

Variety of linguistic parameters of the Estonian text-types

Krista Kerge (Tallinn University) and Hille Pajupuu (Institute of the Estonian Language)

Our goal is to describe natural models of language use to be applied in language teaching and testing of both L1 (mother tongue curriculum, state maturation exams) and L2 (Estonian as a second language curriculum, state exams for students and adults). Arguing with the common in Estonia practice of equalizing natural and normative (standard) language, we consider natural anything widely acceptable in the field in our cultural context (cf. Hayes *et al.* 1980). This principal has to be considered by any high-stakes tests (Abu-Alhija 2007; Shohamy 2001).

For this study, text-types are register-dependable artefacts of different genres in certain field of communication (journalistic, academic, fiction, legal, administrative, personal, etc.; cf. Gee 2008). Using a variety of methods of text and discourse analysis, we have studied linguistic parameters of oral and written text-types, both monologue and dialogue: e.g. Kerge 2003 on syntactic complicity (corpus studies, Michos *et al.* 1996; fuzzy analysis, Meier 2002); Pajupuu *et al.* 2008 on lexical richness (Uber-index, Vermeer 2000); Kerge *et al.* on 2007 contextuality–formality depending on the balance of parts of speech (F-index, Heylighen and Dewaele 2002), etc.

Our results show reliable variety by field and genre. Complicity of written texts seems to relay e.g. on ideology. Formality is clearly growing from oral to written and from dialogue to monologue. Among written types of text studied, the legal ones are the most complex in their syntax; unexpectedly, the most formal texts in Estonian are these of administration (letters between public officers and letters from them to citizens). Lexical richness seems to be less important than vocabulary range. The most important factor of managing the topic is vocabulary range (e.g. 20% of all words belonging to thematic vocabulary of very low frequency).

In our presentation, also some obvious conclusions for linguistic research and applied linguistics (especially language testing) will be made.

- Abu-Alhija, F. N. (2007). Large-scale testing: benefits and pitfalls. *Studies in Educational Evaluation* 33, 50–68.
- Gee, J. P. (2008). *Social Linguistics and Literacies: Ideology on Discourses*. 3rd edition. London, New York: Routledge.
- Heylighen, F., Dewaele, J.-M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science* 7, 293–340.
- Kerge, K. (2003). *Keele variatiivsus ja mine-tuletus allkeelte süntaktilise keerukuse tegurina*. Tallinn: TPÜ Kirjastus.
- Kerge, K., Pajupuu, H. & Altrov, R. (2007). Tekst, kontekstuaalsus ja kultuur. *Keel ja Kirjandus* 8, 624–637.
- Meier, H. (2002). Olulisi aspekte tekstitüübivõrdluses. – R. Kasik (toim.), *Tekstid ja taustad. Artikleid tekstianalüüsist. Tartu Ülikooli eesti keele õppetooli toimetised* 23. Tartu: Tartu Ülikool, 101–114.
- Michos, S. E., Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (1996). Categorising Texts by Using a Three-Level Functional Style Description. *Artificial Intelligence: Methodology, Systems, Applications, Frontiers in Artificial Intelligence and Applications* 35, 191–198.
- Shohamy, E. (2001). *The power of Tests. A critical perspective on the use of language tests*. London: Longman.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. – *Language Testing* 17 (1), 65–83.