Semantic constraints on syntactic analyses of NPs in grammar engineering

Emily M. Bender, Scott Drellishak, Bill McNeill & Laurie Poulson
{ebender,sfd,billmcn,lpoulson}@u.washington.edu
University of Washington
LSA, Oakland CA, 1/7/05

Overview

- *Multilingual grammar engineering*
- *Assumptions*
- *Data: Armenian, Farsi, Swedish*
- *MRS primer and the problem*
- *Proposed solution*
- *Conclusions*

Multilingual Grammar Engineering

- *Monolingual grammar engineering tests linguistic analyses*
  - *for consistency*
  - *against corpora*
- *Multilingual grammar engineering tests cross-linguistic hypotheses*
- *LinGO Grammar Matrix: precision grammar starter-kit* *(Bender et al 2002)*

Cross-linguistic Hypotheses

- *HPSG notions of headedness and valence*
- *Semantic compositionality (so-called "Frege's Principle")* *(Pelletier 2001)*
- *Minimal Recursion Semantics for typed feature structure based semantic composition* *(Copestake et al 2003, Flickinger and Bender 2003)*

Assumption 1

- *A monotonic system for compositional semantics is desirable.*
  - *bidirectionality*
  - *partial interpretation from partial parses*

Assumption 2

- *In an adequate semantic representation, all nominal indices are bound by quantifiers.*

Assumption 3

- *Scopal adjectives exist.*
  - *Type 1:* fake, alleged, former
  - *Type 2:* probable, likely
  - *By contrast, most adjectives are intersective:* good, red, tired

Assumption 4

- *Quantifiers can take scope between scopal adjectives and the nouns they modify.*
  - The most probable winner of every medal was disqualified.

Assumption 4 (cont)

- every > the > probable: *Different probable winners for each medal, all disqualified.*
- the > every > probable: *One super athlete is favored in each competition individually and disqualified.*
- the > probable > every: *No one person dominated all events, but if anyone were to win all the medals, it would X, who was disqualified.*

Assumption 5

- *Determiners combine with nominal expressions to produce quantifier-expressing constituents.*

Assumption 6

- *Markers of (in)definiteness are determiners*
  - the *and* a *are canonical English determiners*
  - *Ghomeshi (2003): Farsi* -i *is a 'quantitative indefinite determiner which heads a QP'.*

Armenian

- *Indefinite article is a separate word, while the definite article appears to be an affix:*

a. dghay měn ē
boy INDEF be.3sg
'He is a boy.' (Bardakjian and Thomson 1977:18)

b. dun-ě medz ē
house-DEF big be.3sg
'The house is big.' (Andonian 1966:22)

Armenian

- *There are also what appear to be scopal adjectives:*

@. . . **hawanagan** k'ayleri masin
. . . **probable** step.PL.GEN about
'. . . concerning probable steps'

Farsi

- -i *marks NPs as indefinite and non-generic*

a. *ketab*: the book

b. *ketab-i*: a book, books in general

Farsi

- -ro/-o *appears on definite direct objects*
- *or, incombination with* -i*, on specific indefinite direct objects*

a. ketab-ro gereft
book-ACC/DEF take.PAST.3SG
'He/she took the book.' (Mace 2003)

b. mi xahænd xane-i-ro bexæræænd
PRES want.3PL house-INDEF-ACC/DEF buy
'They want to buy a house. (A certain house)' (Mace 2003)

Farsi

- *There are also what appear to be scopal adjectives:*

(yek) barandeh-ye **ehtemali** az har
(INDEF) winner-EZAFE **probable** from each
medal mardood shod.
medal disqualify became
'A probable winner of every medal was disqualified.'

Swedish

- *Definite NPs with no adjectives are (usually) marked only with a suffix*
- *With adjectives, both a definite determiner and the suffix are required*

a. hus-**et** är gamla.
house-def is old
'The house is old.'

b.*__det__ hus-**et** är gamla.
the house-def is old

c. **det** röd-**a** hus-**et** är gamla.
the red-DEF house-DEF is old
'The red house is old.'

d.*röd-**a** hus-**et** är gamla.

e.*__det__ röd-**a** hus är gamla.

