



Modeling Information Structure in Discourse and Dialogue Processing

Ivana Kruijff-Korbayová

korbay@coli.uni-sb.de

<http://www.coli.uni-saarland.de/~korbay/>

Nordic Graduate School,
Helsinki, April 3-8 2005

Lecture 4 Outline

- Vallduví's Information Packaging
- File-Change Metaphor for IP Semantics
- Halliday's Thematic Structure
- Daneš's Thematic Progression Types



Vallduví's Information Packaging

Information Packaging

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

- IS-partitioning into *Ground* and *Focus*;
Ground further partitioned into *Link* and *Tail*
- partitioning defined on surface form, not on sentence meaning!
- semantics of IP in terms of operations on file-cards: create, go-to, update,
...
(“file-change” metaphor taken literally)
cf. also (Reinhart, 1995; Erteschik-Shir, 1997)
- (Vallduví and Engdahl, 1996): analysis of IP realization in many languages

Vallduví: Examples

Link-Focus:

- (1) The boss [_F CALLED].
- (2) The boss [_F visited a broccoli plantation in COLOMBIA].
- (3) The boss [_F I wouldn't BOTHER].
- (4) Broccoli the boss [_F doesn't EAT].

Link-Focus-Tail:

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

Vallduví: Examples

All Focus:

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

Focus-Tail:

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

IP and File Change Metaphor

(Vallduví, 1992)

- operations on 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 IS:
 - UPDATE-ADD(I_S) for linkless all-focus sentence
 - UPDATE-REPLACE(I_S , RECORD(fc)) for focus-tail sentence
 - GOTO(fc), UPDATE-ADD(I_S) for link-focus sentence
 - GOTO(fc), UPDATE-REPLACE(I_S , RECORD(fc)) for link-focus-tail sentence

Example(s)

- (11) 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)))
- (12) 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)

- (13) 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))
- (14) H: In the Netherlands I got the president a big Delft china tray that matches the set he has in the living room. Will 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 $X \not\subseteq Y$.

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



IS in Systemic Functional Linguistics

Systemic Functional Linguistics

M. A. K. Halliday (1967, 1970, 1985, . . .)

- initially inspired by the Prague School works
- two independent (though interesting) dichotomies:
 - *Information Structure: Given-New*
 - *Thematic structure: Theme-Rheme*

Close semantic relationship (though they are not the same!):

“[O]ther things being equal, a speaker chooses the Theme from within what is Given and locates *information focus*, the climax of the New, within the Rheme.”

- Information Structure is listener-oriented
- Thematic Structure is speaker-oriented

SFL: Halliday

Information Structure:

- *information unit*
 - not exactly any unit in clause grammar (marked when boundaries overlap)
 - made of two functions/elements:
 - * *Given* (optional; info presented as recoverable)
 - * *New* (obligatory, marked by prominence; info presented as nonrecoverable)
 - Given typically precedes New (cf. CB/NB)
- Halliday discusses information structure in relation to intonation (in English)



SFL: Halliday

Thematic structure:

- Theme is the point of departure of a message;
Rheme is the remainder
- Theme grammaticalized in many languages:
 - e.g., English: first position
 - Japanese: suffix *-wa*
- Theme is a **textual** notion (related to global text-organization strategies; e.g., dates/places in biographies, places in geographical descriptions) (Fries, 1981), locations (e.g., menus, toolbars) or means (e.g., clicking on an icon, mouse button) in software manuals

THEME in “normal” declarative clauses

Definition 1. A THEME in declarative clauses is MARKED \Leftrightarrow it is not Subject.

Subject	nominal group	I	had a little nut-tree.
Subject	nominal group	A wise old owl	lived in an oak.
Subject	nominalization	What I want	is a proper cup of coffee.
Adjunct	adverbial group	Merrily	we roll along.
Adjunct	prep. phrase	On Saturday night	I lost my wife.
Complement	nominal group	A bag-pudding	the King did make.
Complement	nominalization	What they could not eat that night	the Queen next morning fried.
Predicator	(finite?) verb	Forget	it I never shall.

