Corpus-based analysis of Weak Drop and Weak Prop in German Sign Language (DGS)

Rie Nishio

University of Hamburg
Institute of German Sign Language and
Communication of the Deaf

mailto: Rie.Nishio@sign-lang.uni-hamburg.de
http://www.sign-lang.uni-hamburg.de/

Comments on the slide:

Paper presented at the workshop "Sign Language Corpora : Linguistic Issues" (held at University College London on 24-25 July, 2009)

The presentation is based upon…

Nishio, Rie (2008): Korpusbasierte Analyse phonologischer Aspekte der Deutschen Gebärdensprache: Zu Handformendistribution und Optionalität der nichtdominanten Hand. Magisterarbeit, Universität Hamburg.
[Nishio, Rie (2008): Corpus-based analysis of phonological aspects in German Sign Language: On handshape distribution and optionality of the nondominant hand. MA thesis, University of Hamburg.]

- Weak Drop (WD)
  one-handed realization of a two-handed sign

- Weak Prop (WP)
  two-handed realization of a one-handed sign

Data and Program iLex (1)

Currently existing "corpus" of IDGS at Hamburg

- data from research projects (dictionaries on special terminologies, learning materials, free conversations, story(re)telling, classifier constructions)
- commercially or non-commercially available movie materials

Data and Program iLex (2)

iLex (Hanke 2002, Hanke & Storz 2008)

- transcription tool with an integrated lexicon
- information stored in a database
- individual data retrievals by SQL queries possible

The study exploits…

- type-token relations
- HamNoSys notation for types / token deviations
- kinds of data elicitation noted for each movie
- metadata on informants
- different tier kinds

Data and Program iLex (2)

iLex (H
- trar
- info
- indi

The stu
- type
- Har
- kinc

elicitation kinds in the study

- interview
- conversation
- narrative
- story(re)telling
- text translation

→ discourse movies

- vocabulary list → vocabulary movies
- elicitation with flashcards → not used in the study

and others…

Study – part 1

Phonological features which facilitate or block WD occurrences:

- *alternation*
- *contact*
- *symmetricity*

(ASL: Battison 1974, Padden & Perlmutter 1987, Brentari 1998; JSL: Miyahara 2000; NGT: van der Kooij 2001)

➡In DGS?

Restricting the data

- only tokens produced by deaf signers
- only types attested by tokens produced by at least two different deaf signers

Information needed

- presence of the features *alternation*, *contact*, *symmetricity* in two-handed signs
- handedness of the signs' underlying form
- Weak Drop tokens

Alternation, contact

HamNoSys contains…

- symbol for alternation ( ) and/or both point-symmetry symbol (,) and repeat symbol (,)
  ⇒*alternation*
- symbol for contact ( )
  ⇒*contact*

Symmetricity

Combination of three questions…

- Does the sign have the same handshapes on both hands?
- Does the sign have symmetrical, same or different movements on both hands?
- Does the sign have symmetrical, same or different orientations (orientation changes) on both hands?

Comments on the slide:

Extracting information on presence/absence of the feature symmetricity was more difficult, because the symmetry symbols in HamNoSys can be canceled in other parts of the HamNoSys string.

The question if a two-handed sign is symmetrical or not can be answered by combination of the above three questions.

Pattern 1: HamNoSys contains no movement symbols.

⇒ (definitely) There is no movement.

Pattern 2: HamNoSys contains a non-movement symbol followed (optionally with other symbols standing in between) by a symbol for alternation.

⇒ (hopefully) Both hands have the same alternating movement.

Pattern 3: HamNoSys contains a non-movement symbol.

⇒ (hopefully) Only the dominant hand moves.

Pattern 4: HamNoSys contains a two-hands-are-distinct-symbol followed by a movement symbol.

⇒ (hopefully) The nondominant hand moves differently than the dominant hand.

Pattern 5: HamNoSys starts with a symmetry symbol or with a symbol for alternation and contains a repeat symbol followed (optionally with other symbols standing in between) by a movement symbol.

⇒ (hopefully) Both hands have the same/symmetrical/alternating movement, and additionally both hands have the same/symmetrical path movement.

Pattern 6: HamNoSys starts with a symmetry symbol.

⇒ (definitely) Both hands have the same/symmetrical movement.

Pattern 7: HamNoSys contains a symmetry symbol which does not stand at the beginning.

⇒ (hopefully) Only the dominant hand moves, and additionally both hands have the same/symmetrical path movement.

Pattern 8: HamNoSys contains a symbol for alternation.

⇒ (hopefully) Both hands have the same alternating movement.

Pattern 9: HamNoSys exhibits no above patterns.

⇒ (definitely) Only the dominant hand moves.

