

## Digging into Signs: a WEIRD community

Onno Crasborn

Round 3 Netherlands PI, Digging into Signs: Developing Standard Annotation Practices For Cross-Linguistic Quantitative Analysis Of Sign Language Data

## Digging into Signs: developing standard annotation practices for cross-linguistic, quantitative analysis of sign language data

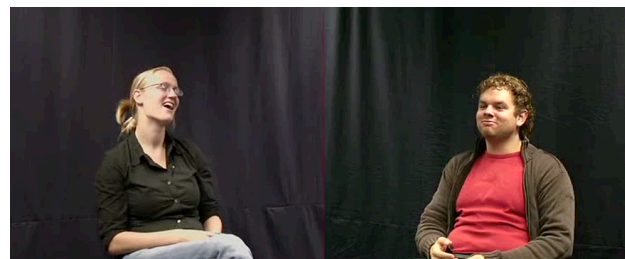


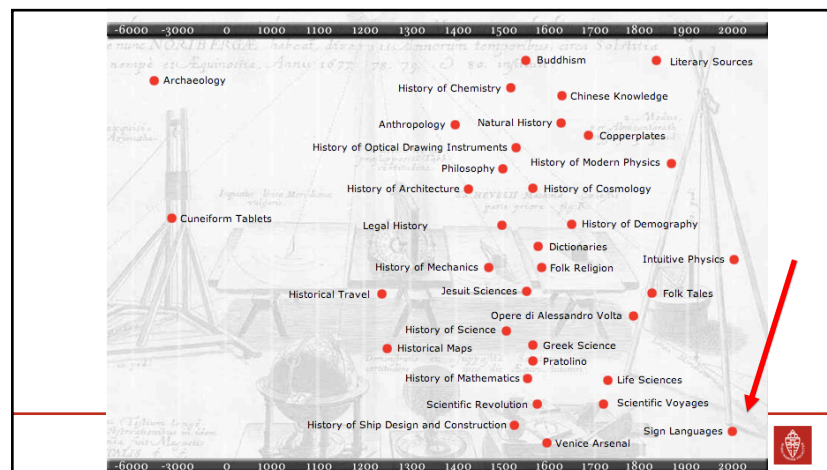
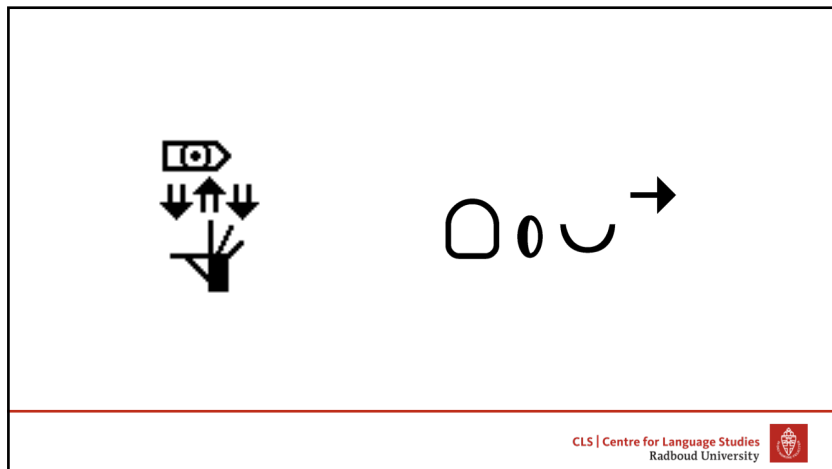
## Digging into Signs: developing standard annotation practices for cross-linguistic, quantitative analysis of sign language data




Radboud University Nijmegen 

## Digging into Signs: developing standard annotation practices for cross-linguistic, quantitative analysis of sign language data








2003-2004

### Sign Linguistics Corpora Network

2008-2010

### Various workshops related to national sign corpus projects


Hamburg 2009, London 2009, Nijmegen 2011, Nijmegen 2012, Hamburg 2014, London 2015