Swedish

- *There are also what appear to be scopal adjectives (which require an overt determiner on definites, like all adjectives)*

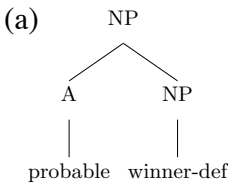
a.@den mest trolig-a orsak-en till hjärtinfarkt
the most probable-DEF cause-DEF of heart attack

b.*mest troliga orsaken till hjärtinfarkt

The Problem

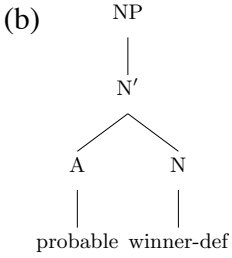
Syntax suggests

(a)



Compositional semantics requires

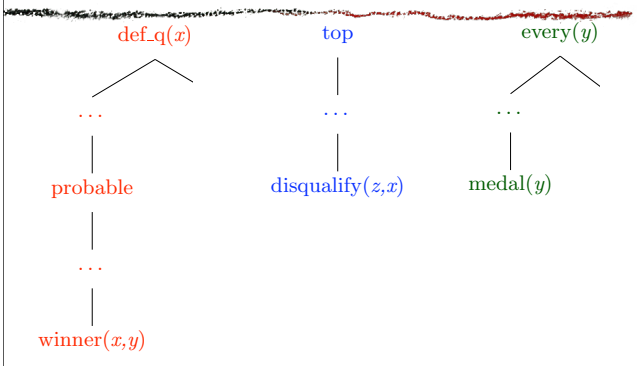
(b)



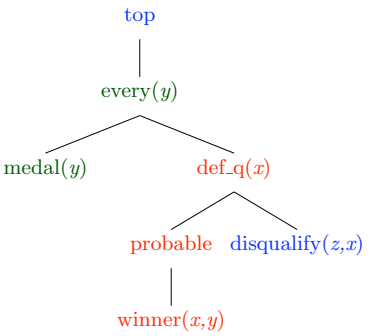
MRS Primer

- *MRS representations are underspecified for scope*
- *Fixed scopal relations are represented as constraints on possible complete scopings*
- *Quantifiers are free to 'float' into any space*

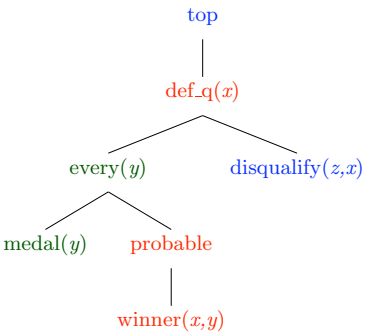
Constraints on Scope



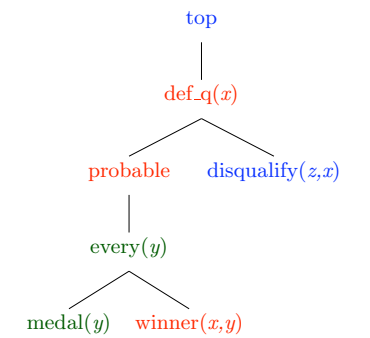
Constraints on Scope



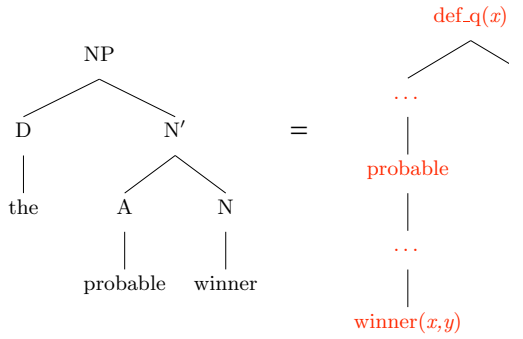
Constraints on Scope



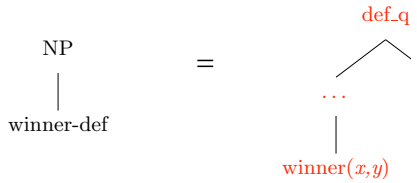
Constraints on Scope



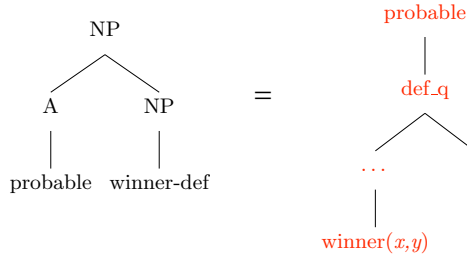
The Problem Again



The Problem Again



The Problem Again



Solutions

- *Loosen the tight coupling between syntax and semantics*
- *Reject the assumptions that require treating the (in)definiteness markers as contributing semantic quantifiers*

