ICT10-049 - Machine Learning Techniques for Modeling of Language Varieties

Zusammenfassung

Language varieties are gaining importance in man-machine interaction. Using them in speech based communication enables computer systems to reflect the socio-cultural identity of users. Current language technology cannot deliver on this, yet. There are a few synthetic voices with localized pronunciation, but language varieties are multi-faceted, involving deviations on various levels.
We will develop algorithms capable of capturing and reproducing all major idiosyncracies displayed by a language variety, be they syntactic, lexical or phonological. The task can be viewed as machine translation with some unique properties: the difficulty posed by the scarcity of available data is counterbalanced by the relative proximity between the varieties and the standard language. Our approach will therefore rely on optimal selection of data and smart use of linguistic knowledge. Standard German and Viennese varieties serve as a test bed for the realization and exploration of our techniques.

Keywords:
language technology, machine learning, language variety, machine translation

---

Principal Investigator:	Harald Trost
Institution:	Austrian Society for Cybernetic Studies
Weitere Projektpartner:innen:	Sylvia Moosmüller (Österreichische Akademie der Wissenschaften) Philipp Koehn (University of Edinburgh)

Status: Abgeschlossen (01.01.2011 - 28.02.2014)
GrantID: 10.47379/ICT10049

Weiterführende Links zu den beteiligten Personen und zum Projekt finden Sie unter
https://www.wwtf.at/funding/programmes/ict/ICT10-049/