Case, Person, Number, and Gender in the Grammar Matrix

Scott Drellishak
University of Washington

August 20, 2007

1

[1] This work is supported by NSF grant BCS-0644097, a gift to the Turing Center by the Utilika Foundation, and the Max Planck Institute for Evolutionary Anthropology.

The LinGO Grammar Matrix (Bender et al., 2002)

- Distill the wisdom of existing broad-coverage grammars
- Provide a typologically-informed foundation for building grammars of natural languages in software
- Syntax-semantics interface consistent with HPSG and Minimal Recursion Semantics (Copestake et al., 2005)
- This talk: description of my upcoming dissertation work that will extend the typological coverage of the Matrix

Matrix Libraries

- Matrix intended to cover all languages, but there exist phenomena that are widespread but not universal
- If not universal, they don't belong in the Matrix
- Solution: divide the Matrix into:
  - The universal or "core" Matrix
  - Matrix "libraries" covering non-universal phenomena
- Libraries are exposed to the user-linguist via a typological questionnaire
- Based on answers, we customize a grammar

Current Libraries

- Word order, sentential negation, matrix yes-no questions, determiners, and coordination
- Most of these were known not to be exhaustive when implemented
- Coordination was based on more thorough typological survey
  - Intended to cover simple *AND*-coordination in all languages
  - Now known to have missed some marking patterns
  - Still, this is how we want to do things: survey linguistic variation first, then implement

Grammar Customization Questionnaire

- Current version updated regularly on the web at: http://www.delph-in.net/matrix/customize/matrix.cgi
- Brief demo: determiners, nouns, verbs, case-marking adpositions

Broadening Coverage

- Current version doesn't handle most marking of verbal arguments, either head- or dependent-marking
- (Exception: the case-marking adposition support)
- Limitations obvious when we try to describe highly inflecting languages
- Solution: more libraries for more phenomena!

Case

- CASE is "a system of marking dependent nouns for the type of relationship they bear to their heads." (Blake, 2001)
- I take this to include whole phrases marked by adpositions, though not everyone does
- Extremely complex phenomenon; first version will only cover case-marking on the selected arguments of verbs
  - Narrowing the range of phenomena simplifies the implementation
  - Excludes, e.g., noun-modifier case concord and possessive cases (unless used to mark a verbal argument)

Case Questionnaire

- What cases does the language have?
- How is case marked on NPs?
  - On nouns or determiners
  - Affixation, suppletion, adpositions on whole NPs
  - In any combination!
- Which verbal arguments (of transitive, intransitive, and possibly ditransitive verbs) get which case?
  - Define classes of verbs for each pattern of marking
  - Allow multiple subclasses in each
  - This allows us to handle straightforward nominative-accusative and ergative-absolutive languages (but split ergativity is on the horizon...)

Verb-Argument Agreement

- Agreement occurs when different parts of a sentence are marked for a grammatical relationship, typically sharing related values for some grammatical feature
- Library will cover agreement in person, number, and gender
  - PERSON is a grammatical feature that marks a discourse role (Siewierska, 2004)
  - NUMBER is a grammatical feature whose associated meaning has to do with number of real-world entities referred to (Corbett, 2000)
  - GENDERS are "classes of nouns reflected in the behavior of associated words." (Corbett, 1991) (citing Hockett)
- As with case, only covering a subset of the broader phenomenon (not, for example, agreement in number between nouns and their modifiers)

Person, Number, and Gender Questionnaire

- What values of person, number, gender does the language have?
- How and where are they marked?
  - Morphologically (affixation, suppletion) or lexically (adpositions/particles)
  - On nouns, noun phrases, determiners (much like case, above)
  - Also, of course, on verbs themselves
- What patterns of agreement are there?
  - Modern Hebrew has sg-du-pl on some nouns, but only sg-pl on verbs (Corbett, 2000, 180)
  - Questionnaire must allow the description of such mismatches, which inform type hierarchy geometry

Factorable Analyses

- Creating grammars based on questionnaires requires a factorable analysis of each phenomenon
- That is, an analysis made up of sub-analyses that can be "snapped together" to describe a language
- Must work with all combinations of answers (unless known to be unattested or impossible)

Analyses Extensible to Future Phenomena

- Ex: we might analyze case as a feature of nominal heads
  - Head nouns have case, case-marked determiners specify nominal case via the SPEC feature
  - Verbs select nominal arguments with particular values of CASE
- Syntax only, no semantic reflex
- Must be extensible to the sorts of case not yet covered
  - Consider an NP that possesses a direct object noun ("I love John's daughters"); what case is marked on the possessive NP?
  - In Latin, genitive. In Armenian, accusative. In Quechua, both:

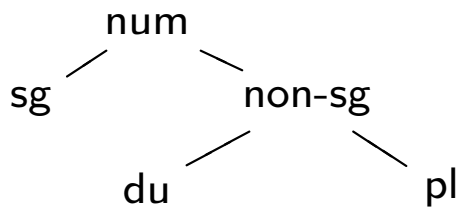
    Hipash-nin-ta *kuya-a* *Hwan-pa-ta*
    daughter-3SG-ACC love-1SG John-GEN-ACC
    'I love John's daughters' (Blake, 2001, 103)
  - This may imply CASE must be list-valued in some languages

