

# Les formes interrogatives totales à travers divers modes de discussion en ligne : quelle connotation de niveau de langue ?

Lehti, Lotta ; Laippala, Veronika

Institut des sciences du langage et de l'interprétation, Université de Turku  
lotta.lehti@utu.fi, veronika.laippala@utu.fi

<b>1 Introduction.....</b>	<b>153</b>
<b>2 Formes interrogatives et niveau de langue .....</b>	<b>153</b>
<b>3 Corpus et méthode .....</b>	<b>155</b>
<b>4 Résultats.....</b>	<b>156</b>
<i>Tableau 1 : la distribution des formes d'interrogation totale .....</i>	<i>156</i>
<b>5 Discussion.....</b>	<b>158</b>
<b>Références bibliographiques .....</b>	<b>159</b>

## 1 Introduction

Dans l'opinion courante, le langage de l'Internet participatif (Web 2.0) est souvent associé aux pratiques non-normatives de l'écriture. Les premières études linguistiques sur la Communication Médinée par Ordinateur (CMO) ont renforcé cette position en se concentrant sur quelques traits micro-linguistiques dans certains modes de la CMO (Bieswanger 2013, Androutsopoulos 2006, Thurlow 2006). Dans la langue française, la présence des traits du langage oral dans la CMO écrite a fait couler beaucoup d'encre (voir par exemple Marcoccia & Gauducheau 2007, van Compernelle & Williams 2007, Pierozak 2003). Aujourd'hui, la recherche sur la CMO est étendue par la notion du « discours numérique », notion référant à la multiplicité et complexité des objets et des approches de recherche linguistique sur l'Internet. La recherche sur le discours numérique prend en compte l'interaction entre les facteurs technologiques, sociaux et contextuels (Thurlow & Mroczek 2011).

Notre étude porte à la fois sur la tradition de la CMO et sur le discours numérique : nous examinons les discussions en ligne du point de vue du degré de formalité du langage, c'est-à-dire le niveau de langue. Dans notre étude antérieure sur le degré de formalité du langage dans les blogs de politiciens (Lehti & Laippala 2014), nous avons combiné l'analyse des traits micro-linguistiques caractéristiques à la CMO à celle des indices sociolinguistiques traditionnels. Dans ce travail, un indice, notamment la forme de l'interrogation totale, s'est avéré complexe quant aux connotations possibles en ce qui concerne le niveau de langue. Le présent travail poursuit l'analyse des formes des interrogations totales dans le discours numérique, en se concentrant sur les discussions en ligne. Les questions que nous posons sont les suivantes : Quelle est la distribution des différentes formes d'interrogation totale employées dans les discussions en ligne ? Comment ces emplois reflètent-ils les connotations de niveau de langue traditionnellement associées à ces formes ?

Notre perspective est surtout quantitative : nous visons à l'analyse d'un corpus relativement vaste, par des outils automatiques. L'identification et le marquage automatique des traits langagiers, syntaxiques dans notre cas, ne se déroulent pourtant pas sans difficultés. L'objectif du présent article est, par conséquent, de présenter les premiers résultats de l'analyse quantitative, mais également de discuter les défis proposés par celle-ci. Au lieu d'essayer de définir le niveau de langue des discussions en ligne, notre étude vise principalement à ouvrir des pistes pour l'étude des formes d'interrogations, et leur éventuelle connotation de niveau de langue dans un environnement numérique.

## 2 Formes interrogatives et niveau de langue

Nous nous intéressons aux formes interrogatives en tant qu'indices possibles de variation de niveau de langue, c'est-à-dire du degré de formalité du langage. Le continuum de variation de formalité est étudié dans la sociolinguistique française de diverses façons : par exemple Gadet (2003) le divise en quatre niveaux, à savoir « soutenu », « standard », « familier » et « populaire ». De même, Arrivé, Gadet et Galmiche (1986) distinguent quatre niveaux différents mais les dénomment « soigné », « courant/moyen », « familier » et « populaire ». Quant à Mougeon, Rehner et Nadashi (2004), ils présentent une division en trois niveaux, dits (en anglais) *formal*, *mildly marked* et *vernacular*. La multiplicité des dénominations manifeste la nature indéfinie des niveaux : ils sont flous et ils peuvent être déterminés de diverses manières.

