



Lecture 6

Information Packaging



Information Packaging

(Chafe, 1976), (Vallduví, 1992; Vallduví, 1994), (Vallduví and Engdahl, 1996)

Focus the part of the sentence that encodes *information* (I_s), i.e., the only augmentation or modification to be made to the hearer's model of the common ground (K_h) ($I_s = \phi_s - K_h$)

Ground the part of the sentence that encodes what is already established and under discussion in K_h at utterance time; ushers I_s to the right location (from the speaker's viewpoint) in K_h ; further subdivided into LINK and TAIL

Link in indicates *where* I_s goes and tail indicates *how* it fits there.

- partitioning of surface form, not of sentence meaning!
- (Vallduví and Engdahl, 1996): analysis of IP realization in many languages



Outline

- Vallduví's information packaging: Link, Tail and Focus
- File change semantics of IP
- Hoffman's database query answering
- Criticism of link as "locator"



Vallduví: Examples

Link-Focus:

- (131) The boss [_F CALLED].
 (132) The boss [_F visited a broccoli plantation in COLOMBIA].
 (133) The boss [_F I wouldn't BOTHER].
 (134) Broccoli the boss [_F doesn't EAT].

Link-Focus-Tail:

- (135) The boss [_F HATES] broccoli.
 (136) The farmers [_F already SENT] the broccoli to the boss.



Vallduví: Examples

All Focus:

- (137) [_F The BOSS called].
 (138) Waiter! [_F There's a fly in my cream of broccoli soup]!
 (139) What doesn't the boss like? [_F BROCCOLI].

Focus-Tail:

- (140) I can't believe this! The boss is going crazy!
 [_F BROCCOLI], he wants now.



IP and File Change Metaphor

- Semantics of IP in terms of operations on file-cards:
 - go to (introduce) a new card
 - go to an existing card
 - access a record on a card
 - add/modify a record on a card
- Four possible instruction types for IP:
 - linkless all-focus sentence: UPDATE-ADD(I_S)
 - focus-tail sentence: UPDATE-REPLACE(I_S , RECORD(f_c))
 - link-focus sentence: GOTO(f_c), UPDATE-ADD(I_S)
 - link-focus-tail sentence: GOTO(f_c), UPDATE-REPLACE(I_S , RECORD(f_c))

= File-change metaphor taken literally;

cf. also (Reinhart, 1995; Erteschik-Shir, 1997)



Example(s)

- (141) a. H: I'm arranging things for the president's dinner. Anything I should know?
 b. S: Yes. The **president** [_F hates the Delft CHINA SET]. Don't use it.
 c. GOTO(125) (UPDATE-ADD(hates the Delft-china-set(125)))
- (142) a. H: In the Netherlands I got the president a big Delft china tray that matches the set he has in the living room. Was that a good idea?
 b. S: Nope. The **president** [_F HATES] the Delft china set.
 c. GOTO(125)
 (UPDATE-REPLACE(hates, { _ : _ Delft-china-set(125) }))



Example(s)

- (143) H: I'm arranging things for the president's dinner. Anything I should know?
 S: Yes. The **president** always uses plastics dishes.
 [_F (He) hates the Delft CHINA SET].
 UPDATE-ADD(hates the Delft-china-set(125))
- (144) H: In the Netherlands I got the president a big Delft china tray that matches the set he has in the living room. Wille the president like it?
 S: Nope. [_F (He) HATES] the Delft china set.
 UPDATE-REPLACE(hates, { _ : _ Delft-china-set(125) }))

Links Without Locations

(Hendriks and Dekker, 1995):

- criticism of the file-change approach
 - links only seem to make sense if we assume files as locations of information
 - what locus of update is to be associated with quantified, negative or disjunctive links?
 - how about multiple links in one sentence?
 - why pronouns as part of focus?
- semantics of information packaging in DRT
- links: non-monotone anaphora

Links Without Locations

(Hendriks and Dekker, 1995):

Non-monotone Anaphora Hypothesis::

Linkhood (marked by L+H* in English) serves to signal non-monotone anaphora. If an expression is a link, then its discourse referent Y is anaphoric to an antecedent discourse referent X such that $X \not\subseteq Y$.

