussi, elle ne donne qu'un portrait statique d'un participant (l'emploi qu'il occupait au moment de l'étude). Cette alternative semble toutefois préférable à celle d'abandonner la dimension professionnelle de l'analyse présente parce qu'il apparaît probable qu'elle exerce une influence sur la variable étudiée puisqu'elle avait généralement un impact sur la variable dans les études sur l'élosion à l'oral. Notons aussi que, pour l'étude présente, le critère de sélection est spécifiquement l'importance de la langue standard écrite pour la profession, contrairement au classement de Sankoff et Laberge qui se basait sur l'importance de la langue standard (vraisemblablement à l'oral et à l'écrit). Pour cette raison, une profession comme « service à la clientèle », qui demande un certain niveau de langue à l'oral mais qui ne demande pas forcément un niveau aussi

élevé à l'écrit, se retrouve dans le champ « autres ». La catégorie « autres » représente la catégorie d'emplois pour lesquels la langue standard écrite n'est pas importante, selon moi, pour accomplir des tâches professionnelles. Ce n'est pas pour autant que les participants qui occupent ces emplois n'accordent pas d'importance à la qualité de leur langue écrite. Ça veut simplement dire que la qualité de la langue écrite n'est pas un obstacle aux tâches professionnelles impliquées par l'emploi. À l'inverse, la catégorie « standard » représente les emplois pour lesquels un niveau de langue standard est implicitement important à cause de la nature même de l'emploi, comme l'exprime ce commentaire : « Je suis une enseignante de français. J'ai donc un peu une réputation à maintenir. [...] » (110500, *Texto4science*).

Tableau 3.6 Niveau de langue requis par l'emploi

Autres			Standard
Agriculture	Administration	Technologie	Éducation
Bar	Événementiel	Étudiants	Traduction
Foresterie	Bureau	Arts	Bibliothèque
Garderie	Économie	Service publique	Communications
Service à la clientèle	Génie	Courrier	Droit
Sports	Santé		Assurances
Vente	Sciences		
n=559 (79.2%)			n=147 (20.8%)

Les étudiants sont traités comme « autres » pour accommoder la disparité dans leurs domaines d'étude. Le questionnaire de Langlais et Drouin (2012) ne précisait pas le domaine d'étude des étudiants alors nous ne pouvons pas savoir s'ils sont tous issus des mêmes programmes et tous les programmes ne requièrent pas forcément le même niveau de langue. Un étudiant en littérature pourrait se classer dans la catégorie « standard » alors qu'un étudiant en administration se classerait dans « autres » selon le découpage proposé. Le découpage standard/autres est justifié par le fait que ce qui semble saillant dans le facteur est le besoin pour la langue standard ou pas. Avoir un aperçu plus graduel de l'impact du niveau de langue sur la variable aurait été souhaitable, mais une telle analyse n'est pas permise par mes données. Ici encore, si

le textage est un reflet de la langue oral, on s'attendrait à ce que les participants occupant un emploi qui demande un niveau de langue standard à l'écrit élide moins leurs ⟨l⟩.

3.5.2.5 Depuis combien de temps le participant texte

Bien qu'on puisse facilement voir le lien entre l'âge et le niveau de compétence linguistique avec la langue standard, il n'est pas forcément corrélé avec la compétence dans le langage SMS. En effet, de la même manière qu'on apprend à mieux écrire à force d'être exposé ou d'employer la langue standard, on apprend à utiliser les textismes à force d'interagir avec ce médium et d'autres texteurs. Ainsi, une autre façon d'opérationnaliser le niveau de compétence des texteurs est d'observer depuis combien de temps ils emploient les textos. Une personne qui texte depuis plus longtemps est présumée être plus compétente en langage SMS et risque d'employer plus de textismes plus souvent et plus librement qu'une personne qui texte depuis moins de temps. Il est donc possible que l'élision du ⟨l⟩ soit corrélée avec plus d'années d'expériences avec les textismes. Dans le cadre de l'étude présente, ce facteur est divisé en trois niveaux : ceux qui textent depuis moins de 2 ans, ceux qui textent depuis 2 à 5 ans et ceux qui textent depuis plus de 5 ans. Ces catégories diffèrent de celles proposées par Langlais et Drouin (2012) largement à cause de problèmes de distribution des données.

3.5.2.6 Nombre de textos envoyés par jour

Parce que les textos répondent à la maxime de rapidité (Thurlow et Brown, 2003), on s'attendrait à ce que les texteurs cherchent à accélérer leur rédaction en employant les textismes qui réduisent le nombre de caractères nécessaires, spécialement s'ils en envoient plusieurs par jours. De plus, une personne qui communique souvent par textos serait plus familière avec le langage SMS et ses pratiques et potentiellement plus encline à l'employer. Je m'attends donc à ce que le nombre de textos envoyés par jour soit positivement corrélé avec le taux d'élision des texteurs, de façon similaire aux nombres d'années d'expérience de textage. Ce facteur est séparé en trois niveaux de facteurs : ceux qui envoient moins de 5 textos par jour, ceux qui envoient entre 6 à 10 textos par jour et ceux qui envoient plus de 10 textos par jour.

3.5.2.7 Capacité de décodage du langage SMS

La capacité de décodage du langage SMS rapportée par les participants est traitée comme une mesure de leur niveau de compétence linguistique en langage SMS, séparée en deux niveaux : ceux pour qui le décodage du langage SMS est facile et ceux pour qui c'est difficile. L'hypothèse est qu'une personne qui

rapporte avoir de la difficulté à décoder le langage SMS élidera probablement moins ses <I> qu'une personne qui a de la facilité à lire le *textese*.

3.5.2.8 Utilisation de dictionnaires

L'emploi d'un dictionnaire ou de fonctions d'écriture prédictive (comme T9) facilite la rédaction de textos en rendant le processus plus ou moins automatique. Normalement, pour rédiger un mot avec un clavier à 12 touches, il faut appuyer plusieurs fois sur une touche pour faire le bon caractère. Avec T9, un texteur peut simplement appuyer sur la touche correspondant à la lettre qu'on veut écrire et le téléphone offre les mots possibles parmi les différentes textonymes⁴². Bien que ces fonctions étaient encore jeunes dans leur développement par rapport aux technologies d'autocorrection moderne, elles étaient déjà capables d'apprendre de l'usage des texteurs et de leur proposer les mots qu'ils employaient le plus. Le texteur pouvait alors écrire sans faute et sans effort. Ainsi, on s'attendrait à ce qu'une personne qui emploie régulièrement ces fonctions ait moins tendances à élider ses <I> puisque son téléphone fait une partie du travail. Pour l'étude présente, ce facteur est divisé en deux niveaux : ceux qui emploient ces fonctions et ceux qui ne les emploient pas. Dans le questionnaire de Langlais et Drouin (2012), les participants pouvaient préciser s'ils utilisaient toujours, souvent, parfois ou jamais ces outils. La distribution des données pour l'étude présente me force à réduire ce facteur à deux niveaux.

3.5.2.9 Type de clavier utilisé

Dans l'étude de Langlais et Drouin (2012), ils faisaient la différence entre les téléphones réguliers (à 12 touches), les téléphones intelligents à clavier QWERTY (comme les téléphones *BlackBerry*), ceux à clavier tactile (comme les *iPhones*) et ceux avec stylet (*Palm*). Pour l'étude présente, la distinction importante est entre les téléphones à 12 touches et ceux à clavier QWERTY (les iPhones employant des claviers QWERTY sur des écrans tactiles, ils sont jugés suffisamment similaires aux téléphones à clavier QWERTY à boutons pour être regroupés en un seul niveau de facteur). Cette distinction est justifiée par le fait que tous les claviers ne sont pas aussi faciles d'usage que les autres (Wood, 2014). En effet, les anciens claviers à 12 touches avec lesquels il faut appuyer plusieurs fois sur une touche pour faire un seul caractère encouragent l'utilisation de textismes pour couper dans le temps de rédaction. C'est en partie ce qui a mené à l'avènement du langage SMS en premier lieu. Par la suite, les nouveaux claviers imitant le clavier

⁴² Des textonymes sont des mots en langage SMS qui sont rédigés en employant les mêmes touches sur un clavier à 12 touches. Par exemple, les mots *good* (bon) et *home* (maison) sont tous les deux écrits en appuyant sur les touches 4-6-6-3.

QWERTY de l'ordinateur ont permis aux auteurs de transposer leur compétence de rédaction à l'ordinateur vers leur téléphone. Il ne faut cependant pas négliger l'impact de la taille de l'écran et des touches qui les rendent tout de même plus difficiles à utiliser que des ordinateurs. Bien que les claviers QWERTY soient plus fonctionnels et rapides que les claviers à 12 touches, le fait que des textos peuvent être rédigés dans des contextes actifs (en mouvement, lors d'autres activités) contribue à l'apparition de textismes. Pour ces raisons, il est probable que les claviers QWERTY défavorisent l'élision par rapport aux claviers à 12 touches.

3.5.2.10 Nombre d'heures de lecture et d'écriture par semaine

Le nombre d'heures consacrées à la lecture et à l'écriture par semaine est traité comme des mesures d'engagement avec la langue standard. Bien que similaires, ils sont traités individuellement dans l'analyse sociologique afin de voir si l'un ou l'autre influencent davantage l'élision du <l> dans les textos. Ils sont traités comme des valeurs continues allant de 0,5 à 20 heures de lecture et de 0 à 12 heures d'écriture. Langlais et Drouin (2012) demandaient aussi aux participants d'ordonner les types de textes⁴³ qu'ils lisaient ou rédigeaient. Dans le cadre de l'étude présente, les types précis étaient moins importants que le simple fait de lire et écrire. De manière similaire avec le niveau de scolarité, je présume ici qu'une personne étant exposée à la langue standard de manière régulière aura un meilleur rapport avec elle qu'une personne qui lit ou écrit moins et que ça se traduirait en meilleure qualité de sa langue. Du moins, c'est généralement la position défendue par les médias et les enseignants. Ces facteurs sont pris en compte ici en partie afin d'essayer de répondre à ces « accusations » selon quoi les jeunes écrivent en langage SMS parce qu'ils ne savent pas lire et écrire.

⁴³ Pour la lecture, les auteurs considéraient les articles de journaux, les bandes dessinées/mangas, les pages de sites Web, les textes de chansons les livres et autres. Pour l'écriture, ils considéraient les petits mots (mémos), les correspondances traditionnelles, les journaux intimes et autres.

CHAPITRE 4

RÉSULTATS

4.1 Taux globaux

Les taux globaux des variantes sont présentés dans les tableaux 4.1-3 ci-bas. Les pronoms qui ont subi le plus d'élision sont les pronoms masculins (38.64 %), puis les pronoms féminins (24.51%) et finalement les POIs (20.39%). Parmi les pronoms sujets masculins, les pronoms personnels singuliers ont subi légèrement plus d'élision, puis les pronoms pluriels et les impersonnels. Comme mentionné plus tôt, les résultats des analyses linguistiques seront présentés en trois temps.

Tableau 4.1 Taux globaux, pronoms sujets masculins

FORMES		N	Proportion (%)
Ensemble des pronoms masculins (38.64% d'élision)		515	
Pronoms personnels singuliers (39.93% d'élision)			
Avec ⟨l⟩	<i>il</i>	179	60.07
Sans ⟨l⟩	<i>i</i>	17	5.70
	<i>y</i>	102	34.23
Sous total		298	100
Pronoms impersonnels (35.61% d'élision)			
Avec ⟨l⟩	<i>il</i>	85	64.39
Sans ⟨l⟩	<i>i</i>	12	9.09
	<i>y</i>	35	26.52
Sous total		132	100
Pronoms météorologiques (60.00% d'élision)			
Avec ⟨l⟩	<i>il</i>	2	40.00
Sans ⟨l⟩	<i>y</i>	3	60.00
Sous total		5	100
Pronoms personnels pluriels (37.50% d'élision)			
Avec ⟨l⟩	<i>ils</i>	50	62.50
Sans ⟨l⟩	<i>i</i>	6	7.50
	<i>y</i>	24	30.00
Sous total		80	100

Tableau 4.2 Taux globaux, pronoms sujets féminins

FORMES		N	%
Pronoms singuliers (24.51% d'élision)			
Avec ⟨l⟩	<i>elle</i>	148	72.55
	<i>al</i>	6	2.94
Sans ⟨l⟩	<i>a</i>	50	24.51
Totaux		204	100

Tableau 4.3 Taux globaux, pronoms objets indirects

FORMES		N	%
Pronoms singuliers (20.39% d'élision)			
Avec ⟨l⟩	<i>lui</i>	82	79.61
Sans ⟨l⟩	<i>y</i>	21	20.39
Totaux		103	100

En comparant les taux d'élision du ⟨l⟩ dans le corpus *Texto4science* avec les taux quasi catégoriques d'élision du /l/ dans les pronoms observés sur l'oral (Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976), on remarque rapidement que l'élision est beaucoup plus rare qu'on aurait pu s'y attendre si le langage SMS était un simple reflet de l'oral comme le prétend le discours populaire sur les textos. Le tableau 4.4 reprend en partie l'information du tableau 2.1 (voir section 2.1) et présente les taux d'élision du /l/ rapportés à l'oral dans les études portant sur les communautés canadiennes et les résultats obtenus pour le corpus *Texto4science*. Comme dans les études mentionnées ci-haut, les pronoms masculins subissent davantage d'élision que les autres pronoms. En revanche, les pronoms féminins subissent plus d'élision que les POIs dans les textos, alors que c'est l'inverse à l'oral. Poplack et Walker (1986), par exemple, rapportent des taux d'élision de 84% pour *elle* et 99% pour *lui*. On peut déjà voir, sur la base des taux obtenus, que l'élision dans les textos ne se manifeste pas de la même façon qu'à l'oral. Les sections 4.2 et 4.3 exploreront plus en détails les différents facteurs linguistiques et sociologiques étudiés afin de voir de quelle manière ils affectent l'élision du ⟨l⟩ dans les textos.

Tableau 4.4 Comparaison des taux d'élision entre les études

Formes	Étude sur les textos	Études sur l'oral	
	Corpus <i>Texto4science</i>	Sankoff et Cedergren, 1976	Poplack et Walker, 1986
<i>Il</i> sujet	39.93%	89%	100%
<i>Il</i> impersonnel	35.61%	98%	100%
<i>Ils</i>	37.50%	92%	99%
<i>Elle</i>	24.51%	63%	84%
<i>Elles</i>	omis	Trop rares	33%
<i>Lui</i>	20.39%	Omis (rare)	99%

4.2 Analyses linguistiques

4.2.1 Pronoms sujets masculins

Le tableau 4.5 ci-bas présente les résultats pour l'analyse des facteurs linguistiques pour les pronoms sujets masculins. On observe le plus d'élision (39.93%) avec les pronoms sujets masculins, ce qui fait écho aux résultats pour les études sur l'oral (Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976). Cependant, si les taux d'élision étaient quasi catégoriques dans ces études, les taux observés dans l'étude présente sont étonnamment bas. De plus, les seuls facteurs qui se sont avérés significatifs sont le contexte phonologique suivant l'occurrence, la catégorie syntaxique du mot suivant et le nombre de caractères. Pour les études sur l'oral, la catégorie grammaticale (personnel vs impersonnel, sujet vs objet) et l'environnement syntaxique étaient généralement significatifs. Cette différence peut s'expliquer en partie par la façon différente de coder les données entre les études sur l'oral et l'étude présente. En effet, mon analyse observe les pronoms sujets et objets séparément pour des raisons de distribution de données et de contraintes imposées par la grammaire du français, notamment la différence de types de mots qui peuvent se retrouver dans l'environnement syntaxique des pronoms. Bien qu'ils aient fait partie d'analyses distinctes et donc qu'ils ne soient pas totalement comparables, les taux globaux révèlent tout de même que les POIs n'étaient élidés que dans 20,39% des cas. Ces résultats s'alignent avec les études citées ci-haut qui rapportaient des taux d'élision du /I/ plus bas pour les POIs. Ceci suggère que la catégorie grammaticale pourrait influencer l'élision de sorte que les pronoms sujets favorisent l'élision du ⟨I⟩ par rapport aux objets. L'étude présente ne peut cependant pas confirmer ou infirmer cette hypothèse parce que l'analyse n'est pas faite de la même façon que dans les études sur l'oral : elle inclut seulement la

différence entre les pronoms personnels et impersonnels. Sankoff et Cedergren (1976) avaient eux aussi dû retirer *lui* de leurs analyses par manque de données. L'étude présente est configurée pour tester si la valeur sémantique des pronoms influence l'élision. Les résultats suggèrent que ce facteur est non significatif. Bien qu'à l'oral les pronoms impersonnels favorisaient généralement l'élision, cet effet semble disparaître dans les textos. Le fait que les pronoms personnels transmettent plus d'information ne les a pas rendus visiblement plus résistants à l'élision que les pronoms explétifs « vide de sens ».