Maximally extended THEME

What if something comes *before* the first experiential element?

Halliday observes only limited set of types of words appearing before the first exp. element. He includes them under the label **THEME**, and classifies them: ¹

Well	but	then	Ann	surely	wouldn't	the best idea	
continuative	structural	conjunctive	vocative	modal	mood-marking	topical	
textual			interpersonal			experiential	
THEME							

	be to join the group
	RHEME

¹This is the *full* classification in the *typical* ordering.

Definitions of parts of THEME

Part of the THEME		Can contain only such an element:
textual	continuative	a member of small set of discourse signallers (<i>yes, no, well, oh, now</i>)
	structural	an obligatory thematic element*
	conjunctive	an conjunctive Adjunct*
interpersonal	vocative	any vocative item (personal name etc.)
	modal	a modal Adjunct*
	mood-marking	finite verbal operator or a <i>WH</i> - interrogative or imperative <i>let's</i>
experiential	topical	the first experiential element

* Defined later.

Structural THEME

OBLIGATORY THEMATIC ELEMENTS are the following expressions:

Class	Type	Examples
conjunctions	co-ordinator	and, or, nor, either, neither, but, yet, so, then
	subordinator	when, while, before, after, until, because, if, although, unless, since, that, whether, (in order) to even if, in case, supposing (that), assuming (that), seeing (that), given that, provided (that), in spite of the fact that, in the event that, so that
relatives	definite	which, who, that, whose, when, where, (why, how)
	indefinite	whatever, whichever, whoever, whomever, whenever, wherever, however

STRUCTURAL THEME contains obligatory thematic elements.

Conjunctive THEME

CONJUNCTIVE ADJUNCTS are the following expressions:

Type	Meaning	Examples
appositive corrective dismissive summative verificative	i.e., e.g. rather in any case in short actually	that is, in other words, for instance or rather, at least, to be precise in any case, anyway, leaving that aside briefly, to sum up, in conclusion actually, in fact, as a matter of fact
additive adversative variative	and but instead	also, moreover, in addition, besides on the other hand, however, conversely instead, alternatively
temporal comparative causal conditional concessive respective	then likewise so (if . . .) then yet at to that	meanwhile, before that, later on, next, soon, finally likewise, in the same way therefore, for this reason, as a result, with this is mind in that case, under the circumstances, otherwise nevertheless, despite that in this respect, as far as that's concerned

CONJUNCTIVE THEME contains conjunctive adjuncts.



Modal THEME

MODAL ADJUNCTS are the following expressions:

Type	Meaning	Examples
probability usuality typicality obviousness	how likely? how often? how typical? how obvious?	probably, possibly, certainly, perhaps, maybe usually, sometimes, always, (n)ever, often, seldom occasionally, generally, regularly, for the most part of course, surely, obviously, clearly
opinion admission persuasion entreaty desirability reservation validation evaluation prediction	I think I admit I assure you I presume how desirable? how reliable? how valid? how sensible? how expected?	in my opinion, personally, to my mind frankly, to be honest, to tell you the truth honestly, really, believe me, seriously please, kindly (un)fortunately, to my delight/distress, regrettably, hopefully at first, tentatively, provisionally, looking back on it broadly speaking, in general, ion the whole, in principle, strictly speaking (un)wisely, understandably, mistakenly, foolishly to my surprise, surprisingly, as expected, by chance

MODAL THEME contains modal adjuncts.

Real examples of extended THEME

- (16) Oh soldier, soldier, won't you marry me.
- (17) Please doctor don't give me any more of that nasty medicine.
- (18) On the other hand maybe on a weekday it would be less crowded.
- (19) So why worry.

Just to remember:

Part of the THEME		Can contain only such an element:
textual	continuative	a member of small set of discourse signallers (<i>yes, no, well, oh, now</i>)
	structural	an obligatory thematic element*
	conjunctive	an conjunctive Adjunct*
interpersonal	vocative	any vocative item (personal name etc.)
	modal	a modal Adjunct*
	mood-marking	finite verbal operator or a <i>WH</i> - interrogative or imperative <i>let's</i>
experiential	topical	the first experiential element



Thematic Progression Types



The Prague School Follow-up

František Daneš et. al (1957, 1970, 1974, 1985 . . .)