- (145) The guys were plying basketball in the rain.
- a. **The fathers** were having fun.
 - b. The fathers were having fun.

Hoffman's Application of IP

- Modeling discourse functions of Turkish word order
 - (Hoffman, 1995b): answers to wh- and yes/no-questions in a DB query task
 - (Hoffman, 1996): translation English → Turkish
- CCG-based grammar formalization
- Approach to IS based on (Vallduví, 1992; Vallduví, 1994):
- Association of sentence positions with discourse functions:
 - sentence initial position tends to be the topic
 - immediately preverbal position tends to be focus
 - elements between topic and focus and postverbal elements are in the ground

IP Representation

(Hoffman, 1995b; Hoffman, 1995a): topic vs. comment (=ground/focus)

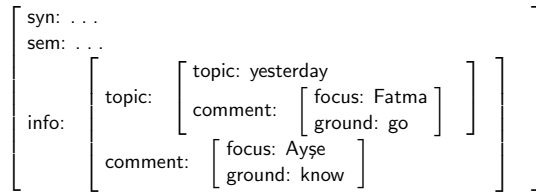
$$(146) \left[\begin{array}{l} \text{syn: } \dots \\ \text{sem: } \dots \\ \text{info: } \left[\begin{array}{l} \text{topic: } \dots \\ \text{comment: } \left[\begin{array}{l} \text{focus: } \dots \\ \text{ground: } \dots \end{array} \right] \end{array} \right] \end{array} \right]$$

- Topic has the value “recoverable” when zero-pronoun or in verb-initial sentences (all-focus)
- T/C structures fully recursive

IP Representation

(Hoffman, 1995b):

(147) *Dün Fatma'nın gittiğini Ayşe biliyor.*
 Yesterday Fatma-Gen go-Ger-Acc Ayşe knows.
 It's AYŞE who knows that yesterday, FATMA left.

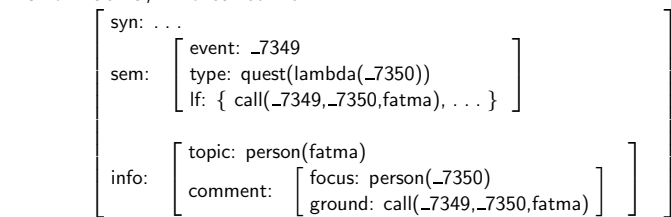


DB Question Answering System

1. Parser determines *syn, sem, info*
2. Planner executes simple plans to handle different types of questions:
 - i. determine question type (*sem : type*): (a) wh-q; (b) yes/no-q: Prop-q (q-morph on verb); Focused-q (q-morph on non-verb); Schedule-q (ability)
 - ii. query DB with *sem : lf*, respecting IP of question
 if success then generate corresponding answer
 else generate a "negative" answer
 - iii. plan answer: copy as much as possible from question, add/modify
 IP: topic of question → topic of answer; info from DB → focus

Example 1

(148) *Fatma'ya kim aradı?*
 Fatma-Acc who call-Past?
 As for Fatma, who called her?

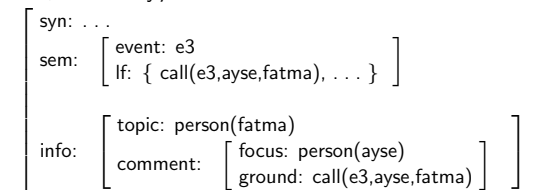


db_file(fatma, person(fatma)).
 db_file(fatma, call(e3, ayse, fatma)).
 db_file(fatma, see(e4, fatma, ahmet)).