Tableau 4.5 Analyse linguistique des pronoms sujets masculins

Pronoms sujets masculins				
N total = 515				
Valeur d'application : sans ⟨l⟩, 38.64% d'élision				
FACTEURS	N	Taux (%)	Poids relatif	Logodds
Contexte phonologique suivant l'occurrence (p = < 0.01)				
Voyelle	168	47.61	0.602	0.414
Consonne	347	34.29	0.398	-0.414
Catégorie syntaxique du mot suivant l'occurrence (p = 0.0343)				
Verbe	375	42.13	0.585	0.343
Autres	140	29.28	0.415	-0.343
Nombre de caractères (valeur continue) (p = < 0.01)				
+1	Logodds = -0.006			
Facteurs non significatifs				
Catégorie grammaticale du pronom				

Le contexte phonologique prévocanique favorise l'élision dans les textos. Environ la moitié des occurrences de pronoms sujets masculins devant une voyelle étaient éliminées. Le contexte préconsonantique, quant à lui, défavorise d'élision. Environ le tiers des occurrences étaient éliminées. Ces résultats viennent à l'encontre des résultats rapportés par Sankoff et Cedergren (1976) et Poplack et Walker (1986) pour les communautés canadiennes et par Ashby (1984) pour la communauté de Tours. Ils rapportaient que le contexte préconsonantique, mais surtout le contexte pré-yod, causait de 30 à 50% plus d'élision que le contexte prévocanique.

Le renversement de la tendance observée dans mes données pourrait à nouveau s'expliquer par la différence dans la manière dont les données ont été codées. Les deux études sur le français canadien (Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976) retenaient le contexte pré-yod comme niveau de facteur (en plus de préconsonantique et prévocalique), alors que je retiens uniquement le contexte préconsonantique et prévocalique dans le cadre de l'étude présente. Ces auteurs ne procurent malheureusement aucun exemple d'occurrence en position pré-yod. Je ne peux donc pas être certain des sortes d'occurrences auxquelles ils font référence. Seul Ashby (1984) donne un exemple de pronom pré-yod ([ijadetryk] *il y a des trucs*) dans son article pour la communauté de Tours. Curieusement, selon Ashby, il y aurait un [i] dans la représentation phonologique du pronom, vraisemblablement le [i] du *il*. Le yod correspondrait quant à lui au y complément. La présence d'un [i] est peut-être une différence entre le français européen et le français laurentien. En effet, je ne dirais pas qu'il y a de [i] pré-yod dans *il y a des trucs* en français québécois selon mon intuition comme locuteur natif de cette variété. Cependant, sans avoir accès aux enregistrements originaux des entrevues faites par Ashby (1984), je ne peux que spéculer sur la représentation phonologique de cet exemple. Cette différence est toutefois importante parce qu'elle mène à des interprétations différentes des données. Selon l'approche d'Ashby, le pronom [i] se retrouve en position pré-yod parce qu'il est suivi du [j] y complément. Au contraire, sans ce [i], le [j] dans [jadetryk] pourrait être considéré comme un *il y* réduit à [j]. Le segment qui suit le pronom [j] serait alors la voyelle [a]. Il s'agirait donc d'un cas prévocalique. Rappelons aussi que, comme mentionné en 3.3, il est également possible que le pronom soit entièrement élide et que le [j] soit le pronom y complément. Cette différence dans la façon de voir la forme sous-jacente du yod [j] pourrait expliquer en partie pourquoi les résultats de l'étude présente diffèrent autant des résultats des études citées ci-haut. On compte 177 occurrences de type *il y a* dans le corpus *Texto4science*⁴⁴. Ces occurrences montrent un taux d'élision de 75.71%. Si ces occurrences avaient été conservées dans l'échantillon et traitées comme des cas pré-yods comme l'avait fait Ashby (1984), ce contexte aurait favorisé l'élision davantage que le contexte prévocalique ($n=168$, 47.6% d'élision). Savoir comment Ashby (1984) codait des cas comme [(i)jave] *il avait*, où il n'y a pas de y complément, aurait pu clarifier la façon de traiter le pronom. Si l'on présume qu'il y a un [i] pré-yod, le yod se serait introduit devant le [a] de *avait* par processus d'épenthèse pour éviter une suite de deux voyelles consécutives une fois le /l/ du pronom élide. Selon cette interprétation, le pronom [i] serait en position pré-yod. À l'inverse, sans ce [i], le [j] pourrait être interprété comme le pronom en soit et donc se retrouve en position prévocalique. Faute d'exemples de ce type de phrase dans les études citées

⁴⁴ Rappelons que ces occurrences ont été retirées de mon échantillon (voir 3.3 pour les raisons derrière ce choix).

plus haut et faute d'avoir accès aux enregistrements originaux de ces études, je ne peux pas confirmer la présence ou l'absence d'un [i] dans la prononciation des participants, et ce pour les deux communautés linguistiques en question. Je ne peux pas non plus confirmer si Ashby (1984) traitait des cas de *il avait* comme étant préyods ou prévocaliques. L'absence totale de cas de *⟨iyavait⟩ *il avait* ou de formes similaires dans le corpus *Texto4science* suggère cependant qu'il n'y a pas de [i] à l'oral, du moins pas en français québécois. En effet, lorsqu'un texteur emploie des constructions de type *il avait* avec un /l/ élidé, il remplace le ⟨i⟩ par un ⟨y⟩, potentiellement pour imiter la prononciation du [j] à l'oral. L'absence de *⟨iyavait⟩ dans le corpus pourrait aussi être causée par le passage de l'oral aux textos, où l'économie de caractère semble primer.

Cette hypothèse semble toutefois moins probable considérant les résultats pour le nombre de caractères. En effet, on observe une tendance inverse où plus un message est long, moins on trouve d'élision. Ce résultat est particulièrement étonnant considérant les maximes de Thurlow et Brown (2003). On s'attendrait à voir, au contraire, une corrélation positive entre la longueur des messages et le taux d'élision puisque l'un des objectifs du langage SMS est d'accélérer la communication et donc vraisemblablement de réduire le nombre de caractères par textos.

Finalement, tout comme Poplack et Walker (1986), qui proposaient que la saillance des verbes causerait davantage d'élision, les pronoms suivis de verbes dans le corpus de texto étaient plus souvent élidés que ceux suivis des « autres ». Bien que cet effet soit similaire dans les deux médiums, les raisons qui le motivent demeurent difficiles à cerner. Il est possible que le fardeau sémantique de la phrase soit porté par le verbe et que le pronom sujet puisse être réduit à sa forme élidée. Lorsque suivis d'un « autre », généralement des éléments moins saillants, le pronom retient sa forme standard.

4.2.2 Pronoms féminins

Tout comme pour les pronoms masculins, le nombre de caractères est significatif pour les pronoms féminins de sorte que le nombre de caractères dans un texto est inversement corrélé avec le taux d'élision des ⟨l⟩ dans les pronoms *elle*. De plus, la corrélation semble plus forte pour les pronoms féminins que masculins. Cette différence est d'autant plus surprenante considérant que, non seulement la corrélation va dans le mauvais sens selon les hypothèses formulées au chapitre 3, mais l'effet est plus fort pour un pronom plus long à écrire. On se serait plutôt attendu à la situation inverse, où les ⟨l⟩ s'élient plus pour

elle parce qu'il y a un plus gros gain à faire en termes de rapidité et d'économie de caractères. Les résultats pour les pronoms féminins sont résumés dans le tableau 4.6.

Tableau 4.6 Analyse linguistique des pronoms féminins

Pronoms féminins				
N total = 204				
Valeur d'application : sans ⟨l⟩, 24.51% d'élision				
FACTEURS	N	Taux (%)	Poids relatif	Logodds
Catégorie syntaxique du mot suivant l'occurrence ($p = 0.0346$)				
Verbe	50	32.00	0.674	0.724
Autres	154	22.08	0.326	-0.724
Nombre de caractère (valeur continue) ($p < 0.01$)				
+1	Logodds = -0.021			

La catégorie syntaxique du mot suivant l'occurrence a le même effet pour les pronoms féminins, où les verbes favorisent l'élision alors que les « autres », qui rappelons-le contient des adverbes de négation, des objets directs et indirects ainsi que des pronoms compléments et réflexifs, défavorisent l'élision. L'effet est cependant plus fort pour les pronoms féminins que masculins.

Malheureusement, le contexte phonologique a dû être omis de l'analyse à cause de problèmes de distribution dans les données. Concrètement, aucun pronom féminin n'a subi d'élision en position prévocallique, ce qui cause des effets *knockouts* qui rendent impossible l'inclusion de ce facteur dans le modèle statistique. De plus, le texto « Et toi? Pourquoi? Elle en a? » (112295, *Texto4science*) contient la seule occurrence d'un pronom féminin suivi par un mot de catégorie « autres » et suivi d'une voyelle, ce qui n'aurait pas permis d'assurer l'indépendance de ces facteurs. Les résultats de Sankoff et Cedergren (1976) et de Poplack et Walker (1986) montraient que le contexte prévocallique causait beaucoup moins d'élision que le contexte préconsonantique. De plus, ils montraient que la forme [æ] est généralement utilisée devant les voyelles /ɑ æ ɔ/ alors que la forme élidée [ɛ] (non attestée à l'écrit dans le corpus *Texto4science*), comme dans [ɛ:tæle] *elle est allée*, se retrouve souvent devant les voyelles antérieures /ɛ e/. Ces réalisations alternatives de *elle* formaient alors des séquences où la voyelle s'allongeait (Poplack et Walker, 1986). Ashby (1984) donne un exemple contraire, où *elle* se réalise comme

un [ɛ] devant /æ/ ([ɛæprepare] *elle a préparé*), mais en général, on retrouverait plutôt [æ:prepare] en français laurentien. Or, comme l’allongement vocalique ne se transpose pas à l’écrit dans les textos, l’élision prévocalique semble interdite dans mes données pour les pronoms féminins.. En ce qui concerne l’absence quasi totale de cas de *elle* devant une voyelle dans un mot de catégorie « autre », les contraintes du français en sont ici encore probablement la cause. En effet, très peu de mots de la catégorie « autres » commencent par une voyelle. Quelques exemples notoires sont *y* et *en*, qui sont plus fréquents avec les *il* impersonnels. Le fait que le contexte phonologique n’ait pas pu être testé pour les pronoms féminins est dommage considérant qu’il était significatif pour les pronoms masculins (et les POIs, voir section 4.2.3) et que, à l’exception des constructions impersonnelles, les pronoms féminins sujets se retrouvent dans des contextes phonologiques identiques aux pronoms masculins sujets.

4.2.3 Pronoms objet indirects

Sankoff et Cedergren (1976) avaient omis les POIs de leurs analyses par manque de données. Le nombre de données dans l’étude présente est aussi très bas. J’ai malgré tout préféré les maintenir dans l’étude afin de comparer mes résultats à ceux de Poplack et Walker (1986). Leur échantillon comptait 610 occurrences, contre 103 occurrences ici. Les résultats obtenus doivent donc être nuancés. Il est probable que ces facteurs aient un effet relativement faible et qu’un échantillon plus grand leur aurait permis d’être détectés par une analyse de régression multiple. Les résultats obtenus se retrouvent dans le tableau 4.7 ci-dessous.

Tableau 4.7 Analyse linguistique des pronoms objets indirects

Pronoms objets indirect				
N total = 103				
Valeur d’application : sans ⟨l⟩, 20.39% d’élision				
FACTEURS	N	Taux (%)	Poids relatifs	Logodds
Contexte phonologique précédant l’occurrence ($p = 0.0174$)				
Voyelle	30	33.33	0.768	1.197
Consonne	73	15.06	0.232	-1.197
Facteurs non significatifs				
Contexte syntaxique suivant l’occurrence				
Nombre de caractères (valeur continue)				

On observe moins d'élision pour ces pronoms que pour les pronoms sujets masculins et féminins. Tout comme pour les pronoms masculins, le contexte phonologique s'est avéré significatif, mais rappelons que le contexte observé est le contexte précédant l'occurrence à cause de la position du ⟨l⟩ dans le mot. Le contexte postvocalique entraîne ici aussi plus d'élision que le contexte postconsonantique. Ces résultats s'alignent avec Poplack et Walker (1986), mais vont à l'encontre des résultats d'Ashby (1984) pour la communauté de Tours. Le nombre de caractères, quant à lui, est étonnamment non significatif pour les POIs. En effet, on aurait pu croire que les maximes de Thurlow et Brown (2003) auraient influencé tous les pronoms de la même façon, de sorte que la longueur des messages soit positivement corrélée avec le taux d'élision par souci de brièveté. En réalité, non seulement le sens de la corrélation va à l'encontre de ce que la maxime de brièveté suggère pour les pronoms masculins et féminins, mais les POIs semblent même y résister. Cette résistance apparente pourrait être due au nombre d'occurrences observées. Plus d'occurrences auraient pu permettre à ce facteur d'émerger. La catégorie syntaxique du mot suivant l'occurrence s'est aussi avérée non significative comme pour les autres pronoms, contrairement aux résultats de Poplack et Walker (1986).

4.3 Analyse sociologique

Les résultats de l'analyse des facteurs sociologiques se retrouvent dans le tableau 4.8. Ultimement, les seuls facteurs sociologiques qui se sont avérés significatifs pour l'élision dans les textos sont l'âge du participant et leur niveau de scolarité.

Plus un participant était âgé, moins il élidait ses ⟨l⟩. Ceci semble contredire l'hypothèse formulée en 3.5.2.1 qui proposait une courbe en cloche plutôt qu'une pente descendante. En effet, les taux d'élision observés dans le corpus ne s'alignent pas avec les taux d'élision attendus si le comportement de textage coïncidait avec l'âge à l'époque du développement du langage SMS et que les comportements des auteurs étaient stables. Nous aurions trouvé une hausse au milieu de la figure qui correspondrait à la cohorte « adolescente », c'est-à-dire ceux qui étaient adolescents et qui ont développé le langage SMS au début des années 90. Ce groupe aurait environ 30 à 40 ans lors de l'étude de Langlais et Drouin (2012). On présumait que ce groupe éliderait davantage leurs ⟨l⟩ que les autres groupes d'âge (les plus vieux et les plus jeunes) parce que les adolescents sont à l'avant-garde des pratiques de textage (Eckert, 2017 ; Grinter et Eldridge, 2003). La figure 1 ci-bas montre comment le taux d'élision est affecté par l'âge des participants. Comme on peut le voir, les « adolescents » ne sont pas le groupe qui élide le plus leurs ⟨l⟩. Bien sûr, n'ayant pas accès à des données sur le comportement de textage des participants au début des années 90, on ne

peut être certain de la nature du changement, à savoir s'il s'agit d'une gradation d'âge ou un changement générationnel. Une deuxième étude moderne avec le même échantillon ou du moins un échantillon similaire serait nécessaire.

