



Modeling Information Structure in Discourse and Dialogue Processing

Ivana Kruijff-Korbayová

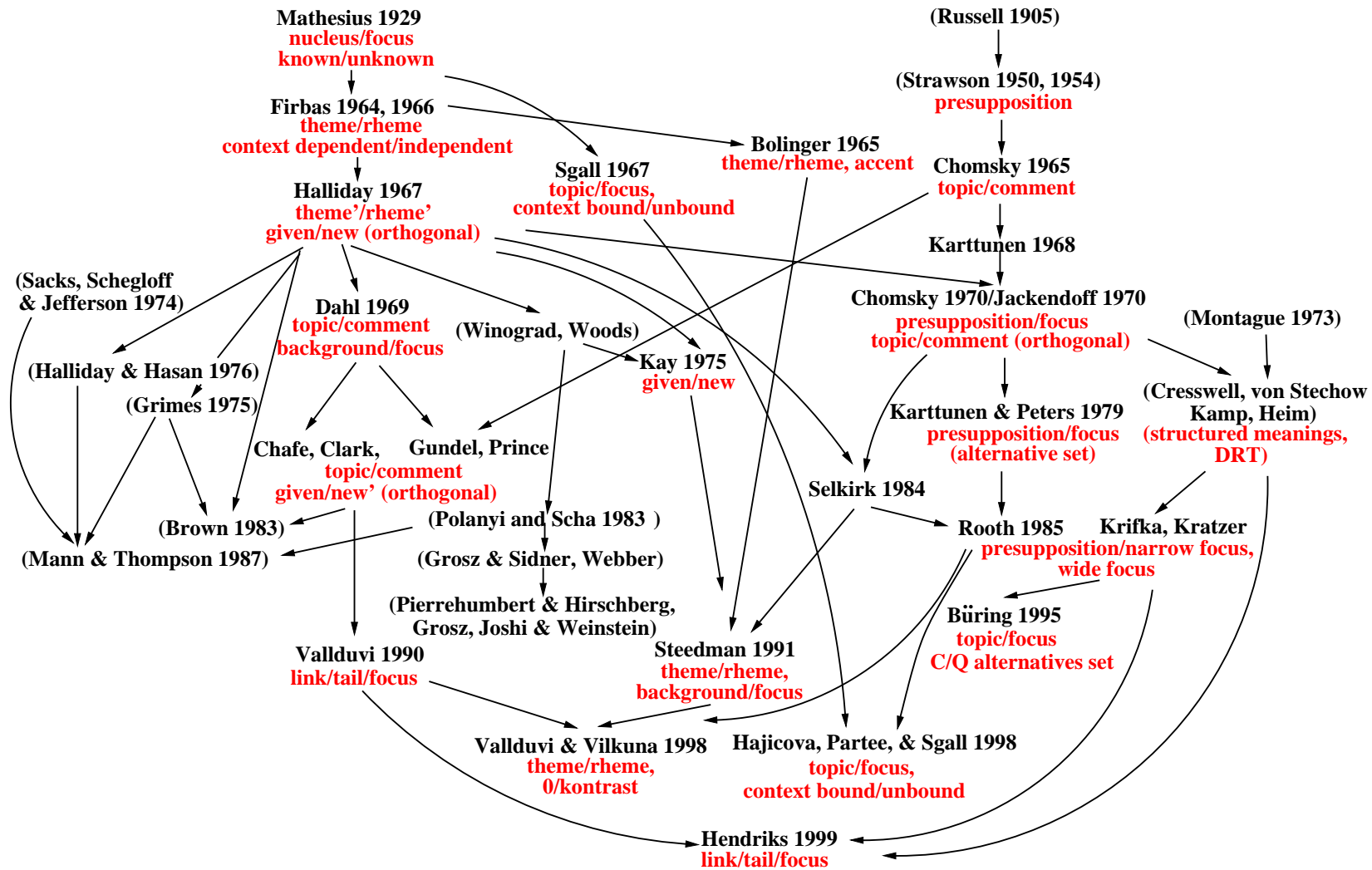
korbay@coli.uni-sb.de

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

Nordic Graduate School,
Helsinki, April 3-8 2005

Lecture 2 Outline

- IS in the Prague School of Linguistics
- Follow-up: Topic-Focus Articulation in Functional Generative Description
- IS-Sensitive Salience Modeling
- Applications: Reference Resolution and Generation
- Comparison with Centering Theory
- Comparison with Prince's Familiarity Taxonomy
- Comparison with Gundel et al.'s Givenness Hierarchy



IS in the Prague School of Linguistics

Vilém Mathesius (1915, 1924, 1929, 1936)

- introduced the IS notions Theme/Rheme into PSL
 - Theme (Cz. *jádro* 'nucleus'): what an utterance is about, point of departure
 - Rheme (Cz. *ohnisko* 'focus'): what an utterance says about the Theme
- structural comparison of English and Czech
- systematic attention to interplay of syntax and IS
- effects of word order variation on interpretation
- awareness of IS-importance for language as a means of communication
- in “free word-order” languages, WO tends to correspond to *communicative dynamism*, i.e., the ordering proceeds from contextually ‘given’/‘assumed’ to contextually ‘new’
- also in languages with “fixed word-order”, some constructions can serve as means of IS; English: WO-change accompanied by passivization

The Prague School Follow-up

Jan Firbas et al. (1957, 1966, 1975, 1992, . . .)

- analyzed different factors that influence *Functional Sentence Perspective* (=IS)
 - linear modification (word order)
 - semantic factor (character of semantic content and relations involved)
 - contextual factor (retrievability of information from preceding context)
- Theme/Transition/Rheme
- analyzed interplay of IS, syntactic structure and word order
- concludes that not only a dichotomy of *Theme-Rheme*, but a whole scale of *communicative dynamism* is concerned
- *degree of communicative dynamism*: the relative extent to which a linguistic element contributes towards the further development of the communication



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) : theme-continuation, rhematization of theme, derivation of theme from hypertheme, etc.
- analysis of complex sentences in terms of condensed Theme-Rheme pairs



The Prague School Follow-up

Petr Sgall (1967, 1979, . . .) with Eva Hajičová (1977, 1980)
and Jarmila Panevová (1986)

also Partee et al. (1998), etc.

- studies of various aspects of *Topic-Focus Articulation* (TFA)
- TFA as part of formal description of syntax and sentence meaning (dependency-based Functional Generative Description, FGD)
- relation between TFA and word order (when “free” WO)
- studies of *systemic ordering* (SO), i.e. neutral surface word order
- question test
- TFA and scope of negation, focusing adverbs and quantifiers
- TFA and presupposition vs. allegation
- TFA and salience of entities in the stock of shared knowledge



IS in Functional Generative Description

Topic-Focus Articulation in FGD

(Sgall et al., 1986; Hajičová et al., 1995b)

Topic (theme, “given” info): the part of the sentence structure that is being presented by the speaker as readily available in the hearer’s memory

Focus (comment, rheme): what is being asserted about the topic.

- Primarily, scope of negation or a “focalizer” adverb is constituted just by the Focus part of the sentence
- This notion of topic has much in common with the concept of background or restrictor, while focus comes close to nuclear scope (Partee et al., 1998)

Status of TFA in the Language System

- In FDG, TFA is considered an inherent aspect of the (underlying) syntactic structure of the sentence:
 - TFA is expressed by grammatical means, e.g., word order, morphemes or their clitic/weak vs. strong shapes, syntactic constructions, position of the sentence intonation center.
 - TFA is semantically relevant, e.g., restrictor vs. scope of quantifiers and other operators (negation, focalizers, e.g., “only”, “even”, “always”); topic tends to have “specific” interpretation.
- ⇒ TFA is a partitioning of a sentence (meaning), not only of utterance (meaning)

TFA Examples

(1) Q. What about dogs?

A. Dogs must be CARRIED.
Topic *Focus*

(2) Q. What must be carried?

A. DOGS must be carried.
Rheme *Theme*

(3) Q. What must we do in order take the metro?

A. DOGS must be carried.
Rheme

TFA Examples

Difference in broad/narrow focus, and hence in presuppositions:

- (4) a. They arrived by car at the LAKE.
b. They arrived at the lake by CAR.
- (5) a. They moved from Boston to CHICAGO.
b. They moved to Chicago from BOSTON.
- (6) a. Last year John came from Cambridge to STANFORD.
b. John came from Cambridge to Stanford last YEAR.
- (7) a. John made a canoe out of every LOG.
b. John made a CANOE out of every log.

TFA Examples

Difference in quantifier scopes:

- (8) a. Everybody in this room knows at least two languages.
b. At least two languages are known to everybody in this room.
- (9) a. John talked to everyone about a problem.
b. John talked about a problem to everyone.
- (10) a. John talked to few girls about many problems.
b. John talked about many problems to few girls.

TFA Examples

Difference in scope of negation:

(11) Harry nezpůsobil naše VÍTĚZSTVÍ.

Harry_{nom} not-cause our victory.

Harry didn't cause our VICTORY.

$\neg \text{cause}(\text{harry}, \text{win}(\text{speaker}_+))$ or also $\text{cause}(\text{harry}, \neg \text{win}(\text{speaker}_+))$

(12) Naše vítězství nezpůsobil HARRY.

our victory not-cause Harry_{nom}.

HARRY didn't cause our victory.

