

Emprunts lexicaux dans un corpus de textos français au Canada

Christian Guilbault
Université Simon Fraser

Patrick Drouin
Université de Montréal
Observatoire de
linguistique Sens-Text

Messages

POURQUOI LES TEXTOS ?

Edit

- Emprunts: TIC favorisent les formes familières et, donc , les échanges liés aux contacts des langues; les textos en sont un bel exemple, langue spontanée
- À cause de l'ampleur du phénomène...
- À cause des affirmations faites sur la langue (jeunes et qualité) sans vérification empirique; besoin d'un corpus
- Comment résister à ceci :
 - « Hey jo! Je vais te faire signe demain ca serait cool de chiller ds la journee. »
 - « ps. sorry g pas reply pour ton auto »



Sherbrooke, juin 2012

Send

Messages

PLAN DE CETTE PRÉSENTATION

Edit

1. Origines du projet
2. Texto4science.ca
 1. Corpus « Canada » français
 2. Question de recherche et hypothèses
 3. Échantillon pour cette analyse
3. Résultats
4. Conclusion



Sherbrooke, juin 2012

Send

SMS4science.org

- Depuis 2004, Belgique (CENTAL : Centre de Traitement Automatique du Langage de l'Université catholique de Louvain) Cédric Fairon, Louise-Amélie Cougnon, etc.
- Antennes et sous corpora en France notamment (sud4science, alpes4science, lareunion4science, etc.), Grèce, etc.; 15 universités partenaires
- But : *décrire, comprendre, théoriser (sur) la langue utilisée dans les textos*
- Analyse séparées mais également comparaisons/contraste des différents corpora dans des régions/dialectes/langues différents



Partenaires au Canada :

- P. Drouin (UdeM)
- Philippe Langlais, Recherche Appliquée en Linguistique Informatique (RALI)
- Annie Desnoyers, Groupe québécois pour la modernisation de la norme du français (GQMNF)

Organismes et institutions impliqués :

- Recherche appliquée en linguistique informatique (RALI)
- Centre de recherche interdisciplinaire sur les technologies émergentes (CITÉ)
- Centre de Traitement Automatique du Langage (CENTAL)
- Département d'informatique et de recherche opérationnelle

Universités : U de Montréal, U d'Ottawa, université Simon Fraser

Compétences : terminologues, informaticiens, linguistes, communications, traducteurs



Sherbrooke, juin 2012



Objectif général :

- colliger et analyser le plus grand nombre de textos possible (en deux phases) au Canada et mise à la disponibilité des chercheurs en base de données
- Mieux comprendre l'utilisation de la langue
- comparer cette version locale avec d'autres langues et dialectes du français
- faire un corpus de référence pour les analyses linguistiques (diachronie)
- Développements d'applications liées à ce nouvel usage de la langue (correcteurs, traduction automatique, etc.)

Méthodologie : similaire aux autres projets liés à SMS4science (Fairon et coll. 2007)



Sherbrooke, juin 2012



Messages

3. RÉSULTATS DE LA CUEILLETTE

Edit

Aperçu

- Effectuée de 2009-2010
- Textos mais aussi questionnaire d'information sociolinguistique
- Total environ 7302 messages (très majoritairement en français, 6870 en français, 420 anglais, 12 d'autres langues)
- Environ 13 135 phrases
- Environ 106 785 mots
- Collecte terminée officiellement (mais text4science.ca...)
- Représentativité ? ...



Sherbrooke, juin 2012

Send

Messages

3. ANALYSE

Edit

Observation : les textos semblent contenir un bon nombre d'alternance codique et d'emprunts :

1. Ac tt ki mtomb dessus. G oublié d't dmandé, WHAT HAPENED WITH THE CUTIE?
2. Jmen vo worker et suis vrm dead ...
3. si tu veux chiller tranquilos.
4. Um dude j'ai dit dme dire avant de partir lol im home. Ill meet u there et c chill g ossi un cadenas

Q. : Comment sont effectués les emprunts ?



Send

Messages

3. ANALYSE

Edit

Questions de recherche :

- Quelle est l'ampleur du phénomène d'emprunt lexical ?
- De quels types sont ces emprunts ?
- Y a-t-il des règles d'assimilation ?
- Sont-ils similaires chez tous les locuteurs ? Sinon, quels sont les profils des locuteurs qui effectuent différents types d'emprunts ?



