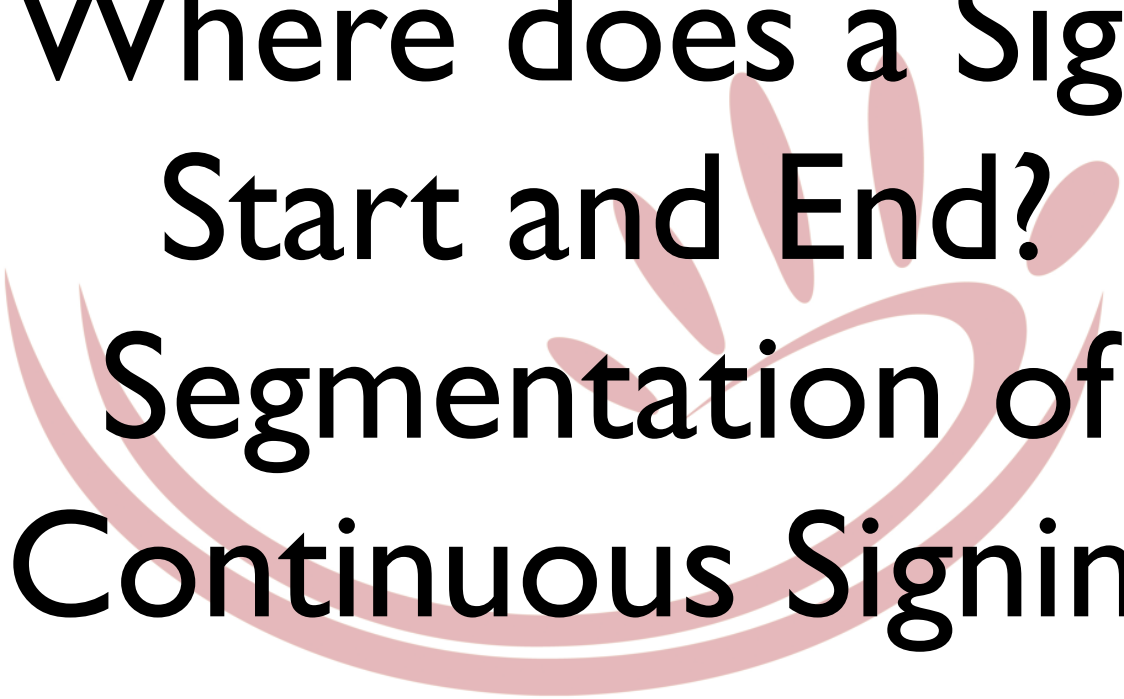


# Where does a Sign Start and End? Segmentation of Continuous Signing



LREC May 2012, Istanbul

Thomas Hanke, Silke Matthes, Anja Regen, Satu Worseck  
& Patricia Barbeito Rey-Geißler

AKADEMIE DER  
WISSENSCHAFTEN  
IN HAMBURG



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG



# Segmentation

- Tokenising and Lemmatising are the basic steps to annotate signing:
  - Where does a tag start?
  - Where does it end?
  - What is its “label”?



# Two approaches to segmentation

- Wide segmentation
  - In fluent signing, there are no gaps between two signs
  - (e.g. T. Johnston)
- Narrow segmentation
  - Identifies the “nucleus” of the sign, not the transitions into and out of the sign



# Wide segmentation

- Identifies borders between signs
- Less work: Only one cut between two signs, not two.
- More in line with speech segmentation
- No ambiguity between transition and pause
  - tag = signing, no tag = pause

