Towards Automatic Sign Language Annotation for the ELAN Tool
Philippe Dreuw and Hermann Ney
Human Language Technology and Pattern Recognition, RWTH Aachen University, Aachen, Germany

Introduction
- automatic annotation for video-based benchmark databases
- analysis and evaluation of:
  - linguistic problems
  - automatic sign language recognition
  - machine translation systems
- data-driven systems rely on
  - adequately sized corpora
  - rich annotation of the video data
- currently available sign language video databases
  - for linguistic purposes
  - gesture recognition using small vocabularies
  - video annotation is very time consuming
- experimental results for the ELAN annotation software

Annotations
- storing and processing sign language
- textual representation of the signs is needed
- several notation systems covering different linguistic aspects
- here: we focus on the so called gloss notation
- example: JOHN GIVE IX SOMETHING-ONE WOMAN BOOK

Automatic Sign Language Recognition
- similar to speech recognition:
  - temporal sequences of images
  - important features:
    - hand-shapes, facial expressions, lip-patterns
    - orientation and movement of the hands, arms or body
  - HMMs are used to compensate time and amplitude variations of the signers
- goal: find the model which best expresses the observation sequence
- recognition output gives information about:
  - word time boundaries
  - word confidences
- recognizer output file can be converted by the sclite tool from the NIST Scoring Toolkit
- creation of a tab-delimited text file for ELAN 3.4.0

Statistical Machine Translation of Sign Language
- processing of the intermediate representation of the recognized signs
- create a spoken language translation
- noisy-channel approach: the source language is interpreted as an encryption of the target language (decoding)
- system accounts for the different grammar and vocabulary of sign language
- goal: find the best decoding

Application Scenario – Speech-To-Speech System

Possible Automatic Annotations
- glosses and spotting
- sentence boundary detection
- body part descriptions: e.g. hands, face, eyes, and shoulders
- pronunciation detection
- speaker identification using face detection and recognition

Automatic Annotation Example imported to ELAN 3.4.0
- http://www.lat-mpi.eu/tools/elan/
- ELAN is a professional tool for the creation of complex annotations on video and audio resources
- since version 3.4.0 CSV import possible
- richness of gloss annotation can be defined by different user needs
  - allows for easy searching of high confident transcriptions
  - fast manual correction of annotations with low confidences

Human Language Technology and Pattern Recognition http://www-i6.informatik.rwth-aachen.de/~dreuw/ dreuw@cs.rwth-aachen.de