Scott Drellishak & Emily M. Bender
Coordination Modules for a Crosslinguistic Grammar Resource
HPSG 2005, Lisbon, Portugal
August 24, 2005

Overview

- This talk will describe a module in the LinGO Grammar Matrix that supports parsing and generating sentences with coordination.
- Five parts:
  - A description of the Matrix and Matrix modules.
  - A brief overview of the typology of coordination.
  - The details of our implementation of coordination.
  - A live demonstration.
  - Theoretical implications and future work.

1. **The Matrix and Matrix Modules**
2. Typology of Coordination
3. Coordination in the Matrix
4. Demonstration
5. Theoretical Implications and Future Work

The LinGO Grammar Matrix (1/2)

- Attempts to distill the wisdom of existing broad-coverage grammars and document it in a form that can be used as the basis for new grammars.
- Goals:
  - Semantic representations and a syntax-semantic interface consistent with other work in HPSG.
  - Represent generalization across linguistic objects and across languages.
  - Allow for quick start-up when analyzing new languages.

The LinGO Grammar Matrix (2/2)

- Currently, the Matrix includes:
  - Definitions of basic features and technical devices (e.g. list manipulation).
  - Types associated with Minimal Recursion Semantics (MRS). (Copestake et al. 2003)
  - Types for lexical and syntactic rules.
  - Hierarchy of lexical types for language-specific lexical entries.
- Compatible with the LKB grammar development environment. (Copestake 2002)

Modules (1/2)

- A problem facing the Matrix: The wide variety of phenomena in the world's languages.
- Writing even a rudimentary grammar requires many (parameter-like) choices in order to parse non-trivial sentences.
- Furthermore, there are recurring patterns across the world's languages that are not universal.
- Solution: In addition to rules and definitions, provide bootstrapping tools that allow grammar writers to create a functional starter grammar very quickly.

Modules (2/2)

- We call these tools “modules”. Each consists of:
  - Rules associated with a particular grammatical phenomenon.
  - Some software code (currently accessed through a web interface) that asks a series of questions, then outputs a starter grammar.
  - This grammar is designed to be scalable.
- Modularity allows us to share the work more easily: linguists with knowledge in a particular area can write a module for that area.

1. The Matrix and Matrix Modules
2. **Typology of Coordination**
3. Coordination in the Matrix
4. Demonstration
5. Theoretical Implications and Future Work

Typology of Coordination

- The module described in this talk covers coordination.
- There are phenomena called “coordination” (or “conjunction”) in most (all?) of the world’s languages.
- What we mean by “coordination” is structures that combine several sentence elements of like or similar category into a single larger element.
- However, different languages mark it with a wide variety of coordination strategies.

Kinds of Marking (1/3)

- Lexical marking: e.g. the conjunction *and* in English (and its cognates in the other I-E languages).
- Juxtaposition: coordinands simply occur in sequence with no additional material.
- Example from Abelam (Sepik-Ramu, New Guinea):

  wʌny balə wʌny acʌ waryʌ.bər

  that dog that pig fight

  'that dog and that pig fight'

  (Laylock 1965:56)

Kinds of Marking (2/3)

- Morphological marking: one or more of the coordinands is inflected into a conjunctive or continuative form.
- Example from Kanuri (Nilo-Saharan):

  kə̀rə̀zə̂ málə̀mrò wálwònò.

  studied.CONJ malam became

  ‘He studied and became a malam.’

  (Hutchison 1981:322)

Kinds of Marking (3/3)

- Phonological marking.
- Example from Telugu (Dravidian):

  kamalaa wimalaa poDugu.

  Kamala Vimala tall

  'Kamala and Vimala are tall.'

  (Krishnamurti and Gwynn 1985:325)

- Juxtaposition might be phonological, often described as having a distinctive "comma" intonation.
- (But this kind of marking can be handled like other morphology.)

