

---

# Je mangerais du junk food, let's go pour une grosse pizza all dress!

**Patrick Drouin<sup>1</sup>, Christian Guilbault<sup>2</sup>, Chantal Hélène Drapeau<sup>1</sup>,  
Philippe Langlais<sup>3</sup>**

<sup>1</sup>Observatoire de linguistique Sens-Texte, Université de Montréal

<sup>2</sup>Département de français, Université Simon Fraser

<sup>3</sup>Département d'informatique et de recherche opérationnelle, Université de Montréal

---

## Section 11 - Linguistique de contact

### Proposition de communication

Cette communication a pour but de présenter une analyse de l'alternance codique réalisée sur un corpus de textos (SMS ou messages textes) dans le cadre du projet *Texto4Science*<sup>1</sup>. Ce projet, qui se greffe au projet international *sms4science*<sup>2</sup>, a pour objectif de mettre sur pied une banque de textos pouvant servir de point de départ à des analyses linguistiques sur des bases scientifiques de ce type d'énoncés. En effet la qualité de la langue des textos fait régulièrement l'objet de jugements de valeur qui ne sont basés que sur des faits anecdotiques.

Lancé en octobre 2009, le projet *Texto4Science* a reçu ses premiers textos au mois de novembre 2009 suite à des appels auprès du grand public. Nous demandions aux participants de nous faire parvenir les messages dont ils étaient les auteurs à un numéro court (202202) spécialement mis en place pour l'étude. Au total, environ 7 300 messages ont été transmis par un peu moins de 300 participants ayant accepté de remplir un questionnaire sociolinguistique. Les textos ont été soumis à une étape entièrement manuelle d'anonymisation et d'annotation qui aura généré près de 35 000 annotations. Une balise spéciale

---

<sup>1</sup><http://www.texto4science.ca>

<sup>2</sup><http://www.sms4science.org>

consacrée au passage d'une langue à une autre nous permet d'identifier les données utilisées pour cette analyse. Au total, nous avons identifié 1 323 instances d'alternance, un message pouvant comporter une ou plusieurs instances.

Notre communication portera plus précisément sur la présence de l'alternance codique observée dans le corpus de français élaboré au Canada. Nous présenterons d'abord le corpus *Texto4Science* pour ensuite y décrire les phénomènes liés à l'alternance codique qui ont été recensés : l'emprunt, l'alternance de code (*code switching*), le mélange de code (*code mixing*) et le transfert. Dans cette dernière catégorie, nous examinerons les transferts portant sur les éléments lexicaux, les syntagmes ou des segments sans statut déterminé. Cette analyse permettra en premier lieu de décrire lesquels de ces phénomènes sont prisés par les auteurs, de même que les langues qu'ils choisissent pour effectuer ces mélanges de langue. La deuxième étape proposera une amorce de réflexion sur la reconnaissance automatique de ces phénomènes d'échanges de langue afin d'identifier des stratégies qui pourraient être exploitées dans d'éventuels outils informatiques.

## Références

- [1] Auer, Peter (1984) : *Bilingual Conversation*. John Benjamins : Amsterdam.
- [2] Auer, Peter (1998) : "Introduction : bilingual conversation revisited". In : id. (ed.) : *Code-Switching in Conversation : Language, Interaction and Identity*. London, Routledge : 1-24
- [3] 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.  
Deumert, Ana, /Masinyana, Sibabalwe O. (2008) : "Mobile language choices – The use of English and isiXhosa in text messages (SMS)". *English World-Wide* 29 : 117-147.
- [4] 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.
- [5] 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.
- [6] 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.

- [7] King, R. (2000). *The Lexical Basis of Grammatical Borrowing : A Prince Edward Island French Case Study*, Amsterdam : Benjamins.
- [8] 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.
- [9] Muysken, Peter (2000) : *Bilingual speech : A typology of code mixing*. Cambridge : Cambridge University Press.
- [10] Pekarek Doehler, Simona (2011) : ‘ Hallo! Voulez vous luncher avec moi hüt?’ Le ”code switching” dans la communication par SMS. *Linguistik Online*, 48 4/2011.
- [11] 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.