The Use of Corpora to Develop Language Teaching Materials

R. David Zorc

McNeil Technologies Language Research Center

Collaboration with Ateneo de Zamboanga University has led to a corpus of five million Chabakano words. This corpus was used in developing a Chabakano Reader (in progress). It will also assist in the development of a Chabakano Textbook, and will culminate in a Chabakano-English Dictionary. Data from the corpus (highest frequency items presented below) will be abstracted and organized to illustrate how such a textbook can be structured for optimum language learning.

The Language Research Center (LRC) of McNeil Technologies has been using corpora to develop bilingual dictionaries since the late 1980's. In 1992, a modestly sized Somali corpus of around 80,000 words, which had been developed to enrich our Somali-English Dictionary, was also utilized to develop the Somali Textbook. If one can impart to the student the highest frequency words (lexemes) and highest frequency grammatical constructions (functors), it is amazing what progress one can make in the language in a relatively short time. Verbs were parsed for tense and mood combinations and nouns for gender and plural affixation, and the resulting frequencies went directly into the organization of chapters. Although the book consisted of 50 chapters, by Chapter 21 the student was grammatically prepared to tackle real world texts. Similarly, Chabakano entries with the highest number of hits will be organized by grammatical class and lexical importance and given priority in the earliest chapters.

ya	127590	hinde	20977
na	107831	le	20658
el	96890	ba	20119
ta	80097	este	20113
yo	63138	o	20006
man	52247	ke (que)	18227
de	50718	pero	17186
lang	49048	kosa (cosa)	17056
si	45809	chene	16113
se	33835	ese	16083
del	33488	y	15501
kay (cay)	30621	kame	15207
tu	29555	tiene	13936
abla	29337	nuway	13496
maga	28062	puede	12864
kel (quel + kel)	27268	aki (aqui)	12666
akel (aquel + akel)	25492		
pa	25244		
di	25231		
mga	24754		
tamen	23674		
gat	23506		
sila	23416		
para	22379		