2014-2015


**LREC workshop series** 'The representation and processing of sign languages'



2002   2004   2006   2008   2010   2012   2014   2016   2018   2020




CLS | Centre for Language Studies  
Radboud University

Sign language corpus community > standardisation





CLS | Centre for Language Studies  
Radboud University










CLS | Centre for Language Studies  
Radboud University

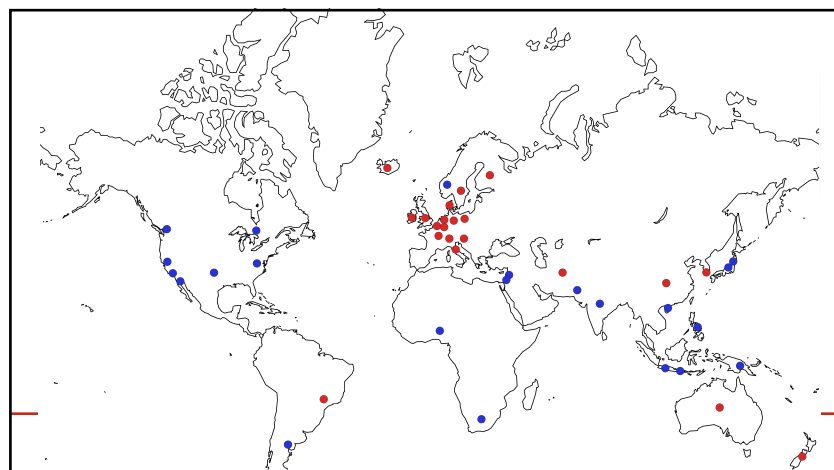
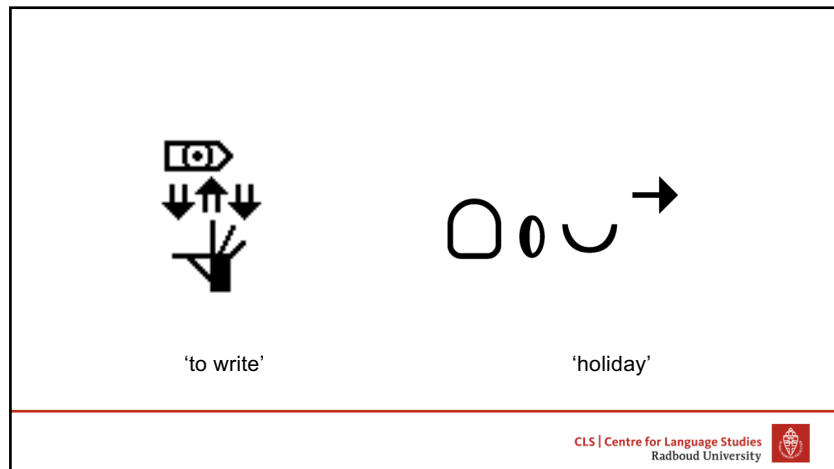
Modern linguistic corpus

	Spoken/text corpora like the BNC	Sign language corpora
Large collection of spoken, written or signed language data, with associated metadata		
Maximally representative (as far as possible) of the language and its users		
Machine-readable form		(not yet!)

Machine-readability requires annotation!



CLS | Centre for Language Studies  
Radboud University



ULRIKA KLOMP

### Conditional Clauses in Sign Language of the Netherlands: A Corpus-Based Study

**Abstract**  
Conditional clauses are underresearched in sign languages, and the research that has been done has mostly employed elicited data. The current study provides a thorough description of conditional clauses in Sign Language of the Netherlands (NGT) based on naturalistic corpus data. Similar to conditionals in other sign languages, conditionals in NGT can be introduced with a manual marker, although it is also possible to only use nonmanual markers such as raised eyebrows, head movement, and head tilt. Different, however, is the striking amount of variation that we found with respect to both manual and nonmanual marking. This might be due to differing methodologies, but we discuss several other potential explanations as well. We further provide evidence that nonmanual markers are less frequent in conditionals with manual markers than without. In addition, we offer an—albeit brief—description of peripheral conditionals in NGT, a conditional type that has not yet been investigated thoroughly for any sign language.

CONDITIONAL CLAUSES are intriguing and interesting to study from multiple perspectives. In its classical form, the conditional sentence type is often formalized as the simple logical formula “if p, q”

### Towards Automated Sign Language to Written Language Translation

Boris Mocalov · Helen Hastie · and Graham Turner  
*Helen Hastie, Graham Turner*

**Keywords:** machine learning, machine translation, sign language

**1 Introduction**  
Very few attempts have been made so far to create end-to-end systems for continuous sign language translation [2]. Translation of sign languages to written form is a significant step towards bridging the gap between deaf and hearing communities as there is no uniform standard for writing sign languages established so far [3]. Our work so far has utilized two datasets. The first is the NGT Corpus<sup>1</sup> for training stacked LSTMs to classify isolated signs and the second is the The ISL Corpus<sup>2</sup> for sign language modelling. In this paper we show the work done so far that will become integrated into a system that will translate from The British Sign Language into English.

