

# Towards more Linguistic Modelling in a Sign Language Transcription Environment

Thomas Hanke, Susanne König,  
Reiner Konrad, Gabriele Langer  
University of Hamburg

Institute of German Sign Language...

<mailto:dgs-korpus@sign-lang.uni-hamburg.de>

<http://www.sign-lang.uni-hamburg.de/dgs-korpus/>

UH



# iLex– Why another Transcription Environment?

- Database approach instead of tons of XML files
- Not general-purpose, but specific to sign language
  - Specific support to compensate the major difference between sign languages and many spoken languages: The lack of a writing system with a standard orthography
  - The essential step is lemmatisation, i.e. identification of the type a token is related to
  - Transcription and growing the lexical database is one single process

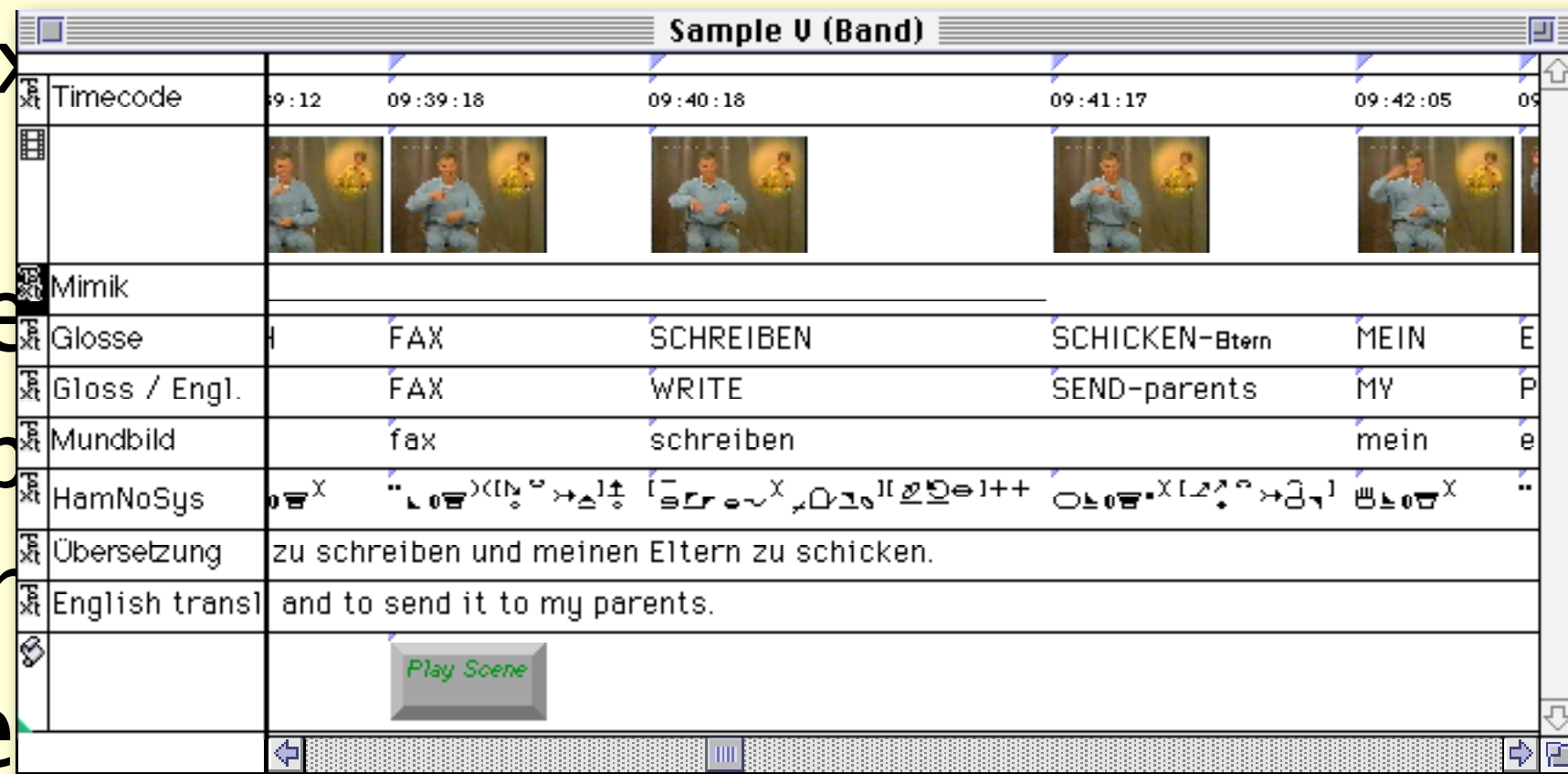


# Building on Past Experience

- 1994 syncWRITER for signed discourse transcription
  - multi-tier text annotation to points in time of a video
- 1996 GlossLexer
  - Lexical database
  - Transcription limited to short phrases
- Project started in 2000
  - Lexical database & flexible transcription

# Building on Past Experience

- 1994 syncWRITER for signed discourse transcription
  - multi-tier text of a video
- 1996 GlossLex
  - Lexical database
  - Transcription
- Project started
  - Lexical database & flexible transcription



The screenshot shows the 'Sample U (Band)' window in syncWRITER. It displays a multi-tier transcription of a video scene. The top row shows timecodes for different segments. Below are video thumbnails for each segment. The main table contains the following rows:

Timecode	09:12	09:39:18	09:40:18	09:41:17	09:42:05	09:42:05
Mimik						
Glosse		FAX	SCHREIBEN	SCHICKEN-Eltern	MEIN	E
Gloss / Engl.		FAX	WRITE	SEND-parents	MY	P
Mundbild		fax	schreiben		mein	e
HamNoSys		ⓧ	ⓧ	ⓧ	ⓧ	ⓧ
Übersetzung	zu schreiben und meinen Eltern zu schicken.					
English transl	and to send it to my parents.					

At the bottom of the window, there is a 'Play Scene' button and a progress bar.

# Building on Past Experience

The screenshot displays a software interface with the following components:

- Top Bar:** Apple logo, menu items 'Ablage', 'Bearbeiten', 'Daten', time '21:03', and system icons.
- TLex/Datenbank2 (Left Panel):**
  - Icons of a sun, a lightbulb, and a moon.
  - Fields: 'Transkribend: Thomas Hanke', 'Bandbezeichnung: 14', 'Interviewte Person: Apel', 'Händigkeit: Rechts'.
  - 'Lexikon' section with buttons 'Neu', 'Alias', 'Löschen'.
  - Table with columns 'Glosse', 'HamNoSys', 'Verweis'.
- Glosse (Right Panel):**
  - Header: 'AUFHÄNGEN' with checkboxes for 'OK' and 'Thomas Hanke'.
  - Text: 'Mit, 26. Mär 1997'.
  - Table with columns: 'Begriff', 'Band', 'Interviewer', 'Timecode', 'HamNoSys'.
  - Buttons: 'Vorkommen anzeigen', 'Abbrechen', 'OK'.
- Vorkommen-Eingabe (Bottom Panel):**
  - Fields for '57 Bankhaken', 'TC-Beginn', 'TC-Ende', 'Thomas Hanke', 'Apel', and '26.03.1997'.
  - Table with columns: 'Glosse', 'Grundform', 'HamNoSys', 'Mundbild'.
  - Buttons: 'problematisch', 'Abbrechen', 'OK', 'OK & Weiter'.

discourse

nts in time

rases

cription

# iLex Development

- 2000–2008 inhouse tool for various projects
- 2009–2023 further development secured by means of the DGS corpus project
  - resources available to make the tool available to other research groups
    - 2 persons working on iLex
- 2009–2012 one of the target platforms of the Dicta–Sign EU project
  - video recognition to be integrated

# In-house Use of iLex

- DGS Corpus project
- Special terminology dictionary projects
- Individual research & exam paper work
  
- peak-time user count > 20
- 26700 transcripts (for 360 hours of video) in the database with some 700000 tags
- nothing gets lost on someone's harddisk
- central backup

# Transcription in iLex

- Building the corpus and (extending) the lexicon at the same time
- Type-token matching is the essential step
  - Relational database!



# iLex: Type-token matching is the essential step in transcription

The screenshot shows the iLex software interface. On the left, a video window displays a sign language interpreter in a red shirt. Below the video is a control panel with a play button, a progress bar, and various settings like 'Segment Name' (Aufgaben), 'Theme' (Integrationsamt), and 'Transcript' (Integrationsamt 1). On the right, a table titled 'Aufgaben from Integrationsamt 1' displays transcription data.

Timecodes	Komple...	Gebärde	HamNoS...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERL	⊖ ⊥ 0 <sup>l</sup> ⊕ ~	behindert			behindert
00:00:01:17 00:00:01:22							
00:00:01:22 00:00:02:16		MENSCH21	⊕ ~ 0 <sup>l</sup> ⊕ +	Mensch	+		mensch
00:00:02:16 00:00:03:05							
00:00:03:05 00:00:03:24		ARBEITEN1	⊖ ⊥ 0 <sup>l</sup> ⊕ ⊕	Arbeit	-		arbeit
00:00:03:24 00:00:04:03							
00:00:04:03 00:00:05:18		TOLL1	⊖ ⊕ 0 <sup>l</sup> ⊕ +	toll		AB zusamm	
00:00:05:18 00:00:06:15							
00:00:06:15 00:00:07:09	Integration	INTEGRATI	⊖ ⊕ ⊕ 0 <sup>l</sup> ⊕ ⊕	Integration	-		integration
00:00:07:09 00:00:08:13		AMT1A	⊖ ⊕ ⊕ 0 <sup>l</sup> ⊕ ⊕	Amt	-		
00:00:08:13 00:00:09:03							
00:00:09:03 00:00:10:24		HELFEN11	⊖ ⊕ ⊕ 0 <sup>l</sup> ⊕ ⊕	helfen			
00:00:10:24 00:00:11:19							
00:00:11:19 00:00:11:19							

2004-12-18 (Thomas Hanke) , 2008-04-14 (Thomas Hanke)

# iLex: Type-token matching is the essential step in transcription

The screenshot displays the iLex software interface. On the left, a video player shows a sign language interpreter. The main window is titled "Aufgaben from Integrationsamt 1" and contains a table with columns: Timecodes, Komple..., Gebärde, HamNoSys, Bedeutung, Pl., Mimik, and Mund. A pop-up window titled "ARBEITEN1A" is open, showing a detailed analysis of a specific time segment (00:00:03:05 - 00:00:03:24). This window includes fields for Gloss (ARBEITEN1A), HamNoSys ([O\_0\_0\_0\_0] 1 G X +), mod (+rept), HamNoSys (+), dir, loc, Meaning (Arbeit), Plurality (Singular), and Interpretation (Identität). The main table shows entries for "behindert" and "arbeit".

Timecodes	Komple...	Gebärde	HamNoSys...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERL	O_0_0_0_0	behindert			behindert
00:00:03:05 00:00:03:24							mensch
							arbeit
							zusamm
							integration

ARBEITEN1A  
Tokens Stills  
Timecode: 00:00:03:05 - 00:00:03:24 M  
Dominant Hand  
Gloss: ARBEITEN1A  
HamNoSys: [O\_0\_0\_0\_0] 1 G X +  
mod: +rept  
HamNoSys: +  
dir:  
loc:  
Meaning: Arbeit  
Plurality: Singular  
Interpretation: Identität  
Nondominant Hand Analysis  
2003-05-07 (Marisa Braasch), 2006-03-31 (Thomas Hanke)

# iLex: Type-token matching is the essential step in transcription

The screenshot displays the iLex software interface, which is used for linguistic analysis of sign language. It consists of several overlapping windows:

- Video Window:** Shows a sign language interpreter in a red shirt. The video player includes a progress bar and playback controls.
- Table Window:** Titled "Aufgaben from Integrationsamt 1", it contains a table with columns for Timecodes, Komple..., Gebärde, and HamNo... The table lists time intervals and associated sign language terms like "behinderte" and "BEHINDERU...".
- Form Window:** Titled "ARBEITEN1A", it shows the linguistic analysis for a specific token. It includes fields for:
  - Gloss:** ARBEITEN1A
  - HamNoSys:** [Oɽθ, Oɽθ] 11 C X +
  - Mouth:** arbeiten
  - Description:** (empty)
  - Parent:** Gloss: MACHEN1A-SSAM
- Analysis Window:** Shows detailed parameters for the token:
  - mod:** +rept
  - HamNoSys:** #
  - dir:** (empty)
  - loc:** (empty)
  - Meaning:** Arbeit
  - Plurality:** Singular
  - Interpretation:** Identität
- Metadata Window:** Located at the bottom, it lists dates and names: "2004-12-18 (Thomas Hanke), 2008-04-14 (Thomas Hanke)", "2003-05-07 (Marisa Braasch), 2006-03-31 (Thomas Hanke)", and "1999-05-03 (Dolly Blanck), 2005-12-05 (Lutz König)".



# iLex: Type-token matching is the essential step in transcription

The screenshot displays the iLex software interface, which is used for transcribing sign language. It consists of several main components:

- Video Player:** On the left, a video window shows a sign language interpreter in a red shirt. Below the video are playback controls and a segment name field containing "Aufgaben".
- Timecodes Table:** A central table lists timecodes and associated data. The visible portion includes:
 

Timecodes	Komple...	Gebärde	HamNo...
00:00:00:00			
00:00:00:18			
00:00:00:18	behinderte	BEHINDERU	0_00[...
00:00:01:17			
- Token Analysis Window (ARBEITEN1A):** A detailed window on the right provides linguistic information for a specific token. It includes:
  - Form:** ARBEITEN1A
  - HamNoSys:** [0\_00, 0\_00] 1 0 X
  - mod:** +rept
  - HamNoSys:** ++
  - dir:** (empty)
  - loc:** (empty)
  - Meaning:** Arbeit
  - Plurality:** Singular
  - Interpretation:** Identität
- Token List:** A table below the analysis window shows 1150 entries with columns for Hand, Film, Timecode, Bedeutung, and HamNoSys.
 