$\text{cause}(\neg \text{harry}, \text{win}(\text{speaker}_+))$

Similarly with “focussing adverbs”, e.g, *only*, *even*, *always*.

TFA Examples

Difference between presupposition and allegation:

(11) Harry nezpůsobil naše VÍTĚZSTVÍ.

Harry_{nom} not-cause our victory.

Harry didn't cause our VICTORY.

$\neg \text{cause}(\text{harry}, \text{win}(\text{speaker}_+))$ or also $\text{cause}(\text{harry}, \neg \text{win}(\text{speaker}_+))$

(12) Naše vítězství nezpůsobil HARRY.

our victory not-cause Harry_{nom}.

HARRY didn't cause our victory.

$\text{cause}(\neg \text{harry}, \text{win}(\text{speaker}_+))$

$\text{win}(\text{speaker}_+)$ is a presupposition in (12), but in (11) only an allegation (Partee, 1995): \approx global vs. local accommodation

Question Test and Systemic Ordering

Question test can be used to compare sentences with different order of dependent elements (arguments and free modifiers):

(13) a. John talked to few girls about many PROBLEMS.

b. John talked about many problems to few GIRLS.

(14) What do you know about John?

(15) How does John behave towards few girls?

(16) To whom does John talk about many problems?

(13a) can answer questions (14) and (15), whereas (13b) can only answer (16).

⇒ The ordering in (13a) is more basic than that in (13b). (13a) adheres to *systemic ordering*.

Systemic Ordering

Studies using the question test enable to determine *systemic ordering* for any given language: a language specific basic (neutral, primary) ordering of types of dependency roles

Sample SO for Czech (also Russian, Bulgarian):

Actor < Time < Purpose < Location < Means < Addressee < Patient < Source < Destination

Sample SO for English:

Time < Actor < Patient < Origin < Effect < Manner < Dir.from < Means < Dir.to < Location

Sample SO for German:

Actor < Time < Location < Means < Addressee < Patient < Source < Destination < Purpose

SO is one of the factors relevant for word order and for the placement of the intonation center. SO provides a “default”.

Communicative Dynamism (Underlying WO)

- The scale of CD (partial ordering) corresponds to the “dynamic” progression from topic-proper through intermediate parts to focus proper (carrying the intonation center).
- CD is relevant for quantifier scopes: more dynamic → narrower scope
- “Ideally”, surface word order respects CD. Deviations are due to:
 - speaker’s discourse strategy
 - grammar restrictions, e.g., verb-secondness (in German, Czech), placement of clitics (in Czech), placement of adjectives or other modifiers before/after their head (in German, Czech etc. vs. French), placement of intonation center (in Hungarian or Turkish) etc.

SO and CD: Rule 1

If sentence parts A and B are in the Focus of sentence S, and A precedes B under SO, i.e., $\text{DepRole}(A) <_{SO} \text{DepRole}(B)$,

then A precedes B in the CD (i.e., underlying word order) of S, i.e. $A <_{CD} B$.

Example:

SO: Temporal < Actor < Patient < Origin < Effect < Manner < Dir.from < Means < Dir.to < Locative

(17) What about John?

(John)_{Act} went (by car)_{Mann} (from Paris)_{Dir.from} (to NANCY.)_{Dir.to}

(18) Where did John go from Paris?

(John)_{Act} went (from Paris)_{Dir.from} (by car)_{Mann} (to NANCY.)_{Dir.to}

(From Paris,)_{Dir.from} (John)_{Act} went (by car)_{Mann} (to NANCY.)_{Dir.to}.

SO and CD: Rule 2

The boundary between Topic and Focus can be drawn between any two elements following the verb (in CD), provided that those belonging to the focus are arranged (in CD) in accordance with SO.

(19) (John)_{Act} went (by car)_{Mann} (from Paris)_{Dir.from} (to NANCY.)_{Dir.to}

(20) (John)_{Act} went (from Paris)_{Dir.from} (by car)_{Mann} (to NANCY.)_{Dir.to}

Contextual Boundness/Non-boundness

- TFA is a partitioning of the meaning of the sentence as a whole
- The individual lexico-semantic items (nodes in dependency tree) are considered either *contextually bound* (CB) or *contextually non-bound* (NB)
- CB items are those that the speaker treats as easily accesible (salient) in the hearer's memory (\rightarrow *stock of shared knowledge*)
- So, CB vs. NB is a primitive opposition, from which TFA is derived
- How can we determine whether an element is CB or NB?

SO, CD and CB/NB: Rule 1 Used Backwards

It is assumed that

- ordering of NB elements adheres to SO (Rule 1)
- deviation of CD from SO indicates an element is (treated as) CB

If $\text{DepRole}(A) <_{SO} \text{DepRole}(B)$ & $B <_{CD} A$ then A is CB.