Patterns of Marking

- Monosyndeton: mark one coordinand (“A B and C”)
- Asyndeton: no marking (“A B C”)
- Polysyndeton: more than one coordinand marked.
  - Both “A and B and C” and “and A and B and C”.
  - These are handled differently; to distinguish them, we call the former *polysyndeton* and the latter *omnisyndeton*.
- Two possible positions: before or after the coordinand (e.g. Latin *et* is before, while -*que* is after).

Different Phrase Types

- In addition to characterizing strategies by method of marking, marking pattern, and position of the mark, what phrase types are covered?
- In most or all I-E languages, one coordination strategy covers many phrase types: e.g. English *and*.
- In many languages, this is not true: some strategies can only be used with a subset of the parts of speech in the language.

Summary of Typology

- So, each strategy can vary along several dimensions:
  - Kind of Marking: lexical, morphological, none.
  - Pattern of Marking: a-, mono-, poly-, or “omni-” syndeton.
  - Position of Marking: before or after the coordinand.
  - Phrase types covered: one or more.
- The coordination module’s web interface asks for this information about the language being described, then outputs an appropriate grammar.

Comitative Coordination

- Following Stassen (2000), the world's languages can be classified as either AND- or WITH-languages.
- AND-langs have the familiar syntactic coordination.
- WITH-languages mark coordination asymmetrically: one coordinand unmarked, the others marked by a particle or morpheme meaning "with".
- The syntax (and possibly the semantic representation) is that of an adjunct.
- Not rare, but a distinct phenomenon, and not covered by this module.

1. The Matrix and Matrix Modules
2. Typology of Coordination
3. **Coordination in the Matrix**
4. Demonstration
5. Theoretical Implications and Future Work

Coordination in the Matrix

- Based on the coordination implementation of the English Resource Grammar (ERG). (Flickinger 2000)
- Borrowed the basic coordination structure and semantics.
- Simplified somewhat, and also generalized to handle non-English structures.
- Handles same-category coordination. HEAD values are constrained for phrase and coordinands, but not identified.

Coordination Structures (1/2)

- Problem: any number of items can be coordinated.
- This seems to imply an infinite number of rules (and semantic relations):

  XP → XP *conj* XP

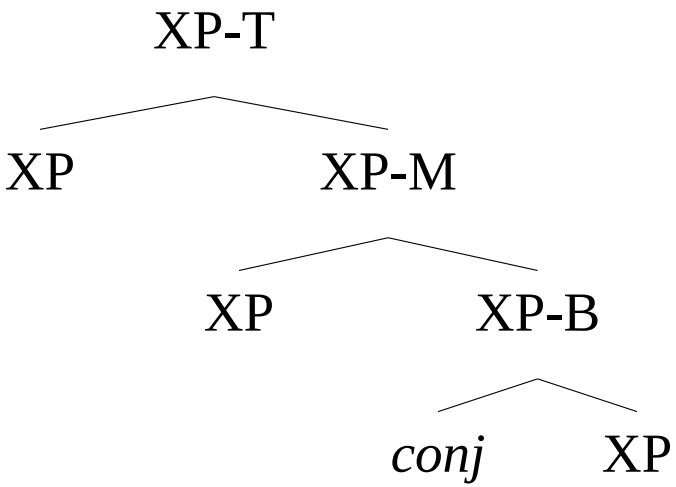
  XP → XP XP *conj* XP

  XP → XP XP XP *conj* XP

  ...
- However, the LKB does not allow rules with an underspecified number of daughters.