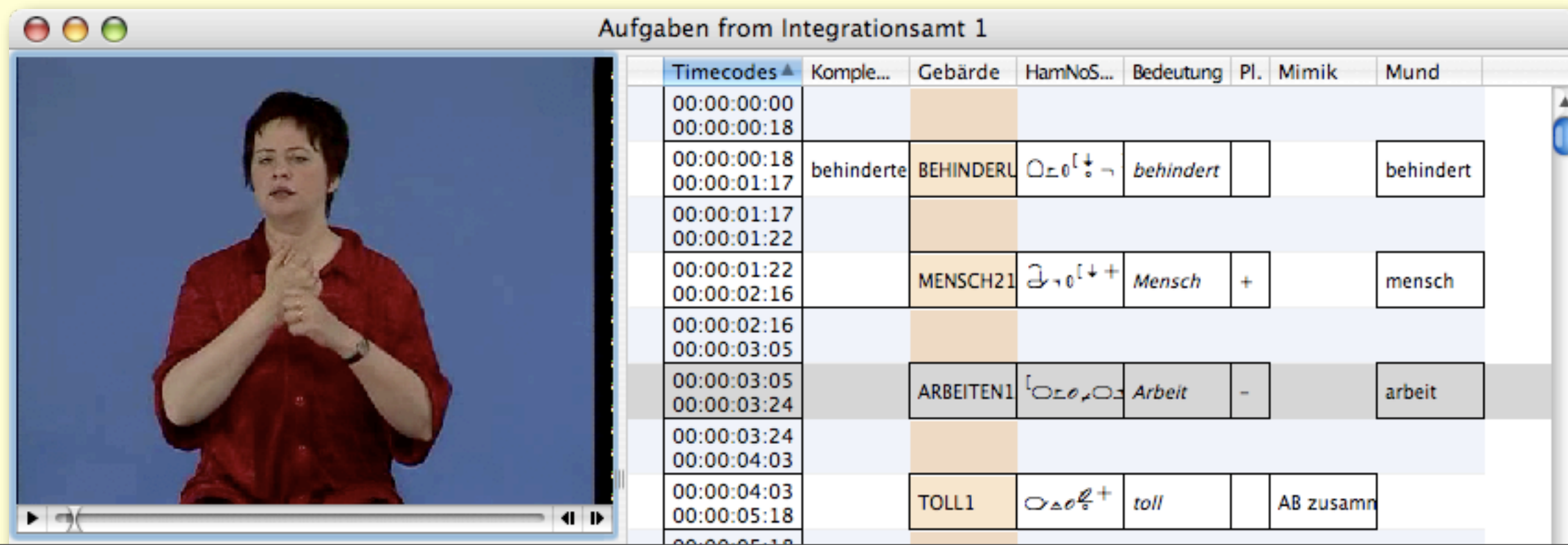
Hand	Film	Timecode	Bedeutung	HamNoSys
d	A6670111	00:00:00:17	Arbeit	
d	A6670113	00:00:00:15	Arbeit	(+)
d	A6673098	00:00:00:13	Arbeit	
d	A6673099	00:00:00:13	Arbeit	
d	A6673101	00:00:00:15	Arbeit	
d	A6673103	00:00:00:15	Arbeit	
d	A6674839	00:00:01:00	Arbeit	
d	ARBEITEN	00:00:00:00	arbeiten	
d	BB_G0439	00:00:00:24	Arbeit	
d	Blaue Büc...	00:06:51:00	Arbeit	
d	Blaue Büc...	00:06:51:12	arbeiten	
d	CLex6191	00:36:07:04	arbeiten	
d	CLex6191	00:36:30:19	arbeiten	
d	CLex6191	00:36:57:07	arbeiten	
d	CLex6191	01:19:21:07	Verarbeitung	
d	CLex6192	00:26:58:14	Verarbeitung	

# Type & Token

- The identified type is dragged into the transcript to become a token
- If no type is found: A new type is created and then assigned

# Two Views on Transcripts

- Time flows from top to bottom
  - A tag is a row in a tabular view
  - Row height independent of tag's duration
  - Focus on intervals of interest



The screenshot shows a window titled "Aufgaben from Integrationsamt 1". On the left is a video player showing a woman in a red shirt signing. On the right is a table with the following columns: Timecodes, Komple..., Gebärde, HamNoS..., Bedeutung, Pl., Mimik, and Mund.

Timecodes	Komple...	Gebärde	HamNoS...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERL	Q <sub>±</sub> 0 <sup>l</sup> ↓ <sub>±</sub> ~	behindert			behindert
00:00:01:17 00:00:01:22							
00:00:01:22 00:00:02:16		MENSCH21	Q <sub>±</sub> 0 <sup>l</sup> ↓ <sub>±</sub> +	Mensch	+		mensch
00:00:02:16 00:00:03:05							
00:00:03:05 00:00:03:24		ARBEITEN1	Q <sub>±</sub> 0 <sup>l</sup> ↓ <sub>±</sub> +	Arbeit	-		arbeit
00:00:03:24 00:00:04:03							
00:00:04:03 00:00:05:18		TOLL1	Q <sub>±</sub> 0 <sup>l</sup> ↓ <sub>±</sub> +	toll		AB zusamm	
00:00:05:18 00:00:05:18							

# Two Views on Transcripts

- Time flows from top to bottom
  - A tag is a row in a tabular view
  - Row height independent of tag's duration
  - Focus on intervals of interest
- Time flows from left to right
  - A tag's length is proportional to its duration

Aufgaben from Integrationsamt 1

Segment Name: Aufgaben

Theme: Integrationsamt

Transcript: Integrationsamt 1

00:00:00:00 2004-12-18 (Thomas Hanke), 2008-04-14 (Thomas Hanke)

00:01:37:05	00:01:37:20	00:01:38:10	00:01:39:00	00:01:39:15	00:01:40:05	00:01:40:20	00:01:41:10	00:01:42:00
Erhebung	Arbeitsleben		*(Grundvokabular)		Integration			*(C
Komplexbedeutung	Arbeitsleben							SN
Gebärde	ARBEITEN1A	LEBEN1D	SNUM-LIST-ERSTENS1A:1		INTEGRATION1			
Syntax Cat								
HamNoSys	[O_εθ_μO_εθ] 1 0 X +	- ε_εθ_μ_ε_ε [μ_ε_ε_ε_ε_ε_ε]	[ε_εθ_μ_ε_ε_ε] 1 0 X		- ε_εθ_μ_ε_ε_ε [μ_ε_ε_ε_ε_ε_ε]			
Bedeutung	Arbeit	Leben			Integration			
Pl.								
Mimik								
Mund	arbeitsleben				integration			

ration

- Time flows from left to right
- A tag's length is proportional to its duration



# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

A tag may consist of one or two tokens (dom+nondom hand) or  
A tag consists of one token only, and you have two or three token tiers per informant

# Token Tiers

- System can make sure that cotermporal tokens for one informant take a maximum of two hands
  - Or only one hand once video processing is able to tell us
  - What then?
    - Wrong type assigned to token?
    - Forgot to mark as weak drop?

# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

Tags group several token tags to a phrase or whatever with a certain meaning

# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

Describe the prompts presented, allows access to multimedia data presented to user

# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

References to types that do not count as tokens, such as corrections

# Tier Kinds in iLex

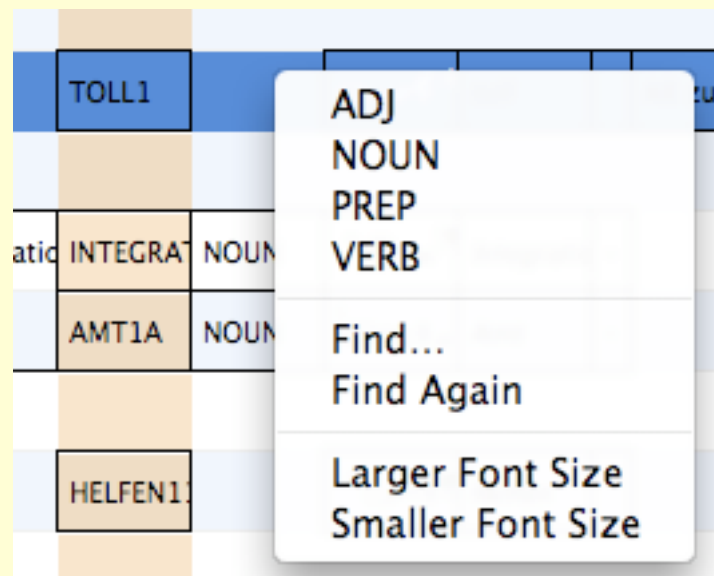
- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

As you know them from ELAN etc.

Can have open or closed vocabularies associated

# Text Tiers with Vocabularies

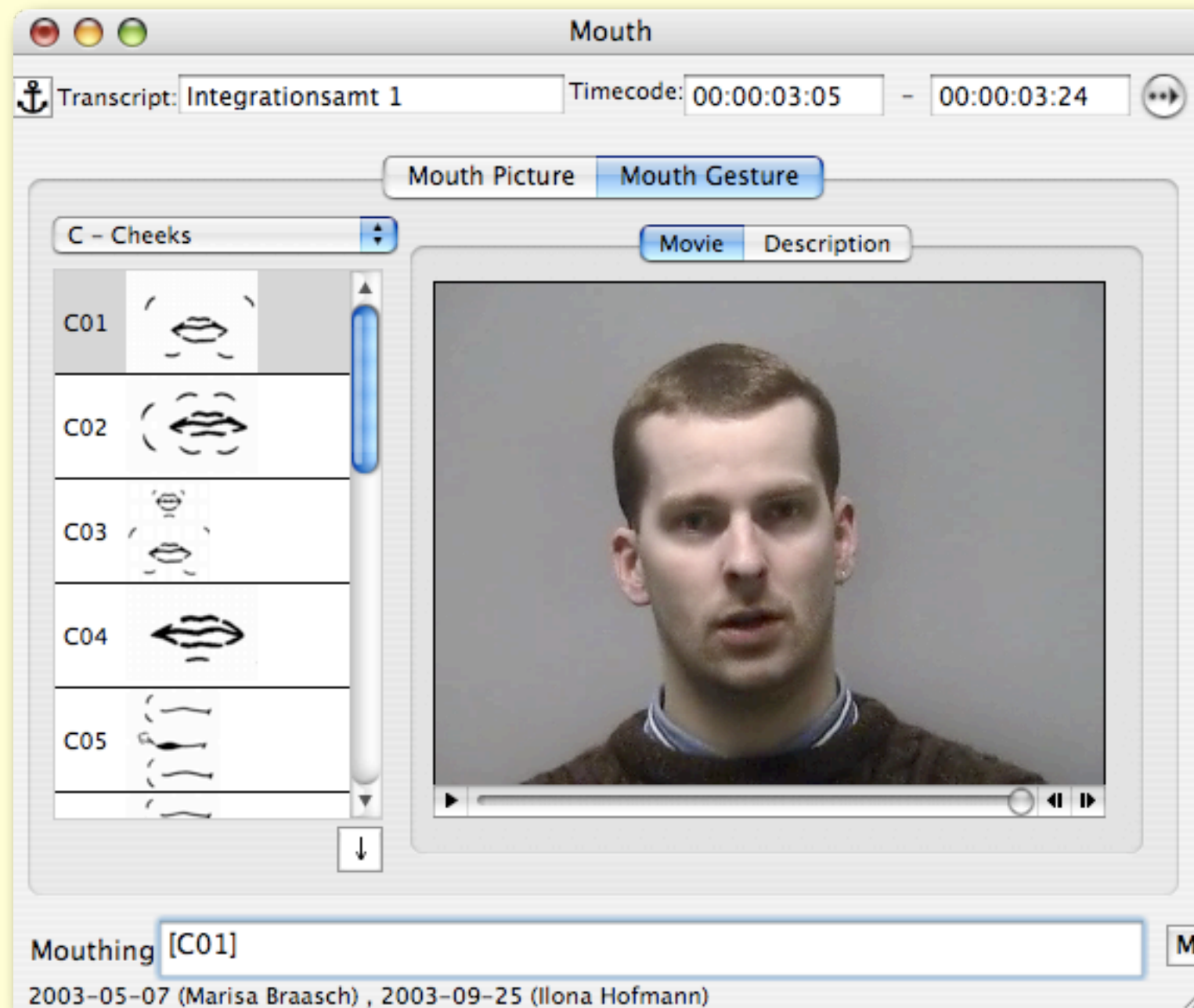
- Closed or open: Texts to choose from



- Mouthings: default mouthing associated to type plus signer's sign/mouthing combinations
- Built-in vocabularies with graphical editors



# Text Tiers with Special Editors, e.g. Mouth gesture encoding



# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

E.g. to describe positions in the video

# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

E.g. to relate anaphoric elements to their referents

# Tier Kinds in iLex

- Tokens tiers
- Phrasal units
- Elicitation tiers
- Type tiers
- Text tiers
- Numerical tiers
- Cross-reference tiers
- Value tiers

To reflect other features of complex tag data than displayed in the respective tier, e.g. citation form in HamNoSys of the type assigned to a token

# Value Tiers: Copy values from complex objects in other tiers

Aufgaben from Integrationsamt 1

Timecodes	Komple...	Gebärde	HamNoS...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERU	0 2 0 1 2 1	behindert			behindert
00:00:01:17 00:00:01:22							
00:00:01:22 00:00:02:16		MENSCH21	2 1 0 1 2 1	Mensch	+		mensch
00:00:02:16 00:00:03:05							
00:00:03:05 00:00:03:24		ARBEITEN1	0 2 0 1 2 1	Arbeit	-		arbeit
00:00:03:24 00:00:04:03							
00:00:04:03 00:00:05:18		TOLL1	0 2 0 1 2 1	toll		AB zusamn	
00:00:05:18 00:00:06:15							
00:00:06:15 00:00:07:09	Integration	INTEGRATI	1 2 3 4 5 6	Integration	-		integration
00:00:07:09 00:00:08:13		AMT1A	0 1 2 0 1 0	Amt	-		
00:00:08:13 00:00:09:03							
00:00:09:03 00:00:10:24		HELFEN11	1 0 1 0 1	helfen			
00:00:10:24 00:00:11:19							
00:00:11:19							

2004-12-18 (Thomas Hanke) , 2008-04-14 (Thomas Hanke)

# Value Tiers: Copy values from complex objects in other tiers

Aufgaben from Integrationsamt 1

Timecodes	Komple...	Gebärde	HamNoS...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERU	0 2 0 1 2 3 4 5 6 7 8 9	behindert			behindert
00:00:01:17 00:00:01:22							
00:00:01:22 00:00:02:16		MENSCH21	2 3 4 5 6 7 8 9	Mensch	+		mensch
00:00:02:16 00:00:03:05							
00:00:03:05 00:00:03:24		ARBEITN1	0 2 0 4 0 5	Arbeit	-		arbeit
00:00:03:24 00:00:04:03							
00:00:04:03 00:00:05:18		TOLL1	0 2 0 4 5 6 7 8 9	toll		AB zusamn	
00:00:05:18 00:00:06:15							
00:00:06:15 00:00:07:09	Integration	INTEGRATI	0 2 0 4 5 6 7 8 9	Integration	-		integration
00:00:07:09 00:00:08:13		AMT1A	0 1 2 0 4 0	Amt	-		
00:00:08:13 00:00:09:03							
00:00:09:03 00:00:10:24		HELFEN11	0 1 2 0 4	helfen			
00:00:10:24 00:00:11:19							
00:00:11:19							

2004-12-18 (Thomas Hanke) , 2008-04-14 (Thomas Hanke)



# Value Tiers: Copy values from complex objects in other tiers

The screenshot shows a software interface for managing video content. On the left is a video player showing a woman in a red shirt signing. Below the video player are controls for playback and segment management. On the right is a table with the following columns: Timecodes, Komple..., Gebärde, HamNoS..., Bedeutung, Pl., Mimik, and Mund. The table contains several rows of data, with two rows circled in red: the row for 'ARBEIT' and the row for 'Integration'.