Quels que soient le nombre et la définition des niveaux, ils sont signalés dans le langage par des indices qui connotent un sens social mais qui n'influencent pas le sens dénotatif de l'énoncé (Coupland 2007) : des exemples de ces indices en français sont l'omission du terme négatif *ne*, les phrases pseudo-clivées et un vocabulaire familier (Blanche-Benveniste 2000, Gadet 2003, Coveney 1996). La forme de l'interrogation totale, qui est l'objet du présent

travail, est reconnue en sociolinguistique comme un indice éminent de variation du niveau de langue (Coveney 1996, Gadet 2003, Riegel et al. 2009 [1994]) : dans le français européen, l'inversion du sujet est considérée comme une marque du langage soutenu alors que l'interrogation marquée par la seule intonation – ou point d'interrogation à l'écrit – est associée au langage familier, ce qui est souvent le cas également de l'interrogation avec *est-ce que*.

La sociolinguistique variationniste traditionnelle considère certains choix langagiers et le niveau de langue qu'ils connotent comme des indices directs de l'appartenance à un groupe social donné. Cette perspective est critiquée entre autres par Rickford & Eckert (2001 : 1) qui posent que la sociolinguistique d'aujourd'hui est plus consciente des fonctions subtiles de différents styles :

In spite of the centrality of style, the concerted attention that has been paid to the relation of variation to social categorizations and configurations has not been equaled by any continuous focus on style. In other words, we have focused on the relation between variation and the speaker's place in the world, at the expense of the speaker's strategies with respect to this place. But as social theories of variation develop greater depth, they require a more sophisticated, integrative treatment of style that places variation within the wider range of linguistic practices with which speakers make social meaning.

De la même manière, Arrighi (2007 : 48) constate que l'association traditionnelle d'une forme interrogative aux caractéristiques sociales du locuteur ne permet pas de comprendre la variation intrapersonnelle, soit l'alternance de formes interrogatives chez un même locuteur, dans une même situation de communication.

Notre questionnement n'est pourtant pas sociolinguistique (quel groupe social emploie quelle forme) mais au cœur de notre conception des choix langagiers est le genre de discours au sein duquel les choix sont réalisés. Dans le sillage de Miller (1984), nous concevons le genre comme un type d'activité sociale langagière, soit un modèle de situations de communication récurrentes, comportant des éléments tels que les identités des participants, le médium de communication et ses affordances et le but communicatif (voir également Devitt 2004, Giltrow 2013 et Lehti 2013). Ce modèle pose des conventions génériques quant à la forme des textes (écrits ou oraux) appartenant au genre en question, y compris les formes syntaxiques. Les affordances technologiques sont au cœur du questionnement du présent travail. Selon Miller & Shepherd (2009 : 281, notre traduction), il s'agit des « possibilités et contraintes rhétoriques caractéristiques d'un médium de communication ». Nous considérons que les affordances du chat ou du forum de discussion, par exemple, peuvent influencer le choix de la forme interrogative.

La discussion en ligne correspond à de multiples types d'activité sociale langagière, soit de genres. Les différents genres de discussion en ligne peuvent être classifiés de diverses façons – par exemple selon le thème, les participants ou la nature du site. La définition des genres de discussion en ligne doit, en principe, se faire sans la connaissance des participants car tout ce qui est visible de ceux-ci est le pseudonyme ou le nom indiqué. Branca-Rosoff (2007) considère le chat comme un genre de discours et constate (id. : 137) que dans les chats, ainsi que dans d'autres genres, la socialisation des comportements conduit les interlocuteurs à s'ajuster au discours antérieur. À notre avis, cela peut être le cas également en ce qui concerne le choix des structures langagières, notamment la forme des phrases interrogatives. Pourtant, comme le remarque Branca-Rosoff (2007 : 138), les individus peuvent chercher à se distinguer « des façons de dire des interlocuteurs ». Branca-Rosoff fait référence surtout aux pratiques ludiques en ce qui concerne l'orthographe des mots.

