Modeling Information Structure
for Computational Discourse and Dialog Processing

Ivana Kruijff-Korbayová

korbay@coli.uni-sb.de

http://www.coli.uni-sb.de/~korbay/esslli04/

ESSLLI 2004 Advanced Course

Nancy, 16-20 August 2004

Lecture 4 Outline

- Steedman's two dimensions of IS
- IS and intonation (in English)
- Alternative-set based semantics of IS
- Intonation assignment in answers to questions
- Assignment in monologue generation
- Intonation assignment for TTS
- Intonation assignment in the GoDIS dialogue system
- Gestures, turn-taking and eye-gaze in multimodal interaction

Reading:

- Course Reader: Section 2.6: Steedman's Two Dimensions of IS
- For further reading suggestions see course website

Steedman's Two Dimensions of IS

Steedman's IS Partitioning

(1) I know who proved soundness. But who proved completeness?

     MARCEL *proved* COMPLETENESS.
      H* L   L+H*   LH%

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

     Marcel PROVED   COMPLETENESS.
        L+H* LH%   H*   LL%

(3) What do you know about Marcel?

     Marcel proved COMPLETENESS.
                      H*   LL%

(ToBI intonation notation (Beckman and Hirschberg, 1999).)

Steedman's IS Partitioning

(Steedman, 2000b; Steedman, 2000a) distinguishes two dimensions of IS within a sentence:

Theme-Rheme partitioning reflects an *aboutness* relation, i.e., the Rheme is semantically predicated over the Theme. This dimension connects the utterance to the rest of the discourse.

Background-Focus partitioning within Theme and Rheme reflects an abstract notion of "kontrast" between alternatives available in the discourse context, against which the Theme and Rheme of the actual utterance are cast. Words whose interpretation contributes to distinguishing Theme/Rheme from alternatives belong to Focus, other words belong to Background.

The Semantics of IS

- Semantics of IS in terms of selecting one member from a *presupposed set of alternatives* (Steedman, 2000a), following (Rooth, 1992; **?**)
  - Theme 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
  - 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
- These are pragmatic presuppositions that the relevant alternative set(s) be available in the context. They can get bound or accommodated.
- The systematic recognition of the alternative sets, and their maintenance as a discourse progresses are open research issues.

The Semantics of IS

(2)   Marcel   PROVED   COMPLETENESS.

L+H* LH%   H*   LL%

$\underbrace{\underbrace{Background}\ \ \underbrace{Focus}}_{Theme}\ \ \underbrace{\underbrace{Focus}}_{Rheme}$

(4) $prove'\ completeness'\ marcel'$

(5) $\exists x.\ \star\, prove'\ x\ marcel'$

(6) $\{\ prove'completeness'marcel',\ prove'decidability'marcel',$
$prove'soundness'marcel'\ \}$

(7) $\{\ \exists x.\ prove'\ x\ marcel',\ \exists x.\ predict'\ x\ marcel'\ \}$

The Semantics of IS

(8)   I know that Marcel likes the man who wrote the musical.
      But who does he ADMIRE?

Marcel	ADMIRES	the woman who	DIRECTED	the musical.
	L+H* LH%		H*	LL%
Background	*Focus*	*Background*	*Focus*	*Background*
	Theme		*Rheme*	
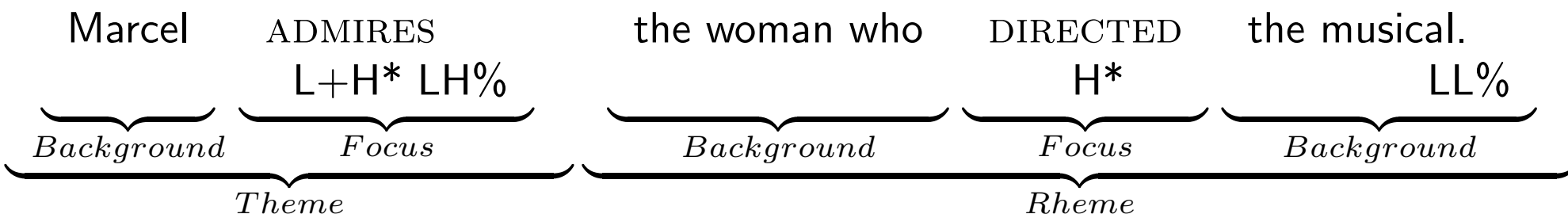
- the Background/Focus partitioning of this Rheme is supported just in case all individuals considered have something to do with the musical, and the property of directing it uniquely identifies one such individual (Prevost and Steedman, 1994; Prevost, 1995; Steedman, 2000b)