(Note that B could be either CB or NB.)

(21) (John)_{Act} went (by car)_{Mann} (from Paris)_{Dir.from} (to NANCY.)^{NB}_{Dir.to}

(22) (John)^{CB}_{Act} went (from Paris)^{CB}_{Dir.from} (by car)_{Mann} (to NANCY.)^{NB}_{Dir.to}

(23) (From Paris,)^{CB}_{Dir.from} (John)_{Act} went (by car)_{Mann} (to NANCY.)^{NB}_{Dir.to}

CB/NB and TFA

Derivation of TFA from the CB/NB assignment:

- The main verb and its immediate dependents belong to the Topic if they are CB, and to the Focus if they are NB
- More deeply embedded elements belong to the Topic (Focus) if their governing element belongs there
- If the main verb and all its immediate dependents are CB, then the Focus consists of the NB elements embedded under the most dynamic CB element



Examples

(24) What trip did John make by car?

$(\text{John})^{CB}_{Act} \text{ went}^{CB} (\text{by car})^{CB}_{Mann} \quad (\text{from Paris})^{NB}_{Dir.from} (\text{to NANCY.})^{NB}_{Dir.to}$
Topic *Focus*

(25) Where did John go from Paris by car?

$(\text{John})^{CB}_{Act} \text{ went}^{CB} (\text{from Paris})^{CB}_{Dir.from} (\text{by car})^{CB}_{Mann} \quad (\text{to NANCY.})^{NB}_{Dir.to}$
Topic *Focus*

(26) By which car did John go from Paris to Nancy?

$(\text{John})^{CB}_{Act} \text{ went}^{CB} (\text{from Paris})^{CB}_{Dir.from} (\text{to NANCY})^{CB}_{Dir.to} (\text{by the car})^{CB}_{Mann}$
Topic
 $(\text{from Avis})^{NB}_{Orig}$
Focus

CB/NB Examples

- (27) How do you find your neighbourhood?
Our new neighbour has stolen my car.
- (28) Which teacher do you mean?
I mean our teacher of chemistry.
- (29) Which teacher of chemistry do you mean?
I mean our FEMALE teacher of chemistry from the first YEAR.
- (30) In the autumn, painters often look for nice sceneries in most varied places.
A painter arrived at a French VILLAGE on a nice September day.
- (31) Which house did John come to?
John came to the house which he wanted to buy.

TFA Identification

For “free” word order languages, (Hajičová et al., 1995b) propose an algorithm based on the following main points:

1. Complementations preceding the verb are CB (belong to the topic)
2. Among the complementations following the verb, those arranged in accordance with SO w.r.t. any other complementation are NB (belong to focus); those complementations that do not respect SO are CB (belong to topic)
3. The verb is generally ambiguous between belonging to topic or focus
4. If the intonation center is placed on a non-final element in the sentence, then the intonation center belongs to the focus, all complementations after the intonation center belong to the topic and for the rest, rules 1 and 2 apply

TFA Identification

For English (Hajičová et al., 1995b) state:

- Surface WO in English is determined by grammatical rules to a large extent, so intonation plays a much more decisive role.
- Rule 2 also applies.
- Otherwise, only certain regularities for simple sentences are incorporated into the algorithm. E.g., the following factors are taken into account:
 - Placement of verb at the end of the sentence
 - Definiteness/indefiniteness of the subject
 - For locative and temporal modification, specific information (Focus) vs. general setting (Topic).



IS and Saliency in Common Ground

The Praguian Model of Salience

The *stock of knowledge assumed by the speaker to be shared by the hearer* (SSK) comprises “not only knowledge in literal sense, but also a wide range of psychological phenomena including beliefs and other attitudes” (Sgall et al., 1986, p.55–56), (Hajičová, 1993, p.70)

- Dynamically changing stock of shared knowledge (SSK):
 - *Set of senses of linguistic expressions* (\approx discourse referents)
 - *Activation*: SSK is partially ordered to reflect the degrees of salience of its elements (i.e., immediate accessibility in the hearer’s memory)
 - *Dynamic character*: the set and the ordering are relativized to the time-points of utterances
- Activation of entities in SSK w.r.t. IS (Topic-Focus Articulation, TFA)

The Praguian Model of Salience

(Hajičová et al., 1990; Hajičová, 1993)

- Degrees of activation (salience):
 - continuous scale from “highly salient” to “faded away”
 - assignment rules take into account TFA partitioning and ling. form:
 1. Reference by expression in Focus → assign top activation
 2. Reference by a noun phrase in Topic → assign high activation (not top)
 3. Reference by a pronoun in Topic → maintain current activation
 4. Not re-accessed referent → decrease activation
 - Referents that remained highly activated (mentioned in Topic) fade away slower than referents only accessed in Focus
 5. Entity *associated* with an activated referent → assign a lowered activation