Example 1

db_file(fatma, person(fatma)).
 db_file(fatma, call(e3, ayse, fatma)).
 db_file(fatma, see(e4, fatma, ahmet)).

(149) *Fatma'ya Ayşe aradı.*
 Fatma-Acc Ayşe call-Past
 As for Fatma, it was Ayşe who called her.



Example 2

db_file(fatma, person(fatma)).
 db_file(fatma, call(e3,ayse,fatma)).
 db_file(fatma, see(e4,fatma,ahmet)).

(150) *Fatma'yn Ahmet mi aradı?*

Fatma-Acc Ahmet Quest call-Past

As for fatma, was it Ahmet who called her?

syn: . . .	
sem: [event: _9041]	
	[type: quest(yes/no,ahmet)]
	[lf: { call(_9041,ahmet,fatma), . . . }]
info: [topic: person(fatma)]	
	[focus: person(ahmet)]
	[comment: [ground: call(_9041,ahmet,fatma)]]

Example 2

db_file(fatma, person(fatma)).
 db_file(fatma, call(e3,ayse,fatma)).
 db_file(fatma, see(e4,fatma,ahmet)).

(151) *Hayır, Fatma'yn Ayşe aradı.*

No, Fatma-Acc Ayşe call-Past

No, as for Fatma it was Ayşe who called her.

syn: . . .	
sem: [event: e3]	
	[lf: { call(e3,ayse,fatma), . . . }]
info: [topic: person(fatma)]	
	[focus: person(ayse)]
	[comment: [ground: call(e3,ayse,fatma)]]

DB Question Answering System: Summary

- Wh-element belongs to focus of question
- “Topic-inheritance” from question to answer
- File-card organization in DB by topics
 - relevance of IP for DB organization?
 - either info must be duplicated or some info not accessible to search
 - does not scale well for multiple topics, or quantified topics, etc.
- cf. question answering system Tibaq (Hajičová and Hnátková, 1984): assign Topic-Focus Articulation to analyzed sentences, and take it into account when retrieving answers: answer only considered exhaustive iff Focus corresponds to question

Target WO in English → Turkish MT

(Hoffman, 1996)

- Determination of Topic and Focus w.r.t. contextual information.
- Using centering, old/new and contrastiveness.
- Not using cues from source language text!
- Topic and Focus determined by algorithms; the rest is Ground.

Topic Determination Algorithm

Given:

- sentence contents,
- list of discourse entities mentioned in text so far,
- C_f lists of current and preceding sentence (cf. Centering (Grosz et al., 1995))

Topic determination:

1. Try to choose most salient discourse-old entity.
2. Else try to choose a situation-setting adverb.
3. Else choose the first item on the C_f list of current sentence (i.e., Subject)

Focus Determination Algorithm

Given:

- the non-topic rest of the sentence contents,
- list of discourse entities mentioned in text so far,

Focus determination:

1. If there are any discourse-new entities, put them into focus.
2. Else determine contrastive focusing of discourse-old information:
For each entity:
 - i. Construct a set of alternatives based on the entity's semantic type
 - ii. If the alternative set is not empty, put the entity into focus

Target WO in Polish → Turkish MT

Contrary to (Hoffman, 1996), (Styś and Zemke, 1995) argue for discourse analysis of the source text in order to preserve its communicative meaning in MT.

- Tracking centers according to Centering Theory (Grosz et al., 1995)
- Additional criteria for center evaluation: special center-poiting constructions, demonstrative pronouns, possessive and demonstrative modifiers, definiteness award, indefiniteness penalty
- Further modifications: gradation of center values, center values for all NPs, composite computation of center values, referential distance, synonyms
- Set of ordering criteria (end weight, given fronting, short before long, specific patterns) and preferences based on statistical models

Summary

- Information packaging: in essence very similar to TFA
- Crosslinguistic comparison of IP realization means
- File-change based semantics: links have an ushering function
- Four basic types of context update instructions based on IP constellation
- IP of question → IP of answer
- Database organization according to IP?