### 3 Corpus et méthode

La composition de notre corpus des discussions en ligne provenant de divers forums est décrite ci-dessous. En plus de notre corpus, nous faisons référence à l'alternance des formes interrogatives dans un corpus de chats étudié par van Compernelle et Williams (2009).

Premièrement, nous avons rassemblé un corpus de commentaires dans les blogs de politiciens : il s'agit de commentaires postés dans 80 blogs de politiciens français pendant le mois de septembre 2007, soit 3316 commentaires. Le nombre de mots pour ces commentaires est de 425 084. Les politiciens auteurs de ces blogs représentent des partis différents (PS, UMP, Verts, MoDem, FN, PRG) et des fonctions différentes (conseillers municipaux, généraux et régionaux, maires, députés, sénateurs, ministres, députés européens). L'identité des commentateurs reste naturellement opaque. Ce corpus a été analysé également dans la thèse de doctorat de Lehti (Lehti 2013).

Deuxièmement, nous avons eu accès au corpus recueilli par l'équipe de recherche IRG (Interactional Research Group)<sup>1</sup> de l'Université de Turku. Ce corpus contient des discussions en ligne, tantôt asynchrones, tantôt synchrones.

Le sous-corpus « Forum Le Monde », contient 14 discussions sur le site du journal *Le Monde* en 2008 (<http://forums.lemonde.fr/perl/wwwthreads.pl>). De chaque discussion examinée, les 50 premiers messages ont été inclus dans le corpus (exceptions : une de ces discussions contient uniquement 45 messages, et 55 messages d'une autre discussion ont été inclus dans le corpus). Ainsi, au total, 700 messages, soit 73610 mots, sont recueillis. Les thèmes de ces discussions sont les suivants : la crise financière, la crise en Géorgie et les élections présidentielles aux États-Unis.

Ensuite, le corpus IRG contient dix discussions concernant la vie des jeunes adultes et étudiants provenant de différentes sources (<http://forums.studyrama.com>, <http://www.reseautudiant.com/forum>, <http://forum-scipo.com/>, <http://forum.hardware.fr/>) concernant désormais le sous-corpus « Forum Étudiants ». Également de ces discussions, les 50 premiers messages ont été recueillis (exceptions : une des discussions ne contient que 44 messages et celle-ci a été sauvegardée dans sa totalité ; une autre discussion contient deux messages complètement blancs et c'est pourquoi les 52 premiers messages de cette discussion ont été inclus dans le corpus). Au total, cette partie du corpus consiste en 494 messages (72 253 mots). Les thèmes traités dans ces discussions sont les suivants : la recherche de l'emploi, le CPE (contrat première embauche) et les questions financières.

Le corpus IRG contient également 10 éditoriaux du journal *Le Monde* (<http://www.lemonde.fr/>) suivis de commentaires des lecteurs (désormais sous-corpus « Éditorial Le Monde »). Les éditoriaux contiennent au total 3 603 mots. Le nombre total des commentaires est de 339 (22 343 mots). Les sujets des éditoriaux sont politiques : la crise financière, les élections présidentielles aux États-Unis et différentes questions liées à la politique française.

Quant aux discussions synchrones, le corpus IRG contient 15 chats des internautes avec des politiciens ou d'autres personnalités publiques ([www.rue89.com](http://www.rue89.com), [www.libération.fr](http://www.libération.fr), [www.20minutes.fr](http://www.20minutes.fr)) désormais sous-corpus « Chat ». Les chats ont été sauvegardés dans leur totalité ; dans l'ensemble, ils contiennent 2 502 messages (234 540 mots).

