



SIGNSPEAK

Scientific understanding and vision-based technological development for continuous sign language recognition and translation

Grant Agreement Number 231424

**Small or medium-scale focused research project (STREP) FP7-ICT-2007-3.
Cognitive Systems, Interaction, Robotics**

Project start date: 1 April 2009
Project duration: 36 months

Deliverable D.1.3: Enriched annotation documents for the existing CNGT corpus containing sentence boundaries units.

Dissemination Level: **Public**

Date: 23 May 2010
Version: 2
Author: Onno Crasborn
Centre for Language Studies
Radboud University Nijmegen

The compressed file **CNGT_EAF_31mar2010.zip** contains the full set of 2375 annotation files of the Corpus NGT, organized in folders per 100 files. The documents contain the following tiers:

GlosL S1	Gloss annotations (present in 200 files)
GlosR S1	
GlosL S2	
GlosR S2	
SignSpeakZin S1	Sentence boundary annotations (present in 43 files)
SignSpeakZin S2	
SS_zinsopmerking	
SS_Head S1	Head movement annotations (present in 2 files)
SS_Head S2	
SS_Mouth S1	Mouth action annotations (see below)
SS_Mouth S2	

With respect to the initial release of the corpus, the documents contain extra tiers (all of the above except the Glos tiers), with better and more glosses, sentence translations, and sentence boundary units.

A mouth tier per signer was added in March 2010 as this was a recurrent wish expressed by the annotators during the glossing process. Especially in cases where the mouth articulation contributes specifying information with respect to the semantics of the manual sign, this makes the glosses more consistent as annotators do not try to change the semantics of the Dutch gloss anymore. The mouth actions will not be systematically annotated, however, as that would be highly time-consuming process and because it is not required by the other Work Packages.

An overview of the present state of annotation is listed in the following table.

	Start of project	M12
Gloss annotations	64.288	91.194
Signs		
tokens	49.463	70.061
types	5.191	7.127
singletons	2.919	4.184
Sentence boundary annotations with translations	0	1276
Sentence boundary annotations	0	758