CENTERING:
application in different languages

Presenter: Kepa Joseba Rodríguez
kepa@coli.uni-sb.de

10. Januar 2003

Lecture plan

1. Data structures in the Centering Theory.

2. Different approaches: an overview.

3. Exercise: apply an algorithm to a fragment of NEGRA.

4. Discussion.

Remember the last presentation....

Data structures (Grosz, Joshi, Weinstein (1995)):

- **C$_b$:** Backward-looking center.
- **C$_f$:** List of forward-looking centers.

But... **how to order the elements of the C_f?**

Different approaches using...

- Grammatical function.
- Linear ordering.
- Information Structure
- $\theta-$roles.
- Information-Status (Givenness).

Grammatical function

- Grosz, Joshi, Weinstein (1995)
  SUBJECT $\prec$ OBJECT $\prec$ OTHERS

- Walker, Prince, Joshi (1998)
  SUBJECT $\prec$ OBJECT2 $\prec$ OBJECT $\prec$ OTHERS

- ....

It seems to work for English, but not for free order languages.

Linear order and centering

An approach for German (Rambow 1993)

Clauses are divided into two parts:

1. **The Vorfeld (VF)** Position in front of the finite verb.
   It must contain one constituent.

2. **The Mittelfeld (MF)** Position between finite and non-finite verbs.
   It contains the remaining arguments and adjuncts in a sentence.
   The MF allows scrambling of arguments and adjuncts based on information status.

Assumption: The order in which the entities appear in the MF reflects their salience in the C_f list.

Evidence given by Rambow.

(1) - Glauben Sie, dass **[eine solche Maßnahme]**$_i$ **[der russischen Wirtschaft]**$_k$ helfen kann?

- Nein, **sie**$_i$ ist zu primitiv.

(2) - Glauben Sie, dass **[der russischen Wirtschaft]**$_k$ **[eine solche Maßnahme]**$_i$ helfen kann?

- Nein, **sie**$_k$ ist zu primitiv.

Information Structure

in Turkish (Hoffman 1998)

Canonical order in Turkish SOV.

IS-partitioning is realized by constituent order as follow (Hoffman 1996):

- **Topic:** Sentence initial position.
- **Focus:** Inmediate preverbal position.
- **Background:** Remainder of the sentence.

Assumption: The topic corresponds to the C_b in most cases.

	SOV	OSV
C_b = Subject	47%	13%
C_b = Object	20%	53%
C_b = Sub. or Obj?	20%	20%
C_b = Sub. or Other Obj?	0 %	7%
No C_b	13%	7 %

The approach of Hoffman doesn't give more intuition about the ranking in the C_f list.

$\theta-$roles

An approach for psychological verbs in Turkish (Turan 1998)

Assumption: Subjects are C_ps in Turkish. But in utterances with psychological verbs *EXPERIENCER* objects seem to rank higher than *THEME* subjects in the C_f list.

(3) 1.

[Türkü]$_i$	**[Karaka**	**Ali'yi]**$_j$	bile	yumuşatmiști.
song	Karaka	Ali.ACC	even	soft.CAUS.PAST.PAST

'The song had softened even Karaka Ali.'

2.

[∅]$_j$	yanindakilere	uzun	uzun	bakti.
	near.POSS.3SG.LOC.PLU.DAT	long	long	look.PAST

'(He) locked at the people next to him for a long time.'

Possible Ranking: AGENT $\prec$ EXPERIENCER $\prec$ THEME

Does it work in German??

(4) 1. **Der Tanz**$_i$ hat **Hans**$_j$ gefallen.

2. **Er**$_j$ fand auch das Theaterspiel ziemlich gut.

But .....

(5) 1. **Der Mario**$_i$ hat **Hans**$_j$ gefallen.

2. **Er**$_{???}$ arbeitet in einer Reisenagentur.

Information Status

The Functional Centering

for German (Strube & Hahn 1999)

Assumption: Hearer-old discourse entities are ranked higher than hearer-new discourse entities.

Familiarity status of discourse entities

Prince (1981), adapted by Strube & Hahn (1999)

- **Evoked (E):** The uttered entity is already in the discourse model. Evoked discourse entities (DE) are resolved anaphoric expressions (pronouns, mentioned proper names, etc).