Creating Type Hierarchies

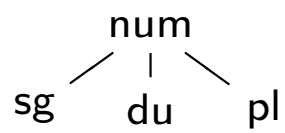
- HPSG implements feature-values like case, person, number, and gender as type hierarchies
- Should the Matrix create per-language hierarchies or use universal ones?
- Both have advantages:
  - Language-particular hierarchies result in more compact, more efficient grammars
  - Universal hierarchies can be shared, and can benefit from cross-linguistic generalizations

Hierarchy Example: Number

- We might propose a universal hierarchy like, in part:



- But doesn't describe all languages
  - Hierarchy above accounts for a language where pl agrees with du (e.g. Modern Hebrew)
  - But if no such mismatches, a flat hierarchy suffices:



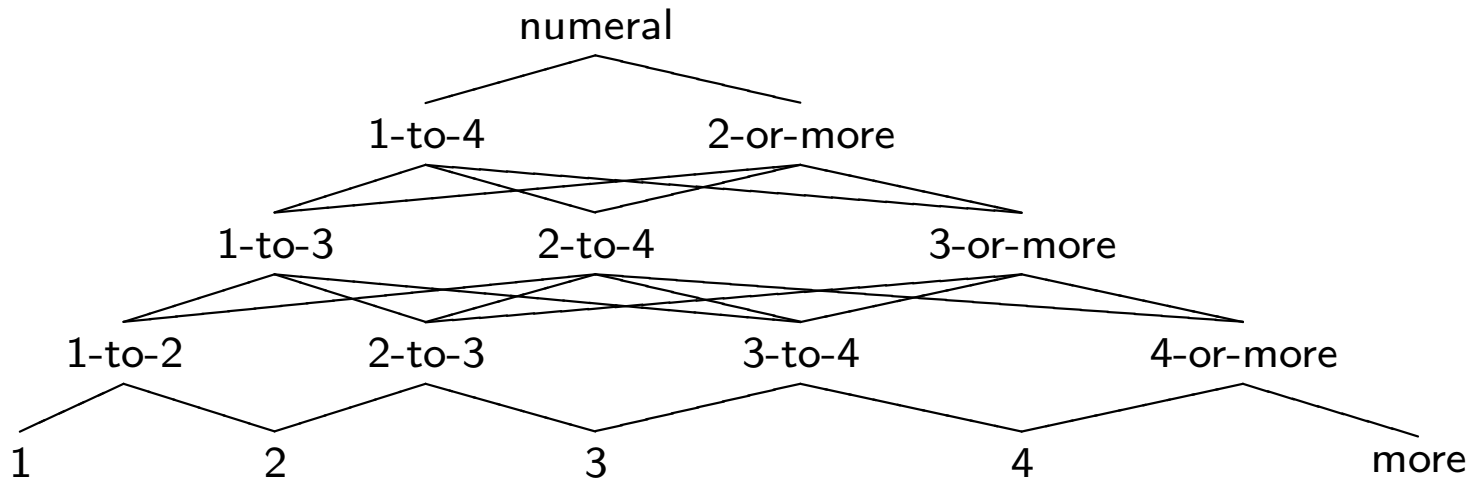
- So language-particular hierarchies seem desirable

Hierarchy Example: Number (cont'd)

- However, we also need a way to relate such hierachies cross-linguistically
- Matrix MT system (demo tomorrow!) based on the LOGON translation machinery (MRS transfer using VPM and transfer grammars)
- I suspect we need **both** language-particular and universal hierarchies

Tentative Proposal for Number

- Map language-particular *number* values onto language-independent *numerals* or *ranges of numerals*: singular → 1, dual → 2, paucal → 3-to-7, pl → 8-or-more
- Arrange these numeral values into a hierarchy:



- But can this be done with a type hierarchy? Consider 3-to-4. If it doesn't inhert from 1-to-3, they won't unify, but if it does then 1-to-3 and 4 will unify.

Web Form User Interface

- Web-based questionnaire is about to become much more complex:
  - Need a UI for morpheme slots: order, optionality, co-occurrence restrictions, and what features the slot specifies
  - Need a UI for inflectional paradigms, including ones that mix affixes and suppletion (and zero-marking)
- Complex interactions: a morpheme slot must be associated somehow with a paradigm, while the shape of the paradigm (the dimensions) depends on what features it marks

Morphology Mockup

- Last year I worked up some web-form mockups of how we might approach the interactions
- Overly simple model of ordering and optionality
- Recently, another researcher (Kelly O'Hara) has been tackling morphology in the general case
- Focusing on the back end: given a set of answers, what types are output?
- Exploring what kind of interactions must be supported
- A UI remains future work

Broader Coverage via New Libraries...Soon!

- Current plan is to have libraries for case, person, number, and gender this year
- Case on selected arguments
- Person, number, and gender agreement between verbal heads and selected arguments
- Analyses must be extensible. Hierarchies will be tuned for the language being described. UI remains an open question.
- Suggestions welcome

References

Bender, Emily M., Flickinger, Dan and Oepen, Stephan. 2002. The Grammar Matrix. In *Proceedings of COLING 2002 Workshop on Grammar Engineering and Evaluation*, Taipei, Taiwan.

Blake, Barry J. 2001. *Case, Second Edition*. Cambridge: Cambridge University Press.

Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan A. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(2–3), 281–332.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.