



Einführung in Pragmatik und Diskurs

Information Structure

Ivana Kruijff-Korbayová

korbay@coli.uni-sb.de

<http://www.coli.uni-saarland.de/courses/pd/>

Summer Semester 2009

Information Structure

- Introduction
- Question test for IS
- IS realization
- IS meaning
- IS in practical applications

Reading: [Ste00]; [KKERK03]

Illustration (1)

- (1) Sign in London underground:
Dogs must be carried.

Hunde müssen getragen werden.
Einen Hund muss man tragen.

Interpretation:

If you have a dog, you must carry it.

Illustration (2)

- (2) Sign in a synagog:
Hats must be worn.
- (3) Man muss einen Hut tragen.
- (4) Es müssen Hüte getragen werden.
- (5) Hüte müssen getragen werden.

Interpretation:

- (6) *You must wear a hat.*

Illustration

- (7) Dogs must be CARRIED.
Hunde müssen GETRAGEN werden.
- (8) # DOGS must be carried.
HUNDE müssen getragen werden.
- (9) HÜTE müssen getragen werden.
- (10) # Hüte müssen GETRAGEN werden.

(Capitals denote intonation center = main stress in the sentence.)

Illustration

Different languages use various means to realize the different meanings:
e.g., Czech: word order

(11) *Psi se musí NÉST.*

Dog refl must carry

Hunde müssen getragen werden.

(12) # *Musí se nést PSI.*

Must refl carry dog

Es müssen Hunde getragen werden.

Illustration (4)

Example from a dialogue system:

- (13) U: What is the status of the stove?
Im welchen Zustand befindet sich die Heizung?
S: The stove is switched ON.
Die Heizung ist ANGESCHALTET.
- (14) U: Which device is switched on?
Was ist angeschaltet?
S: The STOVE is switched on.
Die HEIZUNG ist angeschaltet.

Observations

- Intonation, word order, syntactic structure and other aspects of linguistic form are tightly related
- Variation in these aspects of linguistic form is not arbitrary!
- These aspects of linguistic form reflect the relation of the utterance to the context
- Whether an utterance is appropriate or inappropriate in a given context depends not only on **what is said** but also **how**
- Recall: Grice's Maxim of Manner

Information Structure

- IS concerns a division (partitioning) of an utterance meaning:
 - **Theme** the part which relates it to the purpose of the discourse and anchors the content to the context (i.e., what speaker and hearer are attending to); what the utterance is *about*, the *topic* that the speaker means to address; may also *restrict* the context to a particular type(s) of situation(s)
 - **Rheme** the part which advances the discourse, i.e., adds or modifies some information (i.e., the informative part); what the speaker says about the Theme, i.e., the Rheme is semantically predicated over the Theme
- IS is an inherent aspect of meaning —it is an important factor in establishing coherence with respect to the context in which a sentence is uttered.

Theme-Rheme Partitioning: Examples

- (15) U: What is the status of the stove?
Im welchen Zustand befindet sich die Heizung?

S: The stove is switched ON.

Die Heizung ist ANGESCHALTET.
Theme *Rheme*

- (16) U: Which device is switched on?
Was ist angeschaltet?

S: The STOVE is switched on.

Die HEIZUNG ist angeschaltet.
Rheme *Theme*

Theme-Rheme Partitioning: Examples

(17) Dogs must be CARRIED.

Hunde müssen GETRAGEN werden.
Theme *Rheme*

(18) HATS must be worn.

HÜTE müssen getragen werden.
Rheme

Theme-Rheme Partitioning: Examples

- (19) Das Mädchen kauft eine Bluse.
Theme *Rheme*
- (20) Das Mädchen kauft eine Bluse.
Theme *Rheme*
- (21) Das Mädchen kauft eine Bluse.
Theme *Rheme* *Theme*
- (22) Das Mädchen kauft eine Bluse.
Rheme *Theme*

Question Test for IS

Question-answer pairs are commonly used to indicate or test the context in which a particular IS is appropriate: the question determines the Theme; what is asked for is the Rheme; the answer “fills” the Rheme.

(23) Was macht das Mädchen?