Finalement, nous avons également eu accès au corpus recueilli par Johanna Isosävi<sup>2</sup> contenant des discussions sur Facebook. Ce sous-corpus contient 1262 commentaires sur le profil « Contre Nicolas Sarkozy » en 2010 et 2011, au total 46805 mots. Les thèmes discutés sur ce profil tournent autour de l'opposition à Nicolas Sarkozy ; le registre communicatif est souvent polémique (voir Isosävi à paraître 2014).

<sup>1</sup> Nous remercions Marjut Johansson et les autres membres de l'équipe IRG de nous avoir donné l'accès à ce corpus.

<sup>2</sup> Nous tenons à la remercier.

Tout le corpus est sauvegardé au format .tx et les analyses sont effectuées avec des scripts Unix. Nous avons recueilli les phrases interrogatives du corpus sur la base de l'usage d'un point d'interrogation. Les questions suivies de plusieurs points d'interrogation ont également été collectées. La ponctuation expressive représente un des défis de l'analyse du langage de la CMO : au lieu de l'usage normatif d'un point d'interrogation, on peut en utiliser plusieurs, et les points d'interrogations ne sont pas toujours précédés et / ou suivis d'un espace. Cette variation doit être prise en compte lors de l'analyse automatique pour capturer toutes les occurrences.

## 4 Résultats

Pour des raisons de simplicité, nous nous concentrons sur les interrogations totales et laissons de côté les interrogations partielles, c'est-à-dire portant sur un constituant de la phrase car ces dernières peuvent être réalisées par au moins cinq formes syntaxiques. Ensuite, nous avons divisé les interrogations totales du corpus en quatre classes selon leur forme syntaxique (cf. Riegel et al. 2009 [1994] 670-672) :

- *SV : Interrogation avec l'ordre sujet-verbe signalée par le point d'interrogation, exemple (1)*
- *ESV : Interrogation avec est-ce que, exemple (2)*
- *VS : Inversions simple et complexe, exemples (3) et (4) respectivement*
- *NF : Interrogations sans verbe fini, exemples (5) et (6)*

- (1) *tu as des nouvelles?* (Forum Étudiants, italiques rajoutés)
- (2) *Est-ce qu'on osera encore les menacer de leur supprimer les maigres allocations pour lesquelles ils ont cotisé?* (Éditorial Le Monde, italiques rajoutés)
- (3) *Croyez-vous au come back politique de Villepin?* (Chat Rue89, italiques rajoutés)
- (4) *Le Monde et le journalisme en général sait-il parfois prendre sa part de responsabilité?* (Éditorial Le Monde, italiques rajoutés)
- (5) *Retour au politique?* (Éditorial Le Monde, italiques rajoutés)
- (6) *SARKO???* (Facebook, italiques rajoutés)

Contrairement à notre analyse des blogs de politiciens (Lehti & Laippala 2014), nous avons décidé de laisser de côté les interrogations sans verbe fini même si elles sont fréquentes dans le corpus. La raison derrière cette omission est que la possible connotation du degré de formalité de cette forme nous reste opaque car, à notre connaissance, elle n'est pas explorée dans la sociolinguistique. Ainsi, le Tableau 1 représente la distribution des formes SV, ESV et VS dans notre corpus, ainsi que dans le corpus de chats examiné par van Compernelle et Williams (2009).

	Commentaires dans les blogs de politiciens (n = 979)	Facebook (n= 86)	Forum Le Monde (n=181)	Editorial Le Monde (n= 64)	Forum Étudiants (n= 129)	Chat (n=877)	Chat (van Compernelle & Williams 2009) (n= 167)
SV	35.2 %	62.8 %	61.9 %	37.5 %	51.9 %	20.2 %	97.7 %
ESV	1.8 %	4.7 %	5.5 %	4.7 %	7.8 %	4.7 %	0.6 %
VS	62.9 %	32.6 %	32.6 %	57.8 %	40.3 %	75.1 %	1.7 %

