

Citer ce document :

Antoniadis, G. & Zampa, V. (2013). Manuel du Corpus *smsalps* (*cmr-smsalpes-tei-v1-manuel.pdf*) . In Antoniadis, G. Alpes4science : corpus de SMS réels dans les Alpes, smsalpes. Banque de corpus CoMeRe. Ortolang.fr : Nancy. [cmr-smsalpes-tei-v1]

Nom corpus : smsalpes

Auteurs : Georges Antoniadis et Virginie Zampa

Nombre de sms : 22117 (certains sms, envoyés via Orange, parmi ceux comportant plus de 162 caractères, sont tronqués au-delà du 162^e caractère).

La structure du corpus est évidente. « Id_Envoyeur » permet d'identifier les sms envoyés par la même personne

Contenu : Il s'agit de la version anonymisée des sms. L'anonymisation opérée concerne les :

NOM Nom propre

BLOG Adresse de blog

CB Coordonnées bancaires

URL Adresse de site Internet personnel

MAIL Adresse e-mail

ADR Adresse d'habitation personnelle

RECHARGE Numéros de recharge mobile

TEL Numéro de téléphone

COM Numéro de communication

CODE Codes divers, digicodes, etc.

PRENOM Prénom ou diminutif

SURNOM Surnoms personnels désignant un individu

Les parties anonymisées sont remplacées par : ***étiquette anonymisation_nb de caractères de la séquence anunymisée***

Lorsque l'anonymisation concerne un prénom, on y ajoute le n° d'ordre du lexique des prénoms du corpus : ***PRENOM_n° d'ordre_nb de caractères de la séquence anonymisée***