Timecodes	Komple...	Gebärde	HamNoS...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERU	0 2 0 1 2 3 4 5 6 7 8 9	behindert			behindert
00:00:01:17 00:00:01:22							
00:00:01:22 00:00:02:16		MENSCH21	2 3 4 5 6 7 8 9	Mensch	+		mensch
00:00:02:16 00:00:03:05							
00:00:03:05 00:00:03:24		ARBEIT N1	0 2 0 1 2 3 4 5 6 7 8 9	Arbeit	-		arbeit
00:00:03:24 00:00:04:03							
00:00:04:03 00:00:05:18		TOLL1	0 2 0 1 2 3 4 5 6 7 8 9	toll		AB zusamn	
00:00:05:18 00:00:06:15							
00:00:06:15 00:00:07:09	Integration	INTEGRATI	0 2 0 1 2 3 4 5 6 7 8 9	Integration	-		integration
00:00:07:09 00:00:08:13		AMT1A	0 2 0 1 2 3 4 5 6 7 8 9	Amte	-		
00:00:08:13 00:00:09:03							
00:00:09:03 00:00:10:24		HELFEN11	0 2 0 1 2 3 4 5 6 7 8 9	helfen			
00:00:10:24 00:00:11:19							
00:00:11:19							

# Value Tiers: Copy values from complex objects in other tiers

Aufgaben from Integrationsamt 1

Timecodes	Komple...	Gebärde	HamNoS...	Bedeutung	Pl.	Mimik	Mund
00:00:00:00 00:00:00:18							
00:00:00:18 00:00:01:17	behinderte	BEHINDERU	0 2 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	behindert			behindert
00:00:01:17 00:00:01:22							
00:00:01:22 00:00:02:16		MENSCH21	2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	Mensch	+		mensch
00:00:02:16 00:00:03:05							
00:00:03:05 00:00:03:24		ARBEITN1	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	Arbeit	-		arbeit
00:00:03:24 00:00:04:03							
00:00:04:03 00:00:05:18		TOLL1	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	toll		AB zusamn	
00:00:05:18 00:00:06:15							
00:00:06:15 00:00:07:09	Integration	INTEGRATI	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	Integration	-		integration
00:00:07:09 00:00:08:13		AMT1A	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100				
00:00:08:13 00:00:09:03							
00:00:09:03 00:00:10:24		HELFEN11	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	helfen			
00:00:10:24 00:00:11:19							
00:00:11:19 00:00:11:19							

00:00:03:24 MI →M >|< [play/pause/stop buttons]

Segment Name: Aufgaben

Theme: Integrationsamt

Transcript: Integrationsamt 1

2004-12-18 (Thomas Hanke) , 2008-04-14 (Thomas Hanke)



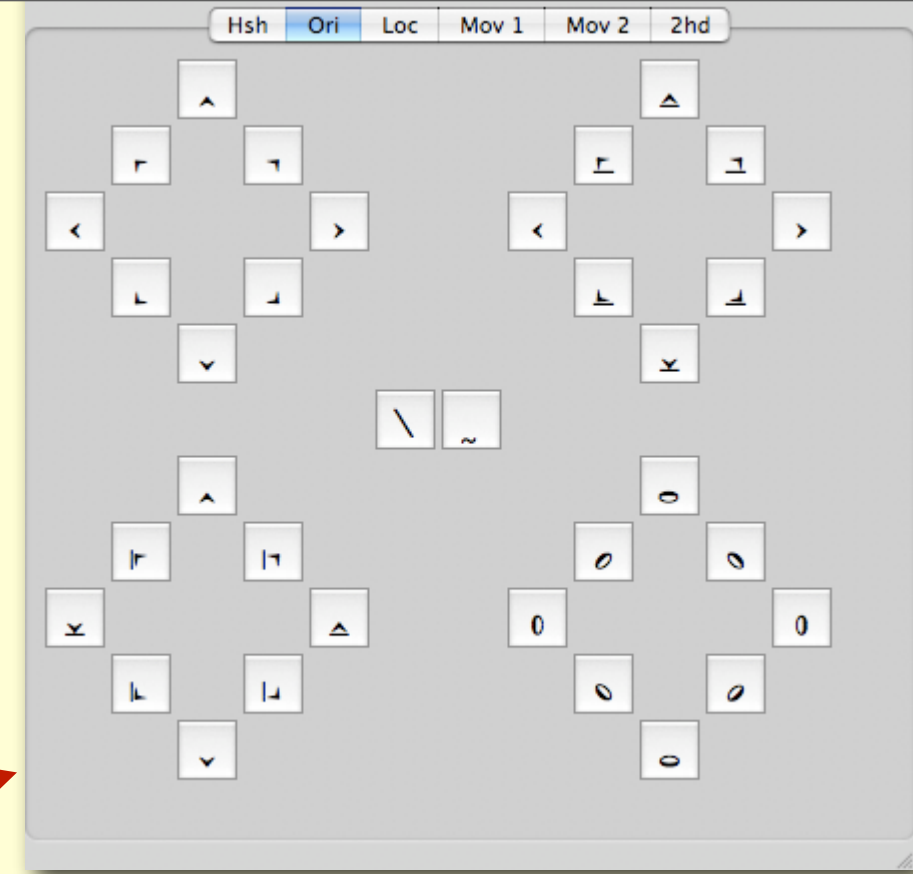
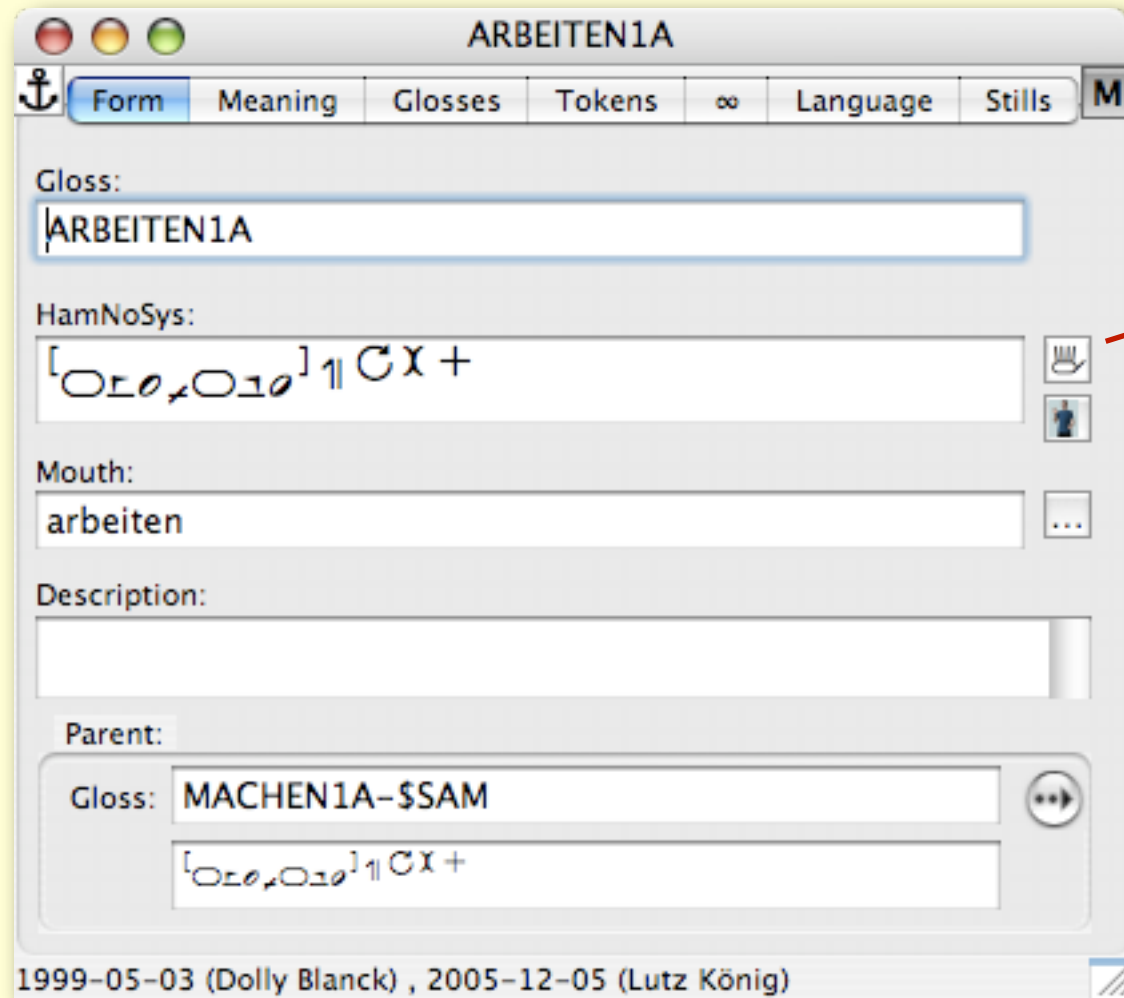
# Immediate input verification: HamNoSys & avatar

The screenshot shows a window titled "ARBEITEN1A" with a menu bar containing "Form", "Meaning", "Glosses", "Tokens", "∞", "Language", "Stills", and "M". The main content area is divided into several sections:

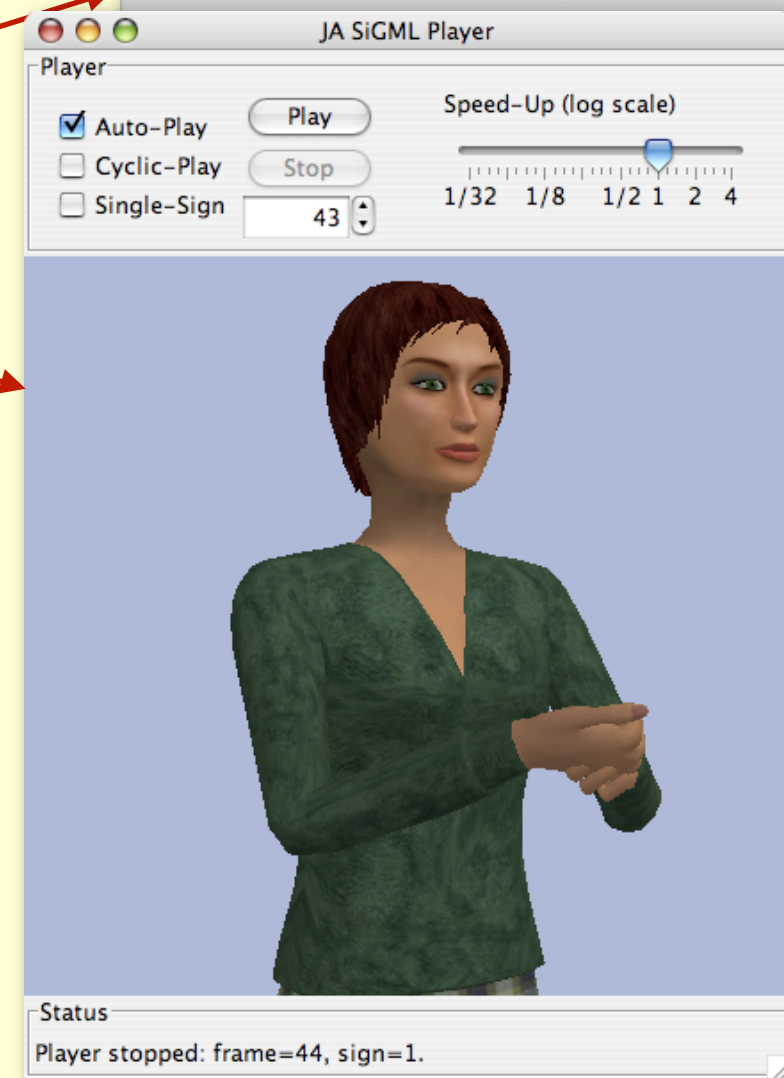
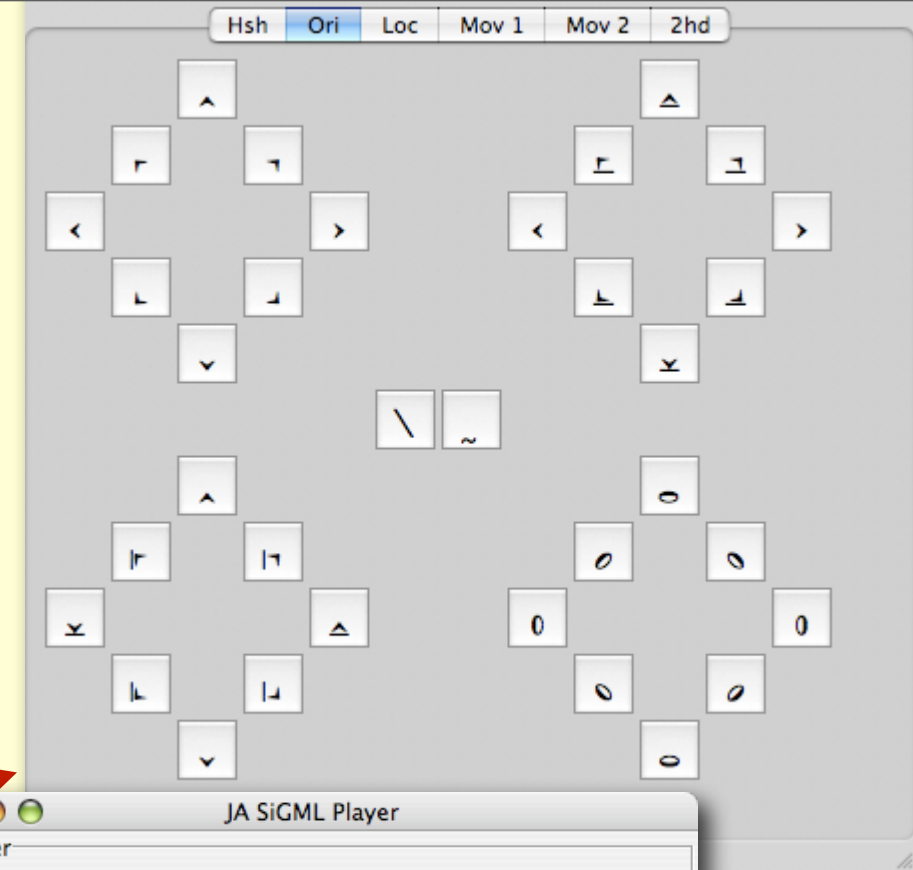
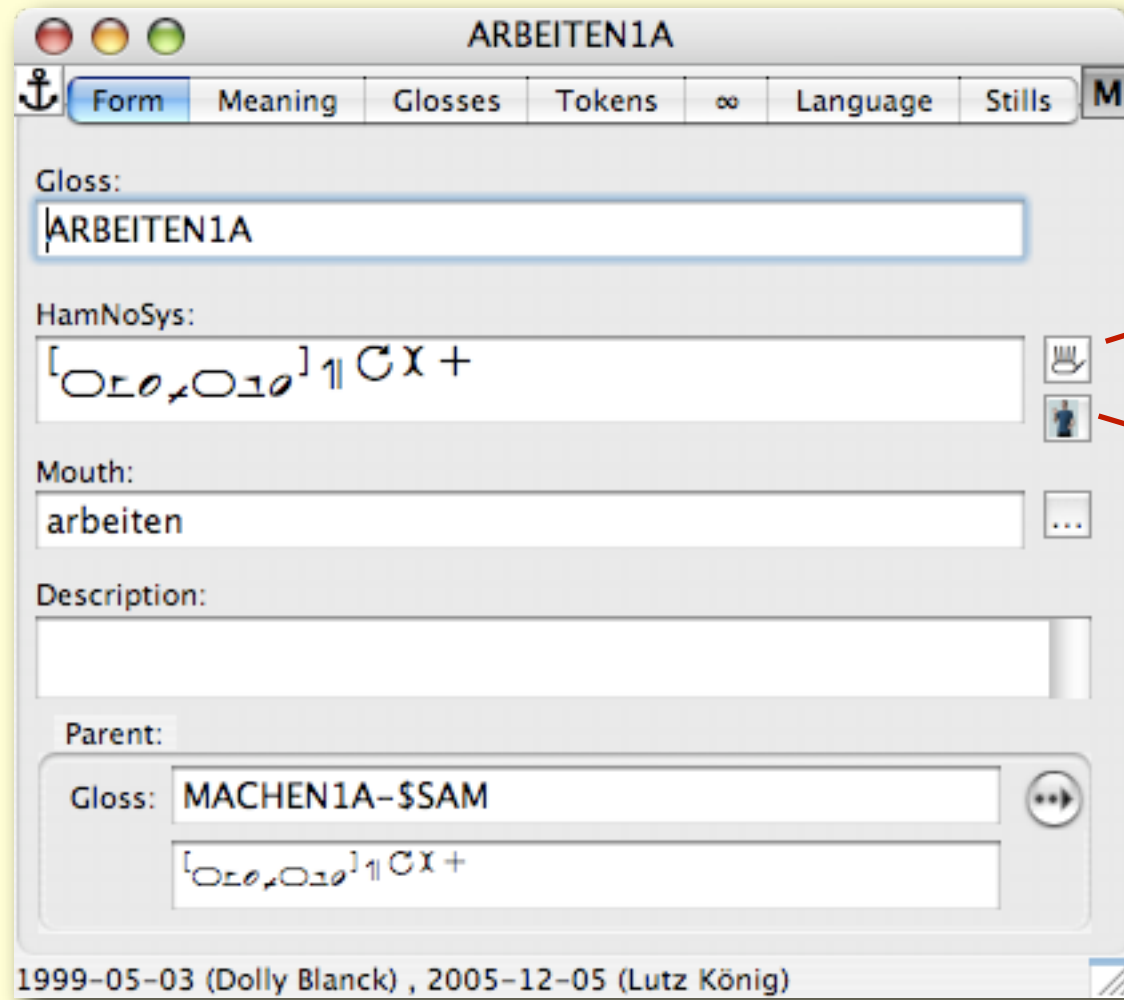
- Gloss:** A text field containing "ARBEITEN1A".
- HamNoSys:** A text field containing the phonetic transcription "[oɛθ,ɔɪθ] 11 ʧX +". To the right of this field are two icons: a hand and a person.
- Mouth:** A text field containing "arbeiten". To the right of this field is a three-dot menu icon.
- Description:** An empty text field.
- Parent:** A section containing a "Gloss:" field with "MACHEN1A-SSAM" and a HamNoSys transcription field with "[oɛθ,ɔɪθ] 11 ʧX +". To the right of these fields is a double-right arrow icon.