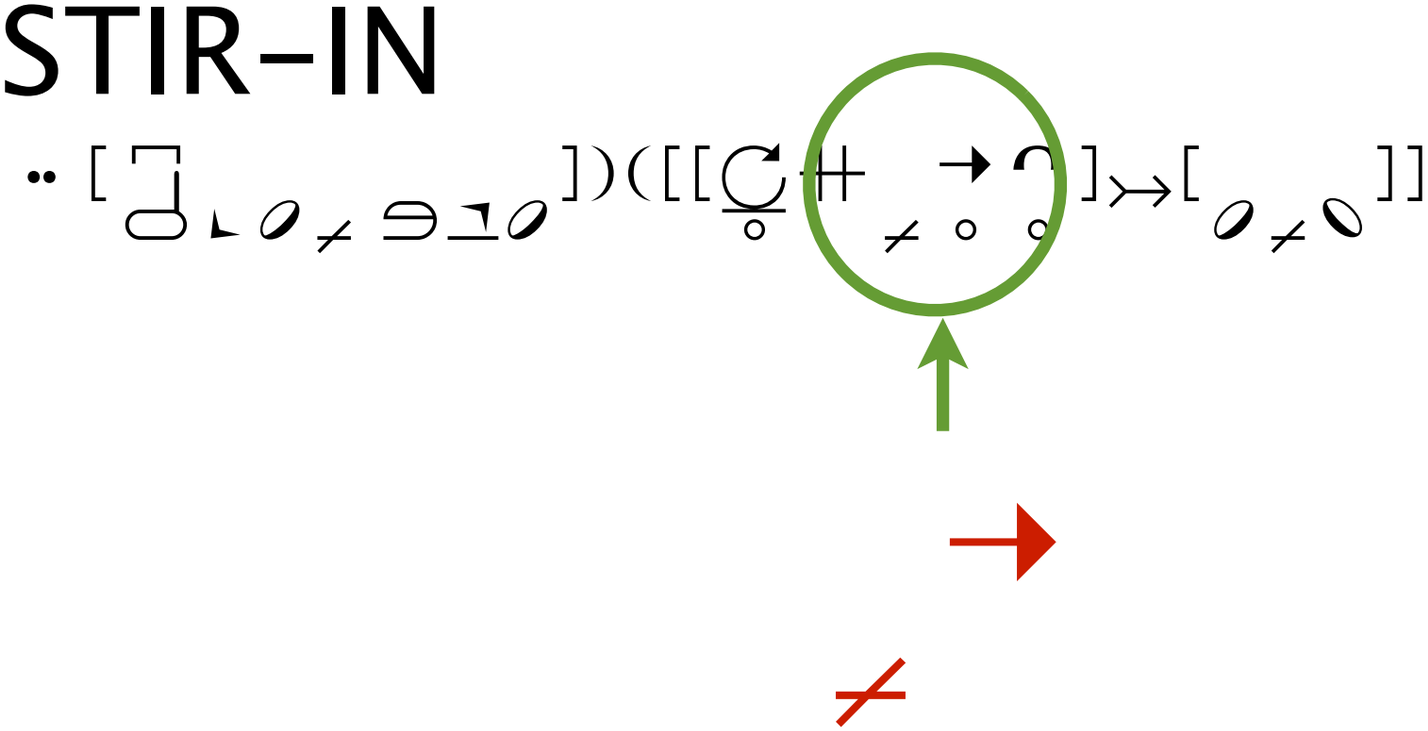
Comments on the slide:

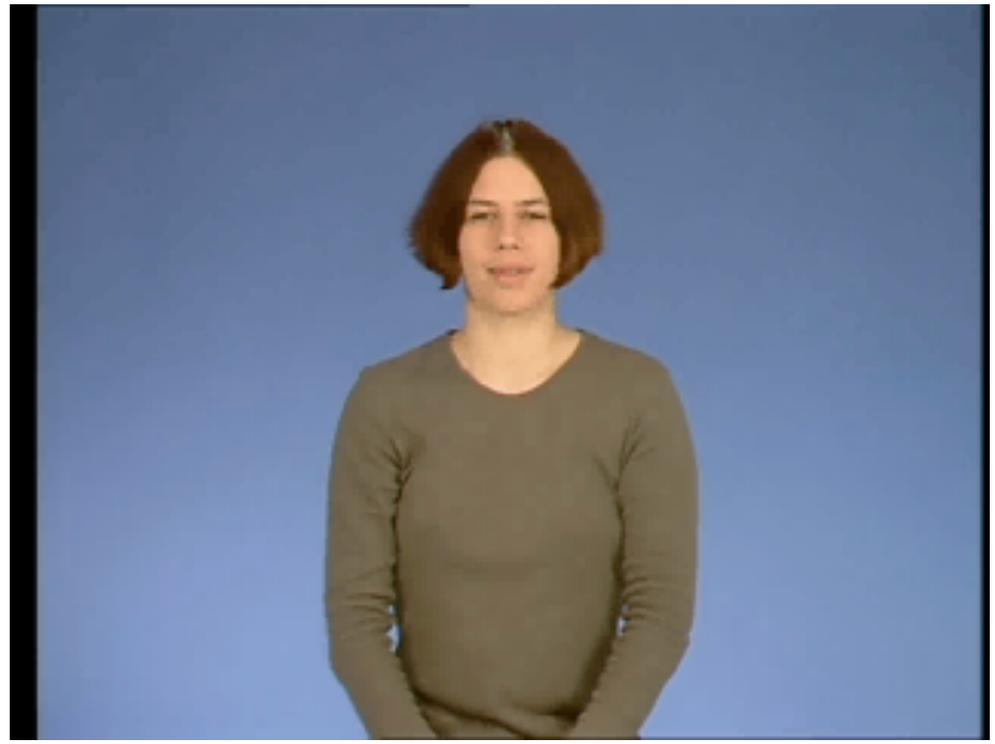
The pattern matching process consisted of 9 HamNoSys patterns ordered in a certain sequence, and it returned an answer as soon as the HamNoSys string matched the pattern. The content of the patterns and their order were both important.

(Some above patterns do not conform to the official HamNoSys syntax but are used in our projects, so they also needed to be taken into account.)

Pattern 4: HamNoSys contains a two-hands-are-distinct-symbol followed by a movement symbol.
⇒ (hopefully) The nondominant hand moves differently than the dominant hand.

STIR–IN





* HamNoSys rule: [dom. H. ≠ nondom. H.]

Pattern 1: HamNoSys contains no movement symbols.
⇒ (definitely) There is no movement.

Pattern 2: HamNoSys contains a non-movement symbol followed (optionally with other symbols standing in between) by a symbol for alternation.
⇒ (hopefully) Both hands have the same alternating movement.

Pattern 3: HamNoSys contains a non-movement symbol.
⇒ (hopeful

Pattern 4: Ha a movement symbol.
⇒ (hopeful and.

Pattern 5: Ha ernation and contains a repe etween) by a movement sym
⇒ (hopefu vement, and additionally

Pattern 6: Ha
⇒ (definitely) Both hands have the same/symmetrical movement.

Pattern 7: HamNoSys contains a symmetry symbol which does not stand at the beginning.
⇒ (hopefully) Only the dominant hand moves, and additionally both hands have the same/symmetrical path movement.

Pattern 8: HamNoSys contains a symbol for alternation.
⇒ (hopefully) Both hands have the same alternating movement.

Pattern 9: HamNoSys exhibits no above patterns.
⇒ (definitely) Only the dominant hand moves.

- Perl with regular expressions
- "hopefully"
  - ➡ 168 of 2845 two-handed types had to be manually inspected and 19 of them corrected

Identifying handedness of underlying form

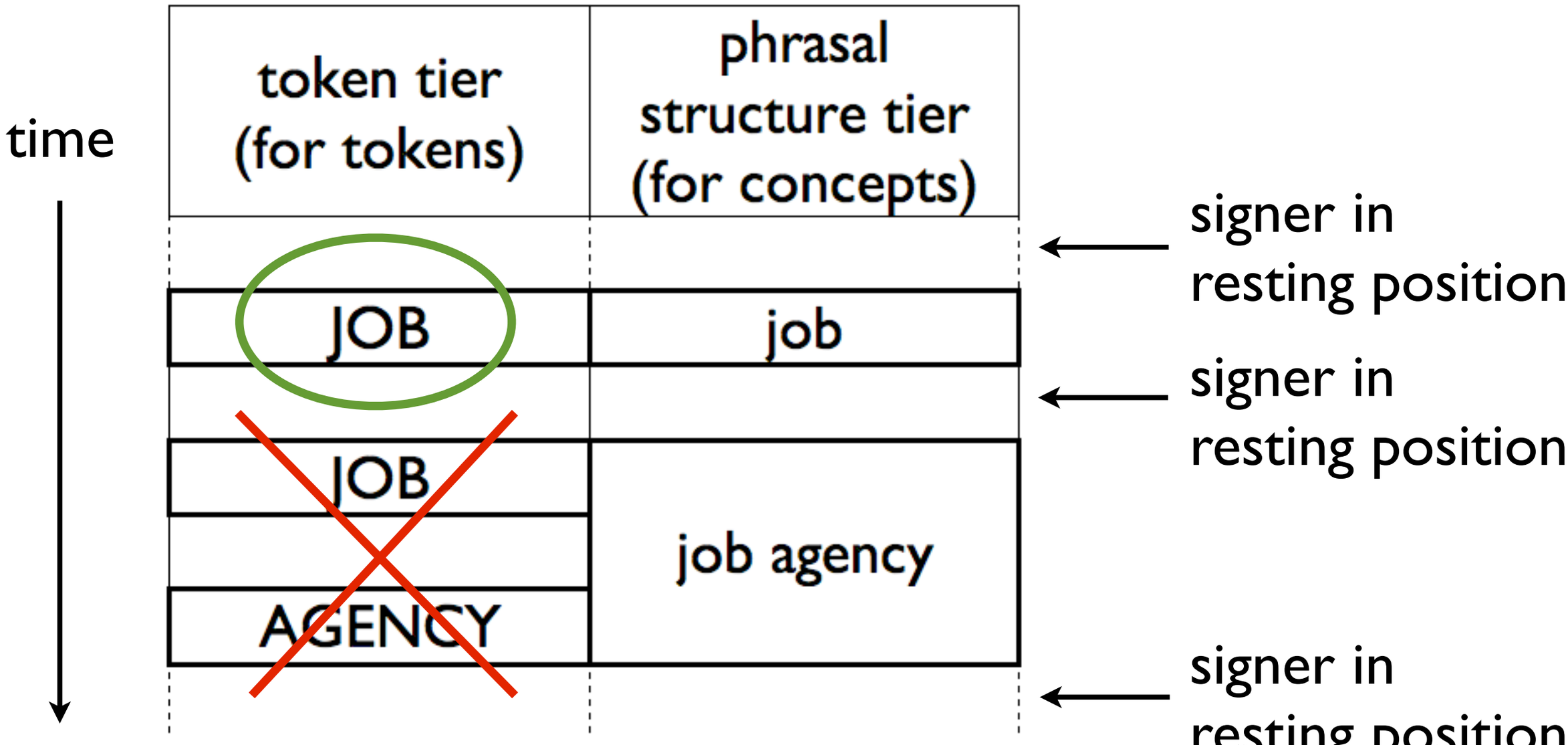
- use of vocabulary movies
- more than 80% of the tokens of a particular type in vocabulary movies are two-handed

$\Rightarrow$ the underlying form of the sign is two-handed

Comments on the slide:

Before talking about WD, one has to be sure that the underlying form of the sign is two-handed. Because this study was supposed to be corpus-based, I decided to use the vocabulary movies in the corpus for the purpose of determining the underlying form.

Excluding compounds etc.



