Representing Information Structure in MRS

Emily M. Bender

University of Washington

DELPH-IN Summit, 6 July 2010, Paris

UNIVERSITY OF
WASHINGTON

Outline

WASHINGTON

What is Information Structure?

INFORMATION STRUCTURE: That component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse context. (Lambrecht, 1996, 5)

WASHINGTON

What is Information Structure?

Our point of departure is the assumption, expressed in e.g. Chafe (1976), Prince (1986), that what underlies the focus-ground distinction is a need to 'package' the information conveyed by a sentence so that hearers can easily identify which part of the sentence represents material that is already subsumed by the information state. (Engdahl and Vallduvı, 1996, 2)

Information Structure v. Information Status

- Closely related, but still distinct.
- Information structure: Relationship between a referent and the utterance.
- Information status (also discourse/cognitive status): Relationship between a referent and the common ground.

WASHINGTON

Categories in Information Structure

It's not yet clear what the core concepts or categories of information structure are, but most people working on it posit some variation on "topic" and "focus", and allow for a third category ("background", "tail", etc.) which is neither.

Molnár (2002) argues for "contrast" as a third, independent (yet cross-cutting) information structure category.

WASHINGTON

Categories in Information Structure

FOCUS

The semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition. (Lambrecht, 1996, 213)

TOPIC

A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e., as expressing information which is relevant to and which increases the addressee's knowledge of this referent. (Lambrecht, 1996, 131)

WASHINGTON

Identifying Topic and Focus

- There are very few operationalizable tests for identifying topic and focus, especially independent of language-specific marking.
    - Focus: Answer to *wh-* questions
    - Topic: Answer to 'Tell me about' questions (Choi, 1999)
- These aren't, however, applicable to identifying topic/focus in running text.

Marking of Information Structure

- Prosody (cf. English; NOT universal)
- Morphology
  - Affixes
  - Function words (clitics, particles, adpositions)
- Syntax
  - Special constructions (e.g. English clefts)
  - Special positions (e.g. preverbal focus position in Turkish, preverbal position in V2 languages)
  - Word order tendencies
  - Word order constraints (e.g. scrambling in Japanese/Korean)

WASHINGTON

Crosslinguistic Studies

There is a strong need for studies of parallel corpora to determine:

- How the information sturctural contrasts marked in different languages relate to each other
- Possible tests that work across languages in monolingual texts to identify topic/focus
- The extent of cross-linguistic consistency in information structure

Cf. Sanghoun's presentation. . .

UNIVERSITY OF WASHINGTON

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Outline

WASHINGTON

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Representing Information Structure: Why?

- It is overtly marked, though differently in different languages.
- Not representing it would lead to spurious synonymity.
- Representations of IS would be useful in applications of generation, including MT and summarization.

However: It's not clear at present whether the best translations at a paragraph level would maintain the same information structure assignments to referents across languages. It may be that other factors, including language-specific paragraph-level information structure strategies intervene.

WASHINGTON

Representing Information Structure: Where?

- Not CAT: This is clearly on the meaning side of things.
- CTXT is meant for pragmatic constraints, which these clearly are.
- However:
    - With current infrastructure, only CONT is available as generator input.
    - Arguably (see below) best represented using semantic indices, already part of MRS.
    - Information structure can interact with truth conditions (Partee, 1991)

Current proposal: Put Information Structure constraints in CONT (i.e. MRS).

Examples from Partee 1991

- MARY always took John to the movies.
- Mary always took JOHN to the movies.
- Mary always took John to the MOVIES.
- John only/even/also introduced BILL to Sue.
- John only/even/also introduced Bill to SUE.

WASHINGTON

Outline

UNIVERSITY OF
WASHINGTON

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Engdahl and Vallduví 1996

Sample constraints:

$$\boxed{1}\begin{bmatrix} \text{PHON} \mid \text{ACCENT} & A \\ \text{INFO-STR} \mid \text{FOCUS} & \boxed{1} \end{bmatrix} \quad \boxed{1}\begin{bmatrix} \text{PHON} \mid \text{ACCENT} & B \\ \text{INFO-STR} \mid \text{GROUND} \mid \text{LINK} & \boxed{1} \end{bmatrix}$$

Disadvantages:

- Circular feature structures
- Mixing syntactic and semantic information

Wilcock 2005

Sample representation:

- What did every dog chase?
- 1:every(x,3,4), 3:dog(x), 7:cat(y), 5:some(y,7,1),
  4:chase(e, x, y)
  TOP-HANDLE:5, LINK:{1}, TAIL:{4}, FOCUS:{5}
- What did every dog do?
- 1:every(x,3,5), 3:dog(x), 7:cat(y), 5:some(y,7,4),
  4:chase(e, x, y)
  TOP-HANDLE:1, LINK:{1}, FOCUS:{4, 5}

