

Towards tagging of multi-sign lexemes and other multi-unit structures

Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer
University of Hamburg, Institute for German Sign Language and Communication of the Deaf

Challenges to the “one token – one tag” relation		General purpose annotation tools such as ELAN	iLex two token tiers	iLex strong & weak token tiers	iLex double-token tier
		<div> <div>text tags</div> <div>token tags</div> <div>type tags</div> </div>			
double-handed signs	<div>Flow of time</div> <div>Flow of time</div>	<div>left hand</div> <div>SIGN</div> <div>right hand</div> <div>SIGN</div> <p>but better avoid two type references:</p> <div>left hand</div> <div>⇒</div> <div>right hand</div> <div>SIGN</div>	<div>left hand</div> <div>SIGN'I</div> <div>right hand</div> <div>SIGN'r</div>	<div>left hand</div> <div>(blocked)</div> <div>right hand</div> <div>SIGN</div>	<div>left hand</div> <div>right hand</div> <div>SIGN</div>
multi-channel signs	<div>Flow of time</div> <div>Flow of time</div>	<div>left hand</div> <div>SIGN</div> <div>right hand</div> <div>SIGN</div> <div>mouth</div> <div>[MG]</div> <p>Same idea as above:</p> <div>left hand</div> <div>⇒</div> <div>right hand</div> <div>SIGN</div> <div>mouth</div> <div>←</div>	<div>left hand</div> <div>SIGN'I</div> <div>right hand</div> <div>SIGN'r</div> <div>mouth</div> <div>←[MG]</div> <p>Nicer, but not yet implemented:</p> <div>left hand</div> <div>SIGN'I</div> <div>right hand</div> <div>SIGN'r</div> <div>mouth</div> <div>SIGN'm</div>	<div>left hand</div> <div>(blocked)</div> <div>right hand</div> <div>SIGN</div> <div>mouth</div> <div>←[MG]</div> <p>Nicer, but not yet implemented:</p> <div>left hand</div> <div>(blocked)</div> <div>right hand</div> <div>SIGN</div> <div>mouth</div> <div>(blocked)</div>	<div>left hand</div> <div>right hand</div> <div>mouth</div> <div>SIGN</div> <div>←[MG]</div> <p>Nicer, but not yet implemented:</p> <div>left hand</div> <div>right hand</div> <div>mouth</div> <div>SIGN</div>
compounds etc.	<div>Flow of time</div>	<div>tokens</div> <div>SIGN1</div> <div>SIGN2</div> <div>or(/and)</div> <div>tokens</div> <div>SIGN1 + SIGN2</div>	<div>tokens</div> <div>SIGN1</div> <div>SIGN2</div> <div>or</div> <div>tokens</div> <div>SIGN1 + SIGN2</div>	<div>tokens</div> <div>SIGN1</div> <div>SIGN2</div> <div>or</div> <div>compound types</div> <div>SIGN1 + SIGN2</div>	<div>tokens</div> <div>SIGN1 + SIGN2</div> <div>or</div> <div>compound components</div> <div>SIGN1</div> <div>SIGN2</div>
holds	<div>Flow of time</div> <div>Flow of time</div>	<div>left hand</div> <div>⇒</div> <div>right hand</div> <div>SIGN</div> <p>Sign and hold separated:</p> <div>left hand</div> <div>⇒</div> <div>↑</div> <div>right hand</div> <div>SIGN</div>	<div>left hand</div> <div>SIGN'I</div> <div>right hand</div> <div>SIGN'r</div> <p>Sign and hold separated:</p> <div>left hand</div> <div>SIGN'I</div> <div>SIGN'I'hold</div> <div>right hand</div> <div>SIGN'r</div>	<div>left hand</div> <div>(blocked)</div> <div>right hand</div> <div>SIGN</div> <p>Looks funny:</p> <div>left hand</div> <div>(blocked)</div> <div>SIGN'hold</div> <div>right hand</div> <div>SIGN</div>	<div>left hand</div> <div>right hand</div> <div>SIGN</div> <div>SIGN'hold</div>

Multi-sign expressions to care about

- ◆ compounds
 - proper compounds
 - loan compounds
- ◆ multi-sign idiomatic expressions
- more than the sum of their components
 - restrictions of use
 - meaning more specific
 - might disambiguate POS attributions of their parts
- entities of their own right
- part of the lexicon
 - lexicographic annotation:
 - e.g. syntactic restrictions, regional use

Tagging necessary to

- ◆ mark, access, list, describe these multi-sign units (not yet general annotation practice)
- ◆ identify them
- ◆ make searching easier (instead of complex search patterns)
- ◆ support lemma revision (document decisions taken: is this an instance of a multi-sign expression or just individual signs?)

Suggested solution inspired by multi-word lexemes treatment in spoken language treebanks:

- ◆ Specify multi-sign expressions as trees (depth≤2) in the lexicon, allowing optional elements where appropriate
- ◆ Dragging such a multi-sign lexeme onto a compatible tags structure makes these tags an instance of the multi-sign lexeme
- ◆ Project tree structures into tag grid as framing basic tags
- ◆ Scales to higher-level structures such as sandwich verb constructs and multi-item list structures
- ◆ Potentially have different framing colours for different types of multi-sign expressions
- ◆ Examples: idiomatic expression and hold, multi-item list structures:

	left hand	right hand	meaning	facial expression
	SIGN-at-me		what you told to me	raised eyebrow
	(hold of SIGN-at-me)	LEAVE-IT	not interesting	„dis-missive“
		OLD	That's an old hat.	„bored“
		SOUP		

	left hand	right hand	meaning
	FIRST		first
	(hold / list buoy)	WATER	water
	SECOND		second
	(hold / list buoy)	TEA	tea
	THIRD		third
	(hold / list buoy)	JUICE	juice