The Semantics of IS

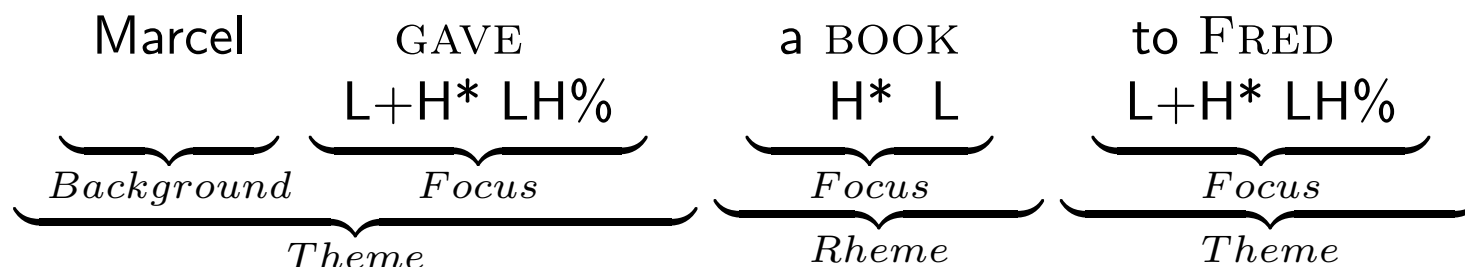| Marcel	ADMIRES	the woman who	DIRECTED	the musical.
	L+H* LH%		H*	LL%

$\underbrace{Background}\ \underbrace{Focus}\ \underbrace{Background}\ \underbrace{Focus}\ \underbrace{Background}$

$\underbrace{Theme}\qquad\qquad\underbrace{Rheme}$
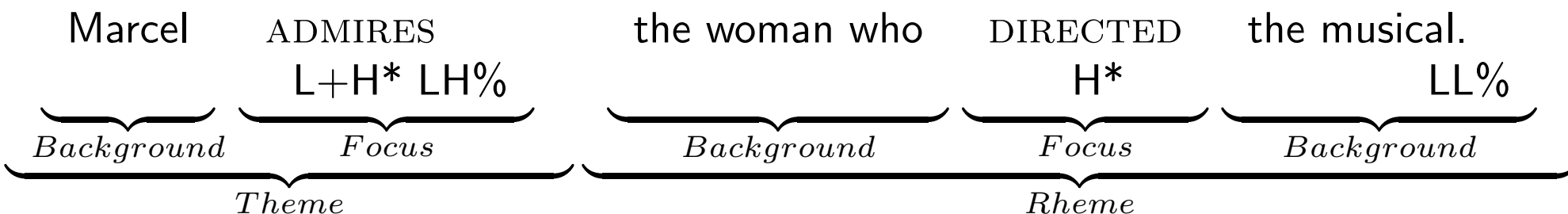
(9) $admire'\ woman_1'\ marcel'$

(10) $\exists x.\ \star\, admire'\ x\ marcel'$

(11) $\{\ admire'woman_1'marcel',\ admire'woman_2'marcel',$
 $admire'man_1'marcel'\ \}$

(12) $\{\ \exists x.\ admires'\ x\ marcel',\ \exists x.\ likes'\ x\ marcel'\ \}$

The Semantics of IS

(13) I know what Marcel SOLD to Harry. But what did he GIVE to FRED?

$$\underbrace{\underbrace{\text{Marcel}}_{Background} \underbrace{\substack{\text{GAVE}\\ \text{L+H* LH\%}}}_{Focus}}_{Theme} \quad \underbrace{\underbrace{\substack{\text{a BOOK}\\ \text{H* L}}}_{Focus}}_{Rheme} \quad \underbrace{\underbrace{\substack{\text{to FRED}\\ \text{L+H* LH\%}}}_{Focus}}_{Theme}$$

