The grammatical description as a mediator between text and typologist

Sebastian Nordhoff

Universiteit van Amsterdam
June 18th, 2009
Workshop on Small Tools for Cross-Linguistic Research
Utrecht University
s.nordhoff@uva.nl

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Formalization of data structure in linguistics

- ▶ texts: good progress in formalization
- ▶ typological data: good progress in formalization
- ▶ how to get from texts to typology
- ▶ → grammatical descriptions
- ▶ little formalization work done up to now

Different types of grammars

- computational grammars
  - formalization of how to generate all syntactically valid sentences in a language
  - $G = (N, \Sigma, P, S)$
- mental grammars
  - The way how humans process language
- grammatical descriptions
  - neither of the preceding
  - grammar as a communicative act (Noonan)
  - interesting for comparatists
    - typologists
    - historical linguists

What are grammatical descriptions?

- proposed definition: **GDs are structured generalizations over data**
- **data**
  - data-free linguistics does not allow the writing of GDs
- **generalizations**
  - The Rosetta stone is not a GD
- **structured**
  - random observations about a language are not a GD

What does "data" mean in the context of a GD

- annotated and translated corpus
  - Toolbox
  - ELAN
  - XML (Bow, Hughes and Bird 2003)
- treated as a given for the following discussion

# What does "generalization" mean?

- regular form-meaning-pair
- in all data of the corpus, we find that
  - *the* is used to mark definiteness in English
  - *the* ↔ "DEFINITENESS"
  - *-s* is used to mark plural in Spanish
  - *-s* ↔ "PLURALITY"
  - ...

More on form-meaning pairs

- one to one
- one to many
    - *-ing*-form used for nominalization: *smoking forbidden*.
    - *-ing*-form used for participle: *He came running*.
- many to one
    - preterit and perfect both express reference to past in German
    - no notional difference between the two
- many to many
    - *-s* used for "plural" and "3s"
    - "plural" marked by *-s* and *-en* (*pigs*, *oxen*)
    - *-en* used for causatives
    - ...
- many-to-many is the normal case

More on form-meaning pairs
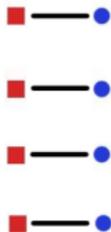
More on form-meaning pairs
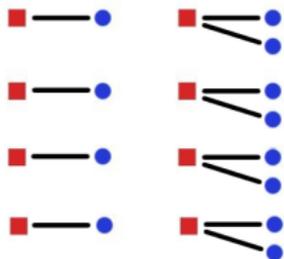
More on form-meaning pairs
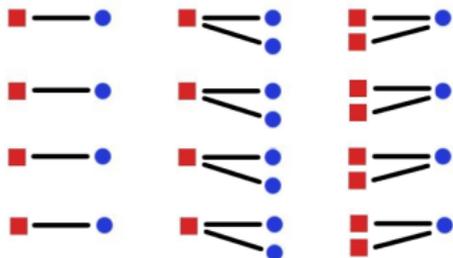
More on form-meaning pairs
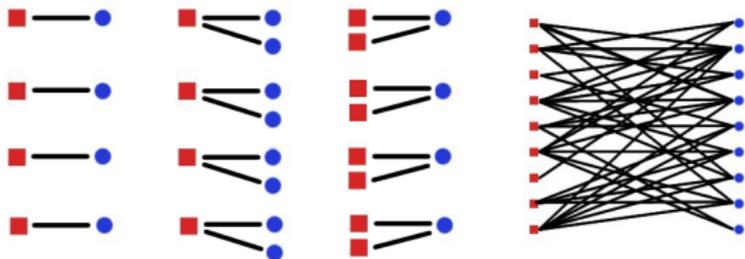
Structure of form-meaning-pairs

- generalizations have no inherent order among them
- every form-meaning-pair is independent from the others
- different structuring principles have been proposed
    - Latinate
    - LDS
    - form-to-function (semasiological)
    - function-to-form (onomasiological)
    - hybrid

The Latinate model

- not applicable everywhere
  - isolating languages
- no place for constructions and idioms
  - certain generalizations tend to end up in random places
  - "instead of the preterit, a temporal adverb can also be used to mark the past"
  - the information contained in the last sentence must be looked up in the morphology chapter
  - not very intuitive placement
- difficult to formalize
- LDS uses the same base