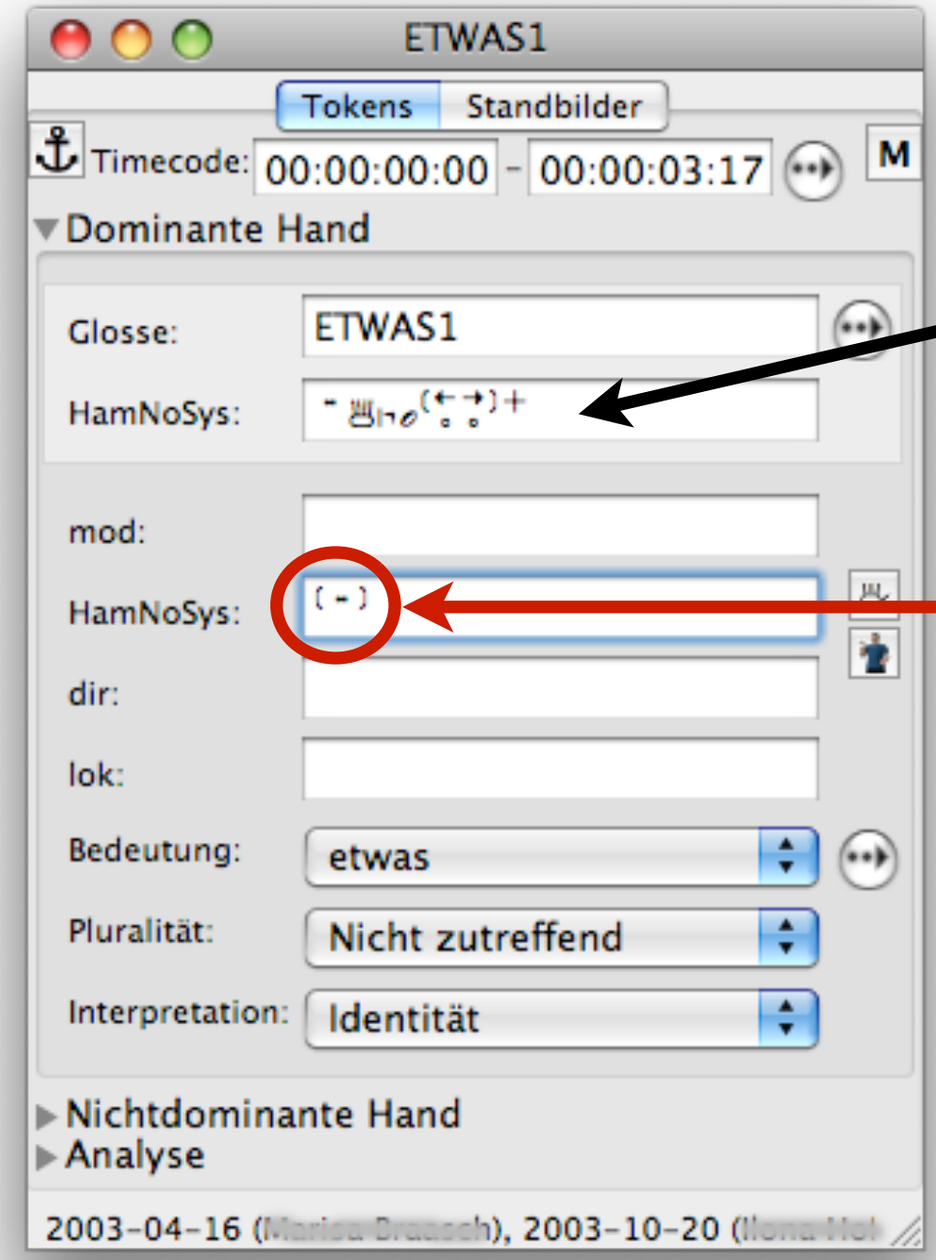
Comments on the slide:

The vocabulary movies sometimes contained not only isolated signs but also compounds and phrasal expressions in which the handedness of the signs might have been influenced by the preceding and/or the following sign. Those items needed to be excluded.

Fortunately, in most cases the transcripts of the vocabulary movies contained a phrasal structure tier whose parent tier was a token tier. Tags in a phrasal structure tier group tokens in the token tier into constituents.

So, tokens were excluded if a tag in a phrasal structure tier covered more than one tag in the token tier.

Weak Drop Tokens



Type HamNoSys

Token HamNoSys (deviation)

(..)

Comments on the slide:

WD tokens could be detected using the HamNoSys notation of token deviations.

(In the above example the type is a two-handed symmetrical sign, but the token is realized one-handed. The deviation is notated using a symmetry symbol in parentheses.)

Signs with WD tokens

	discourse tokens							vocabulary tokens				
gloss	1h (WD)		2h		simult.		total	1h		2h		total
ACHE										12	92%	13
ACHIEVE										6	100%	6
EMPTY										5	100%	5
HEALTH										4	100%	4
HELP										9	100%	9
HOW										7	100%	7
SPECIAL										6	100%	6
SUDDENLY										8	100%	8
WHAT	8	5%	128	86%	13	9%	149	1	7%	13	93%	14

- 156 WD tokens of 46 types found
- discourse movies have more than 15,000 tokens (ca. 1% are WD)

Signs with WD tokens

	discourse tokens							vocabulary tokens				
gloss	1h (WD)		2h		simult.		total	1h		2h		total
ACHE	9	38%	7	29%	8	33%	24	1	8%	12	92%	13
ACHIEVE	3	17%	14	78%	1	6%	18	0	0%	6	100%	6
EMPTY	1	25%	3	75%	0	0%	4	0	0%	5	100%	5
HEALTH	5	33%	10	67%	0	0%	15	0	0%	4	100%	4
HELP	1	1%	162	99%	0	0%	163	0	0%	9	100%	9
HOW	2	11%	16	84%	1	5%	19	0	0%	7	100%	7
SPECIAL	1	20%	4	80%	0	0%	5	0	0%	6	100%	6
SUDDENLY	40	35%	48	42%	25	22%	113	0	0%	8	100%	8
WHAT	8	5%	128	86%	13	9%	149	1	7%	13	93%	14

Comments on the slide:

Example sign WHAT:

13 two-handed tokens and 1 one-handed token were found in the vocabulary movies. So, this sign was determined to be two-handed in its underlying form (cf. the 80% threshold).