Sherbrooke, juin 2012

Send

D'après Cougnon (2011)

- Code switching: « (...) variable length sequences of words of two languages, varieties or registers or languages, juxtaposed in the same verbal exchange, which demonstrate all indications of a monolingual exchange and which respect both languages' grammatical rules (Cougnon 2007). »
- Transfer: « (...) verbal exchange where the switch to a foreign language occurs but is limited to a specific construction followed by a predictable return to the ML. »
- Loan words: « (...) items borrowed from an FL lexicon which have become part of the ML lexicon, and are no longer considered as examples of the use of an FL. »
- Code mixing: « the use of an FL in an ML context with no respect for the FL's grammatical rules. »



- Sur les 7302 messages, 70 % de transferts et 16 % d'emprunts
- Un peu d'alternance (14 %)
- Très peu de mélange (0,3 %)

Importance des *transferts* et *emprunts* !

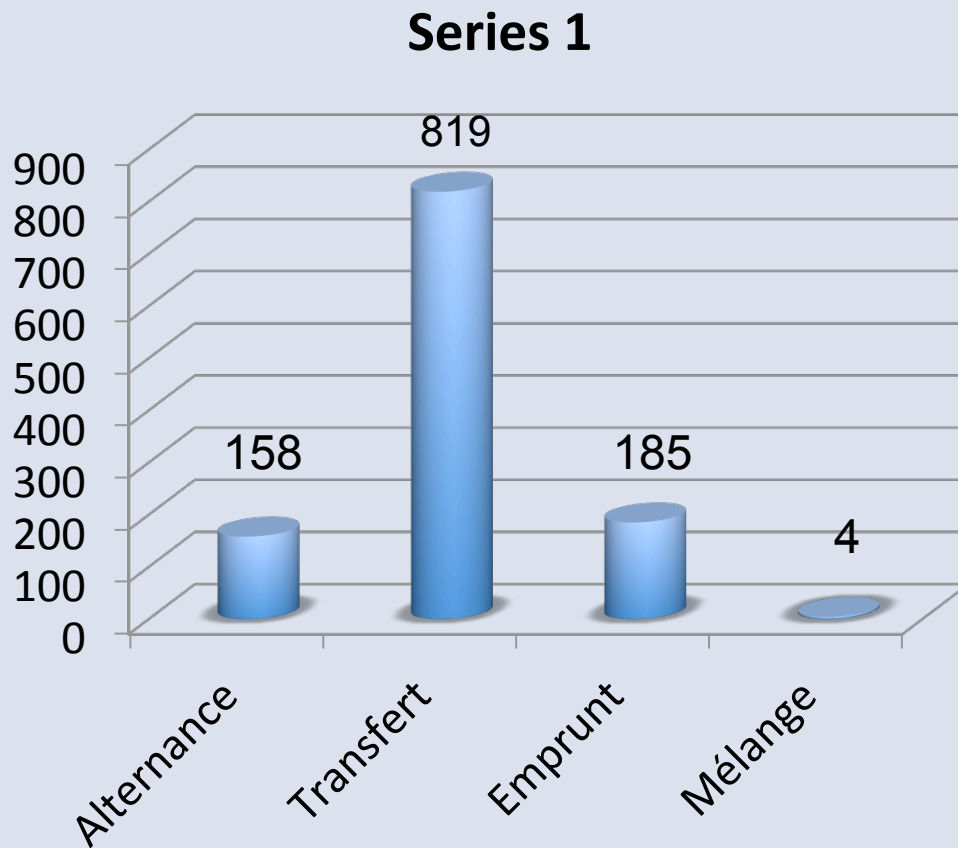
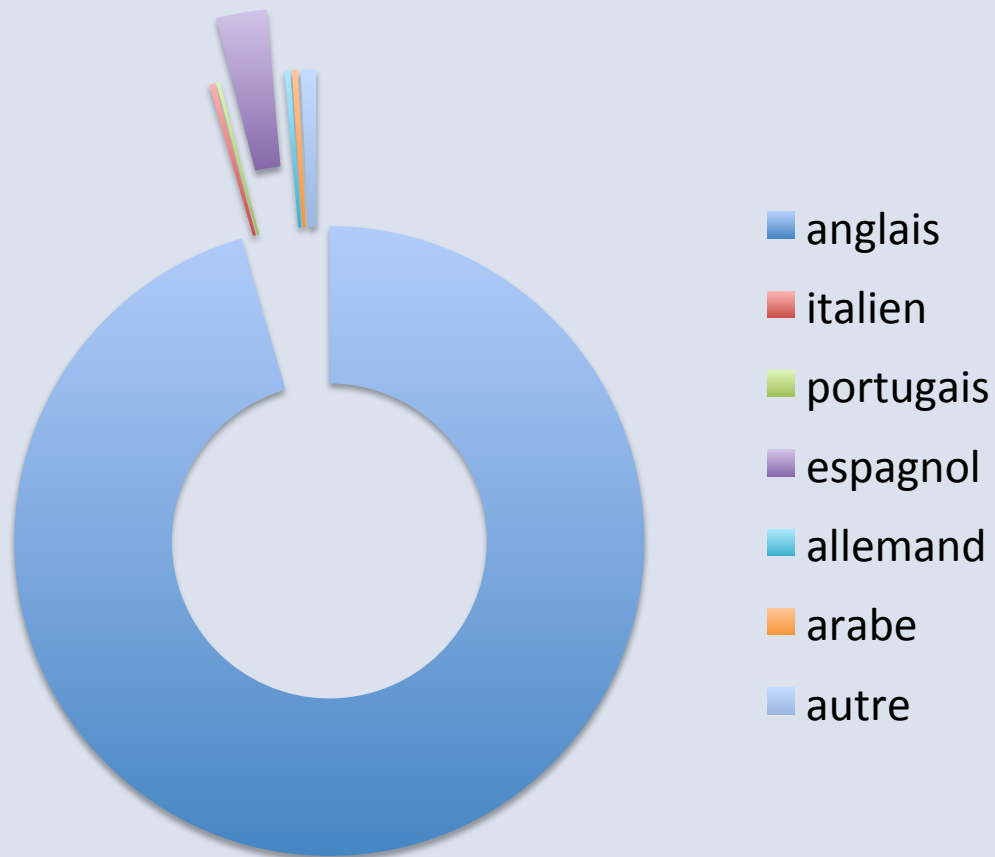