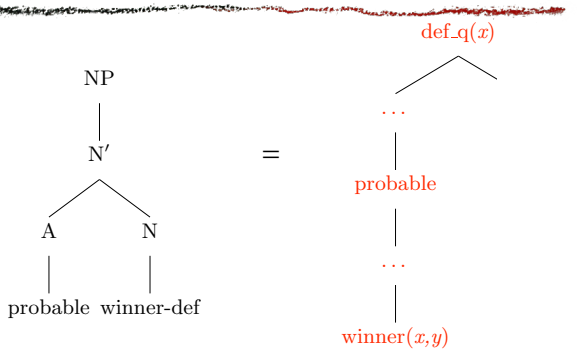
More Flexible Interfaces

- *CLLS (Constraint Language for Lambda Semantics) uses more flexible dominance relations plus semantic types to reduce ambiguity* (Egg et al 2001)
- *GLUE Semantics formalizes 'weak compositionality', semantic representations projected off complete syntactic representations* (Asudeh et al 2002)

(In)definiteness Markers as Non-Quantifiers

- *Definiteness affixes contribute other information about NP semantics (givenness, specificity)*
  (cf. Borthen and Haugereid to appear)
- *Actual quantifier built by non-branching construction*

(In)definiteness Markers as Non-Quantifiers



Language Internal Evidence

- *Some NPs have neither overt determiners nor definiteness inflection*

Armenian:
dinozawr hin gentani mĕn ē
dinosaur ancient animal INDEF be.3SG
'The dinosaur is an ancient animal.' (Hagopian, online)

Farsi:
mærd amæd.
man come.PAST.3SG
'The man/men came.'

Swedish:
Bil-ar är bra att ha.
Car-PL are good to have
'Cars are good to have.'

Language Internal Evidence

- *(In)definiteness markers can co-occur with other determiner-like elements*

Armenian:
ayn kirk'-ĕ/*kirk'
this book-the/book
'this book' (Andonian 1966)

Farsi:
ye ketab-ha-i
INDEF book-INDEF
'some (certain) books'
(Ghomeshi 2003)

Swedish:
det röd-a hus-et
the red-DEF house-DEF
'the red house'

Language Internal Evidence

- *(In)definiteness markers don't pattern with translations of 'every', 'no', etc.*

Armenian:
polor baduhanner-e kots'ets'ek'
all window-PL-DEF shut.2PL.IMP
'Shut all the windows' (Andonian 196687)

Farsi:
hær/hic ketab(-i)
each/no book(-INDEF)

Swedish:
Varje / ingen bok är bra
every / no book is good
'Every/no book is good.'

Typological Evidence

- *Rijkhoff's (2002) survey of the NP in typological perspective doesn't even consider quantifiers*
- *Borthen and Haugereid (to appear) propose an extension to the Matrix to represent definiteness and specificity*
- *English conflation of definitness and quantification is perhaps unusual*

A Prediction

- *Turkish allows free Adj-Det order*
- *In the METU corpus (Oflazer et al 2003), determiners other than* bir *('a') attach outside adjectives.*
  - bir *does not contribute a quantifier*
  - *Adj > Det order is marked, perhaps produced by NP-internal extraction*
- *Prediction: Elements introducing quantifiers attach outside scopal adjectives*

Conclusions

- *MRS makes interesting cross-lingusitic predictions*
- *Grammar engineering supports linguistic hypothesis testing*
- *Computational linguistic resources should be designed to be cross-linguistically applicable*

Acknowledgments

- *Thanks to Markus Egg, Dan Flickinger, Ann Copestake for helpful discussion*
- *Thanks to Arman Maghbouleh, Tom Johnson, and Oskar Gjertsson for assistance with data*