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Wilcock 2005

Pros:
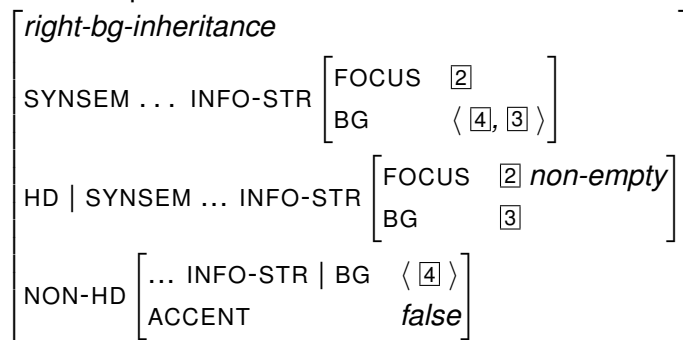
- Represents focus "scope"

Cons:

- Intersective modifiers and their heads required to share information structural properties. (But cf. "No, I said the BLUE flower.")
- Paggio (2009) (cf. Ericsson, 2005) argues that the label refers to the whole ep, including its arguments, but that's not what's focused.
- Wilcock doesn't give a way of cosntructing these representations.

Paggio 2009

- Adds INFO-STR inside CTXT with list-valued features TOPIC, FOCUS, BG.
- Elements of lists are semantic indices, structure-shared with elements in MRS.
- Phrase structure rules constrain the links between the two and propogate INFO-STR.

WASHINGTON
UNIVERSITY OF

Paggio 2009

For example:

$$
\begin{bmatrix}
\textit{right-bg-inheritance} \\[4pt]
\text{SYNSEM} \ldots \text{INFO-STR}
\begin{bmatrix}
\text{FOCUS} & \boxed{2} \\
\text{BG} & \langle\, \boxed{4},\, \boxed{3}\, \rangle
\end{bmatrix} \\[10pt]
\text{HD} \mid \text{SYNSEM} \ldots \text{INFO-STR}
\begin{bmatrix}
\text{FOCUS} & \boxed{2}\ \textit{non-empty} \\
\text{BG} & \boxed{3}
\end{bmatrix} \\[10pt]
\text{NON-HD}
\begin{bmatrix}
\ldots \text{INFO-STR} \mid \text{BG} & \langle\, \boxed{4}\, \rangle \\
\text{ACCENT} & \textit{false}
\end{bmatrix}
\end{bmatrix}
$$

Paggio 2009

Pros:

- Using semantic indices is promising (cf. notion of 'characteristic variable')
- Detailed account of constraints relating word order to information structure in Danish

Cons:

- INFO-STR has to be propogated separately (doesn't "stay with" the semantic index).
- Lists can be cumbersome.
- Difficult to combine constraints on the same index.
    - cf. contrastive v. non-constrastive topic/focus

WUNIVERSITY OF
WASHINGTON

Outline

WASHINGTON

Strategy

- Distinguish focus marking from focus projection
- Record only those information structure contrasts which are morphosyntactically (or prosodically) marked
- Unless something is explicitly marked as background, don't constrain it as such.
    - But cf. idea of a pre-processing step on English text
- Use variable property mapping to take unmarked elements and require lack of marking.
    - But cf. Breton which requires one (marked) element in the preverbal position in any sentence.

WASHINGTON

Information Structure as Variable Properties

```
individual :+ [INFO-STR info-str].


info-str := *top*.
unmarked := info-str.
marked := info-str.
topic := marked.
focus := marked.
```

- Multiple elements can be marked as topic/focus, don't need lists
- Information stays with the indices

WASHINGTON

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Example: Russian question focus clitic
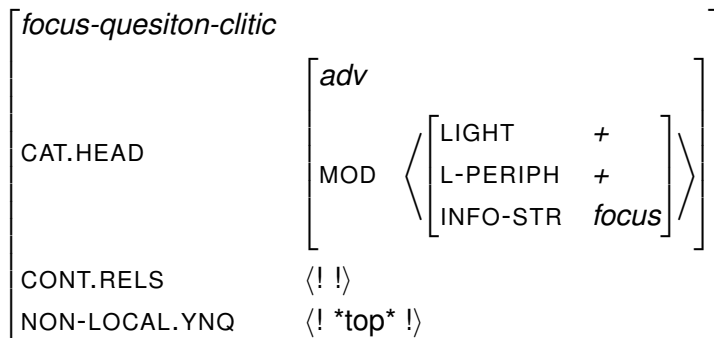
Sobaku li   kupil Ivan?
Dog   FOC buy   Ivan

'Did Ivan by the DOG?' [rus]
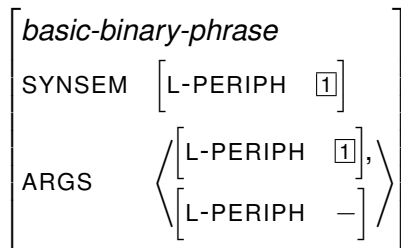
Ivan li   kupil sobaku?
Ivan FOC buy   dog