Tableau 4.8 Analyse sociologique

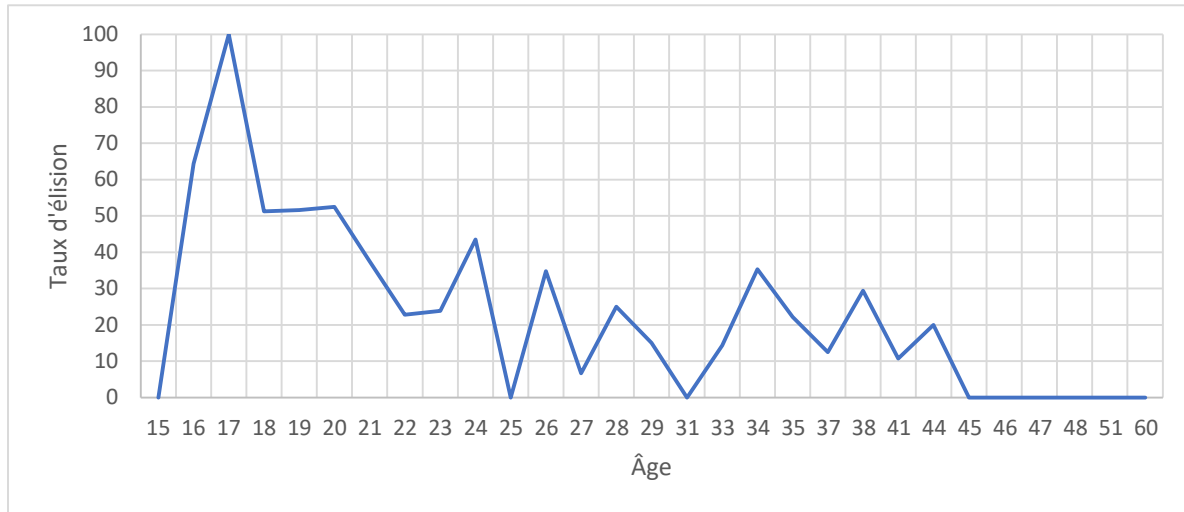
Facteurs sociologiques				
N total = 706				
Valeur d'application : sans ⟨l⟩, 28.75% d'élision				
FACTEURS	N	Taux (%)	Poids relatifs	Logodds
Scolarité (p = 0.0173)				
Préuniversitaire	326	33.44	0.665	0.684
Universitaire	380	24.74	0.335	-0.684
Âge (valeur continue) (p = <0.01)				
+1	Logodds = -0.100			
Facteurs non significatifs				
Genre		Capacité de décodage		
Depuis combien de temps le participant texte		Utilisation de dictionnaire		
Nombre moyen de SMS envoyés par jour		Type de clavier		
		Nombre d'heures de lecture et d'écriture.		
Facteurs omis				
Niveau de langue requis pour l'emploi ⁴⁵				

Il semble cependant plus probable qu'il s'agisse d'un effet de gradation d'âge, ce qui impliquerait que les habitudes de textage des auteurs évoluent à travers le temps. Les taux d'élosion présentés pour chaque âge représentent la moyenne d'élosion de tous les participants d'âge correspondant. À noter qu'il y a une énorme disparité dans le nombre de participants par tranche d'âge et le nombre d'occurrences produites par chaque tranche d'âge. On ne compte, par exemple, qu'un seul participant de 15 ans et ce participant n'a produit qu'une seule occurrence. En comparaison, on retrouve 2 participants de 16 ans, dont un a

⁴⁵ Les raisons pour son omission sont détaillées en 5.6.

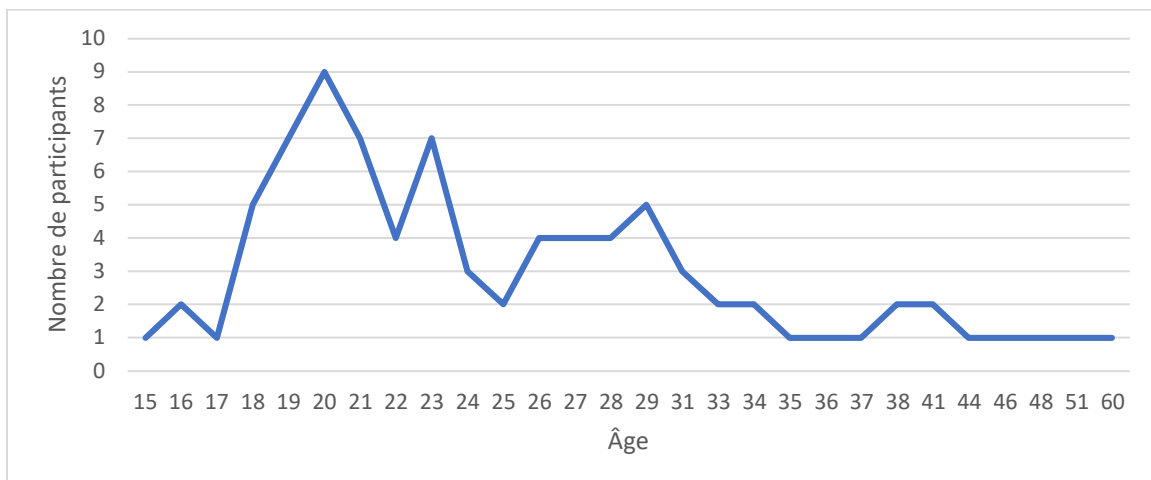
produit 1 occurrence et l'autre 170 occurrences, soit plus de deux fois plus que le deuxième participant le plus productif ($n=69$).

Figure 1 - Taux d'élision par tranche d'âge



La figure 2 montre le nombre de participants par tranche d'âge. On voit que les jeunes sont disproportionnellement représentés dans l'échantillon. Ceci contribue à expliquer pourquoi la courbe est aussi dentelée et pourquoi on retrouve des taux d'élision de 0% et 100% à quelques reprises dans le graphique : plusieurs tranches d'âge ne comptent qu'un seul participant et plusieurs participants n'ont produit qu'une seule occurrence. L'absence apparente de données pour les participants de 45 et plus est due au fait qu'il n'y a qu'un participant par tranche, qu'ils n'ont produit qu'une occurrence chacun et qu'aucun d'entre eux n'a produit de pronoms élidés.

Figure 2 - Distribution des participants par tranche d'âge



Tout comme dans les études sur l'élision du /l/ à l'oral, le niveau de scolarité s'avère significatif pour l'élision dans les textos. Les participants ayant obtenu une éducation préuniversitaire élident davantage leurs ⟨l⟩ que les participants ayant fait des études universitaires, conforme à l'hypothèse émise en 3.5.2.3.

Le fait que les autres facteurs n'aient pas été significatifs est très étrange considérant les résultats des autres études sur l'élision à l'oral. Par exemple, Sankoff et Cedergren (1976), Ashby (1984) et Poplack et Walker (1986) rapportaient tous que les hommes élidait davantage leurs /l/ que les femmes alors que le genre n'avait pas d'impact dans mon étude. Ceci pourrait être dû au fait que la norme est tellement plus flexible dans les textos que l'impact typique du genre dans les études sociales est amoindri. Il est également possible que les problèmes de distribution masquent l'impact réel du genre. Similairement, les facteurs portant sur les habitudes de textage ainsi que le niveau d'exposition à la langue standard écrite n'ont eu aucun effet sur les taux d'élisions des participants.

CHAPITRE 5

DISCUSSION

5.1 De l'oral à l'écrit

Compte tenu du fait que les échantillons pour les pronoms féminins et POIs étaient trop petits, la discussion qui suit portera principalement sur les pronoms sujets masculins. Ils constituent en effet la majorité des occurrences dans le corpus et prennent en compte plus de facteurs (la catégorie grammaticale) que les autres analyses linguistiques. En ce sens, ils offrent un regard plus complet sur l'élision dans les textos.

À la lumière des résultats présentés dans le dernier chapitre, on peut voir que le langage SMS n'est pas simplement le reflet de la langue orale familière, contrairement à ce que le discours de la linguistique populaire prétend. La différence drastique entre les taux d'élision quasi catégoriques à l'oral et les taux beaucoup plus bas observés dans le corpus *Texto4science*, en plus des différences au niveau des facteurs qui influencent l'élision dans les deux médiums, suggèrent que l'élision qu'on retrouve dans les textos est un phénomène similaire en surface à l'élision à l'oral, mais distinct par ses causes et ses influences. Sur le plan linguistique, les facteurs qui influençaient l'élision à l'oral semblent n'avoir que très peu d'effet à l'écrit dans les textos. Sur le plan social, à l'exception de l'âge des participants et leur niveau de scolarité, aucun autre facteur observé dans le cadre de l'étude présente ne s'est avéré significatif. Même les facteurs propres à l'usage des textos comme le nombre de textos envoyés par jour et la capacité de décodage, des facteurs considérés comme étant des mesures de la compétence en langage SMS des participants, n'ont pas eu d'effets significatifs sur l'élision du ⟨l⟩ dans les pronoms. De manière plus importante, le nombre d'heures de lecture et d'écriture ne semble pas avoir d'importance au niveau de l'utilisation du langage SMS. En effet, l'élision du ⟨l⟩ en tant que textisme ne semble pas conditionnée par le niveau de compétence ou d'engagement avec la langue standard. Ceci contredit le discours populaire sur les textos selon quoi les personnes qui écrivent en langage SMS le font parce qu'ils ne savent pas écrire ou qu'ils ne lisent plus. Cependant, l'absence de résultats pour ces facteurs ne révèle pas les mécanismes qui influencent l'élision dans les textos. Plutôt que de concevoir le langage SMS comme une simple représentation écrite de l'oralité vernaculaire, peut-être faudrait-il voir le langage SMS comme une variété ayant ses propres règles et standards écrits.

Si Blondeau et al. (2014) avaient démontré que la morphosyntaxe se manifestait de façon similaire dans le parler des texteurs et leurs textos (pour l’alternance *on* pluriel et *vous*), la phonologie semble quant à elle se réaliser de façon différente entre les deux médiums. En effet, il semblerait que le médium de l’écrit interagisse avec l’oralité perçue dans les textos, de sorte que la rigidité du code écrit, malgré tout plus flexible dans les textos que dans la langue écrite standard, régit les formes employées par les texteurs en favorisant l’emploi de formes standards et en limitant à quel point un texteur peut s’en écarter. En ce sens, les textos sont plus conservateurs que ne le prétend le discours populaire. Les besoins de lisibilité et de transmission de contenu sémantique dans les textos forcent les texteurs à « bien écrire » (Blondeau et al., 2014), du moins suffisamment pour être compris. L’aspect ludique (Blondeau et al., 2014 ; Crystal, 2008, 2009 ; Holtgraves, 2011 ; Tagg, 2009 ; Waldron et al., 2017) est contraint dans une certaine mesure par les règles du français et le besoin d’être compris. Ainsi, le standard, même « affaibli », interfère avec le langage SMS et donc avec l’élision du ⟨l⟩ dans les pronoms.

5.2 Contexte phonologique versus contexte graphique

Dans l’étude présente, les contextes prévocaliques (pour les pronoms masculins) et postvocaliques (pour les POIs) engendraient plus d’élision que le contexte préconsonantique et postconsonantique. Au niveau des pronoms masculins, ces résultats contredisent ceux de Sankoff et Cedergren (1976), Ashby (1984) et Poplack et Walker (1986), pour qui le contexte préconsonantique et postconsonantique génère le plus d’élision. Comme discuté en 4.2.1, cette différence pourrait être en partie attribuable à la façon dont les données ont été codées dans le cadre de l’étude présente. Concrètement, le contexte pré-yod, généralement observé dans les études citées ci-haut, était recodé ici comme prévocalique. Or, bien que le contexte phonologique de l’occurrence exerce une véritable influence sur l’élision à l’oral, son effet à l’écrit dans les textos semble plutôt mitigé. Une autre hypothèse basée sur l’interaction entre le contexte phonologique et le contexte graphique pourrait potentiellement mieux expliquer la différence dans les résultats obtenus, illustrant le caractère hybride des textos.

En effet, l’élision du ⟨l⟩ dans les textos semble en partie conditionnée par le contexte graphique du pronom, c’est-à-dire s’il se retrouve comme caractère seul, connecté à un autre mot comme ⟨qu / k⟩ *qu’* ou collé au mot suivant. Ce facteur n’avait pas été envisagé pour l’analyse linguistique puisque l’idée populaire est qu’on écrit comme on parle dans les textos et donc que les mêmes facteurs que ceux employés dans les études sur l’oral seraient également significatifs dans le cadre de l’étude présente sur l’écrit dans les textos. Or, le médium écrit et les contraintes qui y sont propres semblent aussi déterminants pour le langage SMS

que l’oralité apparente des textos. La distribution des formes élidées ⟨i⟩ et ⟨y⟩ ($n=199$) pour les pronoms sujets masculins, dans le tableau 5.1 ci-dessous, illustre bien ce point. On peut voir que la distribution des formes ⟨i⟩ et ⟨y⟩ répond à la fois au contexte phonologique (suivi d’une voyelle ou d’une consonne) et au contexte graphique (le type de construction dans laquelle se trouve le pronom). Autrement dit, le choix de la variable se produirait en deux étapes. D’abord, le pronom se réaliserait en forme standard ou élidée selon le contexte phonologique, puis le texteur déterminerait la forme élidée appropriée selon le contexte graphique. Dans le tableau 5.1, les occurrences de caractères seuls sont celles où le pronom est isolé du mot qui le précède et qui le suit, comme dans « Y pleut? » (113010, *Texto4science*). Les occurrences de ⟨qu / k⟩ + sont celle qui sont précédés par le complémenteur *qu’*. Les occurrences *collées* sont celles où le pronom est collé au mot qui le suit, comme dans « Ah k scorreect, scorreect! Ya passe son permis? »⁴⁶ (110250, *Texto4science*).

Tableau 5.1 Distribution des pronoms masculins élidés

	⟨i⟩		⟨y⟩		Sous total
Construction	_V %	_C %	_V %	_C %	
Caractère seul	0 0%	5 14.29%	1 0.61%	87 53.05%	
⟨qu / k⟩ +	3 8.57%	27 77.14%	1 0.61%	0 0.00%	31
Collé	0 0%	0 0.00%	75 45.73%	0 0.00%	75
Sous total	35 100%		164 100%		
Totaux	199				

La maxime d’approximation phonologique (Thurlow et Brown, 2003) fait deux prédictions : (1) on retrouverait un ⟨y⟩ dans les constructions prévocaliques pour signaler la prononciation d’un yod comme à l’oral alors et (2) on retrouverait un ⟨i⟩ ailleurs dans les cas où on prononcerait un [i] à l’oral. Ces prédictions ne sont que partiellement confirmées dans mes données. On voit effectivement que le pronom sujet ⟨y⟩

⁴⁶ Ah, okay c’est correct, c’est correct! Il a passé son permis?

est préféré quand il est suivi d'une voyelle et collé. Ceci pourrait être dû au fait que le ⟨y⟩ permet de former une syllabe avec un yod en attaque. La forme ⟨i⟩, quant à elle, ne peut occuper que la position de noyau dans la syllabe et serait donc prononcé comme un [i]. On s'attendrait donc à voir ⟨i⟩ plus fréquemment comme caractère seul ou précédé d'une consonne comme le complémentateur ⟨qu' / k⟩ en attaque et en position préconsonantique, là où on retrouverait un [i] à l'oral. Cependant, lorsque le pronom se retrouve comme caractère seul, les participants montrent une préférence marquée pour la forme ⟨y⟩, spécifiquement lorsqu'elle est suivie d'une consonne. Des 93 occurrences de pronoms seuls, 87 (93.55%) sont des ⟨y⟩ préconsonantiques. La forme ⟨i⟩ (n=35), quant à elle, se retrouve principalement dans des constructions de type ⟨qu k + i⟩ *qu'il* (n=30, 85.71%). Devant une voyelle, on ne retrouve que 3 occurrences de la forme ⟨i⟩ et toutes ces occurrences sont précédées du complémentateur *qu'*. Le lien entre ⟨i⟩ et ⟨qu / k⟩ est si fort qu'on retrouve seulement 5 occurrences de ⟨i⟩ qui ne soient pas précédées du complémentateur et toutes ces occurrences sont suivies de consonnes. Cette préférence pour une forme ou l'autre selon le contexte est remarquablement partagée par les participants, comme le montre le tableau 5.1. La préférence pour ⟨y⟩ comme caractère seul est peut-être due au fait qu'il existe en français écrit standard (comme pronom complément *y*), contrairement à ⟨i⟩. Il serait peut-être plus accessible pour les auteurs que la forme novatrice ⟨i⟩ quand vient le temps de rédiger un texte. La distribution des formes ⟨y⟩ et ⟨i⟩ suggère donc qu'elle n'est pas complètement conditionnée par la phonologie ou un désir d'imiter la prononciation à l'oral. Elle serait aussi influencée par des préférences stylistiques en lien avec le contexte graphique de l'occurrence et le code écrit standard lui-même, montrant à nouveau l'aspect tout de même conservateur des textes et leur caractère hybride.

Aussi, bien que le contexte phonologique n'ait pas pu être testé pour *elle*, on peut tout de même voir une autre manière que la phonologie ne se transpose pas parfaitement à l'écrit. L'allongement de [æ:] ou [ɛ:] pour *elle* à l'oral, causé par l'élision du /l/ devant une voyelle (Laliberté, 1974), ne se traduit pas systématiquement à l'écrit par un diacritique (:) ou un redoublement des caractères dans les textes. Le redoublement de caractères, en tant que textisme, sert davantage à exprimer des émotions comme l'excitation ou l'ennui qu'à reproduire la prononciation éliée des pronoms à l'oral. Ceci suggère à nouveau que cette contrainte phonologique n'opère pas dans les textes de la même façon qu'à l'oral.