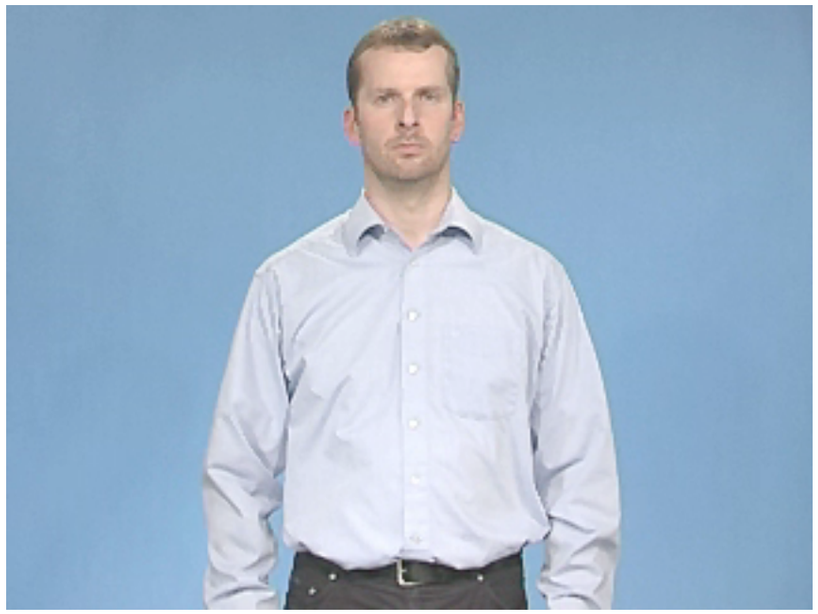
In the discourse movies 8 WD tokens and 128 two-handed “normal” tokens were found. 13 cases turned out to be some kind of simultaneous construction with one-handed WHAT on the dominant hand and, for example, a hold of the previous sign on the nondominant hand.

Result – part 1

	signs total	signs with WD tokens	(%)
symmetrical	700	40	5.7%
non-symmetrical	447	6	1.3%
signs total	1147	46	4.0%
symmetrical and non-alternating	591	35	5.9%
symmetrical and alternating	109	5	4.6%
symmetrical signs total	700	40	5.7%
without contact	596	30	5.0%
with contact	548	16	2.9%
signs total	1147	46	4.0%

Comments on the slide:

The table shows that symmetrical signs, non-alternating signs and signs without contact are more likely to have WD tokens than non-symmetrical signs, alternating signs and signs with contact.



ACHE

+alternation; +contact (body)

SPECIAL

+alternation

ACHIEVE

+contact (hand)

EMPTY

+contact (hand, continuous)

Comments on the slide:

However, there *are* some alternating signs and signs with contact (or even with continuous contact) that turned out to have WD tokens.

Study – part 2

- handedness of surrounding signs of the WD and WP tokens

Surrounding signs are…

- the preceding and the following signs in the same tier with a maximal gap of 7 frames (i.e. 0.28 sec. in PAL)

Comments on the slide:

Setting a threshold was necessary, because in some transcripts not everything had been transcribed, but only those parts or those signs that were relevant to a certain project.

After looking at several transcripts, 7 frames were determined as an interval that two successively produced signs would maximally allow (rather practical decision).

Pattern 1: The tag has a token on the nondominant hand.

$\Rightarrow$ two-handed

Pattern 2: The token HamNoSys contains a set of symbols indicating the presence of perseverence/anticipation on the nondominant hand.

$\Rightarrow$ two-handed

Pattern 3: The token HamNoSys contains a left parenthesis followed by a symmetry symbol or a two-hands-are-distinct-symbol.

$\Rightarrow$ one-handed

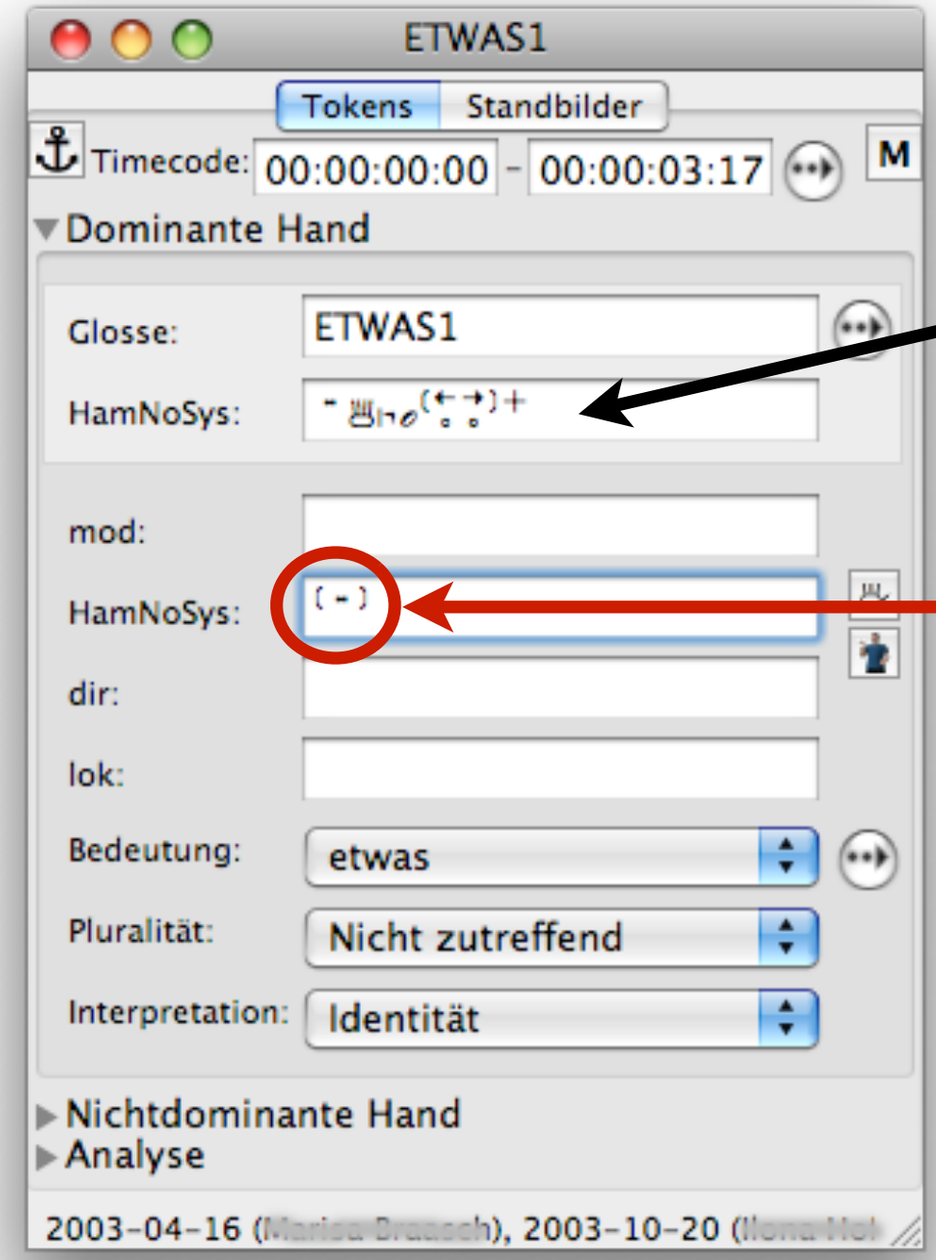
Pattern 4: The token HamNoSys contains a symmetry symbol or a two-hands-are-distinct-symbol.