Theme
Rheme

(24) Wer kauft eine Bluse?

Rheme
Theme

Swapping the questions in the two examples results in incoherent Q-A pairs: the answers become infelicitous, because the IS partitioning then does not match the context set by the question.

Question Test for IS

The linguistic form of an utterance may be compatible with several different IS partitionings (as reflected by different questions)

(25) Das Mädchen kauft eine BLUSE.

a. Was kauft das Mädchen?

Das Mädchen kauft eine BLUSE.
Theme *Rheme*

b. Was macht das Mädchen?

Das Mädchen kauft eine BLUSE.
Theme *Rheme*

c. Was passiert gerade? / Was ist neu?

Das Mädchen kauft eine BLUSE.
Rheme

- (26) Das Mädchen kauft eine BLUSE.
Wer kauft eine Bluse?
Was macht das Mädchen mit eine Bluse?

- (27) Wer kauft eine Bluse?

Das MÄDCHEN kauft eine Bluse.
Rheme *Theme*

- (28) Was macht das Mädchen mit eine Bluse?

Das Mädchen KAUFT eine Bluse.
Theme *Rheme* *Theme*



Question Test for IS

(29) Jan ist aus London nach PARIS geflogen.

- (30) a. Was ist passiert?
b. Was hat Jan gemacht?
c. Wohin ist Jan geflogen?

(29) is good as an answer to (30a-c). It has several possible IS partitionings.

(31) is only a matching answer to (32).

(31) Jan ist nach PARIS aus London geflogen.

(32) Wohin ist Jan aus London geflogen?

(33) is only a matching answer to (34).

(33) Jan ist aus LONDON nach Paris geflogen.

(34) Woher ist Jan nach Paris geflogen?

Neither of the above is a good answer to (35).

(35) Wer ist aus London nach Paris geflogen?

(36) JAN ist aus London nach Paris geflogen.

The Meaning of IS

- Steedman defines the meaning of the IS partitioning in terms of selecting one member from a *presupposed set of alternatives*
 - The Theme-Rheme partitioning presupposes a *Rheme-alternative set*, i.e., a set of alternative propositions that could possibly answer the corresponding question in the given context;
 - Rheme then restricts the Rheme-alternative set to a singleton
- This is a pragmatic presupposition that the relevant alternative set is available in the context.
- The systematic recognition of the alternative sets, and their maintenance as a discourse progresses are open research issues.

IS Meaning

(37) $\underbrace{\text{Das deutsche MÄDCHEN kauft}}_{\textit{Theme}} \underbrace{\text{eine SEIDENE Bluse}}_{\textit{Rheme}} .$

Rheme-alternatives: *das deutsche Mädchen kauft etwas*

{*das deutsche Mädchen kauft eine seidene Bluse, das deutsche Mädchen kauft eine Bluse aus Baumwolle, das deutsche Mädchen kauft eine Bluse aus Nylon, etc. }*

IS: Meaning Differences

- (38) a. Smoke OUTSIDE!
b. SMOKE outside!
- (39) a. STAFF behind counter.
b. Staff behind COUNTER.
- (40) a. *On the Shetlands one speaks ENGLISH.*
b. *One speaks English on the SHETLANDS.*
- (41) a. *Several people in this room know three languages.*
b. *Three languages are known by several people in this room.*
- (42) a. Officers always escorted BALLERINAS.
b. OFFICERS always escorted ballerinas.



IS: Meaning Differences

- (43) *Problém není v tom, že Janouch koupil gamma nůž, ale že gamma nůž koupil Janouch.*
- The problem is not that Janouch bought a gamma-knife, but that the gamma-knife was bought by Janouch.*
 - The problem is not that Janouch bought a GAMMA-KNIFE, but that JANOUCH bought the gamma knife.*
- (44) *Dobrá zpráva je, že Češi udělali revoluci.
Špatná zpráva je, že revoluci udělali Češi.*
- The good news is that the Czechs made a revolution; the bad news is that a revolution was made by the Czechs.*
 - The good news is that the Czechs made a REVOLUTION; the bad news is that the CZECHS made a revolution.*

Second Dimension of IS Partitioning: Background-Focus

(45) Das Mädchen kauft eine seidene Bluse.
Theme *Rheme*

(46) Es gibt eine Bluse aus Baumwolle und eine aus Seide.
Was kauft das Mädchen?

