Lecture 8

Wrapping Up

Outline

- Alligning IS terminologies
- Discussion of differences among approaches
- Empirical verification
- Summary and conluclusions

Wrapping Up

Aligning IS Terminologies

Mathesius, Firbas, Daneš	Theme	vs.	Rheme
Sgall et al.	Topic (CB) topic proper vs. contrastive topic	vs.	Focus (NB) focus proper
Halliday	[Theme] Given vs. New		[Rheme] Given vs. New
Chomsky, Jackendoff, Krifka Rooth	Topic	vs. Comment	
	Presupposition	vs.	Focus
Vallduví	Ground	vs.	Focus
	Tail vs. Link (Kontrast)		Kontrast
Steedman	Theme Background vs. Focus	vs.	Rheme Background vs. Focus

Aligning IS Terminologies

But, be aware of differences concerning:

- Definitions of the IS categories and operational criteria
- Level(s) at which IS distinctions are made, e.g., surface, deep, semantics. . .
- Flexible vs. fixed syntactic constituents, and how do IS components correspond to them
- Multiple themes
- Multiple "foci", discontinuity of IS components
- Degree of recursivity of IS notions (if any)
- IS-boundary at main clause level vs. "deeper"
- IS in complex sentences

Multiple Foci

(181) I know what Marcel gave to Harry.
But what did CHARLES give to FRED?

CHARLES	gave	a BOOK	to FRED
L+H*	LH%	H* L	L+H* LH%
Focus	*Background*	*Focus*	*Focus*
Theme		*Rheme*	*Theme*

(182) I know what Marcel gave to Harry. But who gave what to FRED?

CHARLES	gave	a BOOK	to FRED
H* L		H* L	L+H* LH%
Focus	*Background*	*Focus*	*Focus*
Rheme	*Theme*	*Rheme*	*Theme*

Surface vs. Semantics

(183) U: What is the price of a flight to London?
S:

A flight with	RYANAIR	costs	FIFTY FIVE	Euro.
	L+H*LH%		H*	LL%
	Focus		*Focus*	*Background*
	Theme		*Rheme*	

A flight with	LUFTHANSA	costs	HUNDRED	and fifty five Euro.
	L+H*LH%		H*	LL%
	Focus		*Focus*	*Background*
	Theme		*Rheme*	

(184) I know that THIS car is a PORSCHE.
But what is the make of your OTHER car?

My	OTHER	car	is	ALSO	a Porsche
	L+H*	LH%		H*	LL%
Background	*Focus*		*Background*	*Focus*	*Background*
Theme			*Rheme*		

"Flexible Constituents"

(185) I know which result Marcel predicted.
But which result did Marcel prove?

Marcel PROVED COMPLETENESS.
L+H* LH% H* LL%
Background *Focus* *Focus*
Theme *Rheme*

"Flexible Constituents"

(Steedman, 2000b; Steedman, 2000a): flexible, but not arbitrary!

(186) ⋆ (Three MATHEMATICIANS) (in ten derive a LEMMA).
L+H*LH% H*LL%

(187) ⋆ (Seymour prefers the NUTS) (and bolts APPROACH).
L+H*LH% H*LL%

(188) ⋆ (They only asked whether I knew the woman who CHAIRED) (the *zoning* board).
L+H*LH% H*LL%

Deeper Embedded IS Boundary?

(Sgall et al., 1986), (Partee et al., 1998)(p.135):

(189) Which teacher did you meet yesterday?

(190) (Yesterday)$_{cb}$ (I)$_{cb}$ (met)$_{cb}$ (the teacher)$_{cb}$ (of CHEMISTRY)$_{nb}$.
Topic *Focus*

Steedman:

(191) (I read a book about)$_{Theme}$ (COMPLETENESS)$_{Rheme}$ (Steedman, 2000a)[p.678]

(192) Yesterday I met the teacher of CHEMISTRY .
Theme H* LL%
Focus
Rheme

Deeper Embedded IS Boundary?