Tableau 1 : la distribution des formes d'interrogation totale

Nous considérons que la répartition des formes interrogatives dans le sous-corpus de chat est quelque peu surprenante. La proportion (20.2 %) de la forme SV (dite familière) est la plus faible dans ce sous-corpus et la proportion de la forme VS (dite soutenue) est la plus élevée (75.1 %). Cette répartition diverge considérablement de celle dans les chats étudiés par van Compernelle & Williams (2009) : 97.7 % des interrogations totales produites par les locuteurs natifs du français sont en forme SV alors que seulement 1.7 % sont en forme VS. De plus, van Compernelle et Williams (2009 : 12) constatent que les formes ESV (1) et SV (3) sont toutes produites par un même locuteur et elles peuvent être interprétées comme ludiques.

Nous considérons que la différence entre le chat dans notre corpus et celui dans le corpus de van Compernelle & Williams (2009) résulte des différents participants. Van Compernelle & Williams (2009 : 10) caractérisent le chat en question comme très décontracté et familier. Quant aux chats dans notre corpus, cette interaction est caractérisée par la présence d'un politicien ou autre personne connue en tant qu'invité au chat. Par conséquent, le discours reflète la distance et l'asymétrie de pouvoir entre l'invité et les autres participants comme dans l'exemple (7) :

- (7) **Bonjour Monsieur le ministre. Ne pensez vous pas qu'en enlevant des postes les classes risquent d'être surchargées? De plus, comment allez vous revaloriser les salaires des enseignants du primaire qui ont de l'expérience, qui font plus d'heures que ceux du secondaire, et qui gagnent peu?**

[Réponse] Nous aurons cette année, en moyenne nationale, un taux d'encadrement légèrement meilleur que l'an dernier. Mais il y a évidemment des disparités, notamment dans les villes moyennes et à Paris. La question de la revalorisation des professeurs du premier degré sera résolue; nous avons prévu déjà qu'ils accèdent à des heures supplémentaires et nous revalorisons les indemnités liées à des sujétions spéciales.

(20minutes chat)

Dans (7), l'internaute commence sa question au ministre par une adresse relativement formelle. Les phrases sont complètes et l'orthographe est, dans l'ensemble, conforme à la norme — le niveau de langue de ce message peut être appelé standard, voire soutenu.

Ainsi, les divergences entre les résultats de van Compernelle & Williams et les nôtres suggèrent une corrélation « traditionnelle » entre la forme interrogative et la formalité : dans une situation de communication décontractée, la forme SV domine alors que dans une situation de communication plus formelle, la forme VS est la plus fréquente.

La formalité de la forme VS est pourtant mise en question par une orthographe non standard dans un grand nombre des cas. Les déviations de la norme les plus fréquentes sont : l'absence du tiret entre le verbe et le sujet pronom (voir (8) et (9) ci-dessous) et l'orthographe du *t* de liaison entre « la finale vocalique du verbe et le pronom il(s) ou elle(s) » (Riegel 2009 [1994] : 671). Le *t* de liaison est parfois entouré par d'autres signes que le tiret normatif : par exemple l'apostrophe est parfois employée (voir (10) et (11)) ou le *t* est entouré simplement par des espaces vides (voir (12)).

- (8) *Ne pensez vous pas* que les Parisiens usagers ou travailleurs dans ces structures méritent mieux que ces contrats précaires et les changements incessants de personnel qui en découlent ? (Chat Rue89, italique ajouté)
- (9) *Allez vous virer* tous les profs non-titulaires n'ayant qu'une licence (bac+ 3) lors de la création du master, même si ils enseignent depuis 8,9,10 ans ? (20 minutes chat, italique ajouté)
- (10) Ce concours *sera t'il* conservé ? (20 minutes chat, italique ajouté)
- (11) Concernant le concours de professeur des écoles : *Y aura t'il* 2 fois moins de postes au CRPE 2009 si des Masters 2 arrivent sur le marché en 2010 en même temps que les lauréats 2009!??? (20 minutes Chat, italique ajouté)
- (12) d'ailleurs on pourrait aussi lui poser la question dite « morano » : *pense t il* aussi que quiconque n'est pas d'accord avec le gouvernement est du côté des assassins ? (Chat Rue89, italique ajouté)

Ainsi, il reste à déterminer l'influence de l'orthographe non-normative à la connotation de degré de formalité : il nous semble que la formalité connotée par la forme VS est quelque peu mise en question par l'orthographe non-normative.

