The role of theory in sign language corpus annotation

Kearsy Cormier1, Onno Crasborn2, and Richard Bank2
1 DCAL, University College London, 2 Centre for Language Studies, Radboud University

A modern linguistic corpus
A representative, accessible and machine-readable linguistic corpus requires annotation, which requires ID glossing (type-token matching) for lexical signs (Johnston 2001, 2008). Type-token matching is also needed for the constant parts of semi-lexical material, and non-lexical material needs to be identified as such.

Lexical material
Fairly straightforward – these forms are fully specified phonologically and/or listed in citation form in dictionaries

BSL and NGT

Non-lexical material
Material which is fully unconventionnalised and thus unspecified in the lexicon, e.g. non-lexicalised gestures, manual constructed action

Can we use annotation to test linguistic theory?

Theory-neutral annotation?
• Difficult even with written languages (Myers, 2009)
• More so with languages that lack a written standard (Frischberg, 2012)

Semi-lexical material
Signs which are partly specified but partly unspecified, such as pointing signs, classifier / depicting signs

R: MOVE+B
L: MOVE+B
‘cars bump into each other’ (depicting sign)

Handling constructions and constructed action
• Many lexical signs like KEY involve aspects of handling
• For constructions involving handling that are not lexicalised: Are these semi-lexical (handling constructions, Supalla, 1986, Zwitserlood, 2003) or non-lexical (constructed action, Cormier et al., 2012)?
• By annotating all instances of handling (aside from those like KEY) as potential depicting signs (alternative 1a) and attempting to identify type-like characteristics such as movement and handshape, rather than as constructed action (alternative 1b), we can test how conventionalised these structures are within a corpus.

Conclusions
• Corpus annotation can be used to test theoretical assumptions, and is a useful way to deal with several potential approaches. Try one way to see how well the data fits the theory, and then try another. This is corpus-based linguistics (i.e. corpus as method).
• This approach could later lead to corpus-driven approaches (i.e. corpus as theory) in which linguistic constructs themselves emerge from analysis of a corpus. (See Tognini-Bonelli, 2001 for more on this distinction.)

Summary
By annotating the more categorical of two options, one can look at all tokens of each category to test for evidence of conventionalisation.

Role of theoretical assumptions in categorisation
Categorisation into these types can be difficult, partly because of disagreements about linguistic properties of these constructions (e.g., Emmorey, 1999; Schembri, 2003; Crasborn & De Meijer, 2012; Cormier et al., 2013)

Role of theoretical assumptions in categorisation
• Spreading of non-dominant hand has prosodic functions (Sandler, 1999) and discourse functions (‘fragment buoys’, Liddell, 2003)
• Can prosodic and discourse functions be distinguished?
• By attempting to distinguish between intentionally meaningful spreading (fragment buoys, alternative 2a) vs. non-meaningful phonetic perseveration (alternative 2b) in our annotation scheme, we can test to what extent the distinction can be made and thus the extent to which there is evidence of a linguistic distinction between them.

Fragment buoys and perseveration
1a:

GlossL MeaningL
MOVE+ fist
cat picks up bird cage

GlossR MeaningR
MOVE+ fist
cat picks up case

BE+ fist
bird cage
CASE
MOVE+ fist
5
cat throws case away

2b:

GlossL GlossR
G:CA: CAT-PICK-BIRD-CAGE-AND-MOVE
G:CA: CAT-PICK-CASE-AND-MOVE

Role of theoretical assumptions in categorisation
– Spread of non-dominant hand has prosodic functions (Sandler, 1999) and discourse functions (‘fragment buoys’, Liddell, 2003)
– Can prosodic and discourse functions be distinguished?
– By attempting to distinguish between intentionally meaningful spreading (fragment buoys, alternative 2a) vs. non-meaningful phonetic perseveration (alternative 2b) in our annotation scheme, we can test to what extent the distinction can be made and thus the extent to which there is evidence of a linguistic distinction between them.

Conclusions
• Corpus annotation can be used to test theoretical assumptions, and is a useful way to deal with several potential approaches. Try one way to see how well the data fits the theory, and then try another. This is corpus-based linguistics (i.e. corpus as method).
• This approach could later lead to corpus-driven approaches (i.e. corpus as theory) in which linguistic constructs themselves emerge from analysis of a corpus. (See Tognini-Bonelli, 2001 for more on this distinction.)

Summary
By annotating the more categorical of two options, one can look at all tokens of each category to test for evidence of conventionalisation.

Theory-neutral annotation?
• Difficult even with written languages (Myers, 2009)
• More so with languages that lack a written standard (Frischberg, 2012)

Handling constructions and constructed action
• Many lexical signs like KEY involve aspects of handling
• For constructions involving handling that are not lexicalised: Are these semi-lexical (handling constructions, Supalla, 1986, Zwitserlood, 2003) or non-lexical (constructed action, Cormier et al., 2012)?
• By annotating all instances of handling (aside from those like KEY) as potential depicting signs (alternative 1a) and attempting to identify type-like characteristics such as movement and handshape, rather than as constructed action (alternative 1b), we can test how conventionalised these structures are within a corpus.