Coordination Structures (2/2)

- Solution: Simulate the flat structure like this:



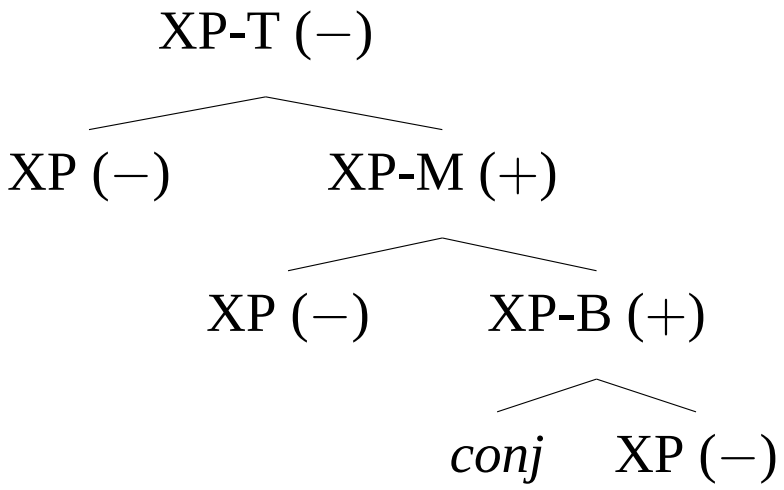
- Three rules: top (binary), mid (binary), and bottom (either binary or unary).
- Structure consists of one top phrase, as many mid phrases as necessary, and one bottom phrase.

The Feature COORD

- The top phrase is a full-fledged XP, but the mid and bottom phrases should not combine with other constituents via ordinary rules.
- Similarly, other kinds of phrases should not appear within these coordination structures.
- To enforce this, we define a new boolean feature COORD, on `local-min` (the type from which LOCAL derives). COORD – is the default.
- The various patterns of marking can now be defined by the COORD values of phrases and their left and right daughters.

Monosyndeton

- XP-T (−) → XP (−) XP (+)
  XP-M (+) → XP (−) XP (+)
  XP-B (+) → *conj* XP (−)

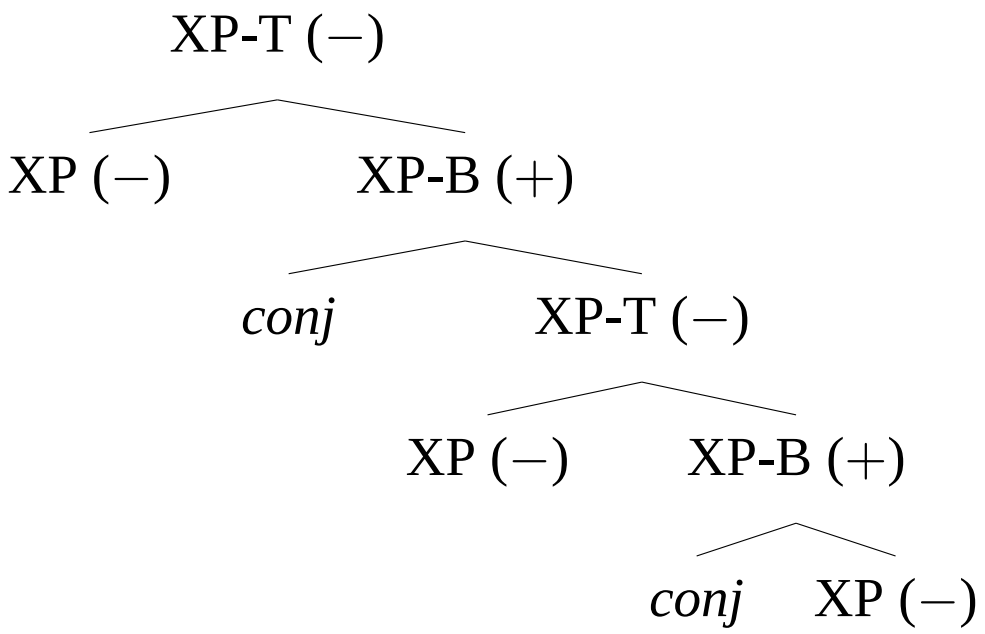


Poly- and Asyndeton

- XP-T (−) → XP (−) XP (+)