At the bottom of the window, the text "1999-05-03 (Dolly Blanck) , 2005-12-05 (Lutz König)" is displayed.

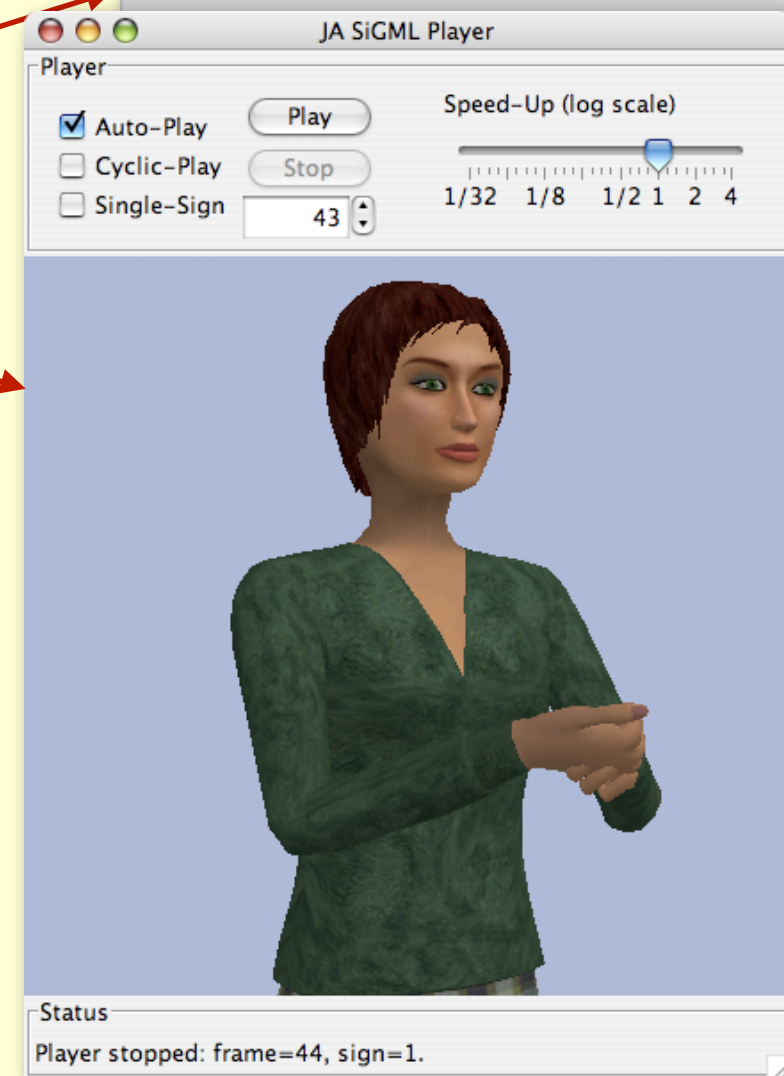
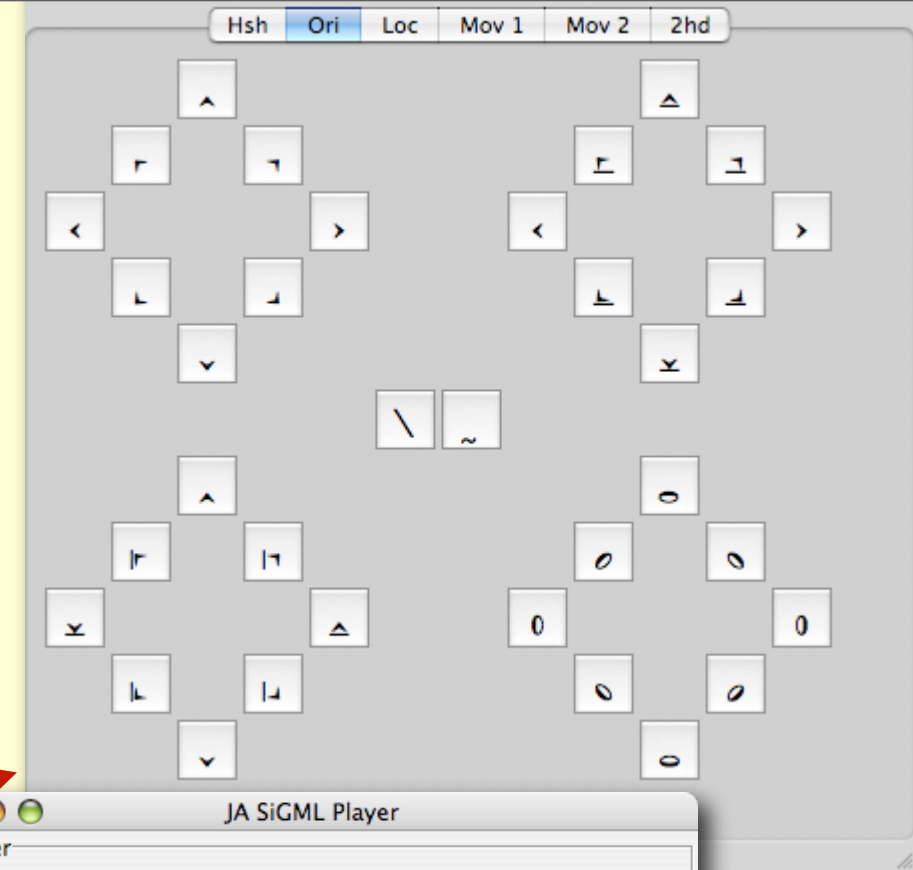
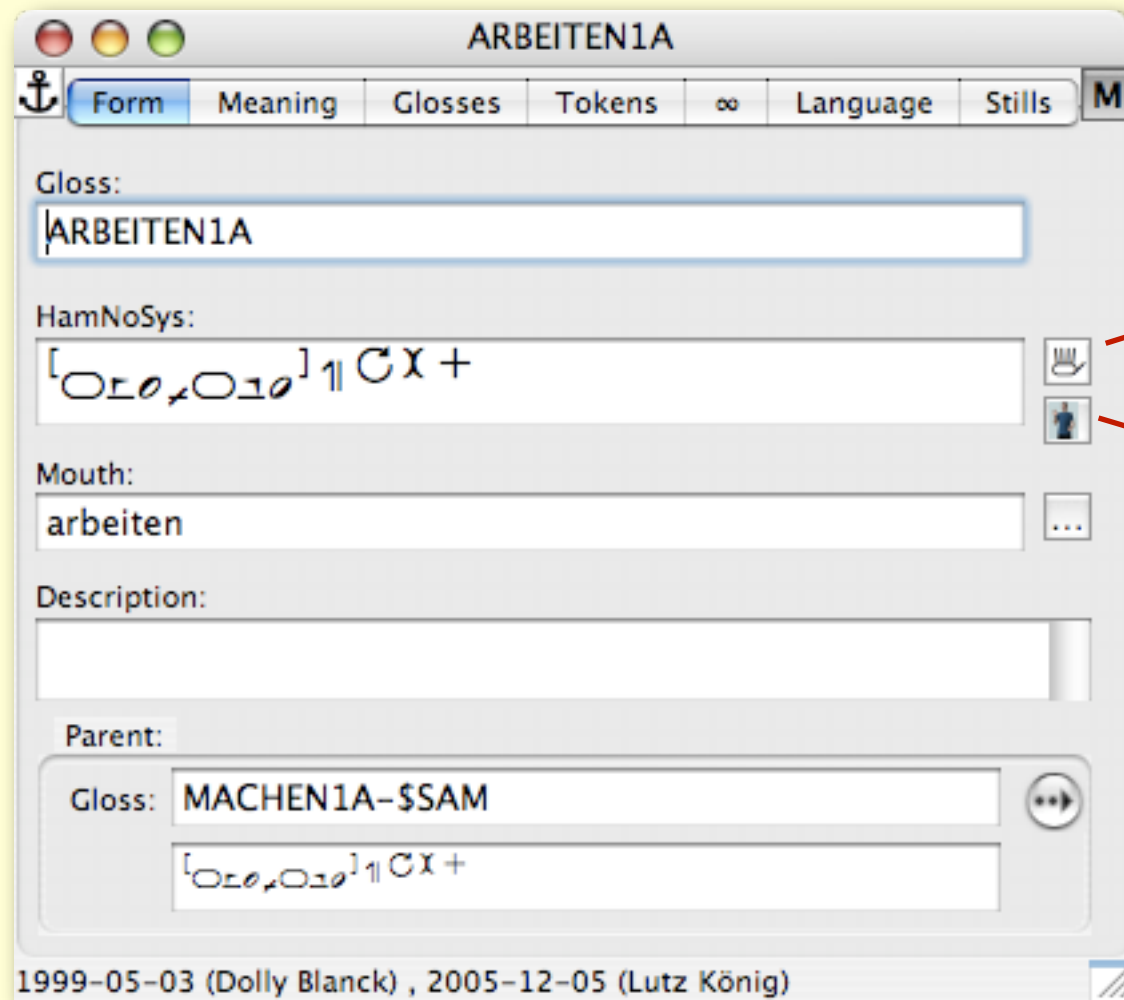
# Immediate input verification: HamNoSys & avatar



# Immediate input verification: HamNoSys & avatar



# Immediate input verification: HamNoSys & avatar



■ Side effect: anonymisation

# How do you annotate the form differences between type and token?

- Just ignore them
  - Refer to the citation form
- Describe the form deviation in the token
- Have separate types to describe citation form and inflected form
- Have a separate text tier describing the kind of modification... you encounter

# Qualifications

- Inflection

  - Agreement

  - Plural repetition

- Modification

  - Extension of the sign's image to convey extra meaning

- Variation

# Qualifiers

- “Inflection paradigms” as controlled vocabularies
- For each type, previously used qualifiers are readily available
- Others require an extra step

# Qualifiers step by step

BESUCHEN1A

$[Q_1 \rightarrow 0, \mu, Q_2 \rightarrow 0] \wedge [S, \sigma \rightarrow \mu, E]$