Das Mädchen kauft eine SEIDENE Bluse.
Theme *Rheme*

(47) Es gibt eine seidene Bluse und eine seidene Jacke.
Was kauft das Mädchen?

Das Mädchen kauft eine seidene BLUSE.
Theme *Rheme*

Second Dimension of IS Partitioning: Background-Focus

Background-Focus partitioning reflects an abstract notion of contrast between alternatives available in the discourse context, against which the actual utterance is cast;

B/F partitioning within Theme and/or Rheme reflects alternative Theme(s) and/or Rheme(s) in the context.

The Focus-Background partitioning in Theme also presupposes a *Theme-alternative set*, i.e. a set of alternative questions; Focus within Theme then restricts the Theme-alternative set to a singleton

IS Partitioning: Background-Focus in Rheme

- (48) Es gibt eine Bluse aus Baumwolle und eine aus Seide.
Was kauft das Mädchen?

Das Mädchen kauft eine SEIDENE Bluse.

Theme *Focus*
Rheme

- (49) Es gibt eine seidene Bluse und eine seidene Jacke.
Was kauft das Mädchen?

Das Mädchen kauft eine seidene BLUSE .

Theme *Focus*
Rheme

IS Partitioning: Background-Focus in Theme

(50) Es gibt ein deutsches und ein amerikanisches Mädchen.

Was kauft das deutsche Mädchen?

Das DEUTSCHE Mädchen kauft eine BLUSE .



Focus
Theme




Rheme

(51) Es gibt eine deutsche Frau und ein deutsches Mädchen.

Was kauft das deutsche Mädchen?

Das deutsche MÄDCHEN kauft eine BLUSE .



Focus
Theme



Rheme

IS Partitioning: Background-Focus in Theme and Rheme

- (52) Ich weiss, dass die deutsche Frau eine Bluse aus Baumwolle kauft.
Aber was kauft das deutsche Mädchen?

Das deutsche MÄDCHEN kauft eine SEIDENE Bluse .



IS Meaning

(53) Das deutsche MÄDCHEN kauft eine SEIDENE Bluse .



Rheme-alternatives: *das deutsche Mädchen kauft etwas*

{ das deutsche Mädchen kauft eine seidene Bluse, das deutsche Mädchen kauft eine Bluse aus Baumwolle, das deutsche Mädchen kauft eine Bluse aus Nylon, etc. }

Theme-alternatives: *jemand kauft etwas*

{ das deutsche Mädchen kauft etwas, die deutsche Frau kauft etwas, der deutsche Mann kauft etwas, der deutsche Junge kauft etwas, etc. }

“Focussing jokes”

- (54) Why do we buy clothes?
Because we can't get them for free.
- (55) Why do we dress girls in pink and boys in blue?
Because they can't dress themselves.
- (56) Why do firemen wear red suspenders?
To keep their pants up.
- (57) Why do you rob banks?
Because that's where the money is!



“Focussing jokes” : Explanation

The question’s “default” IS-partitioning induces a presupposition; the answer should provide the missing information asked for, but the given answers are answers to different questions (moreover, trivial ones), i.e.:

- (58) Why do we buy CLOTHES? (why clothes)
Because we can’t get them for free. (reason for buying)
Why do we BUY clothes?
- (59) Why do we dress girls in PINK and boys in BLUE? (why pink/blue)
Because they can’t dress themselves. (reason for us doing it)
Why do WE dress girls in pink and boys in blue?
- (60) Why do firemen wear RED suspenders? (why red color)
To keep their pants up. (reason for suspenders)

Why do firemen wear red SUSPENDERS?

For 57, the difference is between:

- “broad focus” reading, where *robbing banks* is included in Rheme
- “narrow focus” reading, where only *banks* is included in Rheme, while the fact that the hearer robs something is presupposed

While without any particular context, the broad focus reading is the default, the answers the narrow focus reading of the question, and that’s what creates the surprise effect.