Le taux très faible de la forme ESV dans chaque mode de discussion n'est guère surprenant. Si la forme ESV est fréquente à l'oral (Riegel et al. 2009 [1994] : 672), elle peut paraître laborieuse et lente à produire sur le clavier : par rapport aux formes SV et VS, elle augmente le nombre de frappes. Quant aux formes SV et VS, leur répartition dans notre corpus ne peut pas s'expliquer par les affordances technologiques car en principe, ces deux formes exigent autant de frappes. La rapidité exigée dans la communication synchrone ne semble pas non plus influencer le choix entre SV et VS. Dans notre corpus, la forme SV est plus fréquente dans les discussions moins synchrones – sur *Facebook* et le forum du site de *Le Monde* par exemple – que dans les chats synchrones. Il reste à explorer les raisons pour l'usage fréquent de la forme SV dans ces deux modes : par exemple, quelle est l'influence du registre polémique (*Facebook*) sur le niveau de langue ?

## 5 Discussion

En conclusion, les résultats obtenus dans la présente analyse suggèrent surtout que les affordances technologiques n'expliquent pas toute la variation de niveau de langue dans le discours numérique. La forme VS, traditionnellement conçue comme un indice de formalité, est plus fréquente dans les chats que dans les forums asynchrones. Comme nous l'avons déjà mentionné, il semble que la présence d'un politicien ou autre personne connue dans la discussion en ligne influence les choix langagiers plus que les contraintes techniques de l'interaction. Il serait intéressant d'explorer ce phénomène dans les discussions en ligne en d'autres langues que le français. Par exemple en finnois, l'interaction avec une personne connue n'exige pas nécessairement des signes de formalité, tels que le vouvoiement.

En plus de l'aspect interculturel, notre étude ouvre de nombreuses pistes pour l'exploration des formes interrogatives dans les discussions en ligne. Premièrement, il serait bon d'élargir l'analyse aux interrogations partielles contenant un terme interrogatif. De plus, la ponctuation étant l'objet de beaucoup d'usages non-normatifs, il serait important de développer une méthode pour pouvoir inclure également dans l'analyse les interrogations qui ne se terminent pas par un point d'interrogation. L'analyse automatique morphologique et syntaxique du matériel peut éventuellement contribuer à ces buts. Nous envisageons également de le faire en utilisant l'analyseur syntaxique TALISMANE (Urieli et Tanguy 2013) et l'outil de recherche décrit par Haverinen et al. (2010). Or, comme le TALISMANE est adapté au langage normatif, les déviations fréquentes de la norme dans certains modes de la CMO surtout en ce qui concerne l'orthographe, le lexique et les structures syntaxiques, risquent de compliquer son usage. Par conséquent, nous avons la possibilité d'adapter ce système à des variétés de la même façon que nous avons déjà procédé pour adapter un système similaire, originellement ciblé au finnois normatif, aux variétés utilisées par des docteurs et des infirmières dans les dossiers médicaux (voir Laippala et al. 2013 et Laippala et al. In Press).

Deuxièmement, des corrélations entre différents facteurs co(n)textuels et l'usage d'une forme interrogative devraient être étudiées. Comme le démontrent par exemple Arrighi (2007) et Coveney (1996), les facteurs pragmatiques, notamment l'acte de langage réalisé par la phrase interrogative, influencent le choix syntaxique. La nature du verbe (auxiliaire, modal, ou lexical) et du sujet (lexical ou pronominal) employé est également susceptible d'influencer le choix syntaxique. Ensuite, la variation interpersonnelle au sein d'un sous-corpus devrait également être prise en compte : même si les identités des participants restent, pour la plupart, inconnues à l'analyste, il est possible d'examiner les formes selon les pseudonymes (ou les vrais noms) annoncés par les participants.

