



Universiteit Leiden

AAM-LR **Automatic Annotation of Multi-modal** **Language Resources**

Lou Boves



Het probleem

- Grote hoeveelheden audio- en video-opnamen van veldwerk
 - In Nederland
 - In Frankrijk
 - SOROSORO Project
 - In Engeland
 -
- Annotatie kost orden van grootte meer tijd
 - En geld En menskracht
- Behoeftte aan technische ondersteuning om annotatie minder tijdrovend te maken



Mogelijke Oplossingen

- Gebruik “Diarization” technieken om automatisch aan te geven
 - Waar in een opname gesproken wordt
 - Waar fragmenten van dezelfde spreker zijn
 - Waar “bijzondere” dingen gebeuren
 - Lachen
 - Herkenbaar achtergrondlawaai
 -
- Gebruik automatische spraakherkenning om een globale fonetische transcriptie te maken
 - Voeg globale fonetische features toe
- Integratie in ELAN/ANNEX om zoeken mogelijk te maken



Universiteit Leiden

AAM-LR

Opzet van het project

- Maak de bouwblokken CLARIN-compatible
- Integreer de bouwblokken in ELAN/ANNEX
- Werk samen met onderzoekers in Nijmegen en Leiden om de gebruiksvriendelijkheid te maximaliseren
- Feedback naar CLARIN over de open issues die we onderweg tegenkomen