(61) Why do you rob BANKS? (why robbing)
 Because that’s where the money is! (reason for why banks)

IS and Surface Realization

- IS influences various aspects of linguistic form
 - **intonation** (prosody); predominant in English
 - (word) **ordering**; e.g. predominant in Czech and other “free word order” languages, also to some extent in German (particularly in the “Mittelfeld”)
 - **morphological/grammatical marking**; e.g., particles ‘wa’ and ‘ga’ in Japanese
 - **syntactic constructions**, e.g. it-cleft, wh-cleft, passivization, etc.
 - **ellipsis**
- Different languages employ and combine these aspects of surface realization differently, depending on their typological characteristics;

IS Realization Means: Intonation

- placement and type of accent and type of boundary tone
 - accent placed on Focus part
 - Theme-Focus and Rheme-Focus use different type of accent and boundary tone (tune): $L+H^*LH^0\%$ is a prototypical Theme-tune, $H^*LL^0\%$ is a prototypical Rheme-Tune
- placement of boundaries: boundary between Theme and Rheme

(62) Ich weiss, dass die deutsche Frau eine seidene Jacke anzieht.
Aber was macht das deutsche Mädchen?

(63) Das deutsche MÄDCHEN kauft eine seidene BLUSE .
 $L+H^*LH^0\%$ $H^*LL^0\%$
 Focus Focus
 Theme Rheme

IS Realization Means: Word Order

“Normal” (default) order: Theme before Rheme

(64) Was kauft das Mädchen? Das Mädchen kauft eine BLUSE.
Theme *Rheme*

(65) MP3: Was soll mit den Liedern gemacht werden?
U: Mit den Liedern soll eine Playlist erstellt werden.

(66) U: Bitte suche Titel von Madonna.
MP3: Einen Moment bitte. ... Von Madonna haben wir 1711 Treffer.

“Subjective ordering”: Rheme before Theme (is marked)

(67) Was kauft das Mädchen? Eine BLUSE kauft das Mädchen.
Rheme *Theme*

IS Realization Means: Syntax

Syntactic constructions can be used to mark IS:

Cleft: *It is Rheme (that/who) Theme*

1. What does John hate? It is COMICS John hates.
2. Who hates comics? It is JOHN who hates comics.

Pseudo-cleft: *Who/What Theme is/are Rheme*

1. What does John hate? What John hates are COMICS.
2. Who hates comics? Who hates comics is JOHN.

Passivization: allows opposite ordering than corresponding active sentence

- (68) Who hates comics?
- Comics are hated by JOHN.
 - (JOHN hates comics.)

There-insertion: gets Rheme-subject away from the beginning of the sentence

- (69) a. What is in the garden? There is a TROLL in the garden.
b. Where is a troll? A troll is in the GARDEN.
- (70) a. Es spielt die Tschechische Philharmonie. Es dirigiert Hilary Griffiths.
b. Die Tschechische Philharmonie spielt. Hilary Griffith dirigiert.

IS Realization Means: Ellipsis

Theme can be left implicit when it is “known” (available, retrievable):

- (71) U: What is the status of the stove?
Im welchen Zustand befindet sich die Heizung?
S: It is ON.
Es ist ANGESCHALTET.
- (72) U: Which device is switched on?
Was ist angeschaltet?
S: The STOVE.
Die HEIZUNG.

-
- (73) U: Wieviele Titel von Madonna gibt es?
MP3: Einen Moment. ... 1711.
- (74) U: Bitte suche Titel von Madonna.
MP3: Einen Moment bitte. ... Von Madonna haben wir 1711 Treffer.
MP3: 1711.

Ellipsis example from the Map Task corpus

- (75) G: where are you in relation to the top of the page just now?
F: Uh, about four inches.
G: Four inches?
F: Yeah.
G: Where are you from the left-hand side?
F: About two.