**2 Methodology**

(a) Feature extraction example with opposite shoulder body parts, such as shoulders, phalanges, etc.  
(b) T1-T3 are contours of hand gestures from Figure a) for 3 frames, H1 and H2 are height and width of the minimum bounding box.  
(Figure a) (Figure b)

**3 Results**

Method	BLEU
Stacked LSTM	0.26
Stacked LSTM + Transfer Learning	0.27
Stacked LSTM + Transfer Learning + Feature Extraction	0.30

(Table a) Segmented Video Classification Results (Table b) Transfer Learning Results

**Transfer Learning for Language Modelling:** Stacked LSTMs with adaptive learning rate are used for sign language modelling. Since the data available from The ISL corpus is not sufficiently large for statistical language modelling, the model has been pre-trained on the Penn Treebank (PTB) dataset and then fine-tuned on the sentences, extracted from The ISL corpus.

**Segmentation of Signing Video:** The epochwise detection, extracted from a single video of continuous signing, returns start and end times of the epochwise interval, which is then compared to the annotated ground truth. The method achieves F-score of 0.825.

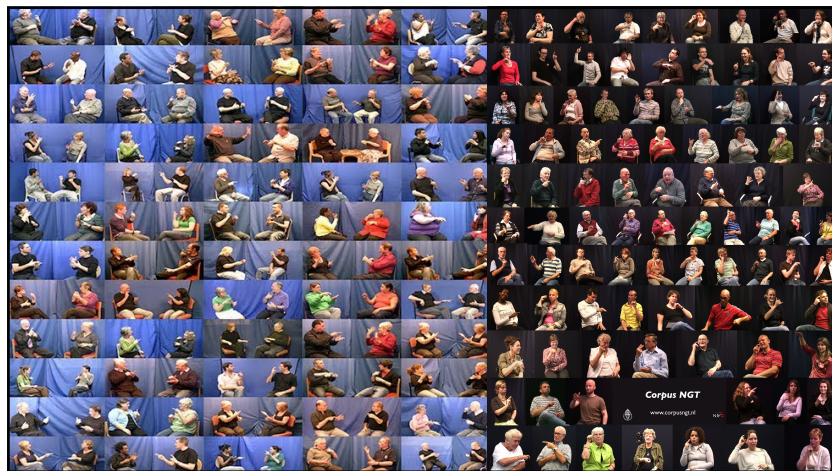
**Segmented Video Classification:** Table a) shows classification results with stacked LSTMs applied on extracted features from opposite library. The accuracy drops down as the number of classes increases.

**Transfer Learning for Language Modelling:** Table b) shows perplexity of the language model, trained on Penn Treebank (PTB) dataset and fine-tuned on the ISL corpus. The transfer learning method gives more than double improvement in perplexity over the baseline.

**Feature Extraction from Video:** Standard camera view features are extracted with the opposite library [1]. On Figure a), the features identify body parts.

**Segmentation of Signing Video:** Hand gestures from Figure a) are used to find trajectory using centroids all hand points for the period of 5 frames (T1-T5).

CLS | Centre for Language Studies  
Radboud University



**Conclusions**

**Research community**

1. Small-scale workshops and Digging-type grants have been highly instrumental for getting WEIRD researchers together, making them converge on methods
2. Expansion to the rest of the world highly urgent for linguistics
3. Small and fragile research groups in each country. Institutional embedding and support varies
4. A long way to go for this community (sign language linguistics)

**Language community**

1. Language communities are important, we study people
2. Ways to give back to these communities should be integrally funded, and not a 'valorization afterthought'

CLS | Centre for Language Studies  
Radboud University

**Outlook: no technology without data**

CLS | Centre for Language Studies  
Radboud University

**How does this apply to today's project presentations?**

**DIGGING INTO EARLY COLONIAL MEXICO**  
Patricia Mulcaugh, Diego Jiménez, Rodolfo and Bruno Wolffs

**Mapping Manuscript Migrations**

**ISEBEL**  
Intelligent Search Engine for Digital Legends

**Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories, 1840-1914**

**Responsible Terrorism Coverage**  
Scott Althaus, University of Illinois at Urbana-Champaign USA  
Hermann Wacker, University of Marburg, GERMANY  
Wouter van Amelselt, Free University of Amsterdam, NETHERLANDS

CLS | Centre for Language Studies  
Radboud University

How does this apply to today's project presentations?

The screenshot shows a slide with a colorful geometric pattern on the left. The main content area is divided into sections: 'Mapping Manuscript Migrations' with a table, 'ISEBEL Intelligent Search Engine for Linked Legends', and 'Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914'. Logos for 'Responsible Terrorism Coverage' and 'CLSV Centre for Language Studies Radboud University' are visible at the bottom.

**Great teams!**

**Communities?**

The screenshot shows a slide with a colorful geometric pattern on the left. The main content area is divided into sections: 'Mapping Manuscript Migrations' with a table, 'ISEBEL Intelligent Search Engine for Linked Legends', and 'Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914'. Logos for 'Responsible Terrorism Coverage' and 'CLSV Centre for Language Studies Radboud University' are visible at the bottom.

**And here?**