$\Rightarrow$ two-handed

Pattern 5: None of above patterns

$\Rightarrow$ inherit the handedness of the type

Token HamNoSys (deviation)



Type HamNoSys

Token HamNoSys (deviation)

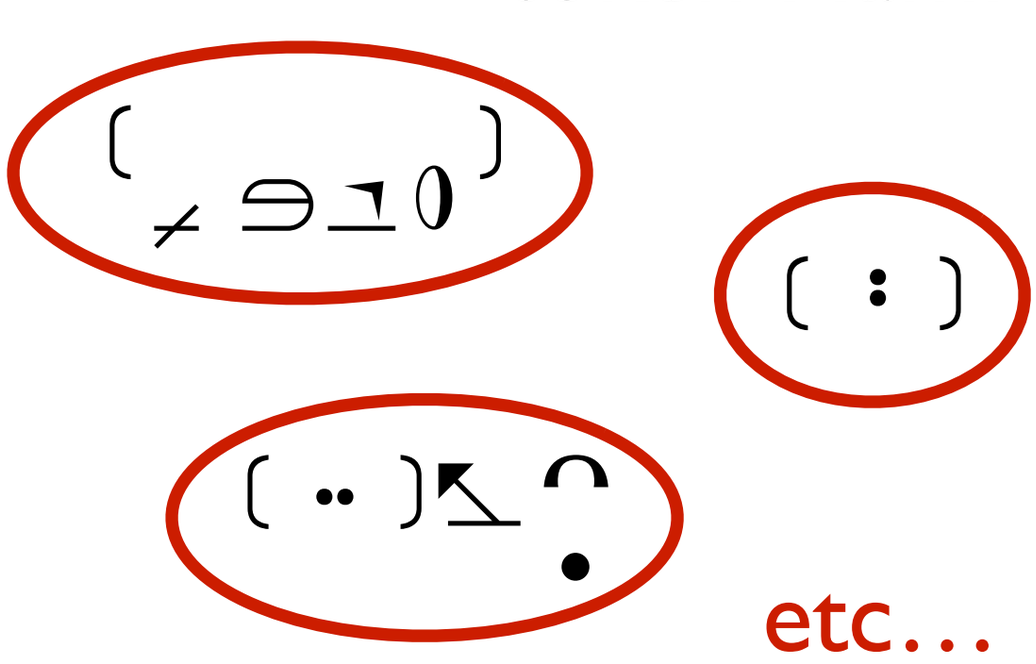
Pattern 3: The token HamNoSys contains a left parenthesis followed by a symmetry symbol or a two-hands-are-distinct-symbol.

⇒ one-handed

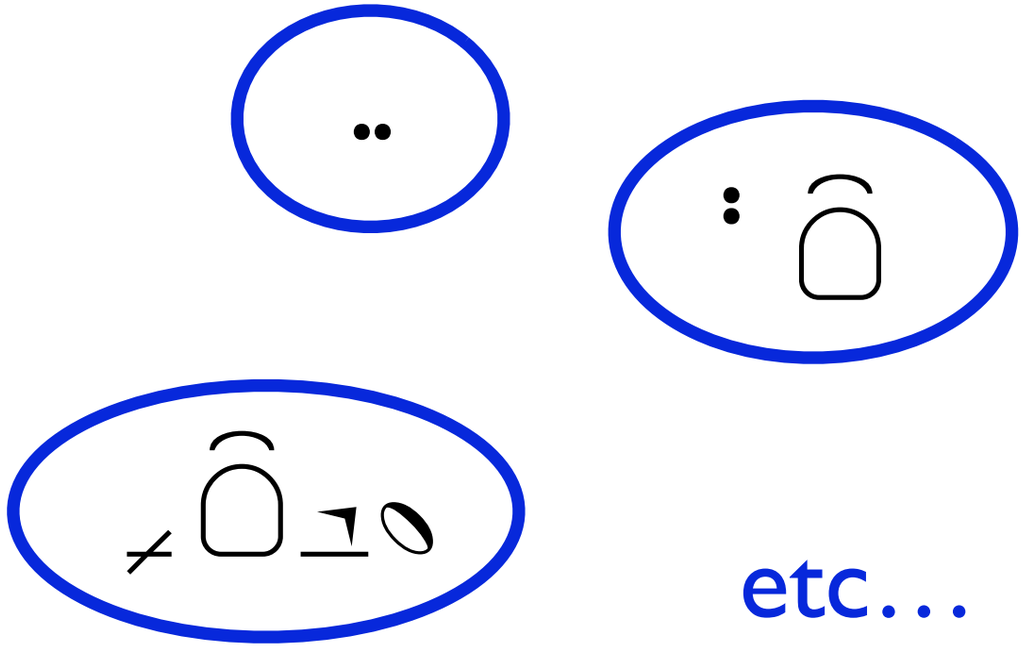
Pattern 4: The token HamNoSys contains a symmetry symbol or a two-hands-are-distinct-symbol.