  (6) He lives in **[a small flat]**$_{BN}$.
  [This flat]$_{E}$ is near the....

- **Unused (U):** The hearer may be assumed to have a corresponding entity in his/her model and places it in the discourse model. Unused DE are proper names, titles, etc.

  (7) In his talk **[George Bush]**$_{U}$ says ....

  (8) We have to speak with **[the governor of the province]**$_{U}$.

- **Inferrable (I):** A discourse entity is inferrable by the hearer using reasoning.

(9) By **[his]** teen-age years, the distorted mentality of anti-Semitism was in full warp.
[The family]$_I$ was expelled to Hungary in 1939....

- **Containing inferrable (I^C):** The inferable NP contains what is inferenced off.

(10) Hey, **[one of this eggs]**$_{I^C}$, is broken.

- **Anchored brand new (BN^A):** The NP representing the new discourse entity is linked to another entity.

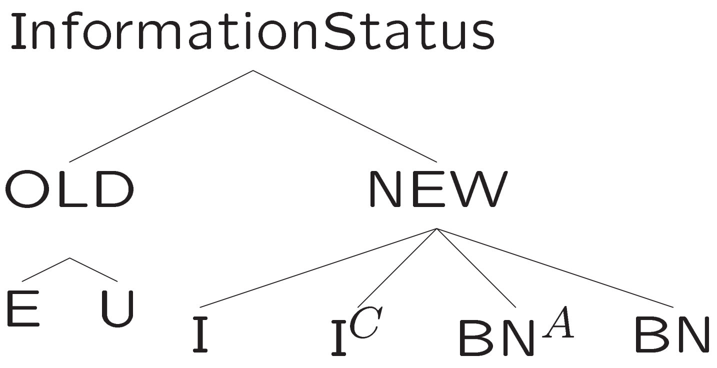
(11) **[She]** has two dogs, a cat and a crocodile.
[[Her]$_E$ flat]$_{BN^A}$ is like a zoo.

- **Brand new (BN):** The speaker creates a new entity. In case of BN proper names they are usually accompanied by a relative clause or apposition.

(12) **[Mary, the sister of John]$_{BN}$**, is coming later.

Basic C_f Ranking

Two different sets of expressions in U_i: **hearer-old** (OLD) and **hearer-new** (NEW).



Basic functional ranking constraints on the C_f list:

1. If $x \in$ OLD and $y \in$ NEW, then $x \prec y$.

2. If $x, y \in$ OLD or $x, y \in$ NEW, then $x \prec y$, if $pos_x < pos_y$.

3. If (1) or (2) do not apply, then x and y are unordered with respect to the C_f-ranking.

(13) a. The sentry was not dead.
b. He was, in fact, showing signs of reviving...
c. He was partially uniformed in a calvary tunic.
d. Mike stripped this from him and donned it.
e. He tied and gagged the man, ...

a	the sentry was not dead. C_b: - C_f: [SENTRY$_E$: sentry]
b	He was, in fact, showing signs of reviving... C_b: SENTRY$_E$: he C_f: [SENTRY$_E$: he, SIGNS$_{BN}$: signs]
c	He was partially uniformed in a calvary tunic. C_b: SENTRY$_E$: he C_f: [SENTRY$_E$: he, TUNIC$_{BN}$: tunic]
d	Mike stripped this from him and donned it. C_b: SENTRY$_E$: him C_f: [MIKE$_E$: Mike, TUNIC$_E$: this, SENTRY$_E$: him]
e	He tied and gagged the man, ... C_b: MIKE$_E$: he C_f: [MIKE$_E$: he, SENTRY$_E$: the man]