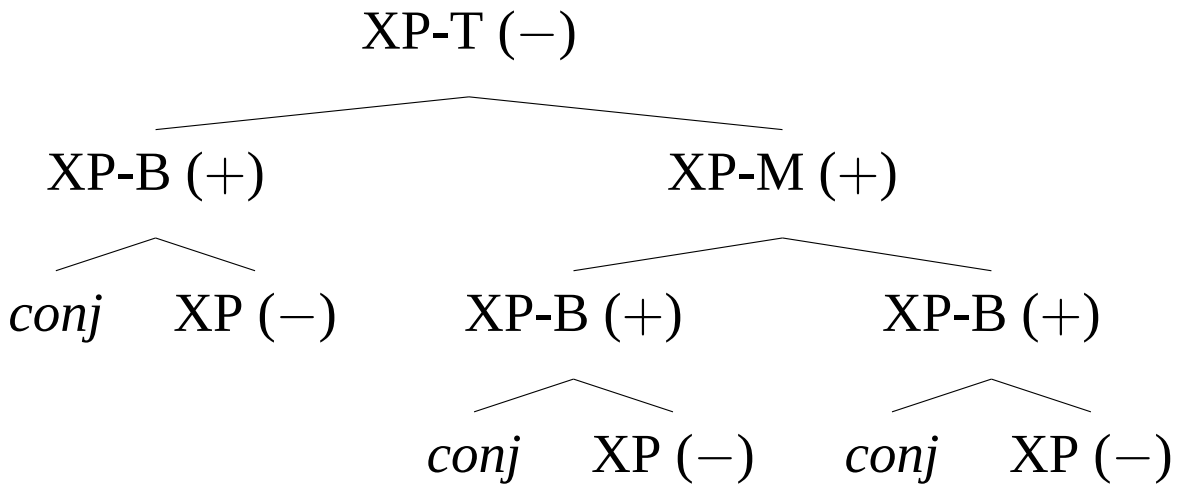
  no mid rule

  XP-B (+) → *conj* XP (−)



“Omnisyndeton”

- XP-T (−) → XP-B (+) XP (+)
  XP-M (+) → XP-B (+) XP (+)
  XP-B (+) → *conj* XP (−)



Semantic Representation (1/3)

- The semantic representation of unbounded coordination is handled in the same way as the syntax.
- We define a relation that coordinates two arguments:

$$\begin{bmatrix} \text{LBL} & \textit{handle} \\ \text{C-ARG} & \textit{coord-index} \\ \text{L-HNDL} & \textit{handle} \\ \text{L-INDEX} & \textit{individual} \\ \text{R-HNDL} & \textit{handle} \\ \text{R-INDEX} & \textit{individual} \end{bmatrix}$$

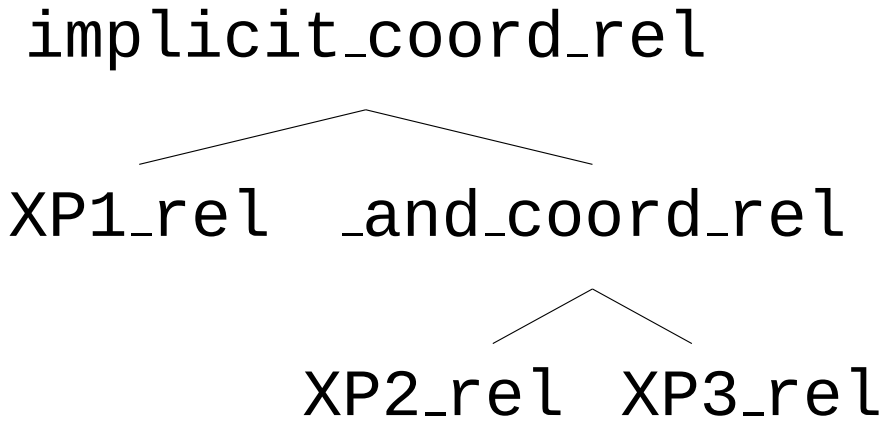
- These binary relations can be strung together like the syntactic rules to represent unbounded coordination.