5.3 Paradoxe de la longueur des textes

Ici encore, contrairement à ce qu'on s'attendrait selon les maximes de Thurlow et Brown (2003), la longueur des textes s'est avérée inversement corrélée aux taux d'élision du ⟨l⟩ dans les textes pour les

pronoms sujets masculins et féminins. Curieusement, l'effet de la longueur des textos était non significatif pour les POIs, mais il est probable que ce soit imputable à la taille de mon échantillon. La maxime de brièveté ne semble pas avoir portée sur les pronoms, même si les résultats de plusieurs auteurs (Androutsopoulos et Schmidt, 2002 ; Dürscheid, 2003 ; Fairon et al., 2006a ; Grinter et Eldridge, 2003) avaient montré que les mots les plus souvent ciblés par des textismes étaient des mots courts ou d'usage courant. De plus, les auteurs ont généralement tendance à effacer les voyelles plutôt que les consonnes, voyant les consonnes comme étant plus importantes pour distinguer les mots (Crystal, 2008). Le taux de compression dans les textos (Langlais et Drouin, 2012) associé à l'élision du ⟨l⟩ pourrait expliquer pourquoi les pronoms dans les textos résistent à l'élision beaucoup plus qu'à l'oral. En effet, économiser un ou deux caractères ne constitue probablement pas un gain suffisamment important pour que les pronoms dans les textos soient systématiquement élidés, contrairement à l'oral où l'élision est presque catégorique. À titre de comparaison, réduire ⟨appartement⟩ à ⟨apt / app⟩ représente un gain beaucoup plus grand pour le texteur que de réduire ⟨il⟩ à ⟨i / y⟩. On aurait donc pu s'attendre à ce que la longueur des textos soit faiblement corrélée avec les taux d'élision dans les pronoms. Cependant, cette hypothèse n'explique pas pourquoi on retrouve une corrélation inverse entre la longueur d'un texto et le taux d'élision.

Une autre hypothèse pour expliquer cette corrélation inverse pourrait se baser sur le fait qu'un message plus long emploie peut-être davantage la langue standard que le langage SMS, ce qui pourrait faire l'objet d'une étude future. En apparence, les messages plus longs contiennent plus de textismes que les messages plus courts. Pourtant, rappelons que selon plusieurs études (Crystal, 2009 ; Tagg et al., 2012 ; Thurlow et Brown, 2003), les textismes ne constituent souvent qu'un maigre 10 à 20% du contenu d'un texto, contrairement à ce qu'on pourrait croire selon le discours de la linguistique populaire qui surestime la proportion de textismes dans les textos (Thurlow, 2006). Ces auteurs ne précisent toutefois pas la méthode qu'ils emploient pour obtenir ces résultats. La méthode employée ici n'a pas été testée pour en assurer la justesse et est donc rudimentaire. Elle consiste à diviser le nombre d'annotations⁴⁷ par le nombre de caractères dans le texto. Les exemples en (25-26) montrent respectivement le message le plus court et le plus long de mon échantillon. Pour la phrase en (25), Langlais et Drouin notent 3 textismes : une abréviation (*ye* pour *il est*), une omission d'accent (*ou* pour *où*) et une erreur de ponctuation (*!* pour *?*).

⁴⁷ Les annotations employées sont celles de Langlais et Drouin (2012). À noter que j'aurais compté l'abréviation comme deux textismes (contraction ou orthographe phonétique pour l'élision du ⟨l⟩ et orthographe phonétique pour *est*). N'ayant pas testé ma méthode, je m'en remets cependant aux annotations originales des chercheurs.

On observe donc que près de 50% du contenu contient des textismes. Par comparaison, le texto le plus long en (26) ne contient que 8.42% de contenu atteint.

(25) Ye ou!⁴⁸ (110 250, *Texto4science*)

(26) Brrr'Quel temps De...!J'ai entendu Parler de Vous ce Am' Ou-eee/Ou-eee! Par un M. Avec tu as eu a faire faire cette semaine! J'ai aussi fait une plainte /demande pour savoir pourquoi in'y avait pu de ces FAMEUSES choix de president double choco Crunch! Et il m'a dit qu'une Jeune Femme etait passée en mi semaine pour la meme cereal' il a sortie son calpin et a prise une note a coté de la tienne!:) Comment allez Vous en ce debut de W.E... Qui tarda commencer a se mettre en place?!Md ce am?!? pour le genou?! Bonne fin de lunch encore QueQues Heures et ByeBye la grande Tour!Aplush;+)OsxxXPluie...euh.⁴⁹ (110 477, *Texto4science*).

La forme normalisée du texto en (26) compte 636 caractères, pour un taux de compression de seulement 4.72%. Dans ce texto, on compte de nombreux accents manquants, quelques abréviations courantes (*AM* pour *matin*, *m.* pour *monsieur*, *W.E.* pour *weekend*, *md* pour *médecin*), des contractions (*calpin* pour *calepin*, *QueQues* pour *quelques*, de la stylisation d'accent (*Aplush* pour *à plus*, *Ou-eee* pour *oui*), de l'hyperc capitalisation (*FAMEUSES*) et une ponctuation et une capitalisation remarquablement libres. Crucialement, par contre, on retrouve 3 occurrences de pronoms masculins sujets dont une seule occurrence élidée ([...] *in'y avait* [...]). En sommes, malgré une pluralité de textismes employés par le texteur, un désir évident de transgresser la norme et un nombre de caractères qui aurait largement justifié l'élision, seulement un des pronoms a été élidé. Ceci pourrait montrer que la longueur des textos a un impact différent pour les différents types de textismes et que l'élision y résiste plus que les autres. Bien sûr, il ne s'agit que de deux exemples sur la base desquels on ne peut tirer de réelles conclusions. Cependant, ces taux s'alignent avec ceux rapportés par Crystal (2009) et Thurlow et Brown (2003). Ces hypothèses pourraient être testées dans une nouvelle étude.

⁴⁸ Il est où?

⁴⁹ Brrr! Quel temps de...! J'ai entendu parler de vous ce matin, oui, oui! Par un monsieur avec qui tu as eu à faire cette semaine! J'ai aussi fait une plainte /demande pour savoir pourquoi il n'y avait plus de ces FAMEUX Choix du président double choco Crunch! Et il m'a dit qu'une jeune femme était passée en mi-semaine pour la même céréale. Il a sorti son calepin et a pris une note à côté de la tienne! :) Comment allez-vous en ce début de week-end... qui tarde à commencer à se mettre en place? Médecin ce matin? Pour le genou? Bonne fin de lunch, encore quelques heures et bye bye la grande Tour! À plus tard. ;+)Osxxx Pluie... euh.

5.4 Valeur sémantique comme obstacle à l'élision

Dans leur étude sur l'élision à l'oral, Sankoff et Cedergren (1976) proposent une analyse basée sur la valeur sémantique pour expliquer pourquoi certains pronoms résistent davantage à l'élision que d'autres. Elles avancent que les pronoms contiennent de l'information sémantique inhérente qui exprime la pluralité (+ ou – pluriel) et le genre (+ ou – féminin) (voir tableau 5.2, repris de Sankoff et Cedergren, 1976, p. 1106). Selon elles, le pronom *il* serait plus susceptible d'être élidé que *ils* ou *elle* parce qu'il contient moins d'information sémantique et serait donc moins saillant et important au niveau de la compréhension. Leur analyse se résume par l'expression « + d'information → - d'élision ».

Tableau 5.2 Valeur sémantique des pronoms

Pronom	Pluriel	Féminin
Il	-	-
Ils	+	-
Elle	-	+
Elles	+	+

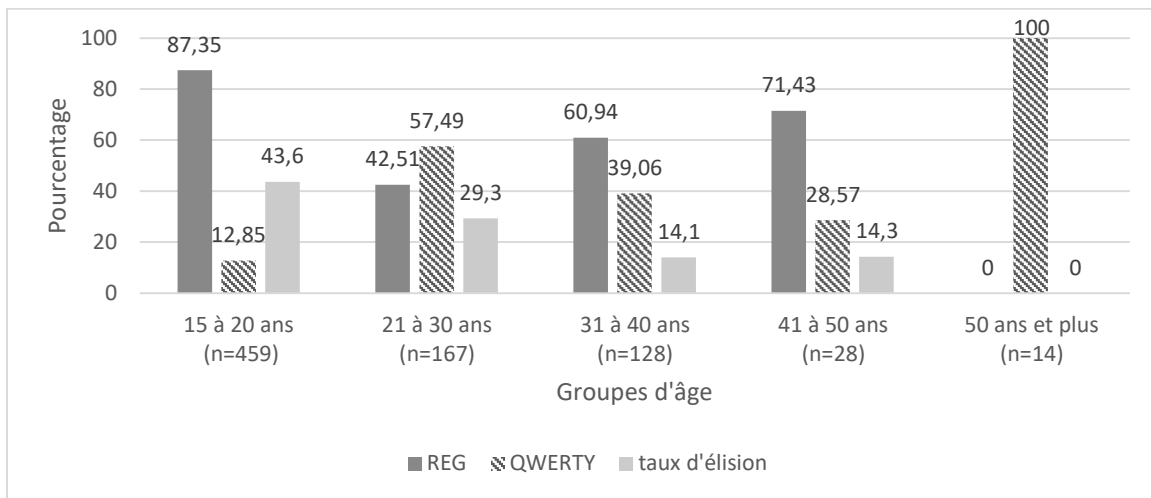
En transposant cette logique aux mots suivant les pronoms, on prédit que les verbes favoriseraient davantage l'élision dans les pronoms que les mots « autres » (des adverbes de négation, des objets directs et indirects ainsi que des pronoms compléments et réflexifs) parce qu'ils portent le fardeau sémantique de la phrase alors que les mots « autres » contiennent moins d'information inhérente. Les mots « autres » laisseraient plus de place aux pronoms pour qu'ils se réalisent avec leur forme standard. On prédit aussi qu'on retrouverait peu de pronoms omis lorsqu'ils sont suivis de mots « autres ». Les taux globaux présentés en 4.1 appuient partiellement cette analyse. Elle explique bien les taux d'élision du ⟨I⟩ obtenus quand on considère la catégorie syntaxique du mot suivant le pronom. Elle est aussi cohérente avec le taux très élevé d'omission de pronoms impersonnels par rapport aux pronoms personnels. Cependant, elle ne saurait expliquer pourquoi les pronoms personnels masculins ont subi (légèrement) plus d'élision que les pronoms impersonnels (39.93% contre 35.61%). En effet, on s'attendrait plutôt au contraire, c'est-à-dire que les pronoms impersonnels, étant essentiellement vides de sens, auraient favorisé l'élision par rapports aux pronoms personnels. Il semble donc possible que l'élision du ⟨I⟩ soit moins affectée par la valeur sémantique. Il est possible que cette différence soit attribuable à l'écart dans le nombre

d'occurrences de pronoms personnels ($n=298$) et impersonnels ($n=132$). Il est également possible que d'autres facteurs n'ayant pas fait part de l'étude puissent influencer l'impact de la valeur sémantique sur l'élision du ⟨l⟩.

5.5 Âge et technologie

L'effet de l'âge s'est avéré complètement différent à l'hypothèse qui avait été formulée en 3.5.2.1 selon laquelle le groupe qui éliderait le plus leurs ⟨l⟩ serait la « cohorte adolescente », c'est-à-dire le groupe qui était adolescent lors de l'apparition des textos et du développement du langage SMS et qui correspond aux participants qui avaient 30 à 40 ans lors de l'étude de Langlais et Drouin (2012). De plus, il ne semble pas y avoir d'interaction entre l'élision et le développement de la technologie. Le type de clavier ou l'usage de dictionnaires sur les téléphones n'étaient pas corrélés avec les taux d'élision. Il faut dire que somme toute peu de participants (20.54%) employaient des dictionnaires électroniques dans mon étude. Cette proportion serait probablement largement plus grande aujourd'hui puisque les technologies d'autocomplétion et d'autocorrection sont beaucoup plus répandues et performantes. Il n'y a pas non plus de corrélation entre l'âge des participants et le type de téléphone. La seule tendance détectable est la diminution de l'élision à mesure que les participants vieillissent. La figure 3, ci-bas, montre, par groupe d'âge, la proportion de participants dans l'étude présente qui détenait un téléphone régulier à 12 touches (REG) et celle qui détient un téléphone à clavier QWERTY ainsi que la moyenne du taux d'élision pour chaque groupe. L'âge est présenté sous forme de catégories plutôt qu'en valeur continue afin de permettre la comparaison avec d'autres facteurs binaires. Ces résultats appuient la conclusion que le langage SMS est la langue des jeunes (Eckert, 2017 ; Grinter et Eldridge, 2003 ; Lenhart et Madden, 2005 ; Pérez-Sabater et Montero-Fleta, 2015 ; Quan-Haase, 2007 ; Verheijen, 2013). Cependant, comme mentionné en 4.3, une nouvelle étude serait nécessaire pour clairement identifier si l'élision subit un effet de changement en cours ou une gradation d'âge et pour voir si l'usage de technologies plus modernes influence les comportements de textage.

Figure 3 Comparaison des groupes d'âge selon les types de clavier et les taux d'élision



5.6 La scolarité et le langage SMS

Le niveau de scolarité des participants a eu l'effet escompté, c'est-à-dire qu'il était négativement corrélé avec leur taux d'élision. Ceci est potentiellement dû à la valeur sociale qu'on accorde à la qualité de la langue et l'éloquence ainsi qu'à l'association entre la langue standard et les études supérieures. Il faut cependant considérer quelques interactions possibles dans les données. Bien que l'indépendance des facteurs étudiés ait été confirmée préalablement lors du codage des données, on peut voir que l'âge, le niveau de scolarité et le niveau de langue requis par l'emploi semblent travailler de concert. Effectivement, il n'est pas surprenant que les plus jeunes rapportent des taux de scolarité généralement plus bas que les plus vieux considérant que certains participants ne sont pas assez vieux pour être rendu à faire des études supérieures. Cependant, ce ne sont pas tous les participants plus âgés qui ont fait des études supérieures. De plus, la majorité de mes participants ont entre 18 et 30 ans et sont donc en âge de faire des études universitaires. Il y a donc une suffisamment bonne distribution des participants pour considérer l'âge et le niveau de scolarité dans la même étude.

Le niveau de langue requis par l'emploi, quant à lui, a dû être omis de l'étude en raison de plusieurs problèmes. Bien que ce facteur ait apparu indépendant lors des vérifications, l'analyse ascendante-descendante présente des résultats contradictoires qui suggéraient qu'il y avait des interactions avec d'autres facteurs. De plus, tel qu'abordé dans la section 3.5.2.4, ce facteur était problématique puisqu'il s'agissait d'une réinterprétation subjective des données de Langlais et Drouin (2012). La distinction « non » et « standard » ainsi que l'appartenance des emplois rapportés à chaque catégorie n'a pas fait l'objet de

tests afin d'en confirmer la validité. Ultiment, le niveau de scolarité des participants était plus robuste comme facteur que le niveau de langue requis par l'emploi et expliquait mieux les taux d'élosion obtenus.

5.7 Les impondérables

Selon Blondeau et al. (2014), l'emploi de formes scripturales non standards est davantage associé à la personnalité, l'état d'esprit du texteur et de la dimension ludique et transgressive du langage SMS qu'au statut social. En effet, dans l'étude présente, à la seule exception de l'âge, aucun des facteurs sociaux retenus n'influence l'élosion. Ceci n'empêche pas que d'autres facteurs sociaux qui n'ont pas été pris en compte puissent influencer l'élosion. Toutefois, tel qu'évoqué en 3.5, il est également possible que d'autres facteurs impondérables puissent jouer un rôle important pour l'élosion (ou pour l'emploi de textisme de façon plus générale). Il semble en effet plausible que différents facteurs situationnels comme la relation entre le texteur et le receveur, le contexte de rédaction (ce que le texteur faisait au moment de rédiger ou son degré d'empressement et d'affairement), le niveau d'attention et d'énergie du texteur lors de la rédaction de son texto, l'état du téléphone lui-même (si l'écran est endommagé, si les touches fonctionnent bien, etc.) ou même d'autres facteurs puissent être à l'origine de la variation dans les taux d'élosion. Compte tenu du fait que ces facteurs sont difficiles ou même impossibles à opérationnaliser convenablement en fonction des données disponibles pour le corpus *Texto4science*, on ne peut que spéculer sur leur importance. En effet, pour des raisons logistiques, plusieurs des facteurs proposés ci-haut pourraient difficilement être mesurés, soit parce qu'il faudrait être présent au moment de la rédaction du message pour observer le texteur dans son environnement ou parce qu'il faudrait que le texteur fournisse des renseignements additionnels au moment de la rédaction de son texto sur son rapport avec l'autre, son état d'esprit ou à quoi il pensait en même temps de rédiger son message. Comme la collecte de donnée se fait après la rédaction des textos, il faudrait compter sur le fait que le participant se souvienne du contexte exact de rédaction et espérer qu'il partage ces informations lors de la collecte de données.

