

# SLTU 2016 Panel Discussion: South Africa

---

*Charl van Heerden*

With Input from: Dr Febe de Wet (CSIR), Prof Marelie Davel (NWU), Prof Gerhard van Huyssteen (Viva, NWU), Prof Justus Roux (NWU), Dr Roald Eiselen (CTexT)

---

# Activities in under-resourced languages in SA

---

- Text processing
    - Spell checkers (CTeXT, NWU)
    - Computational morphological analysis (UNISA)
    - Machine translation (CTeXT)
  - Speech processing
    - ASR, STD (NWU, Meraka CSIR, Stellenbosch)
    - TTS (NWU, Meraka CSIR)
    - Spoken dialogue systems (Univ. Limpopo)
  - Resource collection
    - ASR: AST, Lwazi, SADE (tel. speech), NCHLT (broadband)
    - TTS: Lwazi, NCHLT++, Google NEMO
    - Text: NCHLT corpora
-

# Activities on under-resourced languages in SA (continued)

---

## ● RMA

- Manage & distribute resources developed with gvmnt funding (11 SA langs)
- Vision: becoming hub for resource management in Africa
- Problems:
  - Funding
  - acquisition of third party resources
  - proper licensing of resources (not planned well - data can't be distributed)

## ● DAC

- National Language Service of the Department of Arts and Culture responsible for elevation of status of previously marginalised languages.
- Since 2006, \$5m for HLT projects

## ● VIVA

- Not-for-profit creating, distributing & organizing digital resources for Afrikaans
  - (Android application, online dictionaries, etc)
  - Online assistance via full-time language expert
-

# Obstacles & Solutions

---

## Obstacles

- Lack of text resources
  - Translated material often very different from spoken text
  - Resource generation through usage often not feasible (small # spks)
  - Not commercially viable
- Funding

## Potential solutions

- Availability of content in languages
    - primary school content
    - Newspapers
    - Wikipedia
  - Google translate
    - With subsequent access to many other google apps
-