BESUCHEN1A +

- Source
- Goal
- Repetition

+

Cancel Use



# Qualifiers step by step

BESUCHEN1A

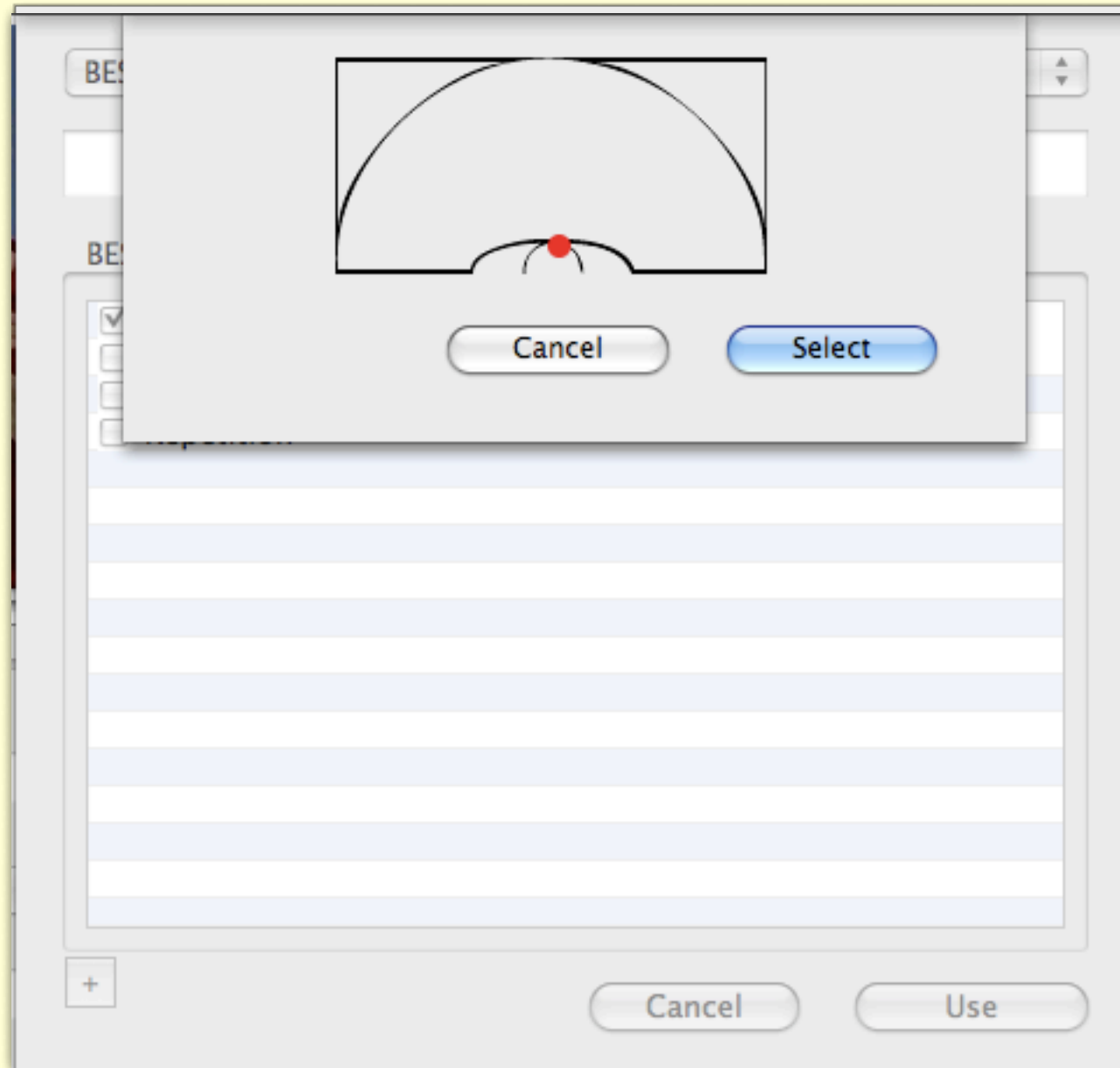
$[Q_1=0, Q_2 > 0] \wedge [S_1 \neq \mu_1]$

BESUCHEN1A +

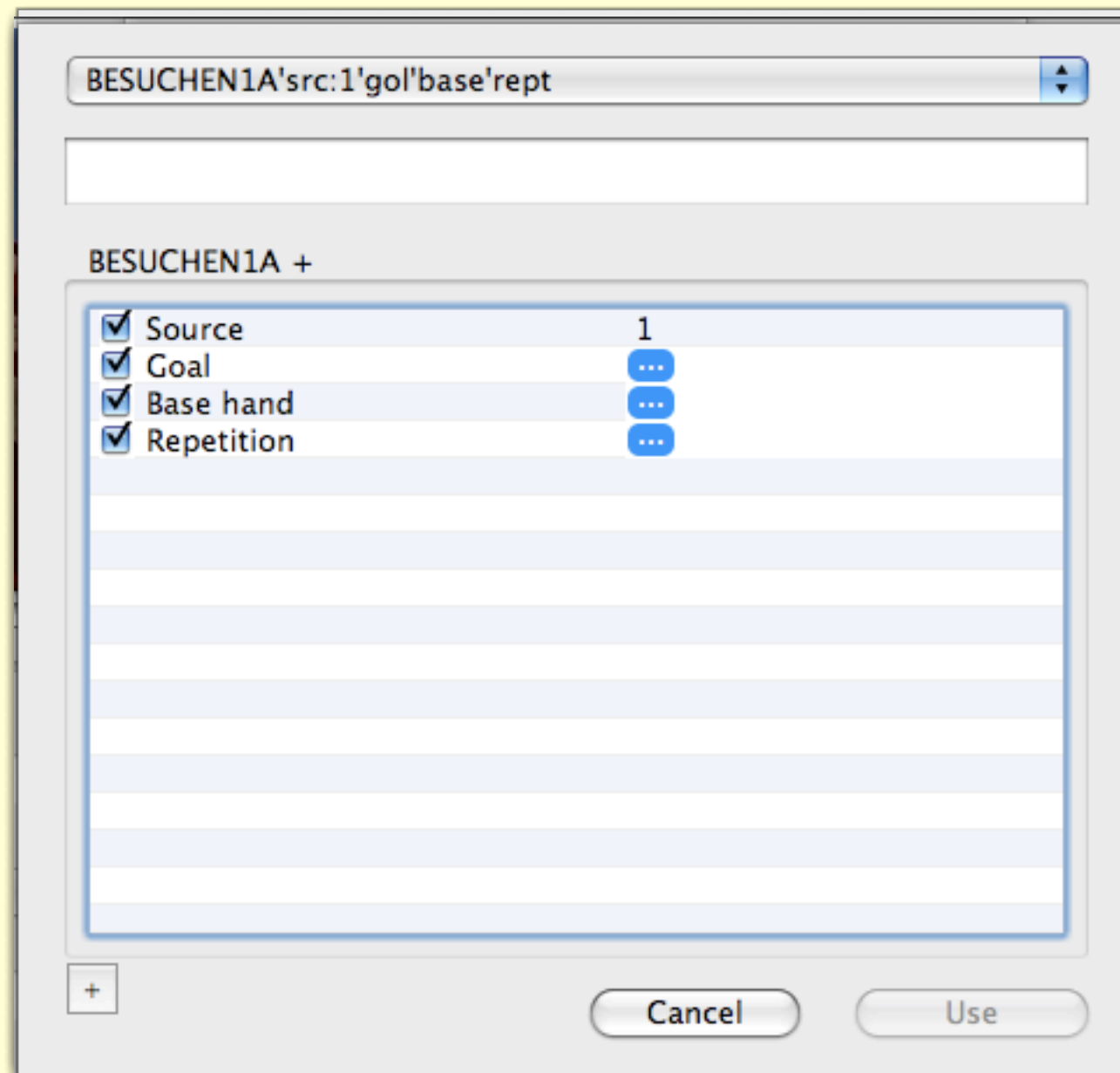
- Source
- Goal
- Base hand
- Repetition

+ Cancel Use

# Qualifiers step by step



# Qualifiers step by step



# Qualified Types: Advantages


- Per-type controlled vocabularies
- Any statistics easily take qualifications into account (“word-form”) or ignore them (“type-level”)
- Lexicographic view supported:
  - Qualified types can “freeze” into lexemes
  - Relations between types can include qualified types

# Cross References

- A new tier kind with tags establishing relations to tags in other tiers
- E.g. anaphora linking to their referents
  - Centering theory
- E.g. source/goal/location linking to where the location was established

# Establishing References

Aufgaben from Integrationsamt 1



00:00:01:22 MI →M >|< [Navigation icons]

Segment Name:  
Aufgaben

Theme:  
Integrationsamt

Transcript:  
Integrationsamt 1

Timecodes	Kompl...	Gebärde	Syntax...	HamN...	Bedeut...	Mimik	Mund	XRef
00:00:00:00 00:00:00:18								
00:00:00:18 00:00:01:17	behindert	BEHINDER	NOUN	⓪_⓪ <sup>l</sup> ↓	behinder		behindert	
00:00:01:17 00:00:01:22								
00:00:01:22 00:00:02:16		MENSCH	NOUN	⓷_⓷ <sup>l</sup> ↓	Mensch +		mensch	
00:00:02:16 00:00:03:05								
00:00:03:05 00:00:03:24		ARBEITEN	VERB	⓪_⓪_⓪_⓪	Arbeit -		arbeit	
00:00:03:24 00:00:04:03								
00:00:04:03 00:00:05:18		TOLL1		⓪_⓷_⓪ <sup>l</sup> +	toll	AB zusam		
00:00:05:18 00:00:06:15								
00:00:06:15 00:00:07:09	Integratio	INTEGRA	NOUN	⓷_⓷_⓷ <sup>l</sup>	Integratio -		integratio	
00:00:07:09 00:00:08:13		AMT1A	NOUN	⓪_⓷_⓪_⓪	Amt -			
00:00:08:13 00:00:09:03								
00:00:09:03 00:00:10:24		HELFEN1		'⓪_⓷_⓪	helfen			
00:00:10:24 00:00:11:19								
00:00:11:19 00:00:12:21	Bundesre	VERBAND		⓷_⓪ <sup>2</sup> _⓪	Bund		bundesre	
00:00:12:21 00:00:13:01								
00:00:13:01 00:00:14:06		REPUBLIK		⓷_⓷_⓪	Republik			
00:00:14:06 00:00:14:09								
00:00:14:09 00:00:15:00		ES-GIBT1		'⓪_⓪_⓪ <sup>l</sup>	gibt		gibt	
00:00:15:00 00:00:15:02								
00:00:15:02 00:00:15:10								



# Establishing References

Aufgaben from Integrationsamt 1

Timecodes	Kompl...	Gebärde	Syntax...	HamN...	Bedeut...	Mimik	Mund	XRef
00:00:00:00 00:00:00:18								
00:00:00:18 00:00:01:17	behindert	BEHINDER	NOUN	⊙_0^⊕	behinder		behindert	
00:00:01:17 00:00:01:22								
00:00:01:22 00:00:02:16		MENSCH	NOUN	⊙_1^⊕	Mensch +		mensch	
00:00:02:16 00:00:03:05								
00:00:03:05 00:00:03:24		ARBEITEN	VERB	⊙_0^⊕	Arbeit -		arbeit	
00:00:03:24 00:00:04:03								
00:00:04:03 00:00:05:18		TOLL1		⊙_0^⊕	toll	AB zusam		
00:00:05:18 00:00:06:15								
00:00:06:15 00:00:07:09	Integratio	INTEGRA	NOUN	⊙_0^⊕	Integratio -		integratio	
00:00:07:09 00:00:08:13		AMT1A	NOUN	⊙_1^⊕	Amt -			
00:00:08:13 00:00:09:03								
00:00:09:03 00:00:10:24		HELFEN1		⊙_1^⊕	helfen			
00:00:10:24 00:00:11:19								
00:00:11:19 00:00:12:21	Bundesre	VERBAND		⊙_0^⊕	Bund		bundesre	
00:00:12:21 00:00:13:01								
00:00:13:01 00:00:14:06		REPUBLIK		⊙_0^⊕	Republik			
00:00:14:06 00:00:14:09								
00:00:14:09 00:00:15:00		ES-GIBT1		⊙_0^⊕	gibt		gibt	
00:00:15:00 00:00:15:02								
00:00:15:02 00:00:15:10								

00:00:01:22 MI →M >|< [Navigation icons]

Segment Name: Aufgaben

Theme: Integrationsamt

Transcript: Integrationsamt 1



# Establishing References

Aufgaben from Integrationsamt 1

Timecodes	Kompl...	Gebärde	Syntax...	HamN...	Bedeut...	Mimik	Mund	XRef
00:00:00:00 00:00:00:18								
00:00:00:18 00:00:01:17	behindert	BEHINDER	NOUN	⊖_0^1↓	behinder		behindert	
00:00:01:17 00:00:01:22								
00:00:01:22 00:00:02:16		MENSCH	NOUN	⊖_1^0↓	Mensch +		mensch	
00:00:02:16 00:00:03:05								
00:00:03:05 00:00:03:24		ARBEITEN	VERB	⊖_0_0_0	Arbeit -		arbeit	
00:00:03:24 00:00:04:03								
00:00:04:03 00:00:05:18		TOLL1		⊖_Δ_0^0	toll	AB zusam		
00:00:05:18 00:00:06:15								
00:00:06:15 00:00:07:09	Integratio	INTEGRA	NOUN	⊖_⊖_0^0	Integratio -		integratio	
00:00:07:09 00:00:08:13		AMT1A	NOUN	⊖_0^1_0_0	Amt -			
00:00:08:13 00:00:09:03								
00:00:09:03 00:00:10:24		HELFEN1		'_0^1_0	helfen			
00:00:10:24 00:00:11:19								
00:00:11:19 00:00:12:21	Bundesre	VERBAND		'_0^2_0	Bund		bundesre	
00:00:12:21 00:00:13:01								
00:00:13:01 00:00:14:06		REPUBLIK		⊖_0_0	Republik			
00:00:14:06 00:00:14:09								
00:00:14:09 00:00:15:00		ES-GIBT1		'_0_0^1	gibt		gibt	
00:00:15:00 00:00:15:02								
00:00:15:02 00:00:15:10								