Figure inutile car :

- L'anglais domine (95,6 %)...
- L'espagnol est aussi présent (2,6 %)



Messages

RÉSULTATS 3 : EMPRUNTS PAR TYPE

Edit

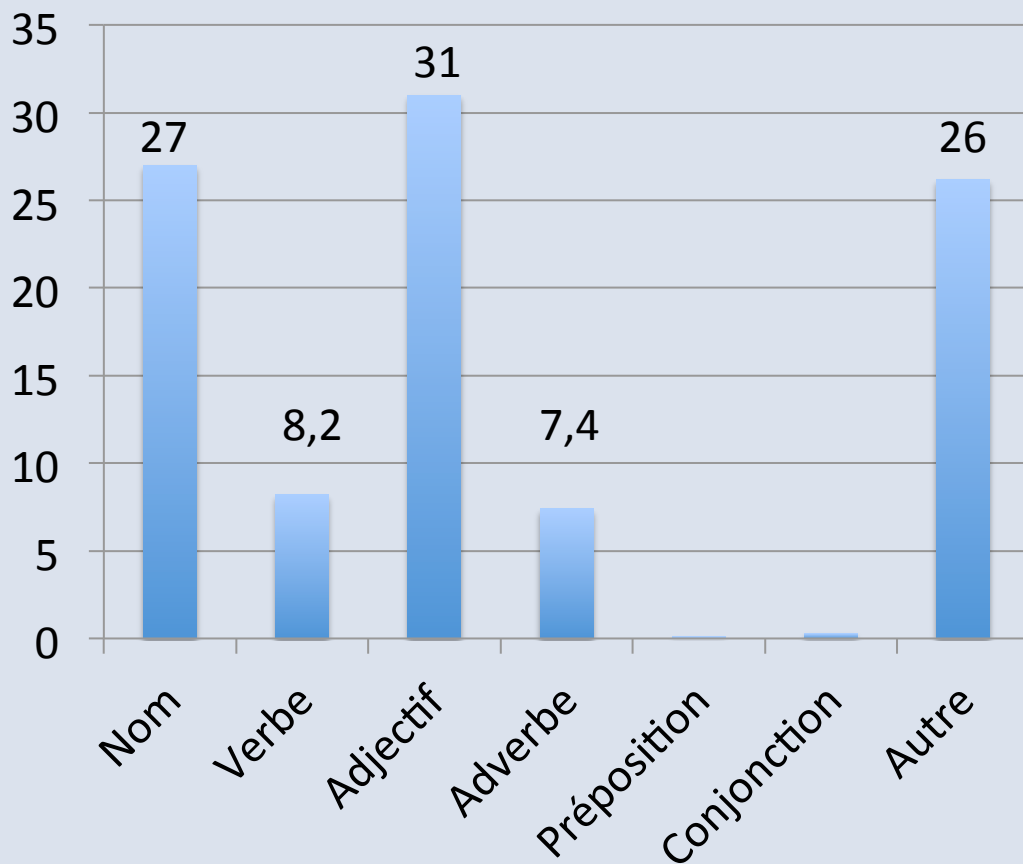
Total : 998 emprunts et transferts

V. : « Tk! Im out. Jvais puke tlm jsuis fatigee »

Adj. : « Et avec un peu dchance jrencontrerai du monde nice. »

Adv. : « Ps mad m'en c'est fucking trop bon. »

Autre : « Ps: tavais raison pr la neige! Check dehors! »

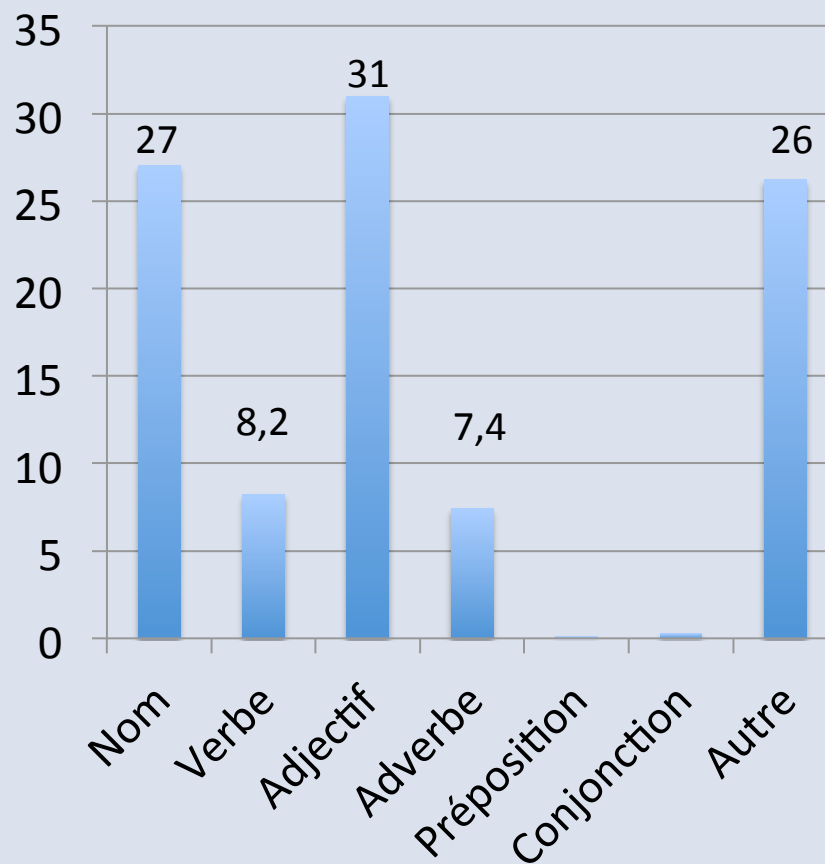


Sherbrooke, juin 2012

Send

Hiérarchie des emprunts : mots à contenu d'abord (Muysken 1981:62)
 N>V>Adj>Prep>Conj>Complém.>P.cli.
 >pron.q>pron.déictiques