(Prevost, 1995; Prevost, 1996)

(193) I know the AMERICAN amplifier produces MUDDY treble.
(But WHAT) (produces CLEAN treble)?
L+H*LH% H* LL%

The BRITISH amplifier produces CLEAN treble.
H* LL% L+H* LH%
Background *Focus* *Background* *Background* *Focus* *Background*
Rheme *Theme*

(194) I know the AMERICAN amplifier produces MUDDY treble.
(But WHICH amplifier) (produces CLEAN treble)?
L+H* LH% H* LL%

Deeper Embedded IS Boundary?

(195) *Which amplifier did you buy? I bought a* BRITISH *amplifier.*
H* LL%

(196) $\underbrace{\text{(I)}_{cb}\ \text{(bought)}_{cb}\ \text{(a}}_{Topic}\ \underbrace{\text{(BRITISH)}_{nb}}_{Focus}\ \underbrace{\text{(amplifier)}_{cb}.}_{Topic}$

(197) $\underbrace{\text{(Amplifier)}_{cb}\ \text{(I)}_{cb}\ \text{(bought)}_{cb}}_{Topic}\ \underbrace{\text{(a BRITISH one)}_{nb}.}_{Focus}$

(198) $\underbrace{\text{(Zesilovač)}_{cb}\ (\emptyset)_{cb}\ \text{(jsem koupila)}_{cb}}_{Topic}\ \underbrace{\text{(BRITSKÝ)}_{nb}}_{Focus}$

cf.

(199) $\underbrace{\text{I bought an amplifier}}_{Theme}\ \underbrace{\text{from BRITAIN}}_{Rheme}$

Recursivity

What is the domain of IS partitioning?

- Sentence?
- Utterance?
- Clause?
- "Basic" proposition? (e.g., cf. Daneš's Theme-Rheme nexus)
- Is IS fully recursive? (e.g., (Partee, 1995))

cf., e.g., (Partee et al., 1998), (Kruijff, 2001), (Komagata 2003)

Recursivity and IS in Complex Sentences

(200) (Where do you buy wine if it's Sunday?)
If it's Sunday, we buy wine over the STATE LINE.
L+H*LH% H*LL%

(201) Although Clyde married BERTHA, he is not HAPPY.

(202) Clyde isn't HAPPY, although he married BERTA.

what's Focus/Kontrast and what's Rheme? Any explicit indicators?

- particles, e.g., Japanese (Komagata, 1999), (Komagata 2003)
- intonation, e.g., (Steedman, 2000a)
- word order in German subordinate clause, e.g., (Günthner, 1996)
- placement of clitics in Czech
- etc. ???

(203) Do you see that old boat next to the *Amsterdam*?

(204) $\underbrace{\text{(I)}_{cb}}_{Topic}\ \underbrace{\text{(can see)}_{nb}\ \text{(a (VERY)}_{nb}\ \text{(old)}_{cb}\ \text{TALLSHIP)}_{nb}}_{Focus}\ \underbrace{\text{(next to the \textit{Amsterdam})}_{cb}}_{Topic}$

(205) $\underbrace{\underbrace{\text{I can see a}}_{Background}}_{Theme}\ \underbrace{\underbrace{\overset{\text{H*}}{\text{VERY}}}_{Focus}\ \underbrace{\text{old}}_{Background}}_{Rheme}\ \underbrace{\underbrace{\overset{\text{L+H*}}{\text{TALLSHIP}}}_{Focus}\ \underbrace{\text{next to the \textit{Amsterdam}}}_{Background}}_{Theme}$

Summary

- There are different proposals in the literature
- How to decide?
- Do the differences actually matter?
- i.e., do they result in different predictions?
- When yes, we can test the predictions! Can't we?

