

Astel.be

Faites don de vos SMS à la science ! ;-)

Le 8 octobre 2004 par Grégoire Bourguignon (modifié le 8 octobre 2004)

Les informaticiens-linguistes de l'UCL tentent de récolter 30 000 SMS. L'objectif est scientifique, l'anonymat garanti, la participation gratuite... et en plus il y a des cadeaux à gagner ! La collecte de milliers de SMS fournira le matériau pour constituer un corpus indispensable à des études linguistiques de grande envergure et permettra de nouveaux développements dans le domaine du Traitement automatique du langage (TAL). Tous les utilisateurs sont donc invités à faire don de leurs SMS... à la science !

Nous le constatons chaque jour, le « langage SMS » est devenu un véritable phénomène de société. Il est aujourd'hui commun et répandu au point de franchir régulièrement les frontières du monde virtuel de la téléphonie mobile pour se retrouver dans la presse, la publicité et même sur des affiches de campagne électorale. Les chiffres sont également très révélateurs : ce sont près de 9.000.000 de SMS qui sont échangés quotidiennement en Belgique, et ce nombre va grandissant.

The logo of the University of Brussels (UCL) is displayed in a black square with the letters 'UCL' in white.

Les linguistes, sociologues, psychologues et autres spécialistes de la communication s'inquiètent ou se réjouissent de ce phénomène. Les plus craintifs voient d'un mauvais oeil la vague SMS en train de balayer grammaire et orthographe, repères déjà peu maîtrisés par les jeunes très friands de nouvelles technologies. Les plus optimistes y voient au contraire des jeux de langue propres à inciter à la fréquentation de l'écrit, un nouveau langage permettant des formes nouvelles d'expression et enrichissant les échanges entre individus.

Pour y voir plus clair et tenter d'objectiver l'analyse, le Centre de traitement automatique du langage - CENTAL - et le Centre d'étude des lexiques romans - CELEXROM - ont décidé de récolter plusieurs dizaines de milliers de SMS pour constituer un corpus de référence. Jusqu'au 15 décembre 2004, ils interpellent tout usager, habitué du SMS ou utilisateur exceptionnel, afin qu'il redirige une copie de ses messages vers un numéro gratuit, et ceci, quel que soit son opérateur et avec la garantie totale que l'anonymat sera respecté.

Attention, il s'agit de n'envoyer que des messages ayant véritablement été échangés dans le cadre normal des communications entre personnes (il s'agit donc d'envoyer des SMS qui sont dans la mémoire des GSM et non de créer des messages de toutes pièces).

Pour participer, il suffit d'envoyer le mot « UCL » par SMS au 3699 (numéro gratuit). Une fois cette inscription réalisée, les participants seront invités à rediriger leurs

messages (sans les modifier !) vers le même numéro. Des informations complémentaires peuvent également être obtenues par email en envoyant par SMS le code « UCL » suivi d'une adresse de courrier électronique. Ces opérations sont gratuites.

Tous les participants sont également invités à répondre, sur le site Web du projet (www.smspouurlascience.be), à quelques questions portant sur leur usage du SMS. Pour remercier ceux qui auront répondu, des cadeaux seront attribués chaque semaine par tirage au sort (GSM/MMS, cartes Pay&Go, etc.). L'opération est réalisée grâce au soutien de Proximus qui joue un rôle important en encadrant les aspects techniques de cette « récolte ».

L'objectif est purement scientifique. Le langage SMS se développe en fonction de différentes contraintes et libertés qui le particularisent : être intelligible en utilisant le moins de caractères possible, exprimer des émotions sans faire appel à des mots, faire abstraction des traditions orthographiques et grammaticales Les "codes" du SMS sont dès lors multiples. D'une part, des phénomènes inhérents au « langage branché » (troncations, néologismes, emprunts, expressions,...), de l'autre, des codes scripturaux, incontestablement la part la plus originale de ce nouveau langage (phonétisation de certaines graphies (ki, "qui" ; kom, "comme" ; alé, "aller"), valeur épellative des lettres (tu mM, "tu m'aimes" ; G, "j'ai"), de chiffres (a12C4, "à un de ces quatre" ; 2vient, "devient") et signes exploités graphiquement pour devenir « émoticons » ;-), etc. En outre, beaucoup d'utilisateurs n'appliquent pas ces procédés de façon systématique, ce qui corse encore l'affaire.

L'ingénierie linguistique, qui consiste à automatiser l'analyse du langage à l'aide d'outils informatiques, pourra prendre en compte les nouvelles formes propres à ce type d'écrit, mais elle doit, pour ce faire, acquérir une compréhension approfondie des mécanismes qui entrent en jeu. L'approche sociolinguistique, quant à elle, pourra mettre le profil de l'utilisateur en rapport avec le type de langage observé.

Déjà de premiers utilisateurs ont accepté de faire don à l'UCL de quelques SMS. Parmi ceux-ci, notons la présence de personnalités issues de divers milieux :

- du monde scientifique et universitaire (Bernard Coulie, Recteur de l'UCL ; André Goose, Secrétaire perpétuel honoraire de l'Académie royale de langue et de littérature françaises ; Marc Lits, Directeur de l'observatoire du récit médiatique de l'UCL ; Serge Flamé, Directeur du Centre audio-visuel de l'UCL) ;
- des chefs d'entreprises (François Schwennicke (Delvaux), Pierre Marcolini, Jacques Malpas (WIP) ;
- des journalistes (Jacques Mercier, Guy Lemaire, Cédric Wauthier) ;
- des écrivains (France Bastia, Présidente de l'Association des écrivains belges de langue française, Vincent Engel) ;

- des artistes (Marka, Plastic Bertrand, Jeff Bodart, l'Enfant Pavé de Starflam) ;
- du monde sportif (Pierre Cornia, Henri Depireux).

En résumé, la collecte de milliers de SMS fournira le matériau pour constituer un corpus indispensable à des études linguistiques de grande envergure et permettra de nouveaux développements dans le domaine du Traitement automatique du langage (TAL). Tous les utilisateurs sont donc invités à faire don de leurs SMS... à la science !