

Afrikaans



Juri Ganitkevitch & Jonny Weese

Demographics & History

- ~6 million native speakers in South Africa & Namibia
- ~20 million speakers total
- Second-most prevalent language in South-African media
- Originated in 17th century Dutch

Linguistics

- West Germanic language family
- Closely related to Dutch: **mutually intelligible**
 - Orthographic simplifications
 - No gender, simple verb morphology
 - Influences from Malay, Portuguese, African languages, and South African English

Examples

Dutch	Afrikaans
ik <u>ben</u> , u <u>bent</u> , het <u>is</u> , ...	ek <u>is</u> , u <u>is</u> , dit <u>is</u> , ...
Ik wil dit <u>niet</u> doen.	Ek wil dit <u>nie</u> done <u>nie</u> .
provinc <u>ie</u> , polit <u>ie</u> , ...	provins <u>ie</u> , polis <u>ie</u> , ...

Afrikaans in MT

Dutch-Afrikaans

- Rule-based text transformation: morphology, orthography, compounds ('09, '11)

Afrikaans-English

- Phrase-based SMT: small parliamentary parallel corpora ('05)
- Google Translate: web data & **probably** rule-based repurposing of Dutch data ('09)

Af-En Parallel Data

URL

words



autshumato.sf.net

~439k



opus.lingfil.uu.se

~700k



af.wikipedia.org

~21k articles

References

1. Wikipedia, 2012
2. Rapid rule-based machine translation between Dutch and Afrikaans. P. Otte & F. Tyers, 2011
3. Processing Parallel Text Corpora for Three South African Language Pairs in the Autshumato Project. H. J. Groenewald & L. du Ploy, 2010
4. Rule-based Conversion of Closely-related Languages: A Dutch-to-Afrikaans Converter. G. van Huyssteen & S. Pilon, 2009
5. Rapid Development of an Afrikaans-English Speech-to-Speech Translator, H. Engelbrecht & T. Schultz, 2005
6. The OPUS corpus - parallel & free. J. Tiedemann & L. Nygaard, 2004

Dankie.

Language Presentations

- Teams of two
- Pick a **language** and a **date**
- Email us: juri@cs.jhu.edu
- First come, first served
- Due: 11:59pm on Sunday, 2/12