Contraire à plusieurs études sur les communautés *bilingues* (Fox et Charbonneau 1995, King 2000) et de Saugera (2007) pour les *courriels*.
 Sur-représentativité dans Adj de « Ok » de « P.S. » et des interjections (« hey », « yeah! », « lol ») dans Autre



Messages

RÉSULTATS : ASSIMILATIONS PHONOLOGIQUES

Edit

Quelles sont les assimilations phonologiques dans les emprunts et transferts ?

- Très peu fréquentes : 7 dans le corpus !
 - « tcheqqq, tcheck, tivi »
 - note: « by thé way »... (correcteur ?)

Quasiment inexistantes !



Sherbrooke, juin 2012

Send

Plus fréquentes (surtout les verbes) : 31 dans le corpus

– « checker, feeler, chiller, skipper, skipperais, booker, watcher, buzzer, chekera, lifter, catcher, choker, skorer, deleter », etc.

– ex. : « Halak Canayen se fait skorer en pochelongation »

- surprise: aucun « -ou » pour les V (« booker »), « -é » pour les adjectifs, etc.?



Messages

RÉSULTATS : ASSIMILATIONS MORPHOLOGIQUES

Edit

NOMS

- « game, download, tatoo, chikita, chainsaw, parking, job, fun, chat, news, party, top, lift, downtown, cell, batch, boss, dealer, shift, show, ride, stocks, pools, freak, mail, casa, etc. »
- noms simples (mais « crazy carpet, offering memo », etc.), sans pluralisation
- certains très fréquents déjà (cell)
- critère d'assimilation phonol./morphol. sera peu productif pour déterminer l'intégration dans cette variété



Sherbrooke, juin 2012

Send

Messages

COMMENTAIRES EN CONCLUSION

Edit

- Transferts > phénomène le plus fréquent
- L'anglais > langue la plus utilisée dans ce corpus
- Adj>N>Autres (interj., sont les emprunts/transferts les plus fréquents ; certaines formes (lol, ok) surreprésentées
- La récolte est assez mince considérant les assimilations; il faudra creuser encore :
 - les emprunts sont-ils fait avec leur propres règles morphologiques?
 - quels liens avec les caractéristiques sociolinguistiques des ps?
- Est-ce que ça sera les mêmes résultats pour le corpus anglais ? (à voir...)



Sherbrooke, juin 2012

Send

Messages

Références

Edit

Cougnon, A. (2011). Tu te Prends Pour The King of The World? Language Contact in Text Messaging Context, *Language Contact in Times of Globalization* (38), 45-59.

Fairon, C., J.-R. Klein et S. Paumier (2007). *Le langage sms : Étude d'un corpus informatisé à partir de l'enquête «Faites don de vos sms à la science»*, Presses universitaires de Louvain • Cahiers du CENTAL.

Fox, C. et L. Charbonneau (1995). Le français en contact avec l'anglais : analyse des anglicismes dans le français parlé à Cohoes, New York. *Revue québécoise de linguistique théorique et appliquée*, 12, 37-63.

Guilbault, C. (2010). *Pratiques et fonctions des textos chez les francophones du Canada: Projet d'analyses et premières observations*, présentation ACFAS au symposium Colloque #307 - TEXTOS : dimensions culturelles, linguistiques et pragmatiques.

King, R. (2000). *The Lexical Basis of Grammatical Borrowing: A Prince Edward Island French Case Study*, Amsterdam: Benjamins.

Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. Dans A. Highfield et A. Valdman (Eds.), *Historicity and variation in Creole studies* (pp.52-78). Ann Arbor: Karoma.

Saugera, V. (2007). *Lexical Borrowing in a French-English Email Corpus: Integration of English Words in the Electronic Discourse of French Immigrants in America*, Bloomington : Indiana University.



Sherbrooke, juin 2012

Send