Segment Name: Aufgaben

Theme: Integrationsamt

Transcript: Integrationsamt 1



# Establishing References

Aufgaben from Integrationsamt 1

Timecodes	Kompl...	Gebärde	Syntax...	HamN...	Bedeut...	Mimik	Mund	XRef
00:00:00:00 00:00:00:18								
00:00:00:18 00:00:01:17	behindert	BEHINDER	NOUN	⊖ <sub>ε</sub> 0 <sup>l</sup> ↓	behinder		behindert	
00:00:01:17 00:00:01:22								
00:00:01:22 00:00:02:16		MENSCH	NOUN	⊖ <sub>ε</sub> 1 <sup>l</sup> ↓	Mensch +		mensch	
00:00:02:16 00:00:03:05								
00:00:03:05 00:00:03:24		ARBEITEN	VERB	⊖ <sub>ε</sub> 0 <sub>ε</sub> ⊖	Arbeit -		arbeit	
00:00:03:24 00:00:04:03								
00:00:04:03 00:00:05:18		TOLL1		⊖ <sub>ε</sub> 0 <sup>l</sup> ⊖	toll	AB zusam		
00:00:05:18 00:00:06:15								
00:00:06:15 00:00:07:09	Integratio	INTEGRA	NOUN	⊖ <sub>ε</sub> ⊖0 <sup>l</sup> ⊖	Integratio -		integratio	
00:00:07:09 00:00:08:13		AMT1A	NOUN	⊖ <sub>ε</sub> 1 <sup>l</sup> ⊖	Amt -			
00:00:08:13 00:00:09:03								
00:00:09:03 00:00:10:24		HELFEN1		⊖ <sub>ε</sub> 1 <sup>l</sup> ⊖	helfen			
00:00:10:24 00:00:11:19								
00:00:11:19 00:00:12:21	Bundesre	VERBAND		⊖ <sub>ε</sub> 0 <sup>2</sup> ⊖0	Bund		bundesre	
00:00:12:21 00:00:13:01								
00:00:13:01 00:00:14:06		REPUBLIK		⊖ <sub>ε</sub> ⊖0	Republik			
00:00:14:06 00:00:14:09								
00:00:14:09 00:00:15:00		ES-GIBT1		⊖ <sub>ε</sub> 0 <sup>l</sup> ⊖	gibt		gibt	
00:00:15:00 00:00:15:02								
00:00:15:02 00:00:15:10								

Segment Name: Aufgaben

Theme: Integrationsamt

Transcript: Integrationsamt 1

# Establishing References

Aufgaben from Integrationsamt 1

Timecodes	Kompl...	Gebärde	Syntax...	HamN...	Bedeut...	Mimik	Mund	XRef
00:00:00:00 00:00:00:18								
00:00:00:18 00:00:01:17	behindert	BEHINDER	NOUN	⊖ <sub>ε</sub> 0 <sup>l</sup> ↓	behinder		behindert	
00:00:01:17 00:00:01:22								
00:00:01:22 00:00:02:16		MENSCH	NOUN	⊖ <sub>ε</sub> 1 <sup>l</sup> ↓	Mensch +		mensch	
00:00:02:16 00:00:03:05								
00:00:03:05 00:00:03:24		ARBEITEN	VERB	⊖ <sub>ε</sub> 0 <sub>ε</sub> ⊖	Arbeit -		arbeit	
00:00:03:24 00:00:04:03								
00:00:04:03 00:00:05:18		TOLL1		⊖ <sub>ε</sub> Δ0 <sup>l</sup> +	toll	AB zusam		
00:00:05:18 00:00:06:15								
00:00:06:15 00:00:07:09	Integratio	INTEGRA	NOUN	⊖ <sub>ε</sub> ⊖ <sub>ε</sub> 0 <sup>l</sup> ⊖	Integratio -		integratio	
00:00:07:09 00:00:08:13		AMT1A	NOUN	⊖ <sub>ε</sub> 0 <sub>ε</sub> ⊖	Amt -			
00:00:08:13 00:00:09:03								
00:00:09:03 00:00:10:24		HELFEN1		⊖ <sub>ε</sub> 0 <sub>ε</sub> ⊖	helfen			00:00:01
00:00:10:24 00:00:11:19								
00:00:11:19 00:00:12:21	Bundesre	VERBAND		⊖ <sub>ε</sub> 0 <sup>2</sup> ⊖	Bund		bundesre	
00:00:12:21 00:00:13:01								
00:00:13:01 00:00:14:06		REPUBLIK		⊖ <sub>ε</sub> ⊖	Republik			
00:00:14:06 00:00:14:09								
00:00:14:09 00:00:15:00		ES-GIBT1		⊖ <sub>ε</sub> ⊖ <sub>ε</sub> 0 <sup>l</sup>	gibt		gibt	
00:00:15:00 00:00:15:02								
00:00:15:02 00:00:15:10								

00:00:09:03 MI →M >|< [Navigation icons]

Segment Name: Aufgaben

Theme: Integrationsamt

Transcript: Integrationsamt 1

# In Combination with Qualifications: Chances for Automatic Linking

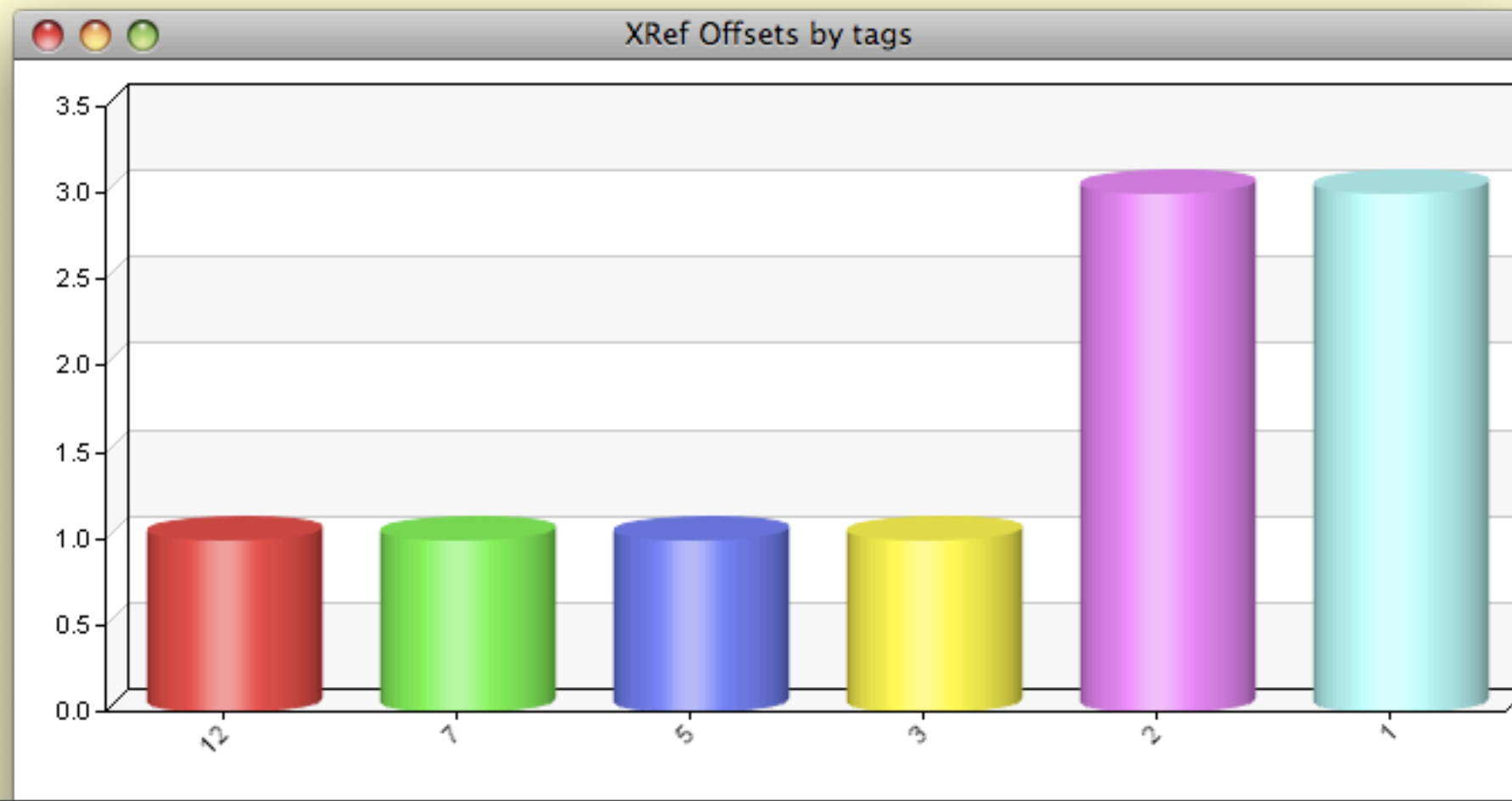
- For tokens with a specific goal:
  - Search for predecessor tags that establish that location in signing space
  - Scripting language available to meet individual needs

# Measures: Distances between Antecedent and Referrer

- by time elapsed between them
- by signs in-between
- Graphing distance distribution

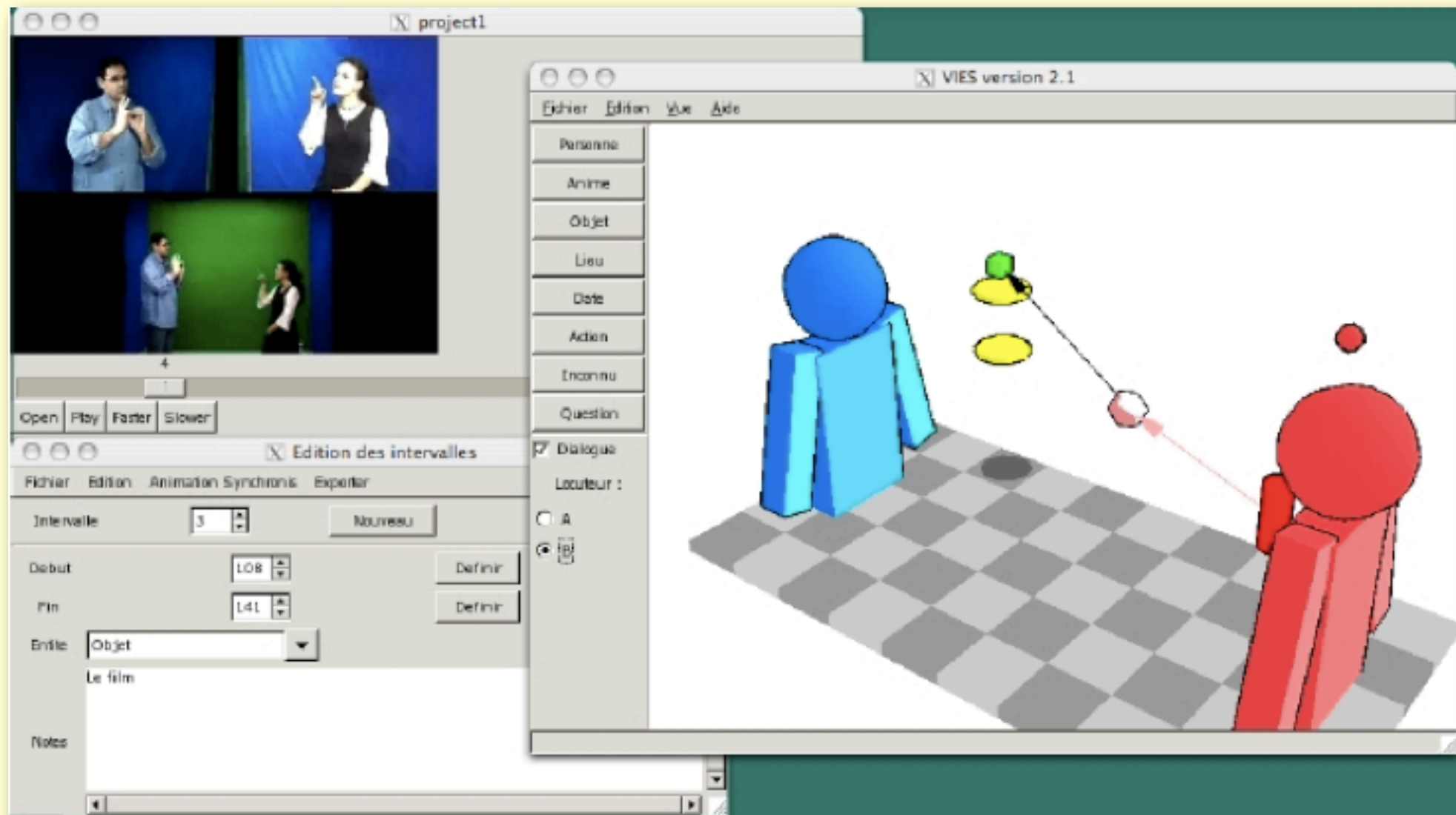
# Measures: Distances between Antecedent and Referrer

- by time elapsed between them
- by signs in-between
- Graphing distance distribution



# Improved visualisation of spatio-temporal data & semantic relations

## ■ IRIT Toulouse



# Applications of iLex

- Documentation
- Lexicography
- Animation
- Teaching
  - eLearning
  - Language lab materials




# Dictionary Production

- Book
- Internet/DVD
- DVD Video

## Soziometrie

DEFINITION




empirische Sozialforschung

Die Soziometrie ist ein [Verfahren](#) der [empirischen Sozialforschung](#), mit dem die Beziehungen zwischen den Mitgliedern einer Gruppe festgestellt, dargestellt und untersucht werden können. Es wird deshalb nach Sympathie und Antipathie unter den Mitgliedern einer Gruppe gefragt, man untersucht die Kontaktwünsche (subjektive Interaktionspräferenzen) und die tatsächlichen Kontakte (faktische Interaktionsbeziehungen). Das erfolgt im Wesentlichen durch die Befragung der Gruppenmitglieder über ihre Beziehungen (Interaktion). Soziometrische Ergebnisse werden in einem Soziogramm grafisch dargestellt. Eine mathematische Darstellung der Beziehungen unter den Mitgliedern einer Gruppe erfolgt in Soziomatrizen. Seltener sind Darstellungen, in denen eine Kennzeichnung der Beziehungen der Gruppenmitglieder informiert. Die Soziometrie kann bei der Analyse kleiner Gruppen, deren Mitglieder sich gut kennen, eingesetzt werden (z. B. in der [Gruppenarbeit](#)). Da die Ergebnisse im Wesentlichen durch die Befragung der Gruppenmitglieder gewonnen werden, sind die Aussagen auch von dem Bewusstsein der Gruppenmitglieder über ihre Beziehungen abhängig und von deren Wunschvorstellungen geprägt.

THEMATISCH VERWANDTE FACHBEGRIFFE [Psychodrama](#)

SACHGRUPPEN

- [Methoden der Sozialarbeit/Sozialpädagogik](#)
- [Psychologie](#)


1 (803)  [VERBAND21A](#)

2 (804)  [SOZIAL2A](#) [MUSTER1A](#)

### Knethaken

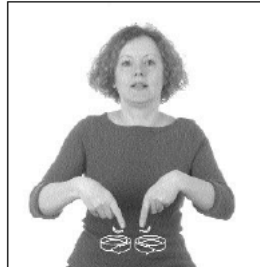
ENGLISCH  
kneading hook

SACHGRUPPE  
Ernährung  
(backen, Teigherstellung)




DEFINITION Knethaken sind Zubehörteile, die in die →Küchenmaschine oder in das →Handrührgerät eingesetzt werden. Sie bestehen aus spiralförmig gebogenen Metallstäben mit einem speziell geformten Ende. Knethaken werden zum →Kneten schwerer Teige (→Teig) wie →Hefeteig und →Knetteig verwendet. Dabei wird keine Luft eingearbeitet (siehe →Rührteig).


748.1: Zwei Knethaken drehen sich.