'Did IVAN buy the dog?' [rus]

Example: Russian question focus clitic

$$
\begin{bmatrix}
\textit{focus-quesiton-clitic} \\[2mm]
\text{CAT.HEAD} \quad
\begin{bmatrix}
\textit{adv} \\[2mm]
\text{MOD} \quad \left\langle
\begin{bmatrix}
\text{LIGHT} & + \\
\text{L-PERIPH} & + \\
\text{INFO-STR} & \textit{focus}
\end{bmatrix}
\right\rangle
\end{bmatrix} \\[10mm]
\text{CONT.RELS} \qquad \langle !\ ! \rangle \\[2mm]
\text{NON-LOCAL.YNQ} \qquad \langle !\ \text{*top*}\ ! \rangle
\end{bmatrix}
$$

- Attaches only to single words (LIGHT)
- Insists that the word it attaches to be the first in the clause (L-PERIPH)
- Marks its presence via the non-local feature YNQ

WASHINGTON

Example: Russian question focus clitic

$$
\begin{bmatrix}
\textit{basic-binary-phrase} \\
\text{SYNSEM} \quad \begin{bmatrix} \text{L-PERIPH} & \boxed{1} \end{bmatrix} \\[2ex]
\text{ARGS} \quad \left\langle \begin{bmatrix} \text{L-PERIPH} & \boxed{1} \end{bmatrix}, \right. \\
\qquad \qquad \left. \begin{bmatrix} \text{L-PERIPH} & - \end{bmatrix} \right\rangle
\end{bmatrix}
$$

- Copies L-PERIPH from left-hand daughter
- Checks that right-hand daughter is comptable with L-PERIPH $-$.

WASHINGTON

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Example: Russian question focus clitic

$$
\begin{bmatrix}
\textit{int-cl} \\[2ex]
\text{SYNSEM} \quad
\begin{bmatrix}
\text{INDEX.SF} & \textit{ques} \\
\text{MC} & \textit{bool} \\
\text{YNQ} & \langle ! \; ! \rangle
\end{bmatrix} \\[4ex]
\text{HEAD-DTR.SYNSEM} \quad
\begin{bmatrix}
\text{MC} & \textit{na} \\
\text{YNQ} & \langle ! \; \text{*top*} \; ! \rangle
\end{bmatrix}
\end{bmatrix}
$$

- Constrains sentential force (SF) to *ques*
- Looks for and discharges non-empty YNQ list

WASHINGTON

UNIVERSITY OF

Emily M. Bender   University of Washington

Representing Information Structure in MRS

Outline

WASHINGTON

Summary

- Information structure is an important part of sentence meaning.
- Representing it can help in various applications.
- Currently, the best solution seems to be to represent it in MRS, even though it is perhaps more properly CTXT than CONT.
- Within MRS, treating Information Structure as a variable property seems promising.

WASHINGTON

Remaining issues

- Focus on single semantic features:
  One book isn't enough, you need to bring bookS.
- Focus on eps contributed by determiners, which don't have their own characteristic variable:
  This isn't just any book, this is THE book.
- Contrast

Bibliography I

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and points of view. In Li, C., editor, *Subject and Topic*, pages 25–56. Academic Press, New York.

Choi, H.-W. (1999). *Optimizing Structure in Context: Scrambling and Information Structure*. CSLI Publications, Stanford, CA.

Engdahl, E. and Vallduvı, E. (1996). Information packaging in HPSG. *Edinburgh Working Papers in Cognitive Science*, 12:1–32.

Ericsson, S. (2005). *Information Enriched Constituents in Dialogue*. PhD thesis.

Lambrecht, K. (1996). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge Univ Press, Cambridge, NY.

Molnár, V. (2002). Contrast – from a contrastive perspective. In HasselgÃrd, H., Johansson, S., Behrens, B., and Fabricius-Hansen, C., editors, *Information Structure in a Cross-Linguistic Perspective*, pages 147–162. Rodopi, Amsterdam.

Paggio, P. (2009). The information structure of Danish grammar constructions. *Nordic Journal of Linguistics*, 32(01):137–164.

Partee, B. H. (1991). Topic, focus and quantification. *Cornell Working Papers in Linguistics*, 10:159–187.

Prince, E. (1986). On the syntactic marking of presupposed open propositions. In *Chicago Linguistic Society*, pages 208–222.

Wilcock, G. (2005). Information structure and Minimal Recursion Semantics. In Arppe, A., Carlson, L., Lindén, K., Piitulainen, J., Suominen, M., Vainio, M., Westerlund, H., and Yli-Jyrä, A., editors, *Inquiries into Words, Constraints and Contexts*, pages 268–277. CSLI Publications, Stanford.

WASHINGTON
UNIVERSITY OF