The Praguian Model of Salience

Predictions Based on Activation:

- A cooperative speaker chooses as CB items (Topic) what is already at/near the top of the activation scale; in contrast, NB items (Focus) can but don't have to be already activated.
- Appropriate use of linguistic forms according to activation.
- *Discourse topic*: an entity at/near top of activation scale (through a stretch of discourse)
- *Discourse segment*: a stretch of discourse with relatively “constant” constellation at/near the top of the activation scale

Salience and Reference Resolution

(Hajičová et al., 1990) propose the following rules for reference resolution:

1. null-subject → preceding subject
2. relative pronoun → head of preceding noun phrase
3. weak pronoun → preceding topic
4. strong or demonstrative pronoun → preceding focus
5. apply saliency in SSK to decide between competitors after (1-4) using saliency
6. if no item activated enough or difference too small, no resolution

Reported precision of 80-85% just with rules (1-4),
SSK is in addition, but no evaluation results presented.

Determination of TFA?

Saliency and Reference Resolution

- (Hajičová et al., 1995a): refinements for different pronominal forms and recursive CB/NB
 - group 1: CB dependents of main verb (in T)
 - group 2: NB dependents of main verb (in F)
 - group 3 and 5: CB dependents deeper below (in T/F)
 - group 4 and 6: NB dependents deeper below (in T/F)
- (Hajičová et al., 1992): algorithm based on the above, also taking simple syntactically encoded associative links, repeated reference and distance into account; set of weighted factors; looking for optimal solution.

Determination of CB/NB?

No evaluation results presented.

Salience and Reference Production

(Hajičová et al., 1990) propose the following rules for deciding between reference by pronominal or nominal expression:

1. if intended referent X not salient above a MIN threshold, then use an NP
else,
2. if X has no (recent) competitor, then use a weak pronoun
else
3. if neither X nor competitor(s) have MAX saliency, then use an NP
else
4. if X was and remains subject, then use a weak pronoun
else if X becoming subject, then use a strong pronoun to refer to non-subject
else use an NP

No evaluation results presented.

Saliency and Reference Production: Example

d_1 : black chihuahua; d_2 : white chihuahua

$ssk(s_0, d_1) = MIN, ssk(s_0, d_2) = MIN$

- (32) i. The white chihuahua was angry. (Rule 1)
 $ssk(s_1, d_2) = 1$
- ii. It viciously attacked the black chihuahua. (Rule 2)
 $ssk(s_2, d_1) = 0; ssk(s_2, d_2) = 1$
- iii. { It / This one / The chihuahua } barked loudly. (Rule 4)

Salience in GRE

(Krahmer and Theune, 2002) propose a modification of the *Incremental Algorithm* for generating referring expressions (Dale and Reiter 1995)

- the IA determines what properties to include in a referring expression to uniquely determine an object X w.r.t. to a context set (set of objects in discourse domain from which intended referent X needs to be distinguished)
- use saliency to determine the context set: set of salient objects
- saliency weighting according to (Hajičová, 1993) and (Grosz et al., 1995) compared
- combined approach: continuous values, subject-preservation preference

Saliency in GRE: Example

d_1 : black chihuahua; d_2 : white chihuahua

- (33) i. The white chihuahua was angry.

$$ssk(s_1, d_2) = 1$$

$$C_f(s_1) = \{d_2\}$$

- ii. It viciously attacked the black chihuahua.

$$ssk(s_2, d_1) = 0; ssk(s_2, d_2) = 1$$

$$C_f(s_2) = \{d_2, d_1\}, C_b(s_2) = d_2$$

- iii. { It / This one / The chihuahua / The black dog } barked loudly.

Other Models of Common Ground

Other Models of Common Ground

- Centering theory: Grosz et al. (Grosz et al., 1995)
- Assumed familiarity taxonomy: Prince (Prince, 1981)
- Cognitive states: Chafe/Lambrecht (Lambrecht, 1994)
- Givenness hierarchy: Gundel et al. (Gundel et al., 1993)

- taxonomy (classification) vs. hierarchy (partial ordering) of statuses
- statuses vs. the dynamics (i.e., how referents acquire and change status)
- incrementality (utterance-by-utterance vs. continuous)
- relation between statuses and linguistic form
- relation between statuses and IS

Centering Theory

(Grosz et al., 1995)

- only a small number of referents can be attended to at the same time
- local vs. global model of attention centering
- Centering Theory concerns the local modelling of attention
 - each utterance has one backward looking center and a partially ordered set of forward looking centers
 - variety of ordering criteria have been proposed: syntactic function, theta-role, surface order, familiarity status (Walker et al., 1998)
 - types of center-transitions depending on whether backward looking center is maintained or changed:
continuation, retaining, smooth shift, rough shift

Centering Theory

- Basic Predictions:
 - preferences for sequences of backward-looking center transitions:
continue > retain > shift
 - preferences for choice of ling. form: if anything is pronominalized, the backward-looking center is
- Comparison with SSK (Kruijff-Korbayová and Hajičová, 1997): backward-looking center \approx least communicatively dynamic Topic-item; but: not only pronominalized, continuous salience values, fading away; differences in ordering
- There exist various implementations, e.g., for NLG and for anaphora resolution.
- Corpus annotation: GNOME project.