(14) *give' fred' book' marcel'*

(15) $\exists x. \star give' \star fred' \ x \ marcel'$

(16) { *give' fred' book' marcel'*, *give' fred' record' marcel'*, *give' fred' biscuit' marcel'*}

(17) { $\exists x. give'\ fred'\ x\ marcel'$, $\exists x. sell'\ fred\ x\ marcel'$, $\exists x. give'\ harry\ x\ marcel'$, $\exists x. sell'\ harry\ x\ marcel'$ }

IS and Intonation

(Steedman, 2000b; Steedman, 2000a) proposes a compositional account of the semantics of tones for English, cast in CCG

- Theme/Rheme partitioning determines overall intonation pattern

  - Theme and Rheme as one intonation phrase each (boundary between)
  - Theme-accents: L+H*, L*+H   (prototypical Theme-tune: L+H*LH%)
  - Rheme-accents: H*, L*, H*+L, H+L* (prototypical Rheme-tune: H*LL%)

- Background/Focus partitioning Determines placement of pitch accents

  - Focus: (words) marked by pitch accent
  - Background: (words) without pitch accent

IS and Intonation

further examples, p. 662 and one, all-theme, ownership by hearer vs. speaker

Intonation Assignment in various Applications

IS-Based Assignment of Intonation in Answers to Questions

- IS used to control intonation synthesized spoken output

- (Prevost and Steedman, 1993) IS of question fully determines the IS of the answer

- Theme/Rheme determination:

  – rheme of the question determines the theme of the answer

- Focus determination:

  – terms focused in question are focused in asnwer

– term instantiating question variable is also focused
– for more complex rhemes, only *new* elements are focused

Example

(18) *I know that widgets contain cogs,*

      but what parts do WODGETS *include?*
        L+H* LH%   H*   LL%

    prop: $s : \lambda x[part(x)\&include(\star wodgets, x)]$
    theme: $s : \lambda x[part(x)\&include(\star wodgets, x)]$ /
                  $(s : include(\star wodgets, x)/np : x)$
    rheme: $s : include(\star wodgets, x)/np : x$

    prop: $s : include(\star wodgets, \star sprockets)$
(19) theme: $s : include(\star wodgets, x)/np : x$
    rheme: $np : \star sprockets$

    WODGETS *include* SPROCKETS.
     L+H*   LH%   H* LL%

More Examples

IS-Based Assignment of Intonation in Answers to Questions

- (Prevost and Steedman, 1994): IS determination in answer not from question alone but also from discourse model (database)

- Theme/Rheme determination:

  - rheme of the question determines the theme of the answer

- Focus determination:

  - terms focused in question's rheme are focused in answer's theme
  - rheme-focus in answer determined from alternative sets in the database

Constructing Rheme-Alternative Sets

(Prevost and Steedman, 1994):

Given database D, object x and a set of properties P that uniquely describe x:

1. construct a set of objects, A, (and their referring properties) which can be considered alternatives to x w.r.t. D

2. restrict A by properties of objects mentioned in theme $\rightarrow A'$

3. mark as contrastive those properties of x in P that exclude some alternatives from A'

Example . . .

IS-Based Assignment of Intonation in Text Generation

(Prevost, 1996)

- Content and text planning: determine a sequence of propositions about an object and the rhetorical relations, segment each proposition into theme/rheme

  - discourse model contains previous themes and rhemes (ISstore)
  - to determine theme, search for most recent match, prefer theme-continuation
  - determine rheme as complement of theme

- Sentence planning: determination of realization, focus assignment

  - each new (not mentioned) property or discourse entity get focus
  - contrasting elements get focus

Assigning Contrastive Focus

Step 1: Determine contrasting propositions

- containing 2 contrasting pairs of entities or 1 pair of contrasting entities and contrasting functors
- discourse entities are contrasting when they are alternatives w.r.t. `isa` in DB

Step 2: Contrastive focus algorithm
Given: object x, properties P, alternatives A:

1. restrict A to objects mentioned in discourse $\rightarrow A'$

2. for each property p in P, include p in set of contrasting propertied P_c iff p excludes some object from A'

