Teenagers' language in Estonia: the use of English in a newly compiled corpus

Virve Vihman, Kadri Koreinik, Liina Lindström, Aive Mandel, Maarja-Liisa Pilvik & Kristiina Praakli
University of Tartu

Teenagers are avid linguistic innovators, and have been shown to be the drivers of some linguistic changes (Eckert 2003; Androutsopoulos 2008; Tagliamonte 2016). Accessing samples of teen language poses a methodological challenge, yet it is crucial for a better understanding of youth language and its role in variation and change. Moreover, most of the research in this area has focussed on the language of L1 English-speaking youth, but their linguistic practices in English are influencing young people's language use much more broadly. English is a *lingua franca* used in entertainment and social media channels, and social media has grown to play a major role in young people's interactions. Hence, examining teenage language outside of the English-speaking world is important for understanding new language contact phenomena that come not from direct spoken communication between interlocutors, but rather from computer-mediated communication (CMC). In this talk, we ask how English affects the language use of young L1 Estonian-speaking youth in communication amongst themselves. We use case studies of lexical usage to identify the extent to which English has affected young people's usage: the rate and lexical diversity of code-switching in the corpora and the functions of the use of *nagu* 'like'.

The paper draws on data from a project investigating teenage language use in Estonia, in which two subcorpora, spoken language and digital messaging (DM), have been compiled through a participatory approach. The dual corpora can offer insights into contact with English as well as the language varieties used in spoken and DM registers. The project aims to identify variation by age, gender and geographic area. Participants (n = 131) range in age from 9 to 18 (mean=13.68, SD=2.40), are 71% female (94 female, 37 male) and represent four regions of Estonia. The language samples include 97 hours of spoken conversations and 204,210 words of DM texts.

We analyse the rate and lexical diversity of English words and phrases and the distribution of items across the dataset to investigate differences across ages and genders. We found that all major parts of speech are represented in the English-language lexical items, as well as categories such as taboo words, interjections and discourse markers. Although we do not find differences in amount of code-switching between genders, based on analysis of a sub-sample of the corpus, preliminary results suggest differences in the lexicon and the functions of code-switching.

We also report on a study of the functions of *nagu*, which is used in similar functions as English 'like', including the quotative 'be like'. The word is used overall much more frequently in the teen corpus than in the Balanced Corpus of Standard Written Estonian and the Phonetic Corpus of Spontaneous Speech. We trace its use across the age groups in our study and functions, to determine whether the use of *nagu* shows ongoing change. Preliminary results indicate an increase in the use of the word across our age groups.

References:

Androutsopoulos, J. 2008. Research on Youth-Language. In U. Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (eds.), *An International Handbook of the Science of Language and Society*, 496-1505. De Gruyter Mouton.
Eckert, Penelope 2003. Language and Adolescent Peer groups. *Journal of Language and Social Psychology*, 22(1): 112-118.
Tagliamonte, Sali A. 2016. *Teen talk: The language of adolescents.* Cambridge: Cambridge University Press.