Title of research proposal
Phonotactic Constraints for Speech Segmentation: The Case of Second Language Acquisition

Sponsor
Netherlands Organization for Scientific Research (NWO)

Grant type
VICI

Host institution
Utrecht Institute of Linguistics - OTS

Period of funding
2005-2010

Researchers
- René Kager (main applicant)
- Johannes Schliesser (postdoctoral researcher)
- Tom Lentz (PhD student, subproject 1)
- Natalie Boll-Avetisyan (PhD student, subproject 2)
- Frans Adriaans (PhD student, subproject 3)

Summary
This research programme aims at clarifying how listeners use their knowledge of sound structure (‘phonotactics’) for segmentation of connected speech and vocabulary acquisition. We focus on second language acquisition, where the speech segmentation system must cope with conflicting knowledge from two languages, native and target. It is known that native language phonotactic knowledge affects speech processing and long-term storage of words in the target language. It is also known that bilinguals use phonotactic knowledge of both languages in word recognition. However, no earlier research has explicitly characterized the phonotactic knowledge underlying second language learners’ growing ability to store words and process speech, and its acquisition.

The hypothesis is that phonotactic knowledge which supports speech processing and word learning is represented by a set of hierarchically ranked constraints, which affects perception of speech at the sub-lexical level of processing. Second language learners start from a copy of their native language constraint ranking, and acquire target language phonotactic knowledge by adding and re-ranking constraints, using generalized shapes of words in their developing vocabularies and feedback from successful segmentation of speech.

The programme features four innovative aspects. First, it models the unconscious phonotactic knowledge underlying speech segmentation by a set of hierarchically ranked constraints in Optimality Theory. Second, it features a constraint-based model of phonotactic learning, to be developed by testing it against natural language and artificial language learning by humans and machines. Third, it monitors second language development of phonotactic knowledge from its initial state until its final state (a bilingual state). Finally, it bridges insights and combines methods from four areas: psycholinguistics, language acquisition, phonological theory, and learnability theory.

The three subprojects address specific issues. First, how is segmentation of connected speech facilitated by knowledge of what are legal and illegal sound sequences (phonotactics) in the target language? Second, how does phonotactic knowledge of the target language assist word learning? Third, how is phonotactic knowledge acquired: do constraints and their ranking originate from the target language lexicon, from a distributional analysis of connected speech, or from both?

Languages involved in the programme include Dutch, Farsi, Hebrew, Mandarin Chinese, Russian, and Spanish.