

Christiane Andersen
University of Gothenburg
Department of Languages and Literatures

Spoken language corpora in German and Swedish (L1) and acquisition of foreign language grammar

Several attempts were made to take special features of spoken language for building a foreign language grammar into consideration. Nevertheless, the authors mostly did not use electronic corpora when building such grammars. German and Swedish spoken language corpora have not been compared so far. We consider what can be achieved when we systematically exploit quantitative data from several spoken language corpora (L1 corpora in German and Swedish) in order to take them into account in teaching of foreign language grammars.

- Are particular grammatical categories more frequent in spoken language than in written language?
- Do particular structures and categories get more preference in spoken language than others?
- How can corpus analyses of spoken language be utilized for applied (i.e. practical) grammars?

We analyzed word form frequencies of digitalized transcripts of the *Mannheimer Datenbank Gesprochenes Deutsch* and the GSLC (*Göteborg Spoken Language Corpus*) of about 1 423 000 tokens. Both the German and Swedish data had been subdivided in subcorpora of communication types like 'formal meeting', 'informal conversation', 'political debate' etc., and quantitatively analyzed in contrast. It is reasonable to believe that the comparison of this substantial quantity of data reflects an authentic language use in German and Swedish. We here demonstrate that there is a gap between the high frequency of several categories and grammatical structures in our investigated corpora and their current description in foreign language grammars of German for Swedish learners. This concerns important structures such as several verbs, adverbs, subjunctions and modal particles, infinitive constructions and subordinate clauses.

References:

Pankow (Andersen), Christiane & Helena Pettersson (Nilsson): Auswertung der Leistung von zwei frei zugänglichen POS-Taggern für die Annotation von Korpora des gesprochenen Deutsch. In: *Göteborger Arbeitspapiere zur Sprachwissenschaft*. 2006. <http://www.sprak.gu.se/publikationer/gup-bibliografi/goteborger/>

Pankow (Andersen), Christiane: Korpora der gesprochenen Sprache und Fremdsprachengrammatik. In: Hall, Christopher & Kirsi Pakkanen-Kilpiä (Hrsg.). *Deutsche Sprache, deutsche Kultur und finnisch-deutsche Beziehungen*. Festschrift für Ahti Jäntti zum 65. Geburtstag. Peter Lang Frankfurt am Main 2007. 197-210.

Andersen, Christiane: Topologische Felder in einem Korpus der gesprochenen Sprache. Probleme zwischen theoretischem Modell und Annotation. In: *Göteborger Arbeitspapiere zur Sprachwissenschaft*. 2008. <http://www.sprak.gu.se/publikationer/gup-bibliografi/goteborger/>