Form-to-function

- all form-meaning pairs are ordered according to form
- useful for diachronic studies
- NEWMAN, P. (2000). The Hausa Language: An Encyclopedic Reference Grammar
- easy to formalize
- a nightmare to use
- neighbours in book are determined by sort order, not by formal or functional similarities
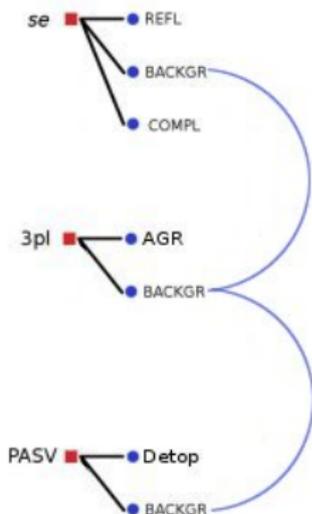
Function-to-form

- inversion of the last model
- all form-meaning pairings are ordered according to meaning/function
- useful for typological research
- sort order unclear
- otherwise easy to formalize
- absence of exhaustive functional structuring principle
- no clear cut examples
  - Frohnmeyer, L. (1913). A progressive grammar of the Malayalam language.
  - Seiler HJ (1977) Cahuilla grammar

Hybrid model

- morphology: form-to-function
- syntax: function-to-form
- impossible to formalize
- Haspelmath (1993) A grammar of Lezgian

Non-linear model

- all preceding models follow a linear-structure
- many-to-many relations are not easy to map on linear structure
- books require linear structure
- web publishing allows for non-linear structure
- most web-publishing is still linear (Hammarström pc)
- some attempts at non-linearity (Nagy, Thieberger, Keegan)
- little theoretical discussion about the principles of non-linear grammar-writing (but see Nordhoff 2008)

New proposal

New proposal

- leave the set of form-meaning pairs unordered
- create a FormPage, which describes a form and links to functions ("-ing")
- create a FunctionPage, which describes how a certain function can be expressed, and link to forms ("completive aspect")
- tag pages for domains to which they pertain
- TagSpatialReference, TagThirdDeclension, TagQuestionFormation, TagInversion
  - aim: restricted set of tags
  - the GALOES platform uses LDS-derived tags
    - LDS1.3.4.1, LDS2.1.1.2

Meaning-bearing items

- what elements can bear meaning?
- (grammatical) morphemes → **morphemicon** (Lehmann)
- constructions (e.g declarative sentence) → **constructicon**
- intonation contours → **contouricon**
    - no metalanguage available as of yet

Non meaning-bearing items

- some linguistic phenomena do not lend themselves to such a treatment
- segmental phonology is treated as a given
  - not linked to function
- noun classes (gender)
- conjugation classes
- these are purely formal phenomena and cannot be treated under the approach presented here

Meaning targets

- structuring of the formal part seems doable
- functional part: comparative concepts (Haspelmath 2007)
- need for precision
- ontology of comparative concepts used in linguistic typology:

Meaning targets

- structuring of the formal part seems doable
- functional part: comparative concepts (Haspelmath 2007)
- need for precision
- ontology of comparative concepts used in linguistic typology: OCCULT

Meaning targets

- ▶ structuring of the formal part seems doable
- ▶ functional part: comparative concepts (Haspelmath 2007)
- ▶ need for precision
- ▶ ontology of comparative concepts used in linguistic typology: OCCULT
- ▶ GOLD ontology is very much on the formal side of the spectrum
- ▶ not so much meaning-based
- ▶ eg nothing called "Reference", "Question", "Quantification" in GOLD

What could a functional ontology look like?

- participants
  - ptcpts of different entity orders (f,x,e,p,U)
  - ptcpt roles (ag, pat, rec, exp, instr)
- predication
  - states
  - events
- refering to time
  - situating (past, present, future)
  - internal (perfective, imperfective)
- refering to space
  - absolute
  - deictic
  - figure-ground
- quantification
- modality
- negation
- possession
- kin

Formalization of the proposal

- important elements

Formalization of the proposal

- important elements
  - sections, possibly nested (called 'annotations' in Good (2004))
  - prose description
  - examples/exemplars
  - links
  - *word*-'gloss'-pairs
  - technical/ontological terms
  - references to exemplars

