



H0688

CONSTRUÇÃO DE CORPORA DE MENSAGENS ELETRÔNICAS PARA SUA CONVERSÃO AUTOMÁTICA EM FALA

Monica Panigassi Vicentini (Bolsista SAE/UNICAMP) e Profa. Dra. Jussara Melo Vieira (Orientadora), Instituto de Estudos da Linguagem - IEL, UNICAMP

Mensagens eletrônicas como SMS e *e-mail* são escritas com a utilização de códigos, abreviaturas e símbolos, como os *emoticons*. Em sistemas computacionais que convertem texto escrito em fala por meio da tecnologia de Conversão de Texto em Fala, esses tipos de escrita precisam ser tratados de forma a serem reescritos por extenso para, então, serem analisados morfossintaticamente, além de passarem por demais processamentos lingüísticos e, finalmente, serem vocalizados. Assim, os objetivos principais desse trabalho são: compilar e tratar dois corpora de mensagens eletrônicas (SMS e *e-mail*). Os objetivos secundários são: contribuir para o aprimoramento de normalizadores de texto, gerar glossários de vocábulos desconhecidos, característicos dessas mensagens, colaborar com desenvolvimentos tecnológicos e com outros pesquisadores. Os corpora foram obtidos através da internet (site [<http://sites.google.com/site/corpusemailesms/>] e e-mail [corpusemail@ymail.com]). Após a compilação dos corpora, eles foram submetidos ao analisador lingüístico *Unitex* para extração dos códigos, abreviaturas e símbolos, que foram submetidos ao normalizador de texto (cedido pela empresa VOCALIZE) para serem escritos por extenso (expandidos) e, depois serem analisados morfossintaticamente (ferramentas: *MacMorpho* e da VOCALIZE).

Linguística de corpus - Linguística computacional - Mensagens eletrônicas