- Empirical evaluation:
  - psycholinguistic experiments
  - corpus-based experiments
  - experiments with practical systems (indirect evaluation)
    - direct evaluation: output-quality judgements
    - indirect, task-based evaluation, e.g., success rate in dialogue system

Example Practical System Evaluation

Controling Intonation of Spoken Dialog System Output

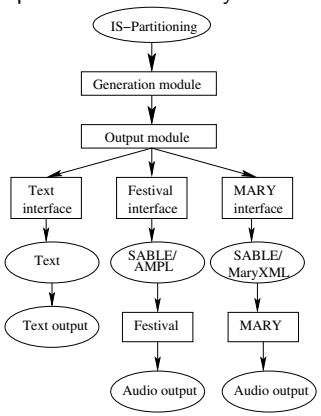
(Kruijff-Korbayová et al., 2003)

- Within the Gothenburg Dialogue System (GoDiS), experimental dialogue system built using the TrindiKit (TRINDI, D'Homme, SIRIDUS projects)

- Determination of Theme/Rheme partitioning according to the QUD (QudTR rule): if QUD corresponds to the result of λ-abstracting over a part of the content, this part becomes the Rheme

- Determination of Focus/Background partitioning within each Theme and Rheme by determining alternatives, i.e., semantically parallell but not identical elements, w.r.t. shared commitments and/or domain knowledge

Producing synthesized output with contextually varied intonation in GoDiS



Experimental Implementation Evaluation

- Using the German TTS system Mary (Schröder and Trouvain, 2001) with various ways of intonation annotation
- Administered through a website (`www.coli.uni-sb.de/cl/projects/siridus`)
- Test of concept

- Experiment 1: default vs. controlled intonation using GToBI or SABLE
  - Dialogue fragments displayed on screen
  - Several turns provide context for target utterance
  - Target utterance synthesized in different versions
  - Subjects judge appropriateness of intonation in the given context
- Experiment 2: only default vs. GToBI controlled intonation
  - Subjects judge intonation without context
  - Subjects judge appropriateness of intonation in the given context

Experimental Evaluation Results

Although the results are not significant,
observed tendencies correspond to expectations:

- overall average judgments worse for default than for controlled intonation

- average judgments per IS pattern also worse for default than for controlled intonation
  (not much difference across patterns, though one would expect it!)

- judgments of default intonation in isolation closer to those where the context is matching with this, then to those where the context does not match

- roughly same results whether looking at absolute values of judgments or taking differences between values in isolation and in context, per subject

Experimental Evaluation Experience

- Proper (standard) evaluation methodology is lacking

- Indirect evaluation through task success / completion time does not seem suitable, because of accumulation of effects through dialogue (moreover, it would have to be Wizard of Oz, because of coverage and robustness issues)

- Direct evaluation is hard to design as a proper experiment:
  - Do subjects really take context into account?
  - Are they judging contextual appropriateness of the intonation pattern and not the quality of the synthesized output as such?
    - Absolute judgments allow comparison of judgments across dialogues
    - Comparative judgments could neutralize relatively low synthesis quality

Annotation of IS in Corpora

Why IS Annotation?

- Diverse and under-formalized terminology
- Lack of intuitions about the interpretation of complex IS, e.g. interaction with clause complexity, quantifiers, discourse relations, thematic structure in texts, . . .
- Can we compare and test the theories against corpus data?

IS Annotation: Problems

- Existing annotations are scarce, disparate, and theory-specific
- Existing theories are empirically inadequate: both too vague and too detailed, i.e., theoretically defined concepts are too brittle to apply to real-life data
- Lack of annotation methodology

We hope IS annotation can contribute to develop better intuitions, identifying critical issues obtain explanatorily more adequate perspectives on the realization & interpretation of IS

Desiderata for Annotation Methodology

- Theory-neutral notions
- Robustness
- Cross-linguistic applicability
- Genre/Register independence
- Multiple layers

IS annotation

- Theory-specific annotation
  - various small corpora, typically not available for reuse
  - Prague Dependency Treebank of Czech (Hajičová et al., 2003)
- Theory-neutral annotation: more basic, IS-relevant features
  - Anaphoric relations: various
  - Familiarity status at Edinburgh Uni (Nissim et al., 2004)
  - Syntactic and semantic features of referring expressions: GNOME project at Edinburgh Uni (Poesio, 2004)
  - MULI project at Saarland Uni (Baumann et al., 2004a; Baumann et al., 2004b)
  - Penn Discourse Treebank Project at UPenn (Miltsakaki et al. 2004)

TFA Annotation in PDT

Annotation of the TFA concepts in the Prague Dependency Treebank (Buráňová et al., 2000; Hajičová et al., 2003)