Finalement, pour pouvoir mieux comprendre le degré de formalité du langage connoté par l'usage d'une forme interrogative, il serait important d'étudier ce trait langagier en même temps que d'autres indices connotant un certain niveau de langue. Par exemple, dans notre travail sur le niveau de langue des billets et commentaires dans les blogs des politiciens français, nous avons étudié la fréquence et l'usage de six indices, tantôt syntaxiques (omission du *ne* de négation et les formes des interrogations totales), tantôt lexicaux (mots familiers et des traits de néographie lexicale (voir Anis 1999, 2007)), tantôt prosodiques (ponctuation répétitive et usage d'émoticônes). Il faudrait développer une méthode d'analyse de fréquence conjointe dans un extrait de texte numérique donné – les résultats obtenus par une telle méthode pourraient être révélateurs de la variation. Naturellement, le sens inverse est également possible : l'analyse peut se dérouler à partir de textes, sans indices prédéterminés, et se baser sur une interprétation holistique par l'analyste. Cette approche ne permet pourtant pas l'analyse d'un corpus vaste, ce qui est l'atout de l'analyse automatique.

### Références bibliographiques

- Androutsopoulos, Jannis K. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics* 10(4), 419–438.
- Anis, Jacques. (1999). Chats et usages graphiques du français. In : Anis, Jacques (ed.) *Internet, communication et la langue française*. Paris: Hermès. 71-90.
- Anis, Jacques. (2007). Neography: Unconventional spelling in French SMS text messages. In: Danet, Brenda & Herring, Susan C. (eds.) *The multilingual internet: Language, culture, and communication online*. Oxford : Oxford University Press. 87-115.
- Arrighi, Laurence. (2007). L'interrogation dans un corpus de français parlé en Acadie. Formes de la question et visées de l'interrogation. *Linx*, 57. 47-56.
- Arrivé, Michel & Gadet, Françoise., & Galmiche, Michel. (1986). *La grammaire d'aujourd'hui: guide alphabétique de linguistique française*. Paris : Flammarion.
- Bieswanger, Markus. (2013). Micro-linguistic structural features of computer-mediated communication. In: Herring, Susan C. & Stein, Dieter & Virtanen, Tuija (eds.) *Handbook of pragmatics of computer-mediated communication*. Berlin/New York: De Gruyter Mouton. 463-485.
- Blanche-Benveniste, Claire. (2000). *Approches de la langue parlée en français*. Gap : Ophrys.
- Branca-Rosoff, Sonia. (2007). Genres et activité langagière : l'exemple des tchats. *Linx*, 56, 127-141.
- Coupland, Nikolas. (2007). *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.
- Coveney, Aidan. (1996). Variability in spoken French: A sociolinguistic study of interrogation and negation. Exeter: Elm Bank Publications.
- Devitt, Amy. (2004). *Writing Genres*. Carbondale : Southern Illinois University.
- Gadet, Françoise. (2003). *La variation sociale en français*. Paris: Ophrys.
- Giltrow, Janet. (2013). Genre and computer-mediated communication. In: *Herring, Susan, Stein, Dieter & Virtanen, Tuija (eds). Pragmatics of Computer-Mediated Communication. Berlin/New York: De Gruyter. 717-737.*
- Haverinen, Katri & Viljanen, Timo & Laippala, Veronika & Kohonen, Samuel & Ginter, Filip & Salakoski, Tapio. (2010). Treebanking Finnish. In : Dickinson, Markus & Müürisepp, Kaili & Passarotti, Marco (eds.) *Proceedings of The Ninth International Workshop on Treebanks and Linguistic Theories (TLT9)*. 79-90.
- Isosävi, Johanna, (à paraître 2014). Le dissensus, la polémique et la violence verbale – quelles sont leurs manifestations dans la communication Facebook ? Exemple du profil « Contre Nicolas Sarkozy ». *Actes du Colloque Dialogic Language Use 3: Miscommunication and Verbal Violence*. Helsinki, 15–17.8.2012.
- Laippala, Veronika & Viljanen, Timo & Airola, Antti & Kanerva, Jenna & Salanterä, Sanna & Salakoski, Tapio & Ginter, Filip. (In press 2014). Statistical parsing of varieties of clinical Finnish. *Artificial Intelligence in Medicine*.
- Laippala, Veronika & Viljanen, Timo & Airola, Antti & Nyblom Jenna & Salanterä, Sanna & Salakoski, Tapio & Ginter, Filip. (2013). Statistical Parsing of Varieties of Clinical Finnish. In: *Proceedings of Louhi2013 - The 4th International Workshop on Health Document Text Mining and Information Analysis with the Focus of Cross-Language Evaluation*. 11-12 February 2013 Sydney, Australia.