Prince's taxonomy of assumed familiarity

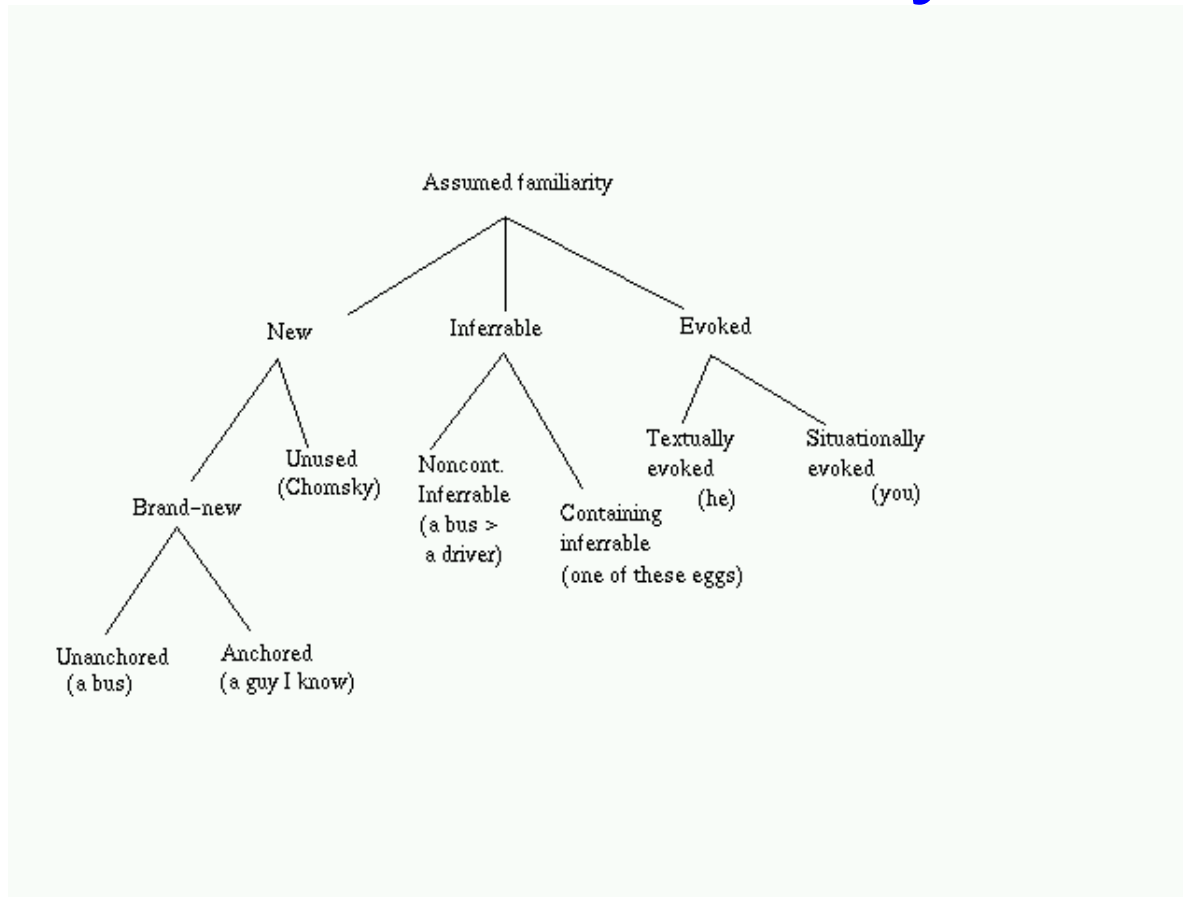
(Prince, 1981; Prince, 1992)

- *brand new*: create a new discourse referent for a previously unknown entity
- *unused*: create a new discourse referent for a known entity
- *inferable*: create a new discourse referent for an inferable entity
- *evoked* (textually or situationally): access an available discourse referent

	Discourse-new	Discourse-old
Hearer-new	brand new	inferable
Hearer-old	unused	evoked



Prince's Taxonomy



Prince's Taxonomy: Examples

Brand new *I bought a dress.* (Hearer creates a new entity in DM)

Unused *Chomsky is famous.* (Hearer moves entity to DM.)