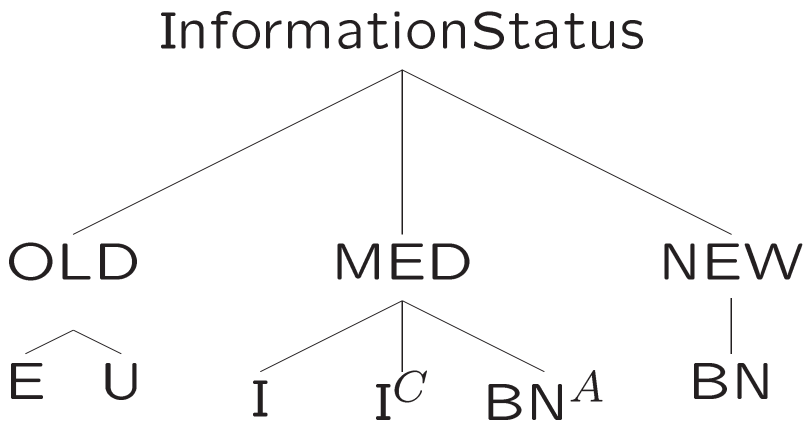
Basic vs. extended C_f-ranking:

Basic C_f-ranking is sufficient for text with a high proportion of pronouns and nominal anaphora (news-paper articles, literary texts, etc).

Expository texts (scientific and technical texts, reports, summaries, etc) contains few pronouns and a large number of inferrables.

Extended C_f-ranking.

A new set of expressions will be distinguished: **mediated discourse entities** MED.



Extended functional ranking constraints on the C_f list:

1. If $x \in$ OLD and $y \in$ MED, then $x \prec y$.
   If $x \in$ OLD and $y \in$ NEW, then $x \prec y$.
   If $x \in$ MED and $y \in$ NEW, then $x \prec y$.

2. If $x, y \in$ OLD or $x, y \in$ MED or $x, y \in$ NEW, then $x \prec y$, if $pos_x < pos_y$.

3. If (1) or (2) do not apply, then x and y are unordered with respect to the C_f-ranking.

Example from the NEGRA Corpus

(14) (...)

a. [[Ihre]$_E$ minimalistischen Klänge]$_I$ sind voller [[Spielfreude]$_{BN}$ und [Experimentierlust]$_{BN}$]$_{BN}$.

b. [Die folkloristischen Elemente]$_I$ beschränken sich nicht allein auf [Brasilien]$_E$,

c. sondern lassen auch [asiatische und afrikanische Akzente]$_{BN}$ einfließen.

d. [Die erklingenden Instrumente]$_I$ sind zum Teil selbstgebaut

e. und [$\emptyset$]$_E$ geben [der Musik]$_E$ [eine ironische Note]$_{BN}$.

f. Dabei besitzt [[Uaktis]$_E$ Werk]$_{BN^A}$ durchaus [klassische Motive]$_{BN}$,

g. [$\emptyset$]$_E$ wird aber gleichzeitig vom starken Hang [zur Kindlichkeit]$_{BN}$ geprägt:

h. [die Melodien]$_I$ sind sehr, sehr eingängig.

(...)

EXERCISE: apply algorithm

Discussion:

Possible requirements in order to apply the Functional Centering algorithm to the NEGRA-Corpus.

- Annotate the information status of the NP's.
- Alternative think how could be used the actual information e.g. Bridging as MED, Anaphoric Relation as OLD, NP's with possesive pronouns as anchored, etc.
- Segmentation of the text according to the Centering Theory.

Literature

Grosz, Joshi, Weinstein (1995). Centering: a framework for modeling the local coherence of discourse. *Computational linguistics, 21*.

Hoffman (1996). Translating into free word order languages. in *Procedings of the 16th International Conference on Computational Linguistics*. Volumen 1. Copenhagen. Denmark.

Hoffman (1998). Word Order, Information Structure, and Centering in Turkish. in Walker, Joshi, Prince (eds) *Centering Theory in Discourse*. Oxford University Pres, 1998.

Prince (1981). Toward a taxonomy of given-new information. in Peter Cole (ed), *Radical Pragmatics. Academic Press, 1981*.

Rambow (1993). Pragmatic Aspects of Scrambling and Topicalization in German. *Workshop on Centering Theory in Naturally-Occurring Discourse*, Institute for Research in Cognitive Science, University of Pennsylvania.

Strube, Hahn (1999). Functional Centering: Grounding referential coherence in information structure. *Computational Linguistics*.

Turan (1998). Ranking Forward-Looking Centers in Turkish: Universal and Language-Specific Properties. in Walker, Joshi, Prince (eds) *Centering Theory in Discourse*. Oxford University Pres, 1998.