Lexical material
Fairly straightforward – these forms are fully specified phonologically and/or listed in citation form in dictionaries

BSL and NGT

Non-lexical material
Material which is fully unconventionnalised and thus unspecified in the lexicon, e.g. non-lexicalised gestures, manual constructed action

Can we use annotation to test linguistic theory?

Theory-neutral annotation?
• Difficult even with written languages (Myers, 2009)
• More so with languages that lack a written standard (Frischberg, 2012)

Semi-lexical material
Signs which are partly specified but partly unspecified, such as pointing signs, classifier / depicting signs

R: MOVE+B
L: MOVE+B
‘cars bump into each other’ (depicting sign)

Handling constructions and constructed action
• Many lexical signs like KEY involve aspects of handling
• For constructions involving handling that are not lexicalised: Are these semi-lexical (handling constructions, Supalla, 1986, Zwitserlood, 2003) or non-lexical (constructed action, Cormier et al., 2012)?
• By annotating all instances of handling (aside from those like KEY) as potential depicting signs (alternative 1a) and attempting to identify type-like characteristics such as movement and handshape, rather than as constructed action (alternative 1b), we can test how conventionalised these structures are within a corpus.

Conclusions
• Corpus annotation can be used to test theoretical assumptions, and is a useful way to deal with several potential approaches. Try one way to see how well the data fits the theory, and then try another. This is corpus-based linguistics (i.e. corpus as method).
• This approach could later lead to corpus-driven approaches (i.e. corpus as theory) in which linguistic constructs themselves emerge from analysis of a corpus. (See Tognini-Bonelli, 2001 for more on this distinction.)

Summary
By annotating the more categorical of two options, one can look at all tokens of each category to test for evidence of conventionalisation.

Role of theoretical assumptions in categorisation
Categorisation into these types can be difficult, partly because of disagreements about linguistic properties of these constructions (e.g., Emmorey, 1999; Schembri, 2003; Crasborn & De Meijer, 2012; Cormier et al., 2013)

Fragment buoys and perseveration
• Spreading of non-dominant hand has prosodic functions (Sandler, 1999) and discourse functions (‘fragment buoys’, Liddell, 2003)
• Can prosodic and discourse functions be distinguished?
• By attempting to distinguish between intentionally meaningful spreading (fragment buoys, alternative 2a) vs. non-meaningful phonetic perseveration (alternative 2b) in our annotation scheme, we can test to what extent the distinction can be made and thus the extent to which there is evidence of a linguistic distinction between them.

Conclusions
• Corpus annotation can be used to test theoretical assumptions, and is a useful way to deal with several potential approaches. Try one way to see how well the data fits the theory, and then try another. This is corpus-based linguistics (i.e. corpus as method).
• This approach could later lead to corpus-driven approaches (i.e. corpus as theory) in which linguistic constructs themselves emerge from analysis of a corpus. (See Tognini-Bonelli, 2001 for more on this distinction.)

Summary
By annotating the more categorical of two options, one can look at all tokens of each category to test for evidence of conventionalisation.

Role of theoretical assumptions in categorisation
Categorisation into these types can be difficult, partly because of disagreements about linguistic properties of these constructions (e.g., Emmorey, 1999; Schembri, 2003; Crasborn & De Meijer, 2012; Cormier et al., 2013)

Fragment buoys and perseveration
• Spreading of non-dominant hand has prosodic functions (Sandler, 1999) and discourse functions (‘fragment buoys’, Liddell, 2003)
• Can prosodic and discourse functions be distinguished?
• By attempting to distinguish between intentionally meaningful spreading (fragment buoys, alternative 2a) vs. non-meaningful phonetic perseveration (alternative 2b) in our annotation scheme, we can test to what extent the distinction can be made and thus the extent to which there is evidence of a linguistic distinction between them.

Conclusions
• Corpus annotation can be used to test theoretical assumptions, and is a useful way to deal with several potential approaches. Try one way to see how well the data fits the theory, and then try another. This is corpus-based linguistics (i.e. corpus as method).
• This approach could later lead to corpus-driven approaches (i.e. corpus as theory) in which linguistic constructs themselves emerge from analysis of a corpus. (See Tognini-Bonelli, 2001 for more on this distinction.)

Summary
By annotating the more categorical of two options, one can look at all tokens of each category to test for evidence of conventionalisation.
The role of theory in sign language corpus annotation

Kearsy Cormier1, Onno Crasborn2, and Richard Bank2
1 DCAL, University College London, 2 Centre for Language Studies, Radboud University

References


Contact:

DCAL: Deafness, Cognition & Language Research Centre, University College London, 49 Gordon Square, London, WC1H 0PD Tel: +44 (0)20 76798679 Fax: +44 (0)20 76798691 Textphone: +44 (0)20 76798693 Email: dcal@ucl.ac.uk Website: http://www.dcal.ucl.ac.uk

CLS: Centre for Language Studies, Radboud University, Erasmusplein 1, 6525 HT Nijmegen Fax: +31 (0)24 36 11070 Textphone: +31 (0)24 36 11516 Email: o.crasborn@let.ru.nl Website: http://www.ru.nl/sign-lang/