Inferrable *I went to the postoffice and the clerck sold me a stamp.*

(Hearer infers entity from an entity in DM.)

Containing Inferrable *One of these eggs is rotten.*

(Hearer infers entity from an entity denoted by containing NP.)

Textually evoked *Sue went to her grandma and the sweet lady baked a cake.*

(Entity already introduced into in DM.)

Situationally evoked *The board is dirty.*

(Entity already in DM because it is in the situational context.)



Prince's Taxonomy

Familiarity scale: $E/E^S > U > I > I^C > BN^A > BN$

This scale can give rise to implicatures based on Grice's Maxim of Quantity, i.e., the use of a weaker expression implicates that the stronger expression would not have been appropriate/possible.

- (34)
- a. I
 - b. Ellen
 - c. One of the people that work at Penn
 - d. A person that works at Penn
 - e. A person
bought a Toyota.

The scale also seems to hold for NPs representing anchors within BN^A entities.



Prince's Taxonomy

In informal conversational discourse, a tendency has been observed (in English) to reserve subjects for entities with higher familiarity; in other words, constructions are used which enable to keep entities with low familiarity out of subject position

(35) “run-on”

- a. I had a little boy, black, about ten years old, he . . .
- b. There's some male beauty shops, they . . .

(36) “deletion of subject relative markers”

- a. We got a lot of fancy Cadillac cars \emptyset don't tip.
- b. I had a great-great-great-grandfather or something \emptyset fought that Revolution.
- c. There was a peice of four-inch bone \emptyset never mended.

Prince's Taxonomy

(Prince, 1978) corpus-based findings:

- The presupposed part of a wh-cleft represents information that the speaker can assume the hearer is thinking about (hearer-old)
- In one variety of it-cleft, the presupposed part represents information which the speaker assumes the hearer knows or can deduce but is not presumable thinking about (hearer-new)
- In another variety of it-cleft, the presupposed part represents information which the speaker takes to be a know fact, though definitiely not known to the hearer (hearer-new)

Chafe's Taxonomy

Chafe (1974, 1976), cf. (Lambrecht, 1994)

- “knowing something and thinking something are different mental states”
- cognitive states of concepts in hearer's consciousness at utterance time:
 - active
 - semi-active (accessible)
 - * textual: deactivation
 - * inferential: from cognitive schema
 - * situational: presence in external world
 - inactive
- interested in correlations between cognitive states and verbalization:
 - active: lack of pitch accent, pronominal coding
 - inactive: accentuation, full lexical coding

Accessibility vs. Identifiability

(Lambrecht, 1994)

Accessibility/Activation “awareness”, “easy access” (cf. Chafe)

Identifiability hearer’s ability to pick out a particular referent (“file”) from among all those which can be designated with a part. ling. expression, and identify it as the one the speaker intends

- no one-to-one correspondence between (non)identifiability and (in)definiteness
- other dimensions: specific vs. non-specific indefinite NPs; generic NPs

Identifiability and activation/accessibility are independent but correlated:

- unidentifiable are outside the activation parameter (Prince’s brand new)
- identifiable can be inactive, accessible or active

Gundel's Givenness Hierarchy

(Gundel et al., 1993) combine activation and identifiability in one hierarchy

	Cognitive Status	Ling. Form
1	in focus	<i>it</i>
2	activated	<i>that, this this N</i>
3	familiar	<i>that N</i>
4	uniquely identifiable	<i>the N</i>
5	referential	<i>an N, this N</i>
6	identifiable type	<i>an N</i>

Predictions: The cognitive status of an item is a necessary and sufficient condition for the use of the corresponding ling. form.

Gundel's Givenness Hierarchy: Examples

Identifiable type H knows the meaning of the type being used; she can access a representation of the type described by N.

I couldn't sleep last night. A rabbit kept me awake.

Referential S refers to specific entity. H does not know which.

I couldn't sleep last night. This rabbit in the garden kept me awake.

Uniquely identifiable H can identify the S's intended referent.

I couldn't sleep last night. The rabbit kept me awake.



Gundel's Givenness Hierarchy: Examples

Familiar H uniquely identifies the intended referent because she has a representation of it in memory.

I couldn't sleep last night. That rabbit in the garden kept me awake.

Activated H has a representation of the intended referent in short-term memory.

I couldn't sleep last night. That kept me awake.

"That" = e.g., the rabbit's gnawing on carrots occurring at utterance time.

In focus (center of attention) H has a representation of the intended referent in the center of attention in short-term memory.

I couldn't sleep last night. That rabbit in the garden kept me awake. It gnaws very loudly.