- Lehti, Lotta. (2013). Genre et ethos : Des voies discursives de la construction d'une image de l'auteur dans les blogs de politiciens. Turku : Université de Turku.
- Lehti, Lotta & Laippala, Veronika (2014). Style in French politicians blogs: Degree of formality. *Language@Internet*, 11, article 1. <http://www.languageatinternet.org/articles/2014/lehti/>
- Marcoccia, Michel. (2004). On-line polylogues: conversation structure and participation framework in internet newsgroups. *Journal of Pragmatics*, 36 (1). 115-145.
- Marcoccia, Michel & Gauducheau, Nadia. (2007). L'analyse du rôle des smileys en production et en réception : un retour sur la question de l'oralité des écrits numériques. *Glottopol*, 10, 39-55.
- Miller, Carolyn R. (1984). Genre as social action. *Quarterly Journal of Speech*, 70. 151-167.
- Miller, Carolyn R. & Shepherd, Dawn. (2009). Questions for genre theory from the blogosphere. In: Giltrow, Janet & Stein, Dieter (eds.) *Genres in the Internet: Issues in the theory of genre*. Amsterdam: John Benjamins, 263-290.
- Mougeon, Raymond & Rehner, Katherine & Nadasdi, Terry. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics*, 8(3), 408-432.
- Pierozak, Isabelle. (2003). Le Français tchaté. Une étude en trois dimensions – sociolinguistique, syntaxique et graphiques – d'usages IRC. Thèse de doctorat en sciences du langage, Aix-Marseille I.
- Rickford, John R. & Eckert, Penelope. (2001). Introduction. In: Eckert, Penelope & Rickford, John R. (eds.) *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 1-18.
- Riegel, Martin & Pellat, Jean-Christophe & René Rioul. (1997 [1994]). *Grammaire méthodique du français*. Paris: Presses Universitaires de France.
- Thurlow, Crispin. (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), article 1.
- Thurlow, Crispin & Mroczek, Kristine. (2011). Introduction: Fresh Perspectives on New Media Sociolinguistics. In: Thurlow, Crispin & Mroczek, Kristine (eds.) *Digital Discourse: Language in the New Media*. New York & London: Oxford University Press, xix-xliv.
- Urieli, A. and L. Tanguy. (2013). L'apport du faisceau dans l'analyse syntaxique en dépendances par transitions : études de cas avec l'analyseur Talisman. In: Proceedings of TALN 2013 conference, Sables-d'Olonne, France, 17-21 juin 2013.
- Van Compernelle, Rémy A., & Williams, Lawrence. (2009). De l'oral à l'électronique : la variation orthographique comme ressource sociostylistique et pragmatique dans le français électronique. *Glottopol*, 10, 56-69.
- Van Compernelle, Rémy. A., & Williams, Lawrence. (2009). Patterns of stylistic variation in synchronous computer-mediated French: *Yes/No* questions and *nous* versus *on*. *Studies in Second Language Acquisition*, 31(3), 471-500.