Par chance, un de ces facteurs potentiels, l'attitude du texteur face à la qualité de sa langue écrite, peut tout de même être abordé grâce à certains commentaires émis par les participants dans leurs réponses au questionnaire. Notons que l'attitude des participants ne faisait pas l'objet d'une question explicite dans le questionnaire de Langlais et Drouin (2012). Les participants avaient plutôt la chance de laisser des commentaires généraux sur leur usage du langage SMS (question 18 de l'annexe A). Ce facteur est particulièrement intéressant parce qu'il permet de remettre en question un argument de la linguistique

populaire selon quoi les jeunes ont une attitude de « je-m'en-foutisme » à l'égard des standards écrits (Baron, 2002) et c'est ce qui mènerait à la dégradation de la qualité de leur langue. Somme toute, peu de participants ont laissé des commentaires dans le questionnaire de Langlais et Drouin (2012), et encore moins de commentaires relevaient de l'attitude des participants face au langage SMS. On doit donc prendre garde de ne pas tirer de conclusions hâtives sur l'impact de l'attitude sur l'emploi de l'élision. Aussi, comme mentionné plus tôt, l'attitude présentée dans le tableau ne correspond pas à une réponse précise formulée par le participant. Il s'agit plutôt d'une interprétation subjective de ma part basée sur le propos des commentaires. Ces commentaires se retrouvent dans l'annexe E. Les commentaires qui témoignent de la crainte du participant de « perdre son français » s'il utilise le langage SMS ou d'un sentiment de fierté par rapport à l'usage du « bon français » sont interprétés comme représentant une attitude négative à l'égard de langage SMS. À l'opposé, un commentaire à propos du plaisir de jouer avec la langue ou le plaisir d'exprimer son identité serait interprété comme positif. Finalement, un commentaire à propos de « faire attention à sa langue » sans exprimer de dédain pour le langage SMS serait considéré comme une attitude neutre. À titre d'exemple, le commentaire en (27) témoigne d'une attitude négative à l'endroit du langage SMS. Le commentaire en (28) témoigne quant à lui d'une attitude positive. L'auteure du commentaire en (29) exprime une attitude que je décrirais comme neutre.

- (27) « Je me fais une fierté d'utiliser un français correct dans mes textos et je sensibilise aussi mes interlocuteurs. Le fait d'avoir un clavier tactile iPhone aide beaucoup ; à l'époque du clavier téléphonique, j'utilisais plus souvent les abréviations. » (110990, *Texto4science*)
- (28) « Je trouve que s'appliquer à écrire en bon français et tout corriger tout le temps serait une perte de temps énorme » (112150, *Texto4science*).
- (29) « Le langage texto ne me sert que lorsque mon destinataire et moi avons une "conversation" suivie par texto et que je dois écrire rapidement. Cependant, depuis que j'ai un clavier QWERTY, il m'arrive vraiment moins souvent d'utiliser ce langage » (111610, *Texto4science*).

Bien que le nombre de commentaires et de participants impliqués soit relativement bas, on peut observer l'attitude des participants pour voir si elle se traduit en comportements plus ou moins normatifs. On s'attendrait à ce qu'un participant ayant une attitude négative à l'égard du langage SMS élide moins ses (l) qu'un participant ayant une attitude neutre ou positive parce qu'il fait plus attention à la qualité de sa langue. Le tableau 5.3 ci-bas suggère que l'attitude du texteur pourrait intervenir dans son emploi de l'élision.

Tableau 5.3 Rapport entre l'attitude des participants envers le langage SMS et leur taux d'élision

Participant	Âge	N	Taux d'élision%	Participant	Âge	N	Taux d'élision%
Négative				Neutre			
110495	26	3	33.33	112314	18	3	100.00
111590	19	19	26.31	110678	21	2	100.00
110990	35	9	22.22	112451	19	25	40.00
110375	22	5	20.00	111570	21	8	37.50
111850	24	5	0.00	110637	34	17	35.29
110851	41	1	0.00	112350	22	6	16.66
110500	29	1	0.00	110751	37	3	0.00
112210	20	1	0.00	111530	23	2	0.00
Positive				111610	22	2	0.00
110661	28	8	100.00	112333	60	2	0.00
112150	18	4	75.00	110715	15	1	0.00
112335	22	1	0.00	110290	21	1	0.00
				110530	19	1	0.00
				111035	22	1	0.00

À première vue, l'âge des participants ne semble pas positivement corrélé avec une attitude moins favorable pour le langage SMS, et ce même s'il était corrélé avec un plus bas taux d'élision du ⟨l⟩ selon les résultats présentés dans le dernier chapitre. Les données sont cependant trop peu nombreuses pour en tirer une conclusion. Une nouvelle étude incluant explicitement l'attitude des auteurs comme facteur serait nécessaire pour bien comprendre l'impact de ce facteur sur les taux d'élision du ⟨l⟩ ou l'emploi d'autres textismes. Quand on observe les participants les plus productifs dans les trois catégories, on voit que le participant avec une attitude négative a tout de même éliminé 26.31% de ses ⟨l⟩, alors que le participant avec une attitude neutre a éliminé 40.00% de ses ⟨l⟩. En ce qui concerne les participants avec une attitude positive, le taux d'élision catégorique du participant 110661 pourrait faire croire qu'une attitude positive favorise l'élision, mais l'échantillon compte trop peu de participants et trop peu d'occurrences

pour être concluant. Il semble cependant possible que l'attitude soit positivement corrélée avec le taux d'élision. En effet, aucun participant avec une attitude négative n'avait un taux d'élision supérieur à 50%, contrairement aux participants avec une attitude positive. Les participants avec une attitude neutre avaient des taux d'élision beaucoup plus variable.

Somme toute, les résultats obtenus dans le cadre de l'étude présente auront montré que le discours de la linguistique populaire ne tient pas la route. Les facteurs qui influencent l'élision à l'oral ne sont pas les mêmes dans les textos, ce qui suggère qu'il s'agit d'un phénomène d'élision différent dans les deux médiums. De plus, mis à part l'âge, les facteurs sociologiques « classiques » (genre, niveau de scolarité, statut socioéconomique, etc.) n'ont pas d'effet statistiquement significatif sur l'emploi de l'élision du ⟨l⟩ dans les textos, ce qui suggère que le langage SMS, du moins en ce qui concerne l'élision du ⟨l⟩, est surtout une pratique transgressive et ludique typique aux jeunes. Aussi, les facteurs propres aux textos (capacité de décodage, depuis quand le participant texte, etc.) ne sont pas significatifs non plus. L'attitude des participants face à la qualité de leur langue ne semble pas permettre de prédire le comportement des auteurs. On ne peut donc pas pointer du doigt ces facteurs en les accusant d'être responsables de l'élision du ⟨l⟩ dans le texto, et par proxy du déclin de la qualité de la langue des auteurs. Il semble plutôt probable que l'élision soit conditionnée par le style idiosyncratique de chaque participant (qui est d'ailleurs étonnamment similaire à travers l'échantillon testé) et par une batterie de facteurs impondérables.

CONCLUSION

6.1 Retour sur les questions de recherche

Le projet actuel avait comme objectif d'étudier l'élision dans les textos en français québécois et de comparer ce phénomène avec les résultats d'études sur l'élision à l'oral. Pour ce faire, une analyse du corpus *Texto4science* a été réalisée afin d'étudier le véritable langage SMS employé par les Québécois, et non le langage « barbare » et « puérile » décrit par la linguistique populaire. À la lumière des analyses effectuées, nous sommes maintenant en position de faire un retour sur les questions qui ont guidé le travail présent.

6.1.1 Question 1 : Quels facteurs linguistiques influencent l'élision dans les textos ?

Les trois contextes variables, les pronoms sujets masculins, les pronoms féminins et les POIs, ont montré des comportements légèrement différents entre eux. Pour les pronoms sujets masculins et les POIs, le contexte phonologique s'est avéré être le facteur le plus fort influençant l'élision. Les contextes prévocaux et postvocaux favorisaient l'élision pour les pronoms masculins et les POIs respectivement, alors que les contextes préconsonantiques et postconsonantiques la défavorisaient. Le contexte phonologique n'a malheureusement pas pu être testé pour les pronoms féminins à cause, notamment, du manque de données et de la mauvaise distribution des occurrences causée par des interactions entre les facteurs qui causaient des effets *knockouts*.

De plus, un regard sur la distribution des formes élidées pour les pronoms masculins a aussi révélé qu'un autre facteur, le contexte graphique, exerçait probablement une influence sur l'élision et sur le choix des formes employées. En effet, les pronoms dans les textos sont élidés d'abord en fonction du contexte phonologique dans lequel ils se trouvent. La forme élidée est ensuite largement déterminée par son environnement graphique. Bien qu'il n'ait pas été testé formellement dans le cadre de l'étude présente, ce facteur semble influencer l'élision en imposant certaines contraintes du code écrit standard sur le contenu des textos. En ce sens, l'élision se produit à l'intersection entre l'écrit et l'oral, ce qui montre l'hybridité des textos en action.

L'autre facteur qui était significatif pour l'élision est la longueur des messages textes. Cependant, la direction de l'effet s'est avérée complètement à l'inverse de ce qui avait été supposé initialement en raison de la maxime de brièveté (Thurlow et Brown, 2003). En effet, plus un message était long, moins les

participants avaient tendance à élider leur ⟨l⟩. Curieusement, ce facteur s'est avéré non significatif pour les POIs, mais ceci pourrait être dû en partie à la taille minuscule de l'échantillon.

6.1.2 Question 2 : Quels facteurs sociaux influencent l'élision dans les textos ?

Parmi les facteurs sociologiques proposés pour l'étude comme le genre, le niveau de scolarité et le domaine professionnel, depuis quand le participant texte, le nombre de textos envoyé par jour, la capacité de décodage, l'utilisation de dictionnaires ou de fonction d'autocorrection et d'autocomplétion, le type de clavier et le nombre d'heures de lecture et d'écriture par semaine, l'âge s'est avéré être le seul facteur exerçant une véritable influence sur l'élision dans les textos. Ceci va à l'encontre du discours populaire sur les textos selon quoi les gens qui écrivent en langage SMS le font parce qu'ils ne savent pas écrire ou qu'ils ne lisent plus ou encore à cause d'une attitude de « je-m'en-foutisme » linguistique. En effet, le nombre d'heures consacrées à la lecture et l'écriture en plus des autres facteurs « classiques » dans les études en sociolinguistique comme le genre, le niveau d'éducation ou le statut socioéconomique ne permettaient pas de prédire les taux d'élision des participants dans mon étude. Le fait que seul l'âge influence l'élision suggère plutôt que l'élision est une pratique typique des adolescents.

6.1.3 Question 3 : De quelle manière les résultats de l'analyse des textos se comparent-ils aux résultats de l'analyse de la langue orale ?

La différence principale entre l'élision à l'oral et dans les textos est la fréquence à laquelle elle se produit. À l'oral en français québécois, l'élision se produit presque catégoriquement pour les pronoms sujets masculins et légèrement moins souvent pour les pronoms objet. Les taux d'élision sont si élevés que certains auteurs (Poplack et Walker, 1986) allaient jusqu'à omettre les pronoms masculins de leur étude parce qu'ils jugeaient que ces pronoms avaient été relexicalisés en [i] et donc qu'ils ne variaient plus. Pour eux, toutes manifestations du [l] seraient le résultat d'une règle de réinsertion et une marque d'hyperstyle. Les pronoms féminins, quant à eux, subissent somme toute moins d'élision que les pronoms masculins et objets. À l'opposé, dans les textos, malgré le fait que le discours populaire prétend qu'on puisse y « écrire comme on parle », on retrouve nettement moins de pronoms élidés, ce qui suggère que le médium de l'écrit crée de la résistance à l'élision. Cette conclusion semble appuyée par le fait que le contexte graphique dans lequel le pronom se trouve influence fortement la forme dans laquelle il se réalise.

L'autre différence marquante entre l'élision à l'oral et dans les textos se trouve au niveau des facteurs qui l'influencent. Comme précédemment mentionné, les facteurs sociaux « classiques », à la seule exception

de l'âge, s'avèrent incapables de prédire les taux d'élision des participants. Dans le cadre de l'étude présente, qui est de nature synchronique, on ne peut cependant pas identifier si l'élision subit un effet de gradation d'âge ou un changement en temps apparent.

Tout ceci semble indiquer que l'élision dans les textos et à l'oral sont en réalité deux phénomènes distincts. À l'oral, l'élision est le résultat d'un discours fluide et rapide conditionné par l'environnement phonologique et le statut social des locuteurs. En contrepartie, l'élision dans les textos est un textisme employé de façon potentiellement stratégique par les auteurs en fonction de leur style idiosyncratique et de leur âge. Il est toutefois possible, comme évoqué par Poplack et Walker (1986), que *il* aie été relexicalisé sans /l/ dans sa représentation phonologique et que la présence du [l] soit une marque d'hyperstyle et le résultat d'une réinsertion. Aussi, une personne qui emploie une variété dans laquelle le /l/ est absent de la grammaire ou qui maîtrise moins la norme pourrait aussi ne pas utiliser de [l] dans ses textos parce qu'il n'y en aurait pas dans sa langue orale. Il serait alors peut-être plus approprié de considérer la forme variable des pronoms comme le résultat de réinsertion du /l/ plutôt que d'une élision. Finalement, il est également possible qu'une batterie de facteurs qui n'ont pas pu être testés dans le cadre de l'étude présente puisse avoir une influence sur l'emploi de l'élision. Toutes ces considérations pourraient faire l'objet d'études futures.

6.2 Perspectives de recherche

Une des limitations principales du projet actuel est l'écart de temps entre les corpus comparés. Les corpus oraux (Poplack et Walker, 1986 ; Sankoff et Cedergren, 1976) et le corpus *Texto4science* (Langlais et Drouin, 2012) ont plus de 20 ans d'écart, un laps de temps considérable dans toutes analyses sociolinguistiques. Les taux d'élision observés dans l'étude actuelle pourraient être dû à un changement dans l'usage à l'oral. Malheureusement, mes recherches n'ont fait ressortir aucune étude plus récente sur l'élision du /l/ dans les pronoms en français montréalais. Un autre problème est que étant donné son âge, le corpus *Texto4science* offre un regard vieillissant sur les comportements de textage des auteurs québécois. Ici encore, aucune étude plus récente n'a été faite sur l'élision dans les textos. Aussi, la technologie de téléphonie et de CMO évolue à un rythme effréné et déjà aujourd'hui, la réalité des auteurs n'est plus la même que lors de l'élaboration du corpus. Notamment, les types de téléphones et de claviers employés ont changé. Les claviers à 12 touches ont essentiellement disparu depuis l'avènement des iPhones et des autres téléphones androïdes à clavier tactile QWERTY. Aussi, les technologies d'autocorrection et d'autocomplétion sont aujourd'hui beaucoup plus performantes qu'elles ne l'étaient et largement plus

répandues. Il ne serait pas surprenant que les comportements des texteurs modernes diffèrent de ceux étudiés ici considérant que les contraintes qui motivaient autrefois l'emploi de textismes sont pour la plupart obsolètes : les limites de caractères ne sont plus aussi contraignantes, l'envoi de textos est beaucoup plus abordable et les téléphones sont plus faciles à utiliser que les vieux appareils à 12 touches. Bien que ces motivations aient disparu, l'emploi de textismes ne semble pas en déclin, ce qui suggère que la dimension ludique du *textese* demeure tout aussi importante aujourd'hui et qu'un nouvel ensemble de facteurs pourraient influencer l'emploi du langage SMS. L'enjeu est maintenant d'identifier quels pourraient être ces nouveaux facteurs. Malgré tout, les questions de recherche proposées pour le travail présent demeurent des enjeux importants à étudier puisque, dans une perspective diachronique, ils peuvent servir de points de référence et permettre de voir comment l'avancement de la technologie influence le développement des habitudes des texteurs et de mieux comprendre l'impact de l'âge des participants sur leurs habitudes de textage. Il serait important d'effectuer une nouvelle étude sur les texteurs modernes avec les nouvelles technologies et la réalité d'aujourd'hui, plus de 10 ans après l'élaboration du projet de Langlais et Drouin (2012).