Formalization of the proposal

- important elements
  - sections, possibly nested (called 'annotations' in Good (2004))
  - prose description
  - examples/exemplars
  - links
  - *word*-'gloss'-pairs
  - technical/ontological terms
  - references to exemplars
- Good (2004) contains a formalization of the relation to ontologies
- Nordhoff (2008) contains an XML example
- see paper handout

Applications

- Incremental grammar-writing
  - add generalizations as they are gained

Applications

- Incremental grammar-writing
  - add generalizations as they are gained
  - Language-construction-example triple (Alexis)
  - "Ugabugu uses *wuduwudu* to mark completive aspect'
    - *dfgdfg wuduwudu ghfsa* 'I finished eating.'

Applications

- Incremental grammar-writing
    - add generalizations as they are gained
    - Language-construction-example triple (Alexis)
    - "Ugabugu uses *wuduwudu* to mark completive aspect'
        - *dfgdfg wuduwudu ghfsa* 'I finished eating.'
    - Micropublications? (Michael)

Applications

- ▶ Incremental grammar-writing
    - ▶ add generalizations as they are gained
    - ▶ Language-construction-example triple (Alexis)
    - ▶ "Ugabugu uses *wuduwudu* to mark completive aspect'
        - ▶ *dfgdfg wuduwudu ghfsa* 'I finished eating.'
    - ▶ Micropublications? (Michael)
- ▶ resurrecting old grammatical descriptions (out of copyright)
    - ▶ scan
    - ▶ portion
        - ▶ every paragraph of a book should ideally treat one topic
        - ▶ that topic should be either fo→fu or fu→fo
        - ▶ this content can be formalized in the schema proposed above
        - ▶ muddled sections will be difficult
    - ▶ tag

Outlook

- general repository of GDs
    - unified query form
    - downlinks to archives
    - uplinks to typological querying tools (TDS etc)

Querying

- select languages of interest
- select comparative concepts of interest
- get all pages of the languages' GD's which deal with the comparative concept, broadly construed
- links in the pages allow deepening of inquiry where necessary

Conclusion

- a grammatical description is a collection of form-meaning-pairs
- these can be presented in a non-linear way
- form-to-function and function-to-form description are dissociated
- need for further formalization
  - validation
  - conversion
- possibility to integrate old and new descriptions into the model
- one-stop-shop

References

Bow, Cathy, Baden Hughes & Steven Bird. 2003. Towards a general model for interlinear text. Proceedings of the 2003 EMELD Workshop. http://www.linguistlist.org/emeld/workshop/2003/bowbadenBird-paper.pdf

FROHNMEYER, L. (1913). A progressive grammar of the Malayalam language. Mangalore:Basel Mission Book and Tract Depository. GOOD, J. (2004) "The Descriptive Grammar as a (Meta)Database". Proceedings of EMELD 2004.

HASPELMATH, M. (1993). A grammar of Lezgian. Berlin:Mouton de Gruyter.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. Linguistic Typology, 11(1):119-132.

Lehmann, Christian. 1993. On the system of semasiological grammar - Allgemein- Vergleichende Grammatik, vol. 1. Bielefeld: Universität Bielefeld, Universität München.

Lehmann, Christian. 1998. Ein Strukturrahmen für deskriptive Grammatiken. In Zaefferer, Dietmar, ed. Deskriptive Grammatik und allgemeiner Sprachvergleich.Tübingen: Niemeyer: 39-52.

Lehmann, Christian. 2002. Structure of a comprehensive presentation of a language. In Tsunoda, Tasaku, ed. Basic materials in minority languages, Osaka: Osaka Gakuin University: 5-33.

Lehmann, Christian. 2004. Funktionale Grammatikographie. In Premper, Waldfried, ed. Dimensionen und Kontinua. Beiträge zu Hansjakob Seilers Universalienforschung. Bochum: N. Brockmeyer: 147-165.

S. Nordhoff (2008) Electronic reference grammars for typology: challenges and solutions Language Documentation and Conservation (2), p. 296-324

NEWMAN, P. (2000). The Hausa Language: An Encyclopedic Reference Grammar. New Haven:Yale University Press.

NOONAN, M. (2006). "Grammar writing for a grammar-reading audience". Studies in Language 30(2):351–365.