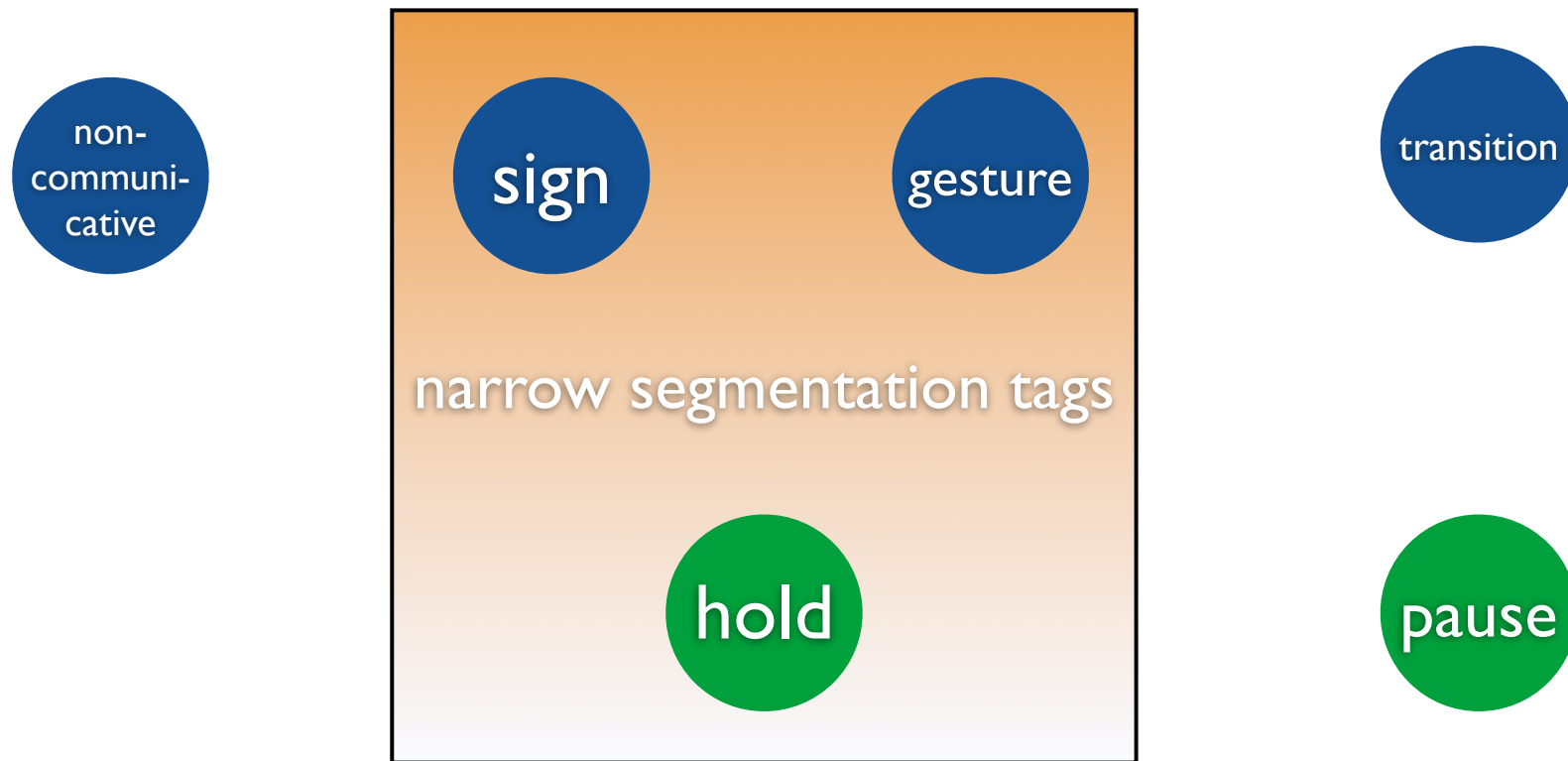


# Narrow segmentation

- Variation between tokens is much smaller
- The segment is exactly what is described by (HamNoSys) token form description
- Recognition: Learn the relevant bit
- Animation: More closely follow the signer's dynamics
- More or less compatible with Johnson & Liddell (2011) phonetic analysis



# Ideally, image processing resolves the ambiguity



# Where does a sign start?

- Jouison 1990: Handshape, location, orientation do not establish simultaneously, but there is a certain pattern:
  - $\text{handshape} < \text{orientation} < \text{location}$



# The general idea

- The sign starts where “all pieces are in place”.
- It ends right before the pieces are in transition to the next sign.





# Insufficient inter-annotator agreement

- “Difficult cases”:
  - Make the rules explicit
- What is the sign?
  - Uses pre-existing knowledge about the sign
  - Mixes top-down and bottom-up



# Making intuition explicit is not always easy

- For signs with an HMH structure in the sense of Liddell & Johnson (1989) the sign starts at the beginning of the initial hold, i.e. as soon as its handshape has been formed and is placed in the right orientation at the starting location of the sign.



# Shared H

- In cases where two signs share a hold (i.e. one sign ends in a hold, and by chance the next sign is beginning with a hold at exactly the same location with the same handshape and orientation), cut the hold in the middle.
- Here it is obvious that there cannot be a gap between the two tags.)



$$HM + MH = HM \times MH$$

- In case of signs without a specific starting location, look for a discontinuity in the movement (e.g. a sudden change in direction) between the end of the previous sign and the end of the target sign.



$$HM + MH = HMH$$

- In case of a continuous movement from the beginning of a sign to the end of the next sign (e.g. DENKEN DU in lax signing), cut in the middle/at the peak of that movement.
- This is then also the end of the previous sign, i.e. there is no gap in-between the two signs.