749: Mundgestik.  
749.1: Zeigefinger zeichnen zwei spiralförmige Knethaken.




750.1 KNETEN1B



750.2 DREHEN22

### AUSPRESSEN1B



VORKOMMEN  
Konventionell verwendet für „auspressen, Presse“; Modifikation von →AUSPRESSEN1B: Zitrone mit Zitronenpresse (linke Hand) auspressen:  
127.1 auspressen 1547.1 Zitronenpresse  
1545.3 Zitronenpresse

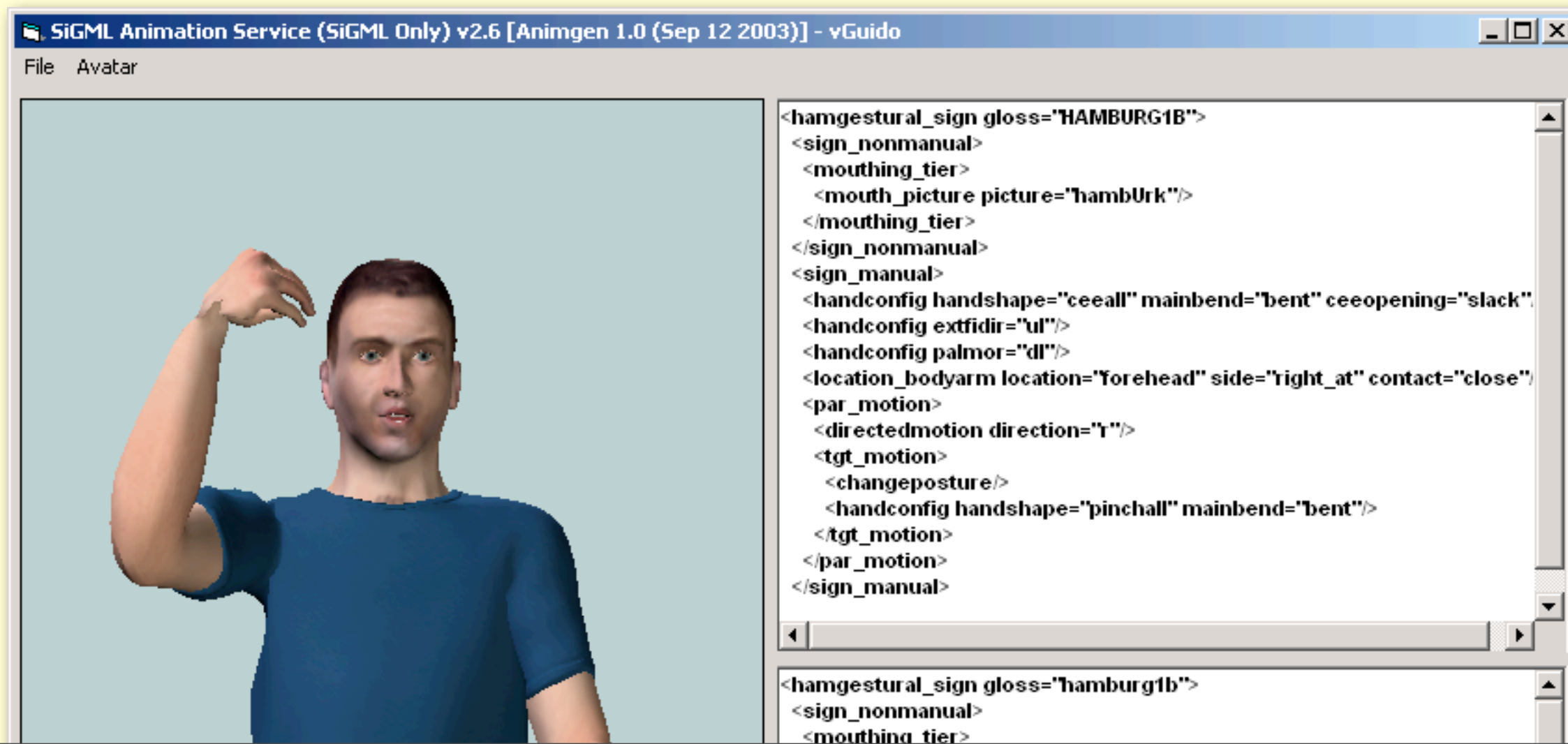
BESCHREIBUNG  
Frucht (z. B. Zitronenhälfte) festhalten und auf einer Zitronenpresse (linke Hand) auspressen.

FORMÄHNLICH  
PFEFFERMÜHLE1A: ähnliche Bilder



# Avatar & Prescription= eSIGN

- Production of signed texts
- Evaluation of transcriptions by re-producing the signing



# Teaching

- Find examples to illustrate what you talk about: Search iLex and have instant access to the data, copy the example into your presentation (if permission available)
- Prepare contact sheets to be handed out to students
- Create (quick & dirty) eLearning materials

# eLearning

- Idea: Transcripts are labour-intensive, so make double use of them:
  - CourseBuilder

The screenshot displays the CourseBuilder interface for 'Stufe 1'. It features a video player on the left showing a man speaking. To the right is a grid of recipe images for 'Geflügelleberwurst'. Below the video is a text box with the instruction: 'ALSO GIBT BROT BESTREICHEN KÄSE BELEGEN WURST BELEGEN.' To the right of the recipe grid is another text box: 'Achte darauf, welche Klassifikatoren für welche Objekte eingesetzt werden.' On the far right, a vertical menu contains buttons for 'Klassifikatoren', 'Raumnutzung', 'Rollenübernahme', 'Beenden', and 'Debug'. At the bottom, a navigation bar includes buttons for 'Linguistik', 'Geschichten', 'Geflügelleberwurst', 'Kampfhund', 'Übungen', and 'Wortschatz'.

# iLex & the Rest of the World

- SignStream

- ELAN

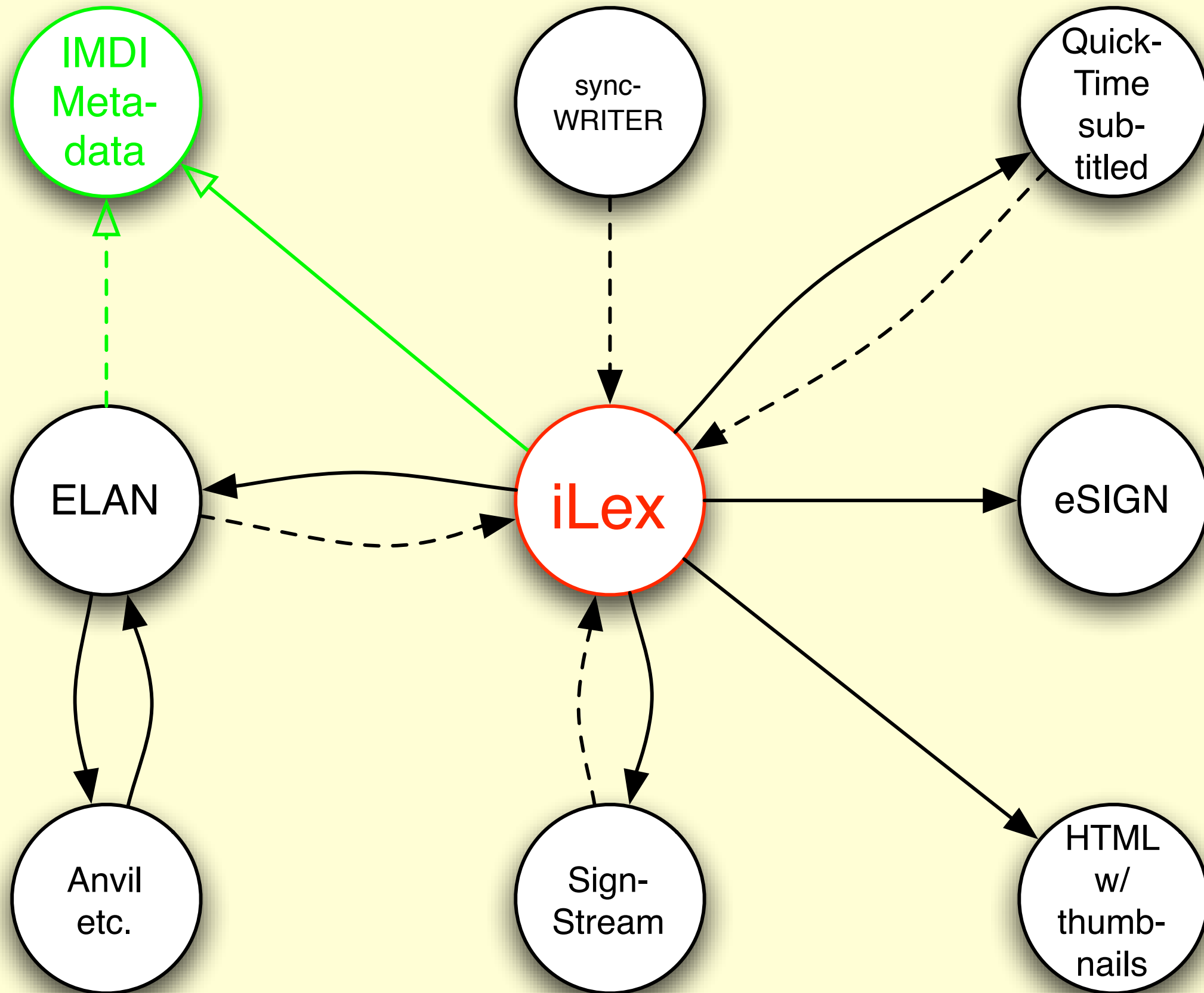
  - Anvil

  - TASX

  - ...

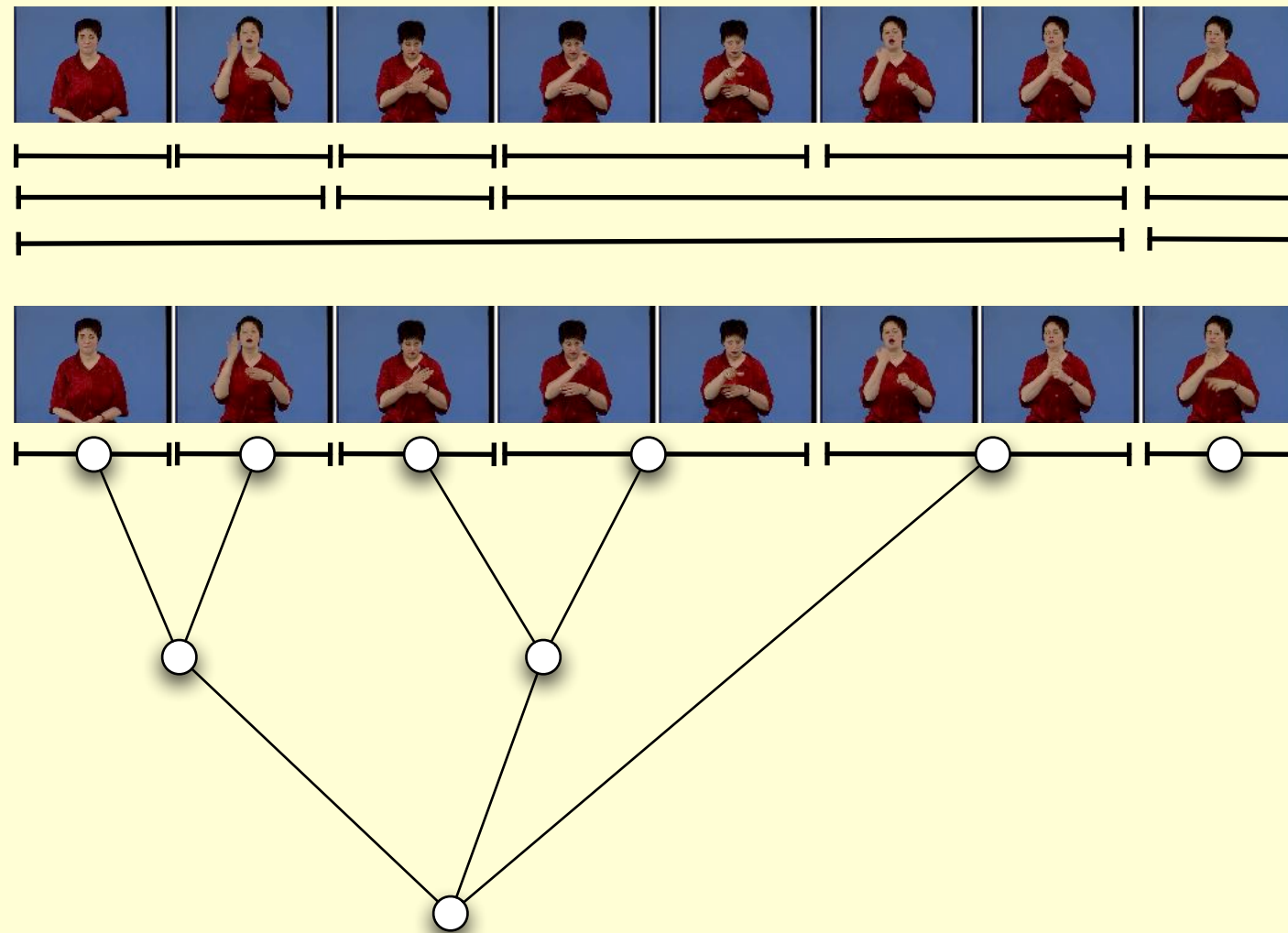
- AnColin

# Data exchange is possible...



# Further Developments?

## ■ Treebanks



# Further Developments?

ANNIS<sup>2</sup>

**Search Form**

AnnisQL: `cat="NP" & pos="VVFIN" & cat="S" & cat="NP" & #3 >[func="OA"] #1 & #3 >[func="SB"] #4 & #1 . #2 &`

Match Count: Calculating...