De plus, à la lumière des résultats obtenus et des problèmes rencontrés lors de l'analyse des données, d'autres études seraient nécessaires pour bien étudier le phénomène de l'élision du <l> dans les pronoms. En effet, comme mentionné à plusieurs reprises, mon échantillon était trop petit pour élaborer un modèle statistique complexe capable de faire ressortir plus de facteurs. Par exemple, l'absence apparente d'effet du contexte phonologique pour les pronoms féminins semble surprenante considérant que ce facteur était significatif pour les pronoms sujets masculins et que les pronoms féminins se retrouvent dans des contextes similaires aux pronoms masculins. Aussi, les données du corpus *Texto4science* étaient insuffisantes pour étudier certains facteurs que j'ai nommés les impondérables. Pourtant, il est possible que ces facteurs exercent une influence sur l'élision et l'emploi du *textese* en général. Ils permettraient d'avoir un meilleur aperçu du style idiosyncratique des texteurs. Une nouvelle étude qui permettrait de prendre en compte notamment l'attitude du texteur face à la langue standard, les contextes de rédaction des textos et le rapport entre les texteurs serait nécessaire pour explorer cette dimension du langage SMS.

Finalement, Il ne faudrait pas se limiter à l'étude de l'élision comme seul phénomène permettant de faire un portrait du langage SMS et de ses pratiques. Si les facteurs retenus pour l'étude présente n'étaient pas significatifs pour l'élision, ça ne veut pas pour autant dire qu'ils ne le seraient pas pour d'autres variables linguistiques comme l'emploi spécifique de certaines formes de textismes. Il serait aussi intéressant de

pouvoir comparer les comportements de textage entre différentes communautés linguistiques. Le langage SMS est un sujet vaste, complexe et en évolution perpétuelle grâce aux avancements des technologies de CMO. Ce mémoire ne constitue qu'un point d'entrée en matière et aura permis d'identifier certains enjeux et besoins d'études futures sur le langage SMS.

ANNEXE A

Questionnaire original pour l'étude *Texto4science*

Quelques remarques

- Les questions auxquelles vous allez répondre sont très importantes pour notre projet. Nous vous remercions pour les quelques minutes que vous consacrerez à compléter ce formulaire.
- Nous vous demandons de répondre aussi précisément que possible, car l'exactitude de ces données sera déterminante pour les études à venir.
- Les informations que vous allez nous confier seront utilisées de manière totalement anonyme. Nous nous y engageons, comme en témoigne la rubrique "protection de la vie privée" dans la section "informations légales" de ce site. Il sera donc impossible à quiconque de faire le lien entre vous et les données entrées ci-dessous. Dans ces conditions, vous pouvez répondre à cette enquête en toute sérénité.

Questionnaire

1. Sexe

- a. Masculin
- b. Féminin

2. Langue maternelle

- a. Anglais
- b. Français
- c. Autre (s)

3. Niveau d'étude

- a. Primaire
- b. Secondaire
- c. CEGEP
- d. Université (1^{er} cycle)
- e. Université (2^e cycle)
- f. Université (2^e cycle)

4. Profession

5. Code postal

- a. De votre domicile
- b. De votre lieu de travail
- c. De votre école

6. Depuis combien de temps utilisez-vous les SMS ?

7. Combien de SMS envoyez-vous, en moyenne, par semaine ?
8. Votre dictionnaire est-il doté d'un dictionnaire d'aide à la rédaction de SMS ?
 - a. Si c'est le cas, utilisez-vous ce dictionnaire ?
9. Combien de touches compte votre téléphone ? (compter uniquement les touches qui permettent d'entrer des caractères)
10. Votre téléphone est-il doté d'un système de reconnaissance de l'écriture (type Palm : Graffiti, Jot, etc.)
11. Envoyez-vous des SMS à partir d'un site Web ?
12. Utilisez-vous d'autres systèmes de communication sur Internet ?
 - a. Les systèmes de clavardage (*chat*) dans Internet ?
 - b. Les forums dans Internet ?
 - c. Le courriel électronique ?
13. À qui envoyez-vous des SMS ?
14. Classez les destinataires par ordre de fréquence en utilisant "1" pour le plus fréquent (notez "0" pour les éventuelles catégories qui ne vous concernent pas) :
 - a. Famille
 - b. Amis
 - c. Collègues
 - d. Concours/Jeux
15. Comprenez-vous aisément les SMS que vous recevez quand ils contiennent des codes ou des abréviations ?
 - a. Oui, sans problème
 - b. Oui, mais je dois parfois "décoder"
 - c. Oui, mais difficilement
 - d. Non, c'est illisible
16. Quand vous rédigez vos SMS, utilisez-vous le français "normal" ou le "langage sms" ?
 - a. Le français « normal » uniquement
 - b. Plutôt le français "normal" avec quelques codes SMS
 - c. Un mélange du français "normal" et du "langage SMS"
 - d. Le "langage SMS" exclusivement
17. Le SMS est-il pour vous un jeu ou un véritable mode de communication ?
 - a. J'utilise le SMS pour communiquer (à la place du téléphone ou de l'email par exemple)
 - b. Un mélange de jeu et de réelle communication

c. J'utilise le SMS par jeu

18. Si vous le souhaitez, vous pouvez faire un commentaire sur la manière dont vous écrivez des SMS.

19. Si vous possédez plusieurs téléphones portables et que vous nous faites parvenir des SMS à partir de ces différents portables, merci d'indiquer ici les autres numéros pour que nous puissions lier tous les messages que vous nous envoyez à ce formulaire:

Vos données seront reprises dans un fichier géré par l'Observatoire de linguistique Sens-Texte (OLST, 3150 rue Jean-Brillant, Montréal, Québec, H3T 1N8, (514) 343-7174 et par le Centre de traitement automatique du langage de l'UCL (CENTAL, Place Blaise Pascal, 1 -1348 Louvain-la-Neuve - 010/473786). Vous pouvez exercer votre droit de regard sur ces données et, le cas échéant, les faire modifier ou supprimer. À ce sujet, vous pouvez contacter Patrick Drouin par téléphone (514) 343-7174 ou par courriel (patrick.drouin@umontreal.ca).

ANNEXE B

Exemple de SMS annoté dans le corpus *Texto4science*

```
<texto ID="2009120927" date="2009-12-09">
  <user_id>user_111250</user_id>
  <orig>
    Salut Florent, je sais pas ou tu es, mais si tu peux connecte toi sur
    Skype, j y serai une partie de la soiree! Ciao
  </orig>
  <transcrip>
    Salut <prenom sexe="masc">Florent</prenom>, je<negat forme="ne"/>
    sais pas <ortho forme="où">ou</ortho> tu es, mais si tu peux
    <ponc forme=","/><typog forme="connecte-toi">connecte toi</typog>
    sur Skype, <typog forme="j'y">j y</typog> serai une partie de la
    <ortho forme="soirée">soiree</ortho>! Ciao<ponc forme="."/>
  </transcrip>
  <norm>
    Salut Florent, je ne sais pas où tu es, mais si tu peux, connecte-toi
    sur Skype, j'y serai une partie de la soirée! Ciao.
  </norm>
</texto>
```


ANNEXE C

Exemples des variantes pour chaque catégorie de pronom

1. Pronoms masculins

a. Personnels singuliers

- i. ⟨il⟩ : Il peut pas non plus! Fuck man. Personne peut! (110250, *Texto4science*)
- ii. ⟨i⟩ : Jmeur..im goss;-j'ai envi dme tuer! (112171, *Texto4science*)
- iii. ⟨y⟩ : Tu sais ye ou mon bulletin de cette annee? (110250, *Texto4science*)
- iv. Absent : Lol. Le boss est parti chercher un laptop... Devrait revenir bientôt. (111850, *Texto4science*)

b. Impersonnels

- i. ⟨il⟩ : il est trop tard... dép fermé... demain j'irai... (112324, *Texto4science*)
- ii. ⟨i⟩ : Eum mai on peu laisser les chiens dehor... Paski fau ke jle sorte! (110451, *Texto4science*)
- iii. ⟨y⟩ : Y vient davoir un accident sur gascon (111030, *Texto4science*)
- iv. Absent : Faudrais vrmt tu check pr la tente (113010, *Texto4science*)

c. Météorologique

- i. ⟨il⟩ : C'est Noel car il neige dans ma tete! (110497, *Texto4science*)
- ii. ⟨y⟩ : Y pleut? (113010, *Texto4science*)

d. Personnels pluriels

- i. ⟨ils⟩ : Des quils arrivent jacourre (110250, *Texto4science*)
- ii. ⟨i⟩ : Kosse qui font caliss? (112350, *Texto4science*)
- iii. ⟨y⟩ : Y vont te laissions rentrer? Haha (112337, *Texto4science*)
- iv. Absent : Laisse faire! Sont pour la semaine prochaine :) (110250, *Texto4science*)

2. Pronoms féminins

- i. ⟨elle⟩ : Elle te pick up chez toi? (110250, *Texto4science*)
- ii. ⟨al⟩ : [...] pis a la juste 8 ans.. [...] (110250, *Texto4science*)
- iii. ⟨a⟩ : Pis la webcam a s'ouvre, a se ferme, a s'ouvre, a se ferme etc (110250, *Texto4science*)
- iv. Absent : Est comment ta robe pr les diplomes deja? (110250, *Texto4science*)

3. Pronoms objets indirects

- i. ⟨lui⟩ : Lol qu'es tu lui a fait? (111250, *Texto4science*)

ii. $\langle y \rangle$: Tu ya pas dit jespere ke jtai dit sa lol (112339, *Texto4science*)

ANNEXE D
Feuille de codage

La variable linguistique

Variable

Avec le ⟨l⟩ du pronom est présent

Sans le ⟨l⟩ est éliminé

Les variables indépendantes sociales

Âge

L'âge tel qu'indiqué dans le questionnaire sociolinguistique. (facteur continu)

Genre

m s'identifie comme un homme dans le questionnaire sociolinguistique

f s'identifie comme une femme dans le questionnaire sociolinguistique

Niveau de scolarité

preuni ayant terminé l'école primaire, secondaire et le cégep

uni ayant terminé des études universitaires (baccalauréat, maîtrise, doctorat)

Niveau de langue requis par l'emploi

non l'emploi ne demande pas de niveau de langue en particulier

Standard l'emploi demande un niveau de langue standard

Depuis combien de temps le participant texte

moins2a texte depuis moins de 2 ans

2a5 texte depuis 2 à 5 ans

plus5 texte depuis plus de 5 ans

Nombre de textos envoyés

moins5 envoie 5 textos ou moins par jour

6a10 envoie 6 à 10 textos par jour

10plus envoie plus de 10 textos par jour

Facilité de décodage

diff rapporte avoir de la difficulté à décoder les textos

easy rapporte avoir de la facilité à décoder les textos

Utilisation de dictionnaire

oui utilise parfois, souvent ou toujours les suggestions du dictionnaire

non n'utilise pas les suggestions du dictionnaire

Type de clavier

reg utilise un clavier à 12 touches

qwerty utilise un clavier QWERTY

Nombre d'heures de lecture par semaine

Le nombre d'heures indiquées dans le questionnaire sociolinguistique. (facteur continu)

Nombre d'heures d'écriture par semaine

Le nombre d'heures indiquées dans le questionnaire sociolinguistique. (facteur continu)

Les variables indépendantes linguistiques

Contexte phonologique

C / C suivi/précédé d'une consonne

V / V suivi/précédé d'une voyelle

Catégorie syntaxique suivant le mot

verbe	le mot suivant est un verbe tensé, un semi-auxiliaire ou un auxiliaire
autres	le mot suivant n'est pas un verbe, une semi-auxiliaire ou un auxiliaire.
inf	le mot suivant est un verbe à l'infinitif

Catégorie grammaticale de l'occurrence

expletif	Occurrence de pronom masculin singulier explétif par ex. Il te reste d'la place ds ton auto?
sujet	Occurrence de pronom masculin sujet. par ex. Des quils arrivent j'acourre

Effets aléatoires

Code pour identifier l'individu

Numéro de 6 chiffres généré aléatoirement

ANNEXE E

Commentaires des participants

110290 : (neg.) Je tente toujours de garder une bonne structure du français si le message est court. S'il est long, je coupe souvent les accords de pluriel ou le fin de syllabe qui ne sont pas pertinents à la compréhension du mot (ex: "apel moi" pour "appelle moi")

110375 : (neg.) Je déteste le langage texto. Parfois j'ai des personnes à qui j'envoie des textos qui ne les reçoivent pas à cause des accents ou des cédilles, je suis alors forcée de "mal écrire". Ça m'enrage :) Ah les smileys... je suis une grande fan. Le lol et le fml ou vdm aussi.

110495 : (neg.) Quelque soit le type de texte que je rédige (courriel, message sur un bout de papier, texto, etc.), j'essaie de faire le moins de fautes possibles. Les fautes m'énervent. J'ai toujours été comme ça. Les gens trouvent que je suis trop rigoureuse, mais je ne veux pas être illettrée

110500 : (neg.) Je suis une enseignante de français. J'ai donc un peu une réputation à maintenir. Je prends souvent la peine d'utiliser la meilleure qualité de langue possible

110530 : (neutre) La plupart du temps j'écris les mots au long sauf pour des mots qui possèdent des abréviations plus évidentes comme : qqn (quelqu'un), qqch(quelque chose), p-e (peut-être), etc. Aussi, j'abrège souvent les pronoms avec des apostrophes (ex: T'arrives, t'es, etc.)</p></div>

110637 (neutre) : Généralement j'ai le souci d'écrire en français correct

110661 : (pos.) Mes textos sont composés d'un mélange de langue usuelle et de langage texto. La composition du mélange varie selon le destinataire mais aussi en fonction du contexte (heure, endroit, jour, événement, émotion, etc.)</p></div>

110678 : (neutre) j'écris en langage texto surtout parce qu'avec un clavier de téléphone régulier à 12 touches c'est super long

110715 : (neutre) Lorsque que je text ma mère je fais toujours attention à l'orthographe, sinon j'écris en sms. :P

110751 : (neutre) La configuration du téléphone y est pour beaucoup. En changeant de téléphone, j'ai modifié de beaucoup ma manière de texter. Sur mon cell 12 touches, avec la fonction t9, j'utilisais les accents et apostrophes couramment. Maintenant, avec mon écran tactile QWERTY, utiliser les accents demande beaucoup de détour, et c'est trop long. Par contre, j'utilise moins les abréviations

110851 : (neg.) J'enrage quand les accents ne sont pas décodés correctement par le récepteur. Je refuse d'écrire ma langue sans accents, ça lui donne l'allure d'une femme chauve.

110990 : (neg.) Je me fais une fierté d'utiliser un français correct dans mes textos et je sensibilise aussi mes interlocuteurs. Le fait d'avoir un clavier tactile iPhone aide beaucoup; à l'époque du clavier téléphonique, j'utilisais plus souvent les abréviations.

92

111035 : (neutre) Je me fais une fierté d'utiliser un français correct dans mes textos et je sensibilise aussi mes interlocuteurs. Le fait d'avoir un clavier tactile iPhone aide beaucoup; à l'époque du clavier téléphonique, j'utilisais plus souvent les abréviations.