Predictions of Gundel's Givenness Hierarchy

- A particular ling. form is inappropriate if the required cognitive status is not met.
- A form corresponding to a weaker cognitive status than the referent actually has can be used (e.g., *the N* for an entity in center of attention).

Tested and mostly verified on naturally occurring discourse for Chinese, English, Japanese, Russian and Spanish (the hierarchy has been tailored to the specifics of each language, e.g., Russian has no articles).

Analysis of the mispredicted cases.

Summary and Conclusions

- TFA: what is being talked about (Topic) and what is said about it (Focus)
TFA is a matter of how speaker constructs the utterance / presents its contents
- CB/NB is correlated with but not identical to salience in SSK:
Cooperative speaker chooses CB from (highly) salient
- Reference resolution/production
- Comparison to other models of common ground
- Correlations between familiarity/activation/givenness status and linguistic form
- Various aspects of linguistic form are relevant, not just NP properties

References

- František Daneš. 1970. Zur linguistischen [a]nalyse der [t]extstructur. *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.
- Jan Firbas. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Studies in English Language. Cambridge University Press, Cambridge.
- Barbara Grosz, Aravind Joshi, and Scott Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2):203–226.
- Jeanette Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, (69):274–307.
- Eva Hajičová, Petr Kuboň, and Vladislav Kuboň. 1990. Hierarchy of salience and discourse analysis and production. pages 144–148.



- Eva Hajičová, Vladislav Kuboň, and Petr Kuboň. 1992. Stock of shared knowledge - a tool for solving pronominal anaphora. pages 127–133.
- Eva Hajičová, Petr Sgall, and Hana Skoumalová. 1993. Identifying topic and focus by an automatic procedure. In *Proceedings of the conference of the European Association of Computational Linguistics, April 1993, Utrecht, the Netherlands*. ACL.
- Eva Hajičová, Tomáš Hoskovec, and Petr Sgall. 1995a. Discourse modelling based on hierarchy of salience. *Prague Bulletin of Mathematical Linguistics*, 64:5–24. Charles University, Prague, Czech Republic.
- Eva Hajičová, Petr Sgall, and Hana Skoumalová. 1995b. An automatic procedure for topic-focus identification (a revised version of (Hajičová et al., 1993)). *Computational linguistics*, 21(1):81–94.
- Eva Hajičová. 1993. *Issues of sentence structure and discourse patterns*, volume 2 of *Theoretical and computational linguistics*. Charles University, Prague, Czech Republic.
- Emiel Krahmer and Mariet Theune. 2002. Efficient context-sensitive generation of referring expressions. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing. Reference and Presupposition in Language and Interpretation*, volume 143 of *Lecture Notes*, pages 223–264. CSLI.
- Ivana Kruijff-Korbayová and Eva Hajičová. 1997. Topics and centers: a comparison of the salience-based approach and the centering theory. *Prague Bulletin of Mathematical Linguistics*, (67):25–50. url.
- Knud Lambrecht. 1994. *Information Structure and Sentence Form*. Cambridge University Press, Cambridge, U.K.
- Vilém Mathesius. 1936. On some problems of the systematic analysis of grammar. In *reprinted in: A Prague School Reader in Linguistics*, J. Vachek (ed.). Bloomington, Ltd.
- Vilém Mathesius. 1975. On information bearing structure of the sentence. In S. Kuno, editor, *Harvard studies in syntax and semantics*. Harvard University Press.



- Barbara H. Partee, Eva Hajičová, and Petr Sgall. 1998. *Topic-Focus Articulation, Tripartite Structures, and Semantic Content*. Kluwer, Amsterdam.
- Barbara Partee. 1995. Allegation and local accommodation. In Barbara H. Partee and Petr Sgall, editors, *Discourse and Meaning*, pages 65–86. John Benjamins Publishing Company.
- Ellen Prince. 1978. A comparison of it-clefts and wh-clefts in discourse. *Language*, (54):883–906.
- Ellen Prince. 1981. Toward a taxonomy of given-new information. In Peter Cole, editor, *Radical Pragmatics*, pages 223–256. Academic Press.
- Ellen Prince. 1992. The ZPG letter: subjects, definiteness, and information status. In Sandra Thompson and William Mann, editors, *Discourse Description: diverse analyses of a fund raising text*, pages 295–325. John Benjamins.
- Petr Sgall, L. Nebeský, A. Goralčíková, and Eva Hajičová. 1969. *A Functional Approach to Syntax in Generative Description of Language*. Elsevier, New York.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Reidel, Dordrecht, The Netherlands.
- Petr Sgall. 1967. Functional sentence perspective in a generative description. *Prague studies in mathematical linguistics*, 2:203–225. Prague, Czechoslovakia.
- Marilyn Walker, Aravind K. Joshi, and Ellen F. Prince, editors. 1998. *Centering Theory in Discourse*. Oxford University Press.