Semantic Representation (2/3)

- Each bottom phrase contributes a coordination relation (with one exception).
- Conjunctions or lexical rules generally contribute explicit coordination relations (e.g. `_and_coord_rel`).
- A phrase's `coordination-relation` is stored in the feature COORD-REL.
- The relation's left and right arguments are specified in the phrase's parent, either a mid or a top rule.

Semantic Representation (3/3)

- A mid phrase contributes an `implicit-coord-rel` that serves to link more-than-two-way coordination. Three-way coordination, for example, is represented:



- (Where branches represent the identification of the left or right argument of the relation.)

"Omnisyndeton" is Exceptional

- Problem: "omnisyndeton" has the same number of bottom phrases as coordinands, and therefore one too many `coordination-relation`s.
- Solution: the bottom rule requires a semantically empty conjunction with the same spelling.
- The rules for "omnisyndeton" now require a new kind of phrase as the left daughter of a mid or a top phrase, that we call a "left" instead of a bottom phrase:

XP-T (−) → XP-L (−) XP (+)

XP-M (+) → XP-L (−) XP (+)

XP-B (+) → conj XP (−)

1. The Matrix and Matrix Modules
2. Typology of Coordination
3. Coordination in the Matrix
4. **Demonstration**
5. Theoretical implications and Future Work

1. The Matrix and Matrix Modules
2. Typology of Coordination
3. Coordination in the Matrix
4. Demonstration
5. **Theoretical Implications and Future Work**

Theoretical Implications (1/4)

- Our implementation makes typological predictions.
- Because the structure is right-branching, we would have trouble with a language that marks coordination only on the first coordinand: "*conj* A B C".
- However, that pattern is apparently unattested (Stassen 2000).
- If it were attested, we could address it by having both left- and right-branching versions of the rules.

Theoretical Implications (2/4)

- Predictions about ambiguity:
  - Monosyndeton languages seem to *always* allow polysyndeton, and our rules reflect that. The semantics will differ, though.
  - Mono-, poly-, and asyndeton can be ambiguous for a given surface string:
    [[A *conj* B] *conj* C] vs. [A *conj* [B *conj* C]]
  - But not, it seems, “omnisyndeton”. That would require:
    [*conj* [*conj* A *conj* B] *conj* C]

Theoretical Implications (3/4)

- We use the feature COORD to separate the syntactic space into two domains: the simulated N-way coordination structures, and everything else (regular syntax).
- This is a powerful tool, but it means that some nodes in the tree do not correspond to constituents.
- We also have rules that require particular *types* of phrases, not just phrases with a particular HEAD type.
- This is usually considered bad (it's certainly not "head-driven"), but we only do it inside of our coordination structures.

Theoretical Implications (4/4)

- Possibly bad prediction:
- We treat right-branching grouping as unmarked, but left-branching grouping as exceptional.
- But surely there are three possibilites:
  - [A and B and C] (flat)
  - [[A and B] and C] (left-branching)
  - [A and [B and C]] (right-branching)

Future Work

- There is plenty of straightforward coordination we still do not cover:
  - Adversative (“but”) coordination, which seems restricted to two-way.
  - Complex conjunctions (e.g. “both...and”).
  - Coordination of different parts of speech.
  - Scary phenomena like gapping and non-constituent coordination.
  - Better interfaces and more flexible scripts.

References

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI.

Copestake, Ann, Daniel P. Flickinger, and Carl Pollard Ivan A. Sag. 2003. Minimal Recursion Semantics. An introduction. Unpublished ms.

Flickinger, Dan. 2000. On building a more efficient grammar by exploiting types. *NLE* 6 (1):15 – 28.

Hutchison, John P. 1981. *A reference grammar of the Kanuri language*. Madison, WI: University of Wisconsin - Madison.

Krishnamurti, BH., and J. P. L. Gwynn. 1985. *A grammar of modern Telugu*. Delhi: Oxford University Press.

Laylock, D. C. 1965. *The Ndu language family (Sepik district, New Guinea)*. Canberra: The Australian National Library.

Stassen, Leon. 2000. And-languages and with-languages. *Linguistic Typology* 4:1–54.