111530 : (neutre) Puisqu'on a un nombre limité de mots permis par texto, j'abrège les mots, par exemple si je veux écrire le verbe arriver dans un texto, je ne mettrais qu'un seul r

111570 : (neutre) Lorsque j'ai du temps, j'essaye d'écrire correctement car le temps passe plus vite. Et quand j'ai peu de temps, jamais dans des occasions formelles, j'écris ce que je veux écrire de manière à me faire comprendre en langage texto. Les conversations sont généralement courtes, mais j'aime bien mieux tout écrire sur les textos, ils ne me coûtent rien

111590 : (neg.) Malgré le fait que mon téléphone est normal,(12 touches), je tape mes textos à une vitesse qui épate beaucoup les gens. Et j'ai peur qu'avec toutes les abréviations, j'oublie l'orthographe française

111610 : (neutre) Le langage texto ne me sert que lorsque mon destinataire et moi avons une "conversation" suivie par texto et que je dois écrire rapidement. Cependant, depuis que j'ai un clavier QWERTY, il m'arrive vraiment moins souvent d'utiliser ce langage

111850 : (neg.) J'inclus même les apostrophes ! C'est long, mais j'aime quand c'est bien écrit. Je communique parfois avec mes clients de cette manière, donc on garde un français formel et on se relit. Très peu de fautes

112150 : (pos.) Je trouve que s'appliquer à écrire en bon français et tout corriger tout le temps serait une perte de temps énorme

112210 : (neg.) J'essaies d'écrire le plus lisiblement possible, même si mes messages textes sont plus longs que la " normale". Je préfère bien écrire et séparer mon message en plusieurs parties que d'écrire en abréviation. J'ai un clavier complet, alors ça me facilite un peu la tâche, mais je n'aime pas massacrer ma langue française

112314 : (neutre) Il faut avouer que mon français n'est pas parfait mais je fait habituellement attention pour ne pas rendre illisible mes textos

112333 : (neutre) Je texte à une seule personne sur cellulaire. Souvent j'utilise des mots à double sens ou j'écris une traduction mot à mot du français en anglais ou de l'anglais en français (ex. ça va ? it goes ?). J'essaye d'écrire en bon français, mais lorsque le temps me manque je peux faire des erreurs ou j'écris sans accents. Par contre j'utilise le langage texto pour clavarder sur les sites d'internet (MSN messenger, facebook

112335 : (pos.) le plus court possible

112350 : (neutre) J'utilise le iPhone. Je dirais que "souvent", je dois me "battre" contre le le dictionnaire d'aide à la rédaction, qui n'est pas super génial pour le "slang" ou est très difficile pour les francophones pour ce qui est des accents

112451 : (neutre) Je ne mets jamais de lettre avec des accents.. puisque souvent les cellulaires ne les transmettent pas et les lettres deviennent des Y .. Ex: Salut comment ça va? .. Salut comment ya va

RÉFÉRENCES

- Androutsopoulos, J., et Schmidt, G. (2002). SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe [SMS communication: Ethnographic genre analysis using the example of a small group]. *Zeitschrift für angewandte Linguistik*, 36, 49-80. <https://www.ids-mannheim.de/fileadmin/sprachvariation/tp/tp7/SMS-Kommunikation.pdf>
- Andujar, A. (2016). Benefits of mobile instant messaging to develop ESL writing. *System*, 62, 63-76. <https://doi.org/10.1016/j.system.2016.07.004>
- Armstrong, N. (1996). Variable deletion of French /l/: Linguistic, social and stylistic factors. *Journal of French Language Studies*, 6(1), 1-21. <https://doi.org/10.1017/S0959269500004956>
- Arthur-Nyarko, E. (2017). Effect of the use of instant messaging applications on writing of English as a second language in higher education. *European Journal of Education Studies*, 3(11), 861-880. <https://doi.org/10.5281/zenodo.1148814>
- Ashby, W. (1984). The elision of /l/ in French clitic pronouns and articles. Dans E. Pulgram (dir.), *Romanitas: Studies in Romance Linguistics* (p. 1–16). Michigan Romance Studies
- Aw, A. T., Zhang, M., Yeo, P. K., Fan, Z. Z. et Su, J. (2005). Input normalization for an English-to-Chinese SMS translation system. Dans *Proceedings of Machine Translation Summit X: Posters* (p. 445-450). Association for Computational Linguistics. <https://aclanthology.org/2005.mtsummit-posters.18.pdf>
- Aw, A. T., Zhang, M., Xiao, J., et Su, J. (2006). A phrase-based statistical model for SMS text normalization. Dans *Proceedings of the COLING/ACL on Main Conference Poster Sessions* (p. 33-40). Association for Computational Linguistics. <https://doi.org/10.3115/1273073.1273078>
- Baron, N. S. (2002, 27-30 décembre). “Whatever.”: A new language model. Dans *Annual Meeting of the Modern Language Association*, New York, New York. https://www.researchgate.net/publication/229005719_Whatever_A_new_language_model
- Baron, N. S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of language and social psychology*, 23(4), 397-423. <https://doi.org/10.1177/0261927X04269585>
- Battestini, A., Setlur, V., et Sohn, T. (2010). A large scale study of text-messaging use. Dans M. De Sá, L. Carriço et N. Correia (dir.), *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (p. 229-238). Association for Computing Machinery. <https://doi.org/10.1145/1851600.1851638>
- Bernicot, J., Goumi, A., Bert-Erboul, A., et Volckaert-Legrier, O. (2014). How do skilled and less-skilled spellers write text messages? A longitudinal study. *Journal of computer assisted learning*, 30(6), 559-576. <https://doi.org/10.1111/jcal.12064>
- Bialystok, E., Craik, F. I., Klein, R., et Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and aging*, 19(2), 290-303. <https://doi.org/10.1037/0882-7974.19.2.290>

- Bialystok, E., et Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, 7(3), 325-339. <https://doi.org/10.1111/j.1467-7687.2004.00351.x>
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12(1), 3-11. <https://doi.org/10.1017/S1366728908003477>
- Blake, C. (2009). Potential of Text-based Internet chats for improving oral fluency in a second language. *The Modern Language Journal*, 93(2), 227-240. <https://doi.org/10.1111/j.1540-4781.2009.00858.x>
- Blondeau, H., Tremblay, M. et Drouin, P. (2014). Hybridité et variation dans les SMS : Le corpus Texto4Science et l'oralité en français montréalais. *Revue Canadienne De Linguistique*, 59(1), 137-165. <https://doi.org/10.1017/S0008413100000189>
- Blondeau, H. et Tremblay, M. (2022). Écrire son vernaculaire: Variation et normes communautaires dans les messages textes en français québécois. *Journal of French Language Studies*, 32(2), 120-144. <https://doi.org/10.1017/S0959269522000096>
- Boeckx, C. (1999). Decomposing French questions. The University of Pennsylvania Working Papers in Linguistics 6:69–80.
- Bougaïeff, A. et Cardinal, P. (1980). La chute du /l/ dans le français populaire du Québec. *La linguistique*, 16(2), 91-102.
- Bourciez, É. (1889). *Précis de phonétique française: ou, Exposé des lois qui régissent la transformation des mots latins en français*. C. Klincksieck.
- Bourdieu, P. et Boltanski, L. (1975). Le fétichisme de la langue. *Actes de la recherche en sciences sociales*, 1(4), 2-32. https://www.persee.fr/doc/arss_0335-5322_1975_num_1_4_3417
- Cameron, D. (2012). *Verbal hygiene*. London: Routledge. <https://doi.org/10.4324/9780203123898>
- Chambers, J. K. (1992). Linguistic correlates of gender and sex. *English World-Wide*, 13(2), 173-218. <https://doi.org/10.1075/eww.13.2.02cha>
- Chen, T. et Kan, M. Y. (2013). Creating a live, public short message service corpus: the NUS SMS corpus. *Language Resources and Evaluation*, 47, 299-335. <https://doi.org/10.1007/s10579-012-9197-9>
- Chomsky, N. (1993). *Lectures on government and binding: The Pisa Lectures*. New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110884166>
- Côté, M-H. (2022). French. Dans C. Gabriel, R. Gess et T. Meisenburg (dir.), *Manual of Romance Phonetics and Phonology*, 27 (p. 669-708). Walter de Gruyter. <https://doi.org/10.1515/9783110550283>
- Craig, David. (2003) "Instant messaging: The language of youth literacy." *The Boothe Prize Essays*, 1 116–133. <https://nl.ijs.si/janes/wp-content/uploads/2014/09/craig03.pdf>

- Crawford, J. (1998, novembre). Ten common fallacies about bilingual education. ERIC Digest. <https://files.eric.ed.gov/fulltext/ED424792.pdf>
- Crystal, D. (2004). *A glossary of netspeak and textspeak*. Edinburgh University Press.
- Crystal, D. (2008). *Texting*. *Elt journal*, 62(1), 77-83. <https://doi.org/10.1093/elt/ccm080>
- Crystal, D. (2009). *Txtng: The gr8 db8*. OUP Oxford.
- De Jonge, S. et Kemp, N. (2012). Text - message abbreviations and language skills in high school and university students. *Journal of Research in Reading*, 35(1), 49-68. <https://doi.org/10.1111/j.1467-9817.2010.01466.x>
- Derakhshan, A. et Kaivanpanah, S. (2011). The impact of text-messaging on EFL freshmen's vocabulary learning. *The EUROCALL Review*, 19, 47-55. <https://doi.org/10.4995/eurocall.2011.16271>
- Devitt, M. et Sterelny, K. (1989). Linguistics: What's wrong with "the right view". *Philosophical Perspectives*, 3, 497-531. <https://doi.org/10.2307/2214279>
- Djuikui Dounstop, C., Auger, J. et Tremblay, M. (2020). A real-time analysis of the variable use of expletive 'il' in Montréal French. *University of Pennsylvania Working Papers in Linguistics*, 26(2), 8.
- Drouin, M. et Davis, C. (2009). R u txtng? Is the Use of Text Speak Hurting Your Literacy? *Journal of Literacy Research*, 41(1), 46-67. <https://doi.org/10.1080/10862960802695131>
- Drouin, M. et Driver, B. (2012). Texting, textese and literacy abilities: A naturalistic study. *Journal of Research in Reading*, 37(3), 250-267. <https://doi.org/10.1111/j.1467-9817.2012.01532.x>
- Dunnewind, S. (2003, 12 avril). Generation text: Teens' IM lingo evolving into a hybrid language. *The Seattle Times*. <https://www.seattletimes.com/life/generation-text-teens-im-lingo-evolving-into-a-hybrid-language/>
- Dürscheid, C. (2003). Syntaktische Tendenzen im heutigen Deutsch. *Zeitschrift für germanistische Linguistik*, 31(3). 327-342. <https://doi.org/10.1515/zfgl.2003.31.3.327>
- Dürscheid, C. et Stark, E. (2011). SMS4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland. *Digital Discourse: language in the new media*, 299-320. <https://doi.org/10.5167/uzh-49872>
- Eckert, P. (2017). Age as a sociolinguistic variable. Dans F. Coulmas (dir.), *The handbook of sociolinguistics* (p. 151-167). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405166256.ch9>
- Ellison, S. (2001, 2 mai) UTLKIN2ME? Can U speak cell phone?. *ZDNET*. <https://www.zdnet.com/article/utlkin2me-can-u-speak-cell-phone/>
- Endong, F. P. C. et Essoh, N. E. G. (2015). Orality versus literacy in the Nigerian SMS discourse. *Khazar Journal of Humanities and Social Sciences*, 18(3), 33-47. <http://hdl.handle.net/20.500.12323/3262>

- Essoh, N. E. G., Odey, V. E. et Endong, F. P. C. (2014). Creativity features of SMS texts in French by Anglophone teachers and learners of the French language. *Journal of Language and Cultural Education*, 2(3), 79-95.
- Fairon, C. et Paumier, S. (2006). A translated corpus of 30,000 French SMS. Dans K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk et D. Tapias (dir.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)* (p. 351-354). European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2006/pdf/270_pdf.pdf
- Fairon, C., Klein, J. R. et Paumier, S. (2006a). Le langage SMS: révélateur d'1compétence. Dans J.-J. Didier, O. Hambursin, P. Moreau et M. Seron (dir.), *Le français m'a tuer. Actes du colloque "L'orthographe française à l'épreuve du supérieur"* (p. 33-42). Presse universitaire de Louvain.
- Fairon, C., Klein, J. R. et Paumier, S. (2006b). *SMS pour la science (licence: 1 utilisateur, manuel+ CD-Rom): Corpus de 30.000 SMS et logiciel de consultation* (Vol. 3). Presses universitaires de Louvain.
- Ferguson, C. A. (1959). Diglossia. *Word*, 15(2), 325-340.
<https://doi.org/10.1080/00437956.1959.11659702>
- Flikeid, K. (1989). Moitié anglais, moitié français? Emprunts et alternance de langues dans les communautés acadiennes de la Nouvelle-Ecosse in Bilinguisme et diglossie. *Revue québécoise de linguistique théorique et appliquée*, 8(2), 177-288.
- François, D. (1974). *Français parlé : analyse des unités phoniques et significatives d'un corpus recueilli dans la région parisienne* (Vol. 2). Paris: Sela.
- Grinter, R. et Eldridge, M. (2003). Wan2tlk? Everyday text messaging. Dans G. Cockton et P. Korhonen (dir.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (p. 441-448). Association For Computing Machinery. <https://doi.org/10.1145/642611.642688>
- Hamlaoui, F. (2011). On the role of phonology and discourse in Francilian French wh questions. *Journal of Linguistics* 47:129–62.
- Hård af Segerstad, Y. (2002). *Use and adaptation of written language to the conditions of computer-mediated communication* [thèse de doctorat, University of Gothenburg]. GUPEA.
<http://hdl.handle.net/2077/15738>
- Holmes, J. (1993). Women's talk: The question of sociolinguistic universals. *Australian Journal of Communication*, 20(3), 125-149. <https://search.informit.org/doi/10.3316/ielapa.940504974>
- Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of research in personality*, 45(1), 92-99. <https://doi.org/10.1016/j.jrp.2010.11.015>
- Howard, M., Lemée, I. et Regan, V. (2006). The L2 acquisition of a phonological variable: The case of /l/ deletion in French. *Journal of French Language Studies*, 16(1), 1–24.
<https://doi.org/10.1017/S0959269506002298>

- Humphrys, J. (2007). I h8 txt msgs: How texting is wrecking our language. *Daily Mail*.
<https://www.dailymail.co.uk/news/article-483511/I-h8-txt-msgs-How-texting-wrecking-language.html>
- Internet Society Foundation. (2023, 15 mai). *What are the most used languages on the Internet?*.
 Internet Society Foundation. <https://www.isocfoundation.org/2023/05/what-are-the-most-used-languages-on-the-internet/>
- James, D. (1996). Women, men, and prestige speech forms: A critical review. Dans V. L. Bergvall, J. M. Bing, et A. F. Freed (dir.), *Rethinking language and gender research : theory and practice* (p. 98-125). Routledge.
- Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed - effects variable rule analysis. *Language and linguistics compass*, 3(1), 359-383. <https://doi.org/10.1111/j.1749-818X.2008.00108.x>
- Kauhanen, I. (2006). Norms and sociolinguistic description. Dans M. Suominen, A. Arppe, A. Airola, O. Heinämäki, M. Miestamo, U. Määttä, J. Niemi, K. K. Pitkänen et K. Sinnemäki (dir.), *SKY Journal of Linguistics* [numéro special], 19, 34-46.
https://www.linguistics.fi/julkaisut/SKY2006_1/1FK60.1.4.KAUHANEN.pdf
- Kling, R. (1996). Hopes and Horrors: Technological Utopianism and Anti-Utopianism in Narratives of Computerization. *Computerization and Controversy: Value Conflicts and Social Choices* (2^e éd.).
<https://doi.org/10.1016/B978-0-12-415040-9.50091-9>
- Koehn, P., Och, F. J. et Marcu, D. (2003). Statistical phrase-based translation. Dans M. Hearst et M. Ostendorf (dir.), *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (p. 48-54).
https://www.pure.ed.ac.uk/ws/portalfiles/portal/25694519/p48_koehn.pdf
- Laberge, S. (1977). *Étude de la variation des pronoms sujets définis et indéfinis dans le français parlé à Montréal* [Thèse de doctorat, Université de Montreal]. DOI:10.13140/RG.2.2.30733.38883
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania press.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language variation and change*, 2(2), 205-254. <https://doi.org/10.1017/S0954394500000338>
- Laliberté, T. (1974). L'élision du 'L' en français québécois. *Lingua*, 33(2), 115-122.
[https://doi.org/10.1016/0024-3841\(74\)90029-1](https://doi.org/10.1016/0024-3841(74)90029-1)
- Langlais, P. et Drouin, P. (2012). Texto4Science: A Quebec French database of annotated text messages. *Linguisticæ Investigationes*, 35(2), 237-259. <https://doi.org/10.1075/li.35.2.07lan>
- Leeman, D. (2006). L'absence du sujet en français contemporain: premiers éléments d'une recherche. *L'information grammaticale*, 110(1), 23-30. https://www.persee.fr/doc/igram_0222-9838_2006_num_110_1_3832