More Corpora

<input type="checkbox"/> Name	Texts	Token
<input type="checkbox"/> b4.muspilli	1	909
<input type="checkbox"/> b4.tatian	268	1502
<input type="checkbox"/> c6.hindi	1	2218
<input type="checkbox"/> d2.2samplesDEU	2	19
<input type="checkbox"/> falko.essay	133	66000
<input checked="" type="checkbox"/> pcc-11	11	1939
<input type="checkbox"/> spec	4	2823

Simple Search | **Query Builder** | Statistics

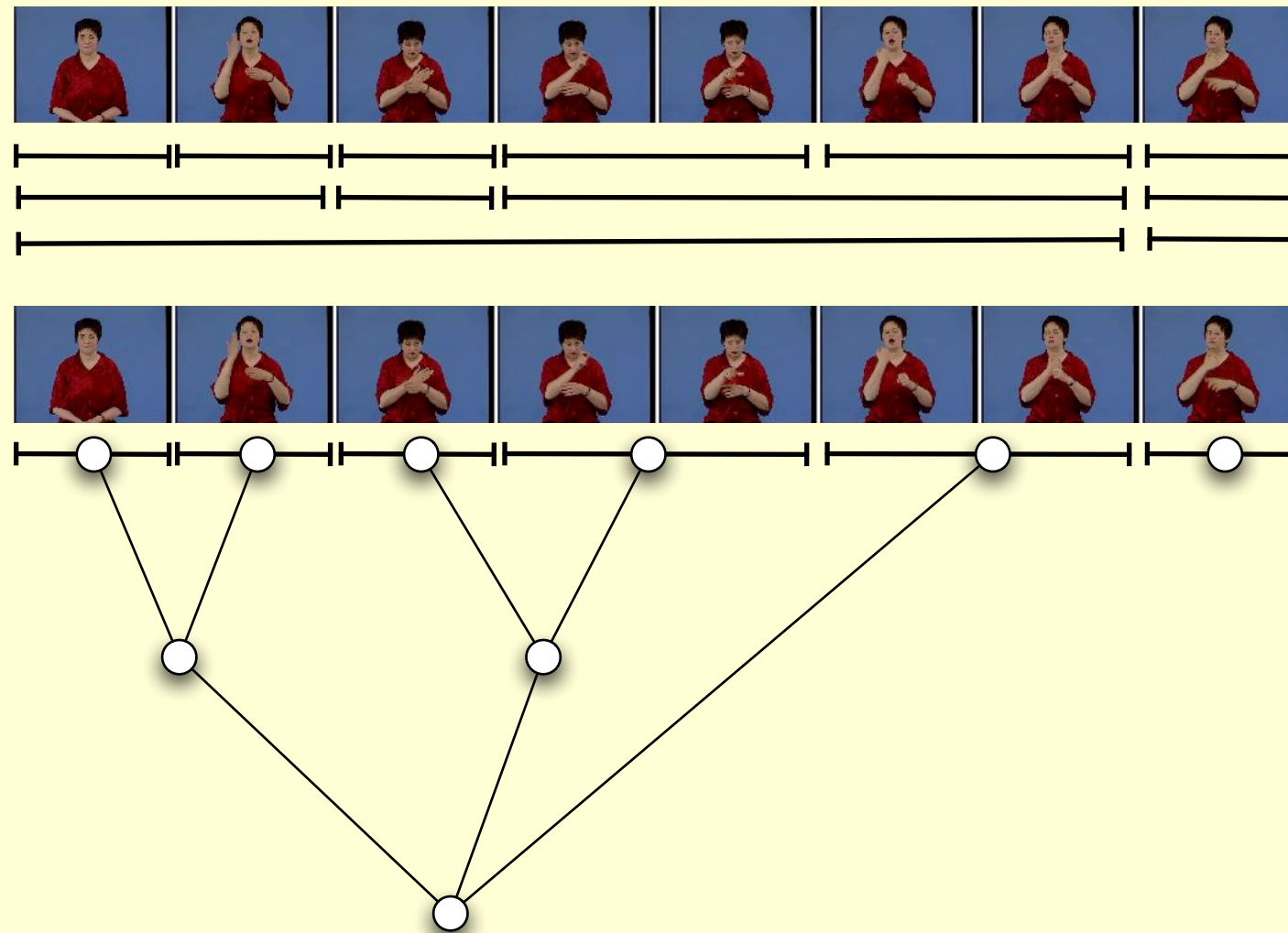
Show Result

**Create Node**

```
graph TD; Root["Edge Add Clear X  
Field op Value  
cat = S"]; Node1[">"]; Node2[">"]; Node3["Edge Add Clear X  
Field op Value  
cat = NP"]; Node4["Edge Add Clear X  
Field op Value  
pos = VVFIN"]; Node5["Edge Add Clear X  
Field op Value  
cat = NP"]; Node6["."]; Node7["."]; Root --- Node1; Root --- Node2; Node1 --- Node3; Node2 --- Node4; Node3 --- Node6; Node4 --- Node5; Node5 --- Node7;
```

# Further Developments?

## ■ Treebanks



## ■ Frame Semantics (Fillmore's FrameNet)



# Statistics

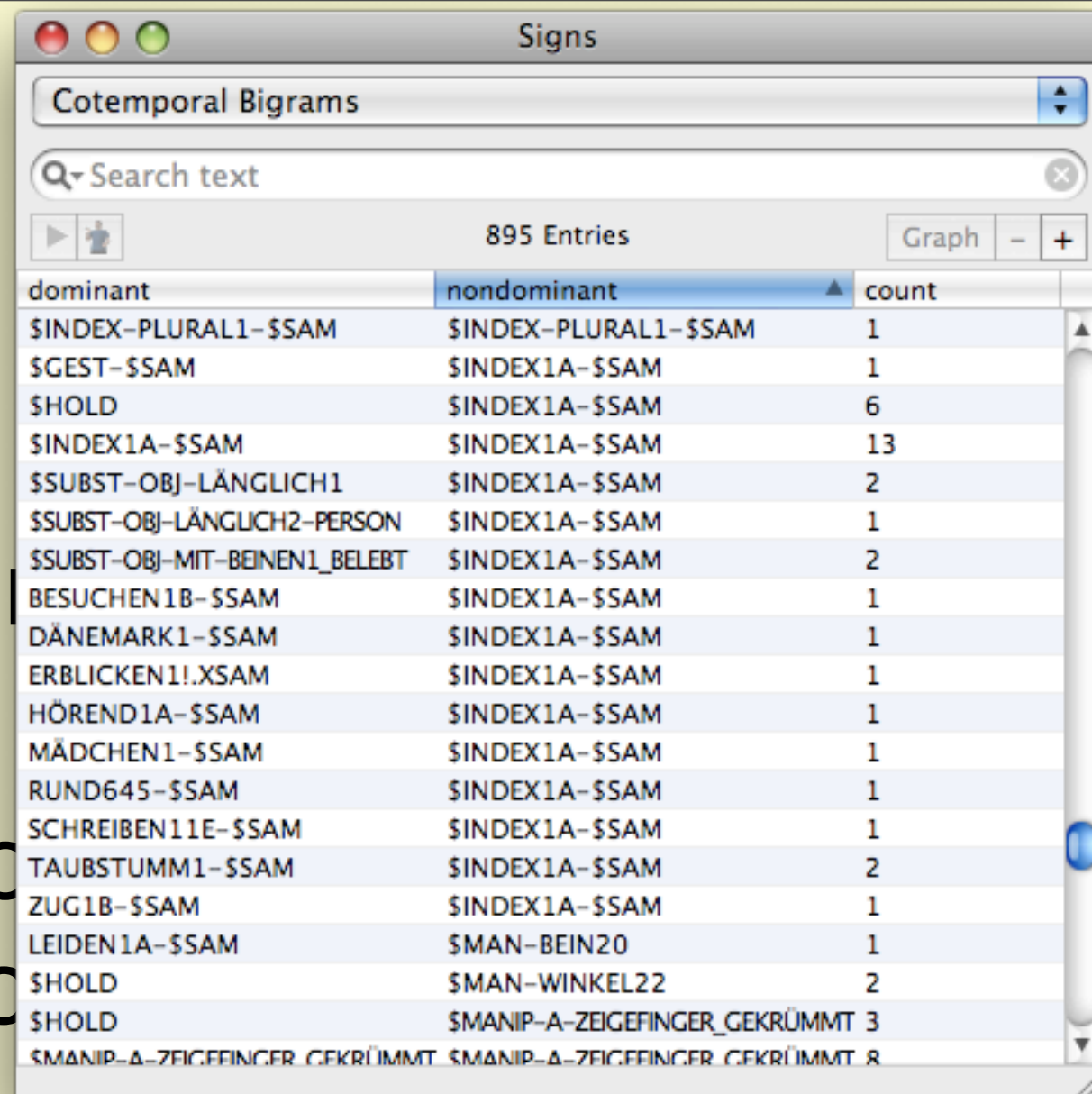
- For the time being, we are quite happy if there is anything worth counting
- How large and balanced should a corpus be before statistics really make sense?
  - Sheer size never an option for sign language

# Counting

- Counting is easy in a relational database
- Views customisable to specific subsets of your database or to show linked data you are interested in
- Bigram analyses adapted to sign language
  - cotemporal bigrams
  - sequential bigrams

# Counting

- Counting is easy in a database
- Views customisable to any part of your database or to any sign you are interested in
- Bigram analyses adapted to sign language
  - coterporal bigrams
  - sequential bigrams



Cotemporal Bigrams

Search text

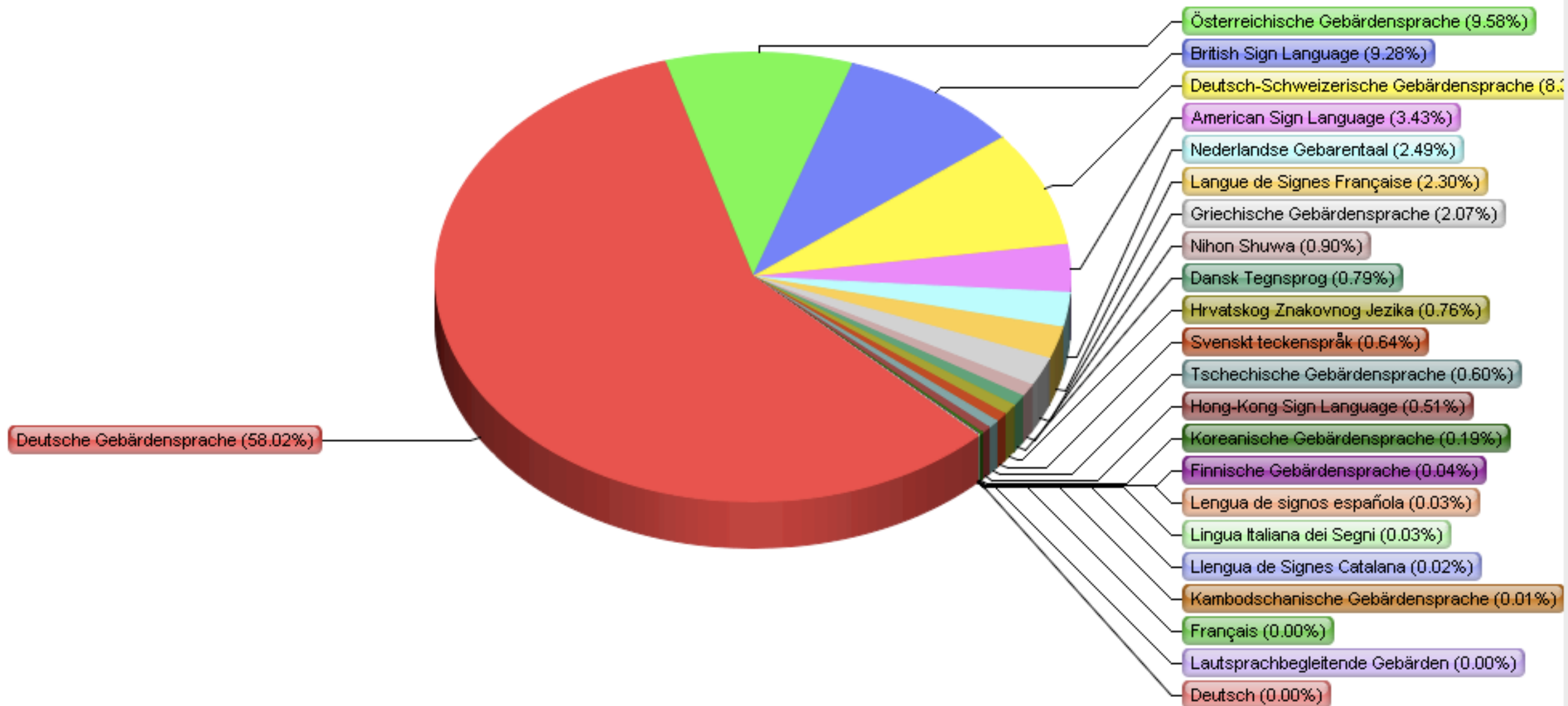
895 Entries

Graph - +

dominant	nondominant	count
\$INDEX-PLURAL1-\$\$SAM	\$INDEX-PLURAL1-\$\$SAM	1
\$GEST-\$\$SAM	\$INDEX1A-\$\$SAM	1
\$HOLD	\$INDEX1A-\$\$SAM	6
\$INDEX1A-\$\$SAM	\$INDEX1A-\$\$SAM	13
\$\$SUBST-OBJ-LÄNGLICH1	\$INDEX1A-\$\$SAM	2
\$\$SUBST-OBJ-LÄNGLICH2-PERSON	\$INDEX1A-\$\$SAM	1
\$\$SUBST-OBJ-MIT-BEINEN1_BELEBT	\$INDEX1A-\$\$SAM	2
BESUCHEN1B-\$\$SAM	\$INDEX1A-\$\$SAM	1
DÄNEMARK1-\$\$SAM	\$INDEX1A-\$\$SAM	1
ERBLICKEN1!.XSAM	\$INDEX1A-\$\$SAM	1
HÖREND1A-\$\$SAM	\$INDEX1A-\$\$SAM	1
MÄDCHEN1-\$\$SAM	\$INDEX1A-\$\$SAM	1
RUND645-\$\$SAM	\$INDEX1A-\$\$SAM	1
SCHREIBEN11E-\$\$SAM	\$INDEX1A-\$\$SAM	1
TAUBSTUMM1-\$\$SAM	\$INDEX1A-\$\$SAM	2
ZUG1B-\$\$SAM	\$INDEX1A-\$\$SAM	1
LEIDEN1A-\$\$SAM	\$MAN-BEIN20	1
\$HOLD	\$MAN-WINKEL22	2
\$HOLD	\$MANIP-A-ZEIGEFINGER_GEKRÜMMT	3
\$MANIP-A-ZEIGEFINGER_GEKRÜMMT	\$MANIP-A-ZEIGEFINGER_GEKRÜMMT	8

# Counts as Graphs

Movies by Languages



# Data Quality

- The deeper the analysis, the higher the costs for inter-transcriber measures
- Consistency
  - Relational database guarantees referential integrity
  - No typos
- Different perspectives on the data
  - Lemma reviewing for a dictionary compilation process reviews transcription data from a different point of view

# Quality

- If we can never reach “big” sign language corpora:
- Other measures about the quality of a corpus needed
  - such as compatibility of results with other methods
  - e.g. cognitive models
- On the lexical level: Which corpus linguistics measures are good predictors for lexical item retrieval time?
  - For spoken language: frequency measures rather uncorrelated

# Thank you for your attention!

- The work described here is partially supported by
  - German Academies of Science programme (DGS-Korpus)
  - German Federal Ministry of Labour & State of Hamburg (GaLex)
  - European Commission, 7th framework IST (Dicta-Sign)