- systematic exploration of the relationship of *Theme* and *Rheme* to word order and intonation, as well as to the structure of text
- thorough analysis of *thematic progression* in text, i.e., textual patterns of thematization (typology of ways in which Themes relate to context)
- analysis of complex sentences in terms of condensed Theme-Rheme pairs

Daneš: Thematic Progression Types

CONTACT THEMATIC SEQUENCES:

Thematic sequence		Notation
thematization of the preceding theme	a <i>repetition</i> of the preceding rheme	$T_{i+1} = R_i$
	a <i>derivation</i> from the preceding rheme	$T_{i+1} \Leftarrow R_i$
continuous theme	a <i>repetition</i> of the preceding theme	$T_{i+1} = T_i$
	a <i>derivation</i> from the preceding theme	$T_{i+1} \Leftarrow T_i$
thematization of preceding utterances	the preceding utterance	$T_{i+1} = U_i$
	a <i>summarization</i> of utterances $U_i \dots U_j$	$T_{i+1} = I_{i,j}$
theme is derived from a <i>hypertheme</i> (the theme of a super-ordinate text unit, e.g. a text paragraph)		$T_{i+1} \Leftarrow T^*$



Thematic Progression Example

0. *Národní muzeum $T_0 \# R_0$ stojí na Václavském náměstí.*
 The National museum $T_0 \# R_0$ stands on the Wenceslas square.
- 1a. *Toto náměstí $T_{1a} \# R_{1a}$ je jedním z nejrušnějších míst v Praze.* $T_{1a} = R_0$
 This square $T_{1a} \# R_{1a}$ is one of the most busy places in Prague.
- 1b. *Horní části tohoto velkého prostranství $T_{1b} \# R_{1b}$ se tak dostalo krásné dominanty.* $T_{1b} \Leftarrow R_0$
 The top part of this large area $T_{1b} \# R_{1b}$ has thus received a nice dominant.
2. *Tato skutečnost $T_2 \# R_2$ je známa snad každému návštěvníkovi Prahy.*
 This fact $T_2 \# R_2$ is known perhaps by every visitor of Prague
- $T_2 = U_0$

- 3a. *Je {to} T_{3a} velmi památná budova.* $T_{3a} = T_0$
{It} T_{3a} is a very memorial building.
- 3b. *Sbírky Národního muzea T_{3b}#R_{3b} představují významnou*
The collections of the National museum T_{3b}#R_{3b} represent an important
národní kulturní hodnotu. $T_{3b} \Leftarrow T_0$
national cultural value.
4. *Jiná mimořádně významná pražská budova, Národní divadlo, T₄#R₄ je*
Another remarkably important Prague building, the National theatre, T₄#R₄ is
umístěna na Smetanově nábřeží. $T_4 \Leftarrow T^*$
situated on the Smetana embankment.

Daneš: T-R in Complex Text Units

Complex utterance		Notation
simple text units	one T-R nexus	$T_1 - R_1$
conjoined (paratactic) text units	conjoined nexuses	$(T_1 - R_1) conj (T_2 - R_2)$
	conjoined topics	$(T_1 conj T_2) - R_1$
	conjoined foci	$T_1 - (R_1 conj R_2)$
condensed (hypotactic) text units	nexus $T_2 - R_2$ incorporated into topic if $T_2 = T_1 \vee T_2 = R_1$, T_2 can be elided	$(T_1 cond (T_2 - R_2)) - R_1$ or equivalently $T * - R$
	nexus $T_2 - R_2$ incorporated into focus if $T_2 = T_1 \vee T_2 = R_1$, T_2 can be elided	$T_1 - (R_1 cond (T_2 - R_2))$ or equivalently $T - R*$

T-R Condensation Example

From (Korbayová and Kruijff, 1996)

1. První autorovi známou prací, $T_1 \# R_1$
The first work known to the author $T_1 \# R_1$
2. která $T_2 \# R_2$ se zabývá strukturálním programováním
which $T_2 \# R_2$ is concerned with structural programming
3. $T_3 \# R_3$ a opírá se o gramatický formalismus (afixové gramatiky),
and $T_3 \# R_3$ relies on a grammar formalism (affix grammars),
4. je práce Silvarberga (1978).
is the work of Silvarberg (1978).

