“Non-tokens”
When Tokens Should not Count as Evidence of Sign Use

Gabriele Langer, Thomas Hanke, Susanne König, Reiner Konrad
University of Hamburg, Institute of German Sign Language and Communication of the Deaf

Introduction
- Lemmatised corpora consist of tokens as instantiations of signs (types).
- Tokens usually count as evidence of the signs use.
- Corpus may contain instances of sign use that do not reflect the sign use of the person uttering them:
  - particularly citing of signs in metalinguistic discussions
  - malformed signs and slips of the hand
  - copying/parroting of signs of the interlocutors or from stimulus
- Such “non-tokens” should be labelled accordingly to be able to search for them specifically or to exclude them from analysis as suitable.

Examples from DGS Corpus
- Multi-purpose language resource for DGS
  - 330 informants from 13 regions
  - Balanced for age, gender, region
  - 20 elicitation tasks among others: free conversation, experiences as being deaf, metalinguistic discussion of young vs. old signs
  - Over 1,154 hrs. of footage
  - About 898 hrs. of filmed signing with 3.5 million tokens
  - Lemmatisation:
    - Running text (without gaps)
    - Still in progress, nearly 50 hrs.lemmatised (May 2016).

Examples from DGS-Feedback
- Answers and comments
  - Collected via online questionnaire on the use of DGS signs
  - Recorded via webcam/recording device
  - Directly referring to the stimuli (sign form, mouthing, meanings)
  - Highly metalinguistic content
  - Contains many instances of various kinds of “non-tokens”.

Consequences for Annotation
- The kind of “non-token” is a valuable piece of information that needs to be made accessible by some kind of annotation.
- In order to make annotations consistent, these annotations need to be reviewed and groups of “non-tokens” may be refined.

Consequences for Analysis
- Making decisions/interpretations of data explicit: Which tokens count as evidence and which do not (“non-tokens”)?
- Retraining (excluding/including) these tokens for analyses like:
  - Frequency count (count as proof)
  - Count of different signers using a certain sign
- Research that takes metadata of individual signers such as age, sex, and region into account.

Examples from DGS Corpus
- Task: Free Conversation

Kinds of “Non-tokens”
1. Metalinguistic reference to existing signs
   a. DGS sign also used by the signer
   b. Existing DGS sign that is normally not used by the signer
2. Un-cases
   a. Non-sign: negation of this particular sign
   b. Non-form: negation of this particular execution of form
   c. Non-use: negation of use with regard to a particular meaning
3. Copy of stimulus
   a. Reference to stimulus by copy of form
4. Non-signs (of SL in concern)
   a. Reference to gestures of hearing persons
   b. False start and aborted articulation
5. Malformed or deviated sign forms
   a. False start and aborted articulation
   b. Slip of the hand
   c. Trying out or searching for signs
   d. Adapted articulation due to limited view of webcam/recording device