⇒ two-handed

token HamNoSys (deviation)



⇒ one-handed



⇒ two-handed

Comments on the slide:

The pattern 3 is processed first, then the pattern 4. Note that the content of the patterns *and* their order are both important in this procedure. If the pattern 4 is processed first, all the results would be two-handed.

Pattern 1: The tag has a token on the nondominant hand.

⇒ two-handed

Pattern 2: The token HamNoSys contains a set of symbols indicating the presence of perseverence/anticipation on the nondominant hand.

⇒ t

Patter arenthesis followed re-distinct-symbol.

⇒ c

- implemented on surrounding tags of WD/WP and on all tags in discourse movies for comparison

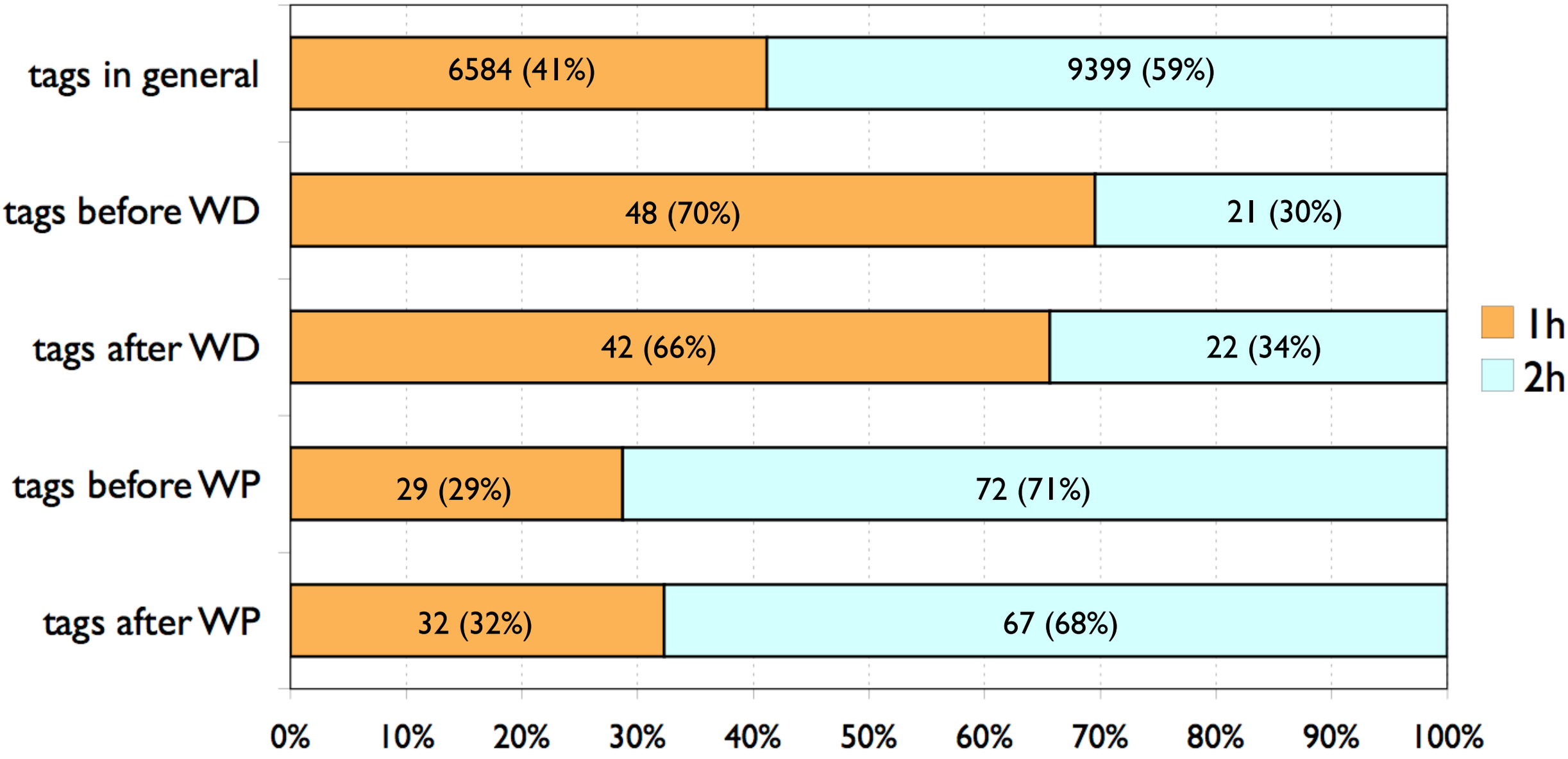
Pattern 4: The token HamNoSys contains a symmetry symbol or a two-hands-are-distinct-symbol.

⇒ two-handed

Pattern 5: None of above patterns

⇒ inherit the handedness of the type

Result – part 2



Comments on the slide:

The handedness distribution of the preceding and the following signs of WD or WP deviate from the general distribution of one-handed and two-handed signs.

The result reveals that WD occurs more likely in a one-handed environment whereas WP occurs more likely in a two-handed environment.

Potential error sources

– Data & design of the study –

Data

- size of the corpus
- naturalness of the data
- transcription
- HamNoSys notation

Study design

- signers
- underlying handedness
  (threshold, use of vocabulary movies)
- automatization

Comments on the slide:

Data:

- size of the corpus
  Today, the corpus is still so small that obtaining reliable results is very challenging.
- naturalness of the data
  Many of the discourse data in the corpus are not really natural data and most of them are monologues. Therefore we do not know what WD and WP look like in natural interactions.
- transcription
  The corpus was not newly compiled for this study. Those who transcribed the data, including the presenter, did not know that someone would be especially interested in handedness in the future. This might have a positive effect, but it might be negative at the same time, because the transcribers did not pay any special attention to the aspects studied.
- HamNoSys notation
  Unfortunately, HamNoSys notation in the corpus often contains some mistakes or does not conform to the official HamNoSys. Although such deviating conventions and typical mistakes were taken into account, erroneous results are still possible. (Another thing about HamNoSys that might be worth mentioning is that even the official notation was sometimes **not** sufficient for investigating important aspects such as distinction between continuous and non-continuous contacts. This is not exactly an error source, but meant restrictions on what could be done by using the notation.)