Computational Modeling of IS in Applications

- word order:
 - database-querying dialogue [Hof95]
 - MT: En. → Turkish [Hof96]; En. → Polish [SZ95]
 - text generation: Cz., En. [KKBK02]
 - anaphora resolution/generation in text [HKK90, HKK92], in dialogue [KT02]
- intonation in dialogue: En. [Pre95], Ger., En. [KKERK03]; En. [MFLW04]
- non-linguistic aspects in modeling embodied conversational agents: gestures, head-nods, gaze [PCB⁺98, CSY00]
- Various researches have observed that distance, posture shifts and other body movements seem to accompany changes in the topic or social relationship

Summary and Conclusions

- IS-partitioning: Theme-Rheme, Focus-Background
- IS realization: intonation, word order, syntactic constructions, ...
- IS is an essential aspect of meaning at the interface between utterance and discourse
- IS is important for accurate NL processing
- formal accounts are emerging, some embodied into practical systems
- many questions concerning IS partitioning and its realization in different languages still open
- Further research topics: [KKS03]
 - further systematization of terminologies
 - formalization and computational modeling&testing
 - empirical and corpus-based studies
 - cross-linguistic investigations and multilingual applications



References

- [CSY00] Justine Cassell, Matthew Stone, and Hao Yan. Coordination and context-dependence in the generation of embodied conversation. In *Proceedings of the INLG Conference*, 2000.
- [HKK90] Eva Hajičová, Petr Kuboň, and Vladislav Kuboň. Hierarchy of salience and discourse analysis and production. In *Proceedings of COLING'90*, pages 144–148, 1990.
- [HKK92] Eva Hajičová, Vladislav Kuboň, and Petr Kuboň. Stock of shared knowledge - a tool for solving pronominal anaphora. In *Proceedings of COLING'92*, pages 127–133, 1992.
- [Hof95] Beryl Hoffman. Integrating free word order, syntax, and information structure. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics, Dublin*, pages 245–252, San Francisco, CA, 1995. Morgan Kaufmann.
- [Hof96] Beryl Hoffman. Translating into free word order languages. In *Proceedings of the International Conference on Computational Linguistics (COLING-96), Copenhagen*, pages 556–561, Copenhagen, Denmark, 1996. Center for Sprogteknologi.
- [KKBK02] Ivana Kruijff-Korbayová, John Bateman, and Geert-Jan M. Kruijff. Generation of contextually appropriate word order. In van Deemter and Kibble [vDK02], pages 193–222.
- [KKERK03] Ivana Kruijff-Korbayová, Stina Ericsson, Kepa Joseba Rodríguez, and Elena Karagjosova. Producing contextually appropriate intonation is an information-state based dialogue system. In *Proceedings of*



the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pages 227–234. ACL, 2003.

- [KKS03] Ivana Kruijff-Korbayová and Mark Steedman. Discourse and information structure. *Journal of Logic, Language and Information: Special Issue on Discourse and Information Structure*, 12(3):249–259, 2003.
- [KT02] Emiel Krahmer and Mariet Theune. Efficient context-sensitive generation of referring expressions. In van Deemter and Kibble [vDK02], pages 223–264.
- [MFLW04] Johanna Moore, Mary Ellen Foster, Oliver Lemon, and Michael White. Generating tailored, comparative descriptions in spoken dialogue. In *Proceedings of the Florida Artificial Intelligence Research Symposium (FLAIRS)*, 2004.
- [PCB⁺98] Catherine Pelachaud, Justine Cassel, Norman Badler, Mark Steedman, Scott Prevost, and Mathew Stone. Synthesizing cooperative conversation. In Harry Bunt and R. J. Beun, editors, *Multimodal Human-Computer Communication - Systems, Techniques and Experiments*, Springer Notes in Artificial Intelligence 1374, pages 68–88. Springer Verlag, 1998.
- [Pre95] Scott Prevost. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. dissertation, IRCS TR 96-01, University of Pennsylvania, Philadelphia, 1995.
- [Ste00] Mark Steedman. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4):649–689, 2000.
- [SZ95] Malgorzata E. Styś and Stefan S. Zemke. Incorporating discourse aspects in english-polish mt: Towards robust implementation. In *Proceedings of the RANLP Conference*, Velingrad, Bulgaria, 1995.

- [vDK02] Kees van Deemter and Rodger Kibble, editors. *Information Sharing. Reference and Presupposition in Language and Interpretation*, volume 143 of *Lecture Notes*. CSLI, 2002.