The complex utterance can be analyzed as $(T_1 \text{ cond } (T_2 - (R_2 \text{ conj } R_3))) - R_1$
where $T_3 = T_2$, and T_3 is elided

Summary and Conclusions

- Information packaging: in essence very similar to TFA
- File-change based semantics: links have an ushering function
- Links without locations?
- Where do topics/themes/links come from, how they relate to one another?
- Textual function of theme in Halliday's sense: scaffolding



References

- Wallace Chafe. 1976. Givenness, contrastiveness, definiteness, subjects, topics and points of view. In Charles Li, editor, *Subject and Topic*, pages 25–55. Academic Press, New York.
- František Daneš. 1970. Zur linguistischen [a]nalyse der [t]extstruktur. *Folia Linguistica*, 4:72–78.
- František Daneš. 1974. Functional sentence perspective and the organization of the text. In František Daneš, editor, *Papers on functional sentence perspective*, pages 106–208. Academia, Prague, Czech Republic.
- František Daneš. 1979. O identifikaci známé (kontextově zapojené) informace v textu [on the identification of known (contextually dependent) information in text]. *Slovo a slovesnost*, 40:257–270. Prague, Czech Republic.
- František Daneš. 1985. *Věta a text [A sentence and a text]*, volume 21 of *Studie a práce lingvistické*. Academia, Prague, Czech Republic.
- Nomi Erteschik-Shir. 1997. *The dynamics of focus structure*. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.
- Peter H. Fries. 1981. On the status of theme in English: Arguments from discourse. *Forum Linguisticum*, 6(1):1 – 38.
- Michael A. K. Halliday. 1967. Notes on transitivity and theme in English — parts 1 and 2. *Journal of Linguistics*, 3(1 and 2):37–81 and 199–244.
- Michael A.K. Halliday. 1970. *A Course in Spoken English: Intonation*. Oxford University Press, Oxford.



- Michael A.K. Halliday. 1985. *Introduction to Functional Grammar*. Edward Arnold, London, U.K.
- Herman Hendriks and Paul Dekker. 1995. Links without locations. information packaging and non-monotone anaphora. In Paul Dekker and Martin Stokhof, editors, *Proceedings of the 10th Amsterdam Colloquium*, pages 339–358. Institute of logic, language and computation, Department of Philosophy, Universiteit van Amsterdam.
- Ivana Korbayová and Geert-Jan Kruijff. 1996. Identification of topic-focus chains. In S. Botley, Glass J., T. McEnery, and A. Wilson, editors, *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC96)*, number 8 in Technical papers, pages 165–179, Lancaster, U.K., jul. University Centre for Computer Corpus Research on Language, University of Lancaster.
- Tania Reinhart. 1995. Topics and the conceptual interface. In Hans Kamp and Barbara Partee, editors, *Proceedings of the workshop on Context-Dependence in the analysis of linguistic meaning, in Prague, February 1995, and in Bad Teinach, May 1995*, volume 1, pages 355–393, Azenbacher st. 12, Stuttgart, Germany. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Enric Vallduví and Elisabet Engdahl. 1996. The linguistic realisation of information packaging. *Linguistics*, 34:459–519.
- Enric Vallduví. 1992. *The Informational Component*. Garland, New York.
- Enric Vallduví. 1994. The dynamics of information packaging. In Elisabeth Engdahl, editor, *Integrating Information Structure into Constraint-based and Categorical Approaches*, pages 3–26. ILLC/Department of Philosophy, University of Amsterdam, The Netherlands.