IS-Based Intonation in TTS

- Previous work: acenting affected by "givenness" (Hirschberg 1990), (Hirschberg, 1993)

- (Hiyakumoto et al. 1997):

  - combine first mention and contrastiveness as reasons for accenting
  - use of WordNet in givenness and contrast determination: to identify sets of synonyms and contrasting words for open-class words (nouns, verbs, adjectives, adverbs)
  - determine theme/rheme in propositional constituents by heuristics applied to pre- and post-verbal and verb-complex material, and considering presence of focus within it

Summary

- IS and intonation (only); formalization in CCG

- Theme/Rheme determined by (a) question, (b) linking in text, (c) heuristics

- Focus is determined from (a) question, (b) discourse newness, (c) contrast w.r.t. alternatives in discourse model

- Discourse model contains entities, propositions, and themes and rhemes

- Theme/Rheme partitioning at proposition level, i.e., recursive in complex utterances

- IS assignment in questions differs from Hoffman

Example Application: Controling Intonation of Spoken Dialog System Output

(Kruijff-Korbayová et al., 2003)

- Motivation

- Deriving IS from the information-state

- Information structure realization through intonation

- Experimental implementation in the GoDiS dialogue system

- Evaluation setup and results

- Conclusions and outlook

Motivation

- Most current systems have limited dialogue flexibility, which enables them to use carefully scripted interactions with predefined and prerecorded output

- Flexible interaction requires output to be dynamically generated

- The realization of dynamically generated output needs to be controlled, to ensure that it is contextually appropriate

- In particular, the intonation of synthesized spoken output needs to be controlled w.r.t. the context

Motivation: Intonation in context

U: Which devices are in the house?

S: There is a STOVE   in the KITCHEN,   a RADIO   in the kitchen
                H*   H* LH%   H*   LH%
           and a radio   in the BATHROOM.
                                H*   LL%

U: What is the status of the devices in the kitchen?

S: The STOVE   in the kitchen   is ON.
          L+H*   L%   H*LL%
     The RADIO   in the kitchen   is OFF.
          L+H*   L%   H*LL%

Information-State Update Approach to Dialogue Modeling

• Dialog moves are modeled as information state update transitions

• Information State represents the current discourse context
  (in a dialogue participant's view)

• e.g. a version of the Dialogue Game Board (Ginzburg, 1996) in GoDIS:

$$
\begin{bmatrix}
\text{PRIVATE} & : & \begin{bmatrix} \text{AGENDA} & : & \text{STACK(ACTION)} \\ \text{PLAN} & : & \text{STACKSET(ACTION)} \\ \text{BEL} & : & \text{SET(PROPOSITION)} \end{bmatrix} \\
\text{SHARED} & : & \begin{bmatrix} \text{COM} & : & \text{SET(PROPOSITION)} \\ \text{QUD} & : & \text{STACK(QUESTION)} \\ \text{LU} & : & \begin{bmatrix} \text{SPEAKER} & : & \text{PARTICIPANT} \\ \text{MOVES} & : & \text{ASSOCSET(MOVE,BOOL)} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

- Utterances push questions onto the QUD stack; resolved QUDs get popped

Information Structure Assignment

Theme/Rheme partitioning: determined according to the QUD

- **QudTR** rule: given an utterance content u to partition, if QUD corresponds to the result of λ-abstracting over a part of u, this part is marked as the Rheme

  If on QUD: $?\lambda x.status(x)$,
  then $status(\underbrace{\text{on}}_{Rheme})$

Information Structure Assignment

Background/Focus partitioning: determined by comparing parallel elements

- **ComFB** rule: if there is an element in the **shared commitments** that is parallel but not identical to an element in the utterance content, the part that is non-identical is marked as the Focus

  If in shared commitments: $\{type(stove)\&location(kitchen);$
  $type(radio)\&location(kitchen); type(radio)\&location(bathroom)\}$
  then $type($ stove $)$ & $location(kitchen)$

  $$\underbrace{\qquad\qquad}_{Focus}$$

Information Structure Assignment

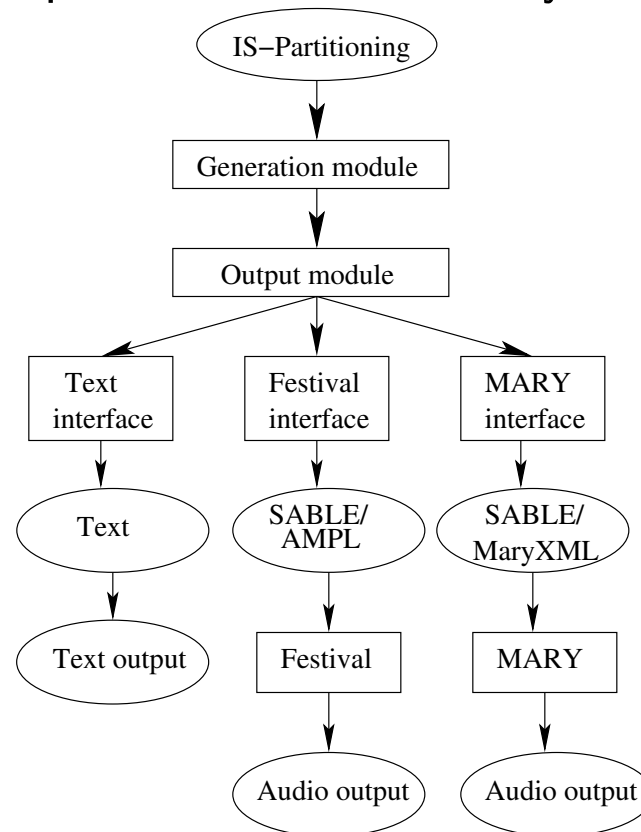Background/Focus partitioning: determined by comparing parallel elements

- **ComFB** rule: if there is an element in the **shared commitments** that is parallel but not identical to an element in the utterance content, the part that is non-identical is marked as the Focus

- **DomFB** rule: if there is an element in the **domain model** that is parallel but not identical to an element in the utterance content, the part that is non-identical is marked as the Focus

Information Structure Assignment
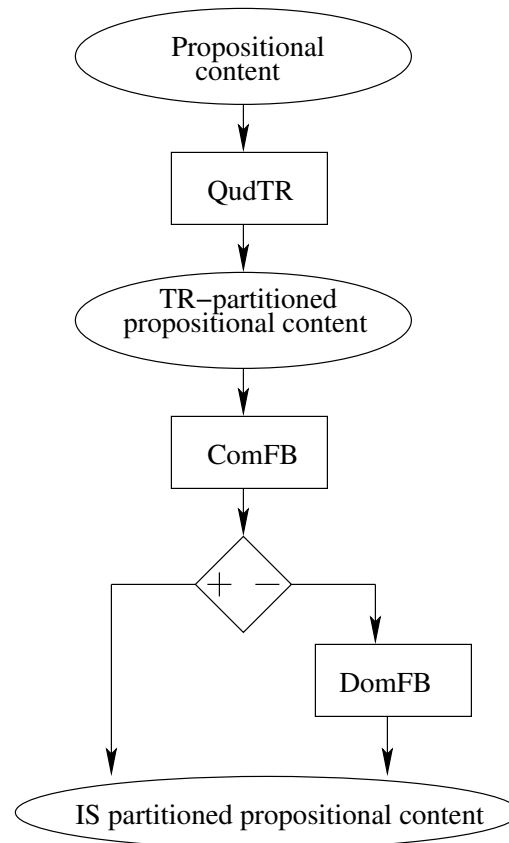
Experimental Implementation

Producing synthesized output with contextually varied intonation in GoDiS

```
                    IS–Partitioning
                          │
                          ▼
                   Generation module
                          │
                          ▼
                     Output module
              ┌───────────┼───────────┐
              ▼           ▼           ▼
           Text       Festival      MARY
         interface    interface   interface
              │           │           │
              ▼           ▼           ▼
           Text        SABLE/      SABLE/
                        AMPL       MaryXML
              │           │           │
              ▼           ▼           ▼
        Text output    Festival     MARY
                          │           │
                          ▼           ▼
                    Audio output  Audio output
```

Experimental Implementation Evaluation

- Using the German text to speech synthesis system Mary (Schröder and Trouvain, 2001) which supports intonation annotation using GToBI (Grice et al., to appear)

- 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 synthesis quality

Conclusions

- Domain- and application-independent rules determining IS partitioning into Theme/Rheme and Background/Focus from the information state

- Domain- and application-independent rules mapping IS partitioning to realization through intonation (in template-based generation)

- Experimental implementation using TTS systems which support ToBI-based intonation determination

- Test-of-concept evaluation

Outlook

- Make more fine-grained decisions about information structure

    - When needed elaborate dialogue context representations
    - Employ more adequate semantic representation

- Replace template-based generation with a generation module that can combine various means to realize information structure

- Use information structure also in interpretation

- What is actually needed in various practical dialogue systems?

IS in Multimodal Interaction

- Appropriate and synchronized speech, intonation, facial expressions and hand gestures (Pelachaud et al., 1998)

- Integrating turn-taking and IS provides better explanation for gaze behavior (Cassell et al., 1999)

- Generation of either speech, gesture or combination of both as a function of IS status and surprise value of a discourse entity (Cassell et al., 2000)

- Various researches have observed that distance, posture shifts and other body movements seem to accompany changes in the topic or social relationship (cf. Cassell et al. 2001)

IS and Gestures

Gilbert and George (Pelachaud et al., 1998)

- Some facial expressions are automatically generated according to intonation (cf. also the COMIC project (**?**))
- Head nods and look-toward listener punctuate accented and emphasized items
- Iconic and metaphoric gestures (i.e., representing something) are generated for
  - rhematic verbal elements (roughly, information not yet spoken about)
  - hearer new references

  provided that the semantic content can receive such a gesture (e.g., spatial)
- Beat gestures are generated
  - otherwise
  - to accompany discourse new definite references
- Duration of intonation phrases is used to time gestures

Summary and Conclusions

- Contextually apropriate system output needed in a variety of applications: dialog, monolog; written, spoken, multimodal; etc.

- Contextually appropriate realization requires account of IS to motivate realization choices by various means in a uniform way

- Modeling of the interplay of linguistic IS-realization choices in practical systems so far largely not done, i.e., concentration on either intonation or word order

- Multimodal systems try to combine linguistic and non-linguistic signals, based on empirically observed correlations

References

Mary Beckman and Julia Hirschberg. 1999. The tobi annotation conventions. Ms. Ohio State University.

Justine Cassell, Obed. E. Torres, and Scott Prevost. 1999. Turn taking vs. discourse structure: How best to model multimodal conversation. In Y. Wilks, editor, *Machine Conversations*. Kluwer.

Justine Cassell, Matthew Stone, and Hao Yan. 2000. Coordination and context-dependence in the generation of embodied conversation. In *Proceedings of the INLG Conference*.

Jonathan Ginzburg. 1996. Interrogatives: Questions, facts and dialogue. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, chapter Chapter 15, pages 385–422. Blackwell Publishers.

M. Grice, S. Baumann, and R. Benzmüller. to appear. German intonation in autosegmental phonology. In S.-A. Jun, editor, *Prosodic typology*. OUP.

Julia Hirschberg. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, (63):305–340.

Ivana Kruijff-Korbayová, Stina Ericsson, Kepa Joseba Rodríguez, and Elena Karagjosova. 2003. 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.

Catherine Pelachaud, Justine Cassel, Norman Badler, Mark Steedman, Scott Prevost, and Mathew Stone. 1998. 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.

Scott Prevost and Mark Steedman. 1993. Generating contextually appropriate intonation. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics, Utrecht*, pages 332–340, San Francisco, CA. Morgan Kaufmann.

Scott Prevost and Mark Steedman. 1994. Specifying intonation from context for speech synthesis. *Speech Communication*, 15:139–153.

Scott Prevost. 1995. *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.

Scott Prevost. 1996. An information structural approach to spoken language generation. In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics, Santa Cruz*.

Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.

Marc Schröder and Jürgen Trouvain. 2001. The german text-to-speech synthesis system MARY: A tool for research, development and teaching. In *The Proceedings of the 4th ISCA Workshop on Speech Synthesis, Blair Atholl, Scotland*.

Mark Steedman. 2000a. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4):649–689.

Mark Steedman. 2000b. *The Syntactic Process*. M.I.T. Press, Cambridge, MA.