# Result

- No substantial improvement in inter-transcriber agreement



# Is our decision tree compatible with Johnson/Liddell?

- Johnson/Liddell 2011:
- sign starts with a (video) frame where all parameters are in rest (not blurred or “fuzzy”)

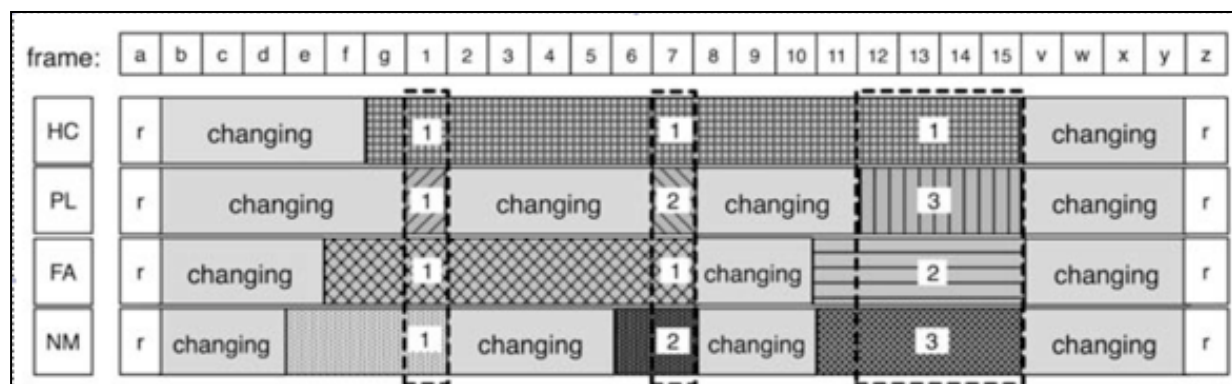


FIGURE 4. The three aligned and definitive postures of the sign CHICAGO and their sequential organization with respect to the four major structural components of the sign.



# Identifying Ps

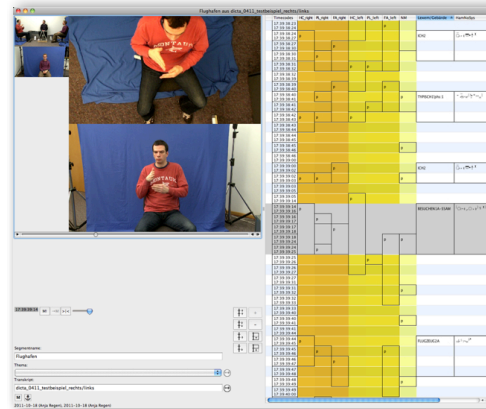
- With 25fps: Often impossible
- With 50fps: Often ok, but many cases where there aren't any all-in-rest frames
- Method depends on technical equipment
  - Does it reflect phonetic reality?






# Experimental transcription

- Based on 3 time-aligned camera views
- 960x540p50



Flughafen aus dicta\_0411\_testbeispiel\_rechts/links



Timecodes	HC_right	PL_right	FA_right	HC_left	PL_left	FA_left	NM	Lexem/Gebärde	HandSys
17:39:38:23									
17:39:38:24									
17:39:38:24									
17:39:38:27	p							ICH2	$\hat{O} = \hat{O} + \hat{O}$
17:39:38:27									
17:39:38:30									
17:39:38:30	p								
17:39:38:31									
17:39:38:31									
17:39:38:32				p	p				
17:39:38:32									
17:39:38:32									
17:39:38:39									
17:39:38:39									
17:39:38:40				p					
17:39:38:40									
17:39:38:41									
17:39:38:41									
17:39:38:42									
17:39:38:42									
17:39:38:42	p	p		p					
17:39:38:43									
17:39:38:43									
17:39:38:44									
17:39:38:44									
17:39:38:45									
17:39:38:45									
17:39:38:46							p		
17:39:38:46									
17:39:39:00									
17:39:39:00									
17:39:39:02				p					
17:39:39:02									
17:39:39:02	p	p						ICH2	$\hat{O} = \hat{O} + \hat{O}$
17:39:39:03									
17:39:39:05									
17:39:39:05									
17:39:39:14				p					
17:39:39:14									
17:39:39:14	p							BESUCHEN1A-SSAM	$\hat{O} = \hat{O} + \hat{O} + \hat{O}$
17:39:39:16									
17:39:39:17									
17:39:39:17									
17:39:39:18									
17:39:39:24							p		
17:39:39:24									
17:39:39:25									
17:39:39:25									
17:39:39:26									
17:39:39:26									
17:39:39:27									
17:39:39:27									
17:39:39:31									
17:39:39:31									
17:39:39:32							p		
17:39:39:32									
17:39:39:32									
17:39:39:33									
17:39:39:33									
17:39:39:40									
17:39:39:40									
17:39:39:41									
17:39:39:41							p		
17:39:39:44									
17:39:39:44									
17:39:39:45	p							FLUGZEUG2A	$\hat{O} = \hat{O} + \hat{O}$
17:39:39:45									
17:39:39:46									
17:39:39:46									
17:39:39:47									
17:39:39:47									
17:39:39:48									
17:39:39:48									
17:39:39:49							p		
17:39:39:49									
17:39:40:00									

17:39:39:14 M -M ><

Segmentname:  
Flughafen

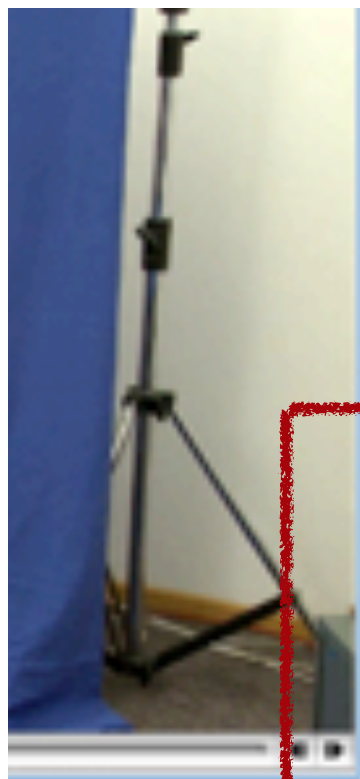
Thema:

Transkript:  
dicta\_0411\_testbeispiel\_rechts/links

M

2011-10-18 (Anja Regen), 2011-10-18 (Anja Regen)





17:39:38:44							
17:39:38:44							
17:39:38:45							
17:39:38:45							
17:39:38:46					p		
17:39:38:46							
17:39:39:00							
17:39:39:00			p			IOHQ	0-0-0-0
17:39:39:02							
17:39:39:02	p	p			p		
17:39:39:03							
17:39:39:05							
17:39:39:05							
17:39:39:14			p				
17:39:39:14	p					BESUCHENIA-SSAM	0-0-0-0
17:39:39:16							
17:39:39:16		p					
17:39:39:17							
17:39:39:17			p				
17:39:39:18							
17:39:39:18					p	p	
17:39:39:24							
17:39:39:24		p					
17:39:39:25							
17:39:39:25					p		
17:39:39:26							
17:39:39:26							
17:39:39:27							
17:39:39:27							
17:39:39:31							
17:39:39:31							
17:39:39:32					p		
17:39:39:32							
17:39:39:33							
17:39:39:33							
17:39:39:40							
17:39:39:40					p		
17:39:39:41							
17:39:39:41							
17:39:39:44							
17:39:39:44	p					FLUGZEUGZA	0-0-0-0
17:39:39:45							

# Side effect

- In some cases, annotators revised their segmentation decisions on the basis of the 50fps data
- Is then our method also dependent on framerates?



# Is our own segmentation also subject to framerate?

- Experiment:
  - Annotate the same performance twice, based on two different camera recordings:
    - 50fps standard half-HD
    - 500fps (stereo) HD



# 500fps capture



# Is our own segmentation also subject to framerate?

- Segmentation requires movement reconstruction from the video frame images as the criteria are in the motion domain, not in the image domain.
- Segmentation stable beyond 50fps, but interesting details nevertheless!



# Conclusions for segmentation

- With 25fps,  $\pm 1$  frame has to be tolerated.
- Except for that, segmentation is well defined.
- Corpus annotation now completely switched to 50fps.
- Automatic segmentation cannot be expected to outperform human annotators.





# The Bonus Material





# Thank you for your attention!



DGS-KORPUS

AKADEMIE DER  
WISSENSCHAFTEN  
IN HAMBURG



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG



This publication has been produced in the context of the joint research funding (DGS Corpus) of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

The research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135 (Dicta-Sign).