- data from the Czech National Corpus
- TFA is annotated in the dependency structures at the tectogrammatical level
- ordering of nodes represents communicative dynamism (deep, underlying order)
- each node is annnotated with the TFA attribute:
  T contextually bound
  F contextually non-bound
  C contrastivelly bound (Partee et al., 1998)
- guidelines in Czech (cca 50 pages)

- PDT version 2.0 with several thousand sentences annotated with TFA as well as coreference will be released in the fall of 2004

Anaphora Annotation

- DRAMA (Passoneau, 1996) (available through the MATE project website)
- MUC-6 and MUC-7 (MUC Coreference Specification)
- the DRI guidelines (Carletta et al., 1997), see (`www.dfki.de/dri`)
- the MATE project (Poesio et al., 1999),
  see (`www.cogsci.ed.ac.uk/~poesio/MATE/coreference.html` or `mate.mip.ou.dk`
- bridging references (Poesio and Vieira, 1998)
- coreference and bridging in the Heidelberg Text Corpus, MMAX tool (Müller and Strube, 2001a)

Familiarity Status Annotation

(Nissim et al., 2004)

- data: the Switchboard section of the Penn Treebank (dialog)
- annotation of referring expressions with familiarity status (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

- tool: the NITE workbench

NP Annotation in the GNOME Project

(Poesio, 2004)

- data from various sources
- Annotation of nominal referring expressions with syntactic and semantic features relevant for NP generation
  - the semantic attributes include, e.g., animacy, ontological status, countability, quantification and generic vs. specific reference
- detailed guidelines available from the GNOME project website

Penn Discourse Treebank Project

(Miltsakaki et al. 2004)

- data: the Penn Treebank
- annotation of explicit and implicit discourse connectives and their arguments (Webber and Joshi, 1998)
- tool: WordFreak

Multilayer IS Annotation in the MULI project

(Baumann et al., 2004a; Baumann et al., 2004b)

- data from the WSJ section of the PTB (English) and from the Negra/TIGER Treebank (German)
- annotation of markables at multiple layers, with layer-specific features (syntax, discourse, prosody)
- English: 7k words/320 sentences; German 3.5k words/250 sentences
- Multi-layer stand-off annotation with shared base data; layer-specific tools, translation into shared XML format

MULI: Syntax Layer Annotation

- IS-relevant aspects of realization:
  - Positioning
  - Ordering (words, phrases, clauses)
  - Marked syntactic constructions
  - Morphological marking (in some languages)
  - Definiteness marking (in some languages)
- Treebank data already available
- annotation of additional features: presence of marked syntactic constructions in clauses
- tool: XMLspy

MULI: Layer Annotation

- recording of read text, just one speaker, only the German corpus
- annotation of intonation following GToBI (**?**)
  - Position and type of boundary tones
  - Position and type of accents
  - Position and size of phrase boundaries cf. (**?**)
- tool: EMU

MULI: Discourse Layer Annotation

- Annotation of expressions with discourse referent properties
  - Semantic type and sort
  - Denotation characteristics:
    - Extensional vs. intensional reference
    - Unique, existential or variable determitation
    - Countability and quantification
  - Familiarity status (Prince, 1981)
  - Anaphoric relations: Coreference and various types of bridging
- guidelines (cca 30 pages)
- tool: MMAX (Müller and Strube, 2001b)

How Much Work is it?

- It's hard: . . .
  - Also designing annotation methodology and schemes
- And quite time-consuming
  - Intonation: 30 min/sent. (incl. discussion)
  - Discourse: 10 min/sent. (create markables and links, assign properties)
  - Syntax: 5 min/sent. –only addt'l annotation (clause segmentation, classification); Treebank available

Conlusions

Summary and Conclusions

- IS is an important aspect of meaning at the interface between utterance and discourse
- a theory relating IS and DS is essential for accurate NL processing
- formal accounts emerging, some embodied into practical systems
- many questions concerning IS partitioning and its realization in different languages still open
- Further research topics:
  - further systematization of terminologies
  - formalization and computational modeling&testing
  - empirical and corpus-based studies
  - cross-linguistic investigations and multilingual applications