Comments on the slide (cont.):

Study design

- signer
  The data used in the study also contained those produced by late learners. So it is possible that native signers would produce no WD tokens for a certain type which was identified as allowing WD. (Restriction of the data only to those produced by early learners would have made the dataset too small for a quantitative study, not because there are not enough early learners in the corpus, but simply because many signers who *are* early learners are today not coded as such. This is due to the fact that the "informants" in the corpus include not only the real informants of research projects in which social data were collected, but also actors in other movie materials of whom no social data exist.)
- underlying handedness
  In determining the underlying form the threshold of 80% was applied. A different threshold might have yielded a different result. The study probably lost some portion of WD cases, because no vocabulary tokens were available for certain types or the number of the vocabulary tokens was sometimes so small that the decision on the underlying handedness was probably not reliable. So, it might have been better to conduct an additional study with native signers to identify or validate the underlying handedness.
- automatization
  The study attempted to automatize as many procedures as possible. This might have caused losing some cases that should have been found.

Potential solutions

- larger corpus consisting of natural data produced by native signers coded thoroughly and consistantly
- consider some revisions on HamNoSys and the convention of its use
- integration of additional studies to the corpus-based analysis for validation

Thank you!

Submitted abstract:

The study discussed here investigates phonological factors related to Weak Drop (one-handed realization of a two-handed sign) and Weak Prop (two-handed realization of a one-handed sign) in DGS by using a currently existing corpus which consists of data from various research projects and commercially or non-commercially available movie materials already transcribed. The program iLex (Hanke 2002, Hanke & Storz 2008) with which the corpus has been compiled allows researchers to conduct individual data retrievals by SQL queries. The present study exploits special characteristics of the data and the data structure of the corpus such as type-token relations, HamNoSys notations for types as well as for token deviations, different kinds of data elicitation noted for each movie and metadata on informants.

The first part of the study concerns phonological features such as symmetricity, alternation and contact that are reported to facilitate or block Weak Drop occurrences (Battison 1974, Padden & Perlmutter 1987, Brentari 1998 on ASL; Miyahara 2000 on JSL; van der Kooij 2001 on NGT). The question here is whether and to what extent DGS signs in the corpus with or without these features have Weak Drop tokens. The presence of the features in signs (types) was detected automatically by taking the HamNoSys syntax into account, by means of regular expressions in SQL queries as well as Perl functions prepared for this purpose. In addition, the handedness of the signs' underlying forms was determined by consulting tokens from vocabulary movies produced in studio settings. Finally, signs (types) with at least one Weak Drop token were identified by the notation of token deviations. In order to exclude idiosyncratic signs from the study, only those signs attested by tokens produced by at least two different deaf informants were taken into account. The result shows that symmetrical signs, non-alternating signs and signs without contact are more likely to have Weak Drop tokens than non-symmetrical signs, alternating signs and signs with contact, though there *are* some signs from the latter group which undergo Weak Drop.

The second part of the study focuses on the handedness of the surrounding signs of the identified Weak Drop or Weak Prop tokens. The preceding and the following signs in the same tier with a maximal gap of 7 frames (i.e. 0.28 sec. in PAL) were taken into account. In order to identify the handedness a control structure was used in which several patterns including the existence of a simultaneous token on the nondominant hand and the notation of token deviations concerning handedness were inspected to return the appropriate handedness. Only when the tokens did not exhibit these patterns, the handedness identified for types was inherited. The distribution of the handedness of the preceding and the following signs turned to deviate from the general distribution of one-handed (41%) and two-handed (59%) signs in the signed texts. The result reveals that Weak Drop occurs more likely in the one-handed environment whereas Weak Prop occurs more likely in the two-handed environment. Signs preceding Weak Drop tokens are especially prominent in this respect as 70% of them turned out to be one-handed.

References:

Battison, Robbin (1974): Phonological deletion in American Sign Language. In: *Sign Language Studies* 3:5, 1-19.

Brentari, Diane (1998): *A prosodic model of sign language phonology.* Cambridge, MA: MIT Press.

Hanke, Thomas (2002): iLex – A tool for sign language lexicography and corpus analysis. In: González Rodriguez, Manuel & Paz Suarez Araujo, Carmen (eds.): *Proceedings of the third International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain.* Paris: ELRA, 923-926.

Hanke, Thomas & Storz, Jacob (2008): iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. In: Crasborn, Onno et al. (eds.): *LREC 2008. 6th International Conference on Language Resources and Evaluation. Workshop Proceedings. W25. 3rd Workshop on the Representation and Processing of Sign Languages. Sunday 1st June 2008, Marrakech – Morocco.* Paris: ELRA, 64-67.

Kooij, Els van der (2001): Weak drop in Sign Language of the Netherlands. In: Dively, Valeriedt et al. (eds.): *Signed languages. Discoveries from international research.* Washington, D.C.: Gallaudet University Press, 27-42.

Miyahara, Maiko – 宮原麻衣子 (2000): 日本手話におけるWeak Dropについて. 日本手話学会第26回大会予稿集, 42-43. [On Weak Drop in Japanese Sign Language. In: *Conference proceedings of the 26th meeting of the Japanese Association of Sign Linguistics*, 42-43.]

Padden, Carol A. & Perlmutter, David M. (1987): American Sign Language and the architecture of phonological theory. In: *Natural Language and Linguistic Theory* 5:3, 335-375.