- Lenhart, A. (2012). *Teens, smartphones & texting* [Rapport]. Pew Research Center.
https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_Teens_Smartphones_and_Texting.pdf
- Lenhart, A. et Madden, M. (2005). *Teens and technology* [Rapport]. Pew Research Center.
<https://www.pewresearch.org/internet/2005/07/27/teens-and-technology/>
- López Rúa, P. (2007). Teaching L2 vocabulary through SMS language: Some didactic guidelines. *ELIA*, 7, 165-188. <https://idus.us.es/handle/11441/34164>
- Martin-Rhee, M. M. et Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: language and cognition*, 11(1), 81-93.
<https://doi.org/10.1017/S1366728907003227>
- Massot, B. (2008). Français et diglossie. *Décrire la situation linguistique française contemporaine comme une diglossie: arguments morphosyntaxiques* (Doctoral dissertation, Université Paris VIII Vincennes-Saint Denis).
- McCulloch, G. (2019). *Because internet: Understanding the new rules of language*. Penguin.
- McSweeney, M. (2016). *Literacies of Bilingual Youth: A profile of bilingual academic, social, and txt literacy* (Publication n° 10151988) [Thèse de doctorat, City University of New York]. Proquest.
<https://www.proquest.com/docview/1835126808/80DAEB021E0D48E2PQ/1?accountid=14719&sourcetype=Dissertations%20&%20Theses>
- Mehl, M. R. et Pennebaker, J. W. (2003). The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology*, 84(4), 857-870. <https://doi.org/10.1037/0022-3514.84.4.857>
- Morin, Y. C. (1979). La morphophonologie des pronoms clitiques en français populaire. *Cahier de linguistique*, 9, 1-36. <https://doi.org/10.7202/800076ar>
- Mphahlele, M. L. et Mashamaite, K. (2005). The impact of short message service (SMS) language on language proficiency of learners and the SMS dictionaries: A challenge for educators and lexicographers. Dans P. Isaías, C. Borg, P. Kommers et P. Bonanno (dir.), *Proceedings of the IADIS International Conference on Mobile Learning* (p. 161-168). <https://www.iadisportal.org/digital-library/the-impact-of-short-message-service-sms-language-on-language-proficiency-of-learners-and-the-sms-dictionaries-a-challenge-for-educators-and-lexicographers>
- Niedzielski, N. A. et Preston, D. R. (2000). *Folk linguistics* (Vol. 122). Walter de Gruyter.
<https://doi.org/10.1515/9783110803389>
- Ostiguy, L. (1979) *La chute de la consonne "l" dans les articles définis et les pronoms clitiques en français montréalais* [Mémoire de maîtrise, Université de Montréal].
- Ostiguy, L. et Tousignant, C. (2008). *Les prononciations du français québécois : Normes et usages*, 2^e édition, Montréal, Guérin.

- Palasis, K. (2013). The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies*, 23(1), 17-35. <https://doi.org/10.1017/s0959269512000348>
- Panckhurst, R., Détrie, C., Lopez, C., Moïse, C., Roche, M. et Verine, B. (2013). Sud4science, de l'acquisition d'un grand corpus de SMS en français à l'analyse de l'écriture SMS. *Episteme*, 9(9), 107-138. <https://hal.science/hal-00923618/>
- Papineni, K., Roukos, S., Ward, T. et Zhu, W. B. (2002). Bleu: a method for automatic evaluation of machine translation. Dans P. Isabelle, E. Charniak et D. Lin (dir.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (p. 311-318). Association for Computational Linguistics. <https://doi.org/10.3115/1073083.1073135>
- Paquin-Drouin, C. (2017). *Analyse de l'élision variable de [l] dans les pronoms « il(s) » et « elle » en français québécois oral*. [Manuscrit non publié]. Université du Québec à Montréal.
- Paveau, M. A. (2007). Les normes perceptives de la linguistique populaire. *Langage et société*, 119, 93-109. <https://www.cairn.info/revue-langage-et-societe-2007-1-page-93.htm>
- Pennebaker, J. W. et King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296-1312. <https://doi.org/10.1037/0022-3514.77.6.1296>
- Pérez-Sabater, C. et Montero-Fleta, B. (2015). A first glimpse at mobile instant messaging: Some sociolinguistic determining factors. *Poznan Studies in Contemporary Linguistics*, 51(3), 411-431. <https://doi.org/10.1515/psicl-2015-0016>
- Plester, B., Wood, C. et Bell, V. (2008). Txt msg n school literacy: Does texting and knowledge of text abbreviations adversely affect children's literacy attainment?. *Literacy*, 42(3), 137-144. <https://doi.org/10.1111/j.1741-4369.2008.00489.x>
- Plester, B., Wood, C. et Joshi, P. (2009). Exploring the relationship between children's knowledge of text message abbreviations and school literacy outcomes. *British Journal of Developmental Psychology*, 27(1), 145-161. <https://doi.org/10.1348/026151008X320507>
- Pope, M. K. (1952). *From Latin to Modern French with especial consideration of Anglo-Norman: Phonology and morphology*. Manchester University Press.
- Poplack, S. et Walker, D. (1986). Going through /l/ in Canadian French. Dans D. Sankoff (dir.), *Diversity and Diachrony* (p. 173–198). John Benjamin Publishing Company.
- Poplack, S. (1988). Language status and language accommodation along a linguistic border. Dans P. Lowenberg (dir.), *Language spread and language policy: Issues, implications and case studies*. Washington, DC: Georgetown University Press, p. 90-118.
- Poplack, S. (2015). Norme prescriptive, norme communautaire et variation diaphasique. Dans K. Kragh et J. Lindschouw (dir.), *Les variations diasystématiques et leurs interdépendances* (p. 293-319). Société de linguistique romane.

https://www.researchgate.net/publication/317039901_Norme_prescriptive_norme_communaux_taire_et_variation_diaphasique

- Preston, D. R. (1993). The uses of folk linguistics. *International Journal of Applied Linguistics*, 3(2), 181-259. <https://doi.org/10.1111/j.1473-4192.1993.tb00049.x>
- Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness*, 5(1), 40-74. <https://doi.org/10.1080/09658416.1996.9959890>
- Preston, D. R. (2011). Methods in (applied) folk linguistics: Getting into the minds of the folk. *AILA Review*, 24(1), 15-39. <https://doi.org/10.1075/aila.24.02pre>
- Pupier, P. et Légaré, L. (1973). L'effacement du /l/ dans les articles définis et les clitiques en français de Montréal. *Glossa*, 7(1), 63-80. Glossa Society.
- Quan-Haase, A. (2007). University Students' Local And Distant Social Ties: Using and integrating modes of communication on campus. *Information, Communication & Society*, 10(5), 671-693. <https://doi.org/10.1080/13691180701658020>
- R Core Team (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.r-project.org/>
- Ross, K. (2007, 1 novembre). Does text messaging hurt student writing skills?: Teachers say text messages r ruining kids' riting skills. *American Teacher*, 92(3). <http://www.mccanntesting.com/docs/articles/myaccess.american.teacher.20071101.pdf>
- Rowlett, P. (2013). Do French speakers really have two grammars? *Journal of French Language Studies*, 23(1), 37-57. doi:10.1017/S095926951200035X
- Ruane, M. E. (1999, 17 décembre) Net spawns a new teen lingo: Noticed some odd spelling when looking over your kids' shoulders when they're online? Here's why. *The Vancouver Sun*. <https://www.proquest.com/docview/242744838?accountid=14719&sourcetype=Newspapers>
- Sankoff, G. et Cedergren, H. (1971). Some results of a sociolinguistic study of Montreal French. Dans R. Darnell (dir.), *Linguistic Diversity in Canadian Society* (p. 61-87). Linguistic Research Inc.
- Sankoff, G. et Cedergren, H. (1976). Les contraintes linguistiques et sociales de l'élision du L chez les Montréalais. Dans M. Boudreault and F. Möhren (dir.), *Actes du XIIIème congrès international de linguistique et philologie romanes* (p. 1101-1116). Presses de l'Université Laval
- Sankoff, D. et Laberge, S. (1978). The linguistic market and the statistical explanation of variability. Dans D. Sankoff (dir.), *Linguistic variation: Models and methods* (p. 239-250). Academic Press <https://web.stanford.edu/~eckert/Courses/ParisPapers/SankoffLaberge.pdf>
- Sax, K. (2003). *Acquisition of stylistic variation by American learners of French* (Publication n° 305333555) [Thèse de doctorat, Indiana University]. ProQuest Dissertations and Theses Global. <https://www.proquest.com/dissertations-theses/acquisition-stylistic-variation-american-learners/docview/305333555/se-2?accountid=14719>

- Shafie, L., Darus, N. et Osman, N. (2010). SMS language and college writing: The languages of the college texters. *International Journal of Emerging Technologies in Learning (IJET)*, 5(1), 26-31.
<https://www.learntechlib.org/p/45202/>
- Soffer, O. (2010). "Silent orality": Toward a conceptualization of the digital oral features in CMC and SMS texts. *Communication Theory*, 20(4), 387-404. <https://doi.org/10.1111/j.1468-2885.2010.01368.x>
- Squires, L. (2012). Whos punctuating what? Sociolinguistic variation in instant messaging. Dans A. Jaffe, J. Androutsopoulos, M. Sebba et S. Johnson (dir.), *Orthography as social action: Scripts, spelling, identity and power* (vol. 3, p. 289-323). De Gruyter Mouton.
<https://doi.org/10.1515/9781614511038.289>
- Standage, T. (1999). *The victorian internet: The remarkable story of the telegraph and the nineteenth century's online pioneers*. The Berkeley publishing group.
- Stark, E. et Robert-Tissot, A. (2017). Subject drop in Swiss French text messages. *Linguistic variation*, 17(2), 251-271.
- Stark, E., Robert-Tissot, A., Frick, K. (2018). Determiner ellipsis in electronic writing-discourse or syntax? Dans A.-K. Gärtig, R. Bauer et M. Heinz (dir.), *Pragmatik – Diskurs – Kommunikation : un Festschrift pour les 65 ans de Gudrun Held* (p. 186-198). Praesens Verlag.
<https://doi.org/10.5167/uzh-152176>
- Stroud, J. (2004, 22 février) 'Thumb talk' : text messaging takes off. *The Morning Journal*.
<https://www.morningjournal.com/2004/02/22/thumb-talk-text-messaging-takes-off/>
- Sutherland, J. (2002). Can u txt? John Sutherland asks what texting is doing to the English Language—and finds it all a bit-(-. *The Guardian*.
<https://www.theguardian.com/technology/2002/nov/11/mobilephones2>
- Swanson, H. L., Sáez, L., Gerber, M. et Leafstedt, J. (2004). Literacy and cognitive functioning in bilingual and nonbilingual children at or not at risk for reading disabilities. *Journal of Educational Psychology*, 96(1), 3-18. APA PsycNet. <https://doi.org/10.1037/0022-0663.96.1.3>
- Tabatabaei, O. et Goojani, A. H. (2012). The impact of text-messaging on vocabulary learning of Iranian EFL learners. *Cross-Cultural Communication*, 8(2), 47-55. CSCanada.
<https://doi.org/10.3968/j.ccc.1923670020120802.1689>
- Tagg, C. (2009). *A corpus linguistics study of SMS text messaging* [Thèse de doctorat, University of Birmingham]. UBIRA ETheses. <https://etheses.bham.ac.uk/id/eprint/253/>
- Tagg, C., Baron, A. et Rayson, P. (2012). "i didn't spel that wrong did i. Oops": Analysis and normalisation of SMS spelling variation. *Lingvisticæ Investigationes*, 35(2), 367-388.
<https://doi.org/10.1075/li.35.2.12tag>
- Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.
<https://www.cambridge.org/us/universitypress/subjects/languages-linguistics/discourse-analysis/analysing-sociolinguistic-variation?format=HB&isbn=9780521771153>

- Tagliamonte, S. A. et Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American speech*, 83(1), 3-34. Duke University Press.
<https://doi.org/10.1215/00031283-2008-001>
- Tagliamonte, S. A. (2016). So sick or so cool? The language of youth on the internet. *Language in Society*, 45(1), 1-32. Cambridge University Press. <https://doi.org/10.1017/S0047404515000780>
- Thurlow, C. et Brown, A. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse analysis online*, 1(1).
<https://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003.html>
- Thurlow, C. (2006). From Statistical Panic to Moral Panic: The Metadiscursive Construction and Popular Exaggeration of New Media Language in the Print Media. *Journal of Computer-Mediated Communication*, 11(3), 667-701. International Communication Association.
<https://doi.org/10.1111/j.1083-6101.2006.00031.x>
- Totanes, B. C. et Lintao, R. B. (2019). Textese categories and textese application in L2 class discussion. *Journal on English Language Teaching*, 9(1), 14-31. i-manager Publications.
<https://doi.org/10.26634/jelt.9.1.15249>
- Tremblay, M. (2020). Le texto: une pratique langagière distincte?. Dans K. Reinke (dir.), *Attribuer un sens. La diversité des pratiques langagières et les représentations sociales* (p. 79-100). Presses de l'Université Laval. <https://corpus.ulaval.ca/server/api/core/bitstreams/9b83630b-ded8-4b38-8c27-17bc964bf18b/content#page=87>
- Tremblay, M., Blondeau, H., & Labeau, E. (2020). Texting the future in Belgium and Québec: Present matters. *Journal of French Language Studies*, 30(1), 73-98.
- Treurniet, M., De Clercq, O., Heuvel, H. et Oostdijk, N. H. J. (2012). Collection of a corpus of Dutch SMS. Dans N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk et S. Piperidis (dir.), *Proceedings of LREC 2012* (p.2268-2273). European Language Resources Association (ELRA). <https://hdl.handle.net/2066/101552>
- Verheijen, L. (2013). The Effects of text messaging and instant messaging on literacy. *English Studies*, 94(5), 582-602. Routledge. <https://doi.org/10.1080/0013838x.2013.795737>
- Villeneuve, A. J., & Auger, J. (2013). 'Chtileu qu'i m'freumereu m'bouque i n'est point coér au monne': Grammatical variation and diglossia in Picardie. *Journal of French Language Studies*, 23(1), 109-133.
- Volckaert-Legrier, O., Goumi, A., Bert-Erboul, A. et Bernicot, J. (2015). Focus on text messages: A review of studies in French, *Encyclopedia of Mobile Phone Behavior*, 24, 1037-1050). IGI Global.
<https://doi.org/10.4018/978-1-4666-8239-9.ch085>
- Waldron, S., Wood, C. et Kemp, N. (2017). Use of predictive text in text messaging over the course of a year and its relationship with spelling, orthographic processing and grammar. *Journal of Research in Reading*, 40(4), 384-402. Wiley Online Library. <https://doi.org/10.1111/1467-9817.12073>
- Walker, D. C. (1984). *The pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

- Weimers, M. (2008). *2 Mix or nt 2 MIX? (Dats d kwestion): The effect of instant messaging on Afrikaans speaking adolescents in South Africa* [Travail dirigé, document non publié]. University of Stellenbosch.
- Winzker, K., Southwood, F. et Huddleston, K. (2009). Investigating the impact of SMS speak on the written work of English first language and English second language high school learners. *Per Linguam: a Journal of Language Learning - Tydskrif vir Taalaanleer*, 25(2), 1-16. University of Stellenbosch. <https://hdl.handle.net/10520/EJC87019>
- Wood, C., Meachem, S., Bowyer, S., Jackson, E., Tarczynski - Bowles, M. L. et Plester, B. (2011). A longitudinal study of children's text messaging and literacy development. *British Journal of Psychology*, 102(3), 431-442. <https://doi.org/10.1111/j.2044-8295.2010.02002.x>
- Wood, N. (2014). *Autocorrect awareness: Categorizing autocorrect changes and measuring authorial perceptions* [Honors Thesis, Florida State University]. FSU Digital Library. <https://diginole.lib.fsu.edu/islandora/object/fsu:204751/datastream/PDF/view>
- Zimmermann, M. et Kaiser, G. A. (2014). On expletive subject pronoun drop in Colloquial French. *Journal of French Language Studies*, 24(1), 107-126. Cambridge University Press. <https://doi.org/10.1017/S0959269513000392>