Producing Contextually Appropriate Intonation in an Information-State Based Dialogue System

Ivana Kruijff-Korbayová[1] **Stina Ericsson**[2] **Kepa J. Rodríguez**[1] **Elena Karagjosova**[1]
[1]University of the Saarland, Germany [2]University of Gothenburg, Sweden
{korbay,kepa,elka}@coli.uni-sb.de stinae@ling.gu.se

Abstract

Our goal is to improve the contextual appropriateness of spoken output in a dialogue system. We explore the use of the information state to determine the information structure of system utterances. We concentrate on the realization of information structure by intonation. We present the results of evaluating the contextual appropriateness of varied system output produced with a text-to-speech synthesis system that supports intonation annotation.

1 Introduction

Most commercial spoken dialogue systems use carefully scripted dialogues. This has the advantage that the system output can be pre-recorded and have high quality. The disadvantage is limited dialogue flexibility, as user-initiative must be restricted to ensure the dialogue adheres to the script. More flexible dialogues need dynamically produced output. As the range of possible system utterances grows pre-recording becomes infeasible, and speech synthesis becomes necessary.

One challenge for systems using synthesized speech is the generation of contextually appropriate intonation. With dynamically produced output, the same sequence of words may appear in different contexts, possibly needing different intonation. For example, the intonation of an answer needs to correspond to the respective question: whereas in (1S) the nuclear intonation center has a "default" placement, in (2S) it does not.[1]

(1) U: What is the status of the stove?
S: The stove is switched ON.
H* LL%

(2) U: Which device is switched on?
S: The STOVE is switched on.
H* LL%

Contextually inappropriate intonation may have negative effect on intelligibility or even lead to confusion; for example, when (1U) is answered with (2S), or (2U) with (1S), a mismatch arises.

The details of relating intonation and other aspects of realization to context are still a research topic. In this paper, we concentrate on one function of intonation, to realize *information structure* (IS). We consider IS as a level of meaning that unifies a range of interacting contextually-dependent aspects of utterance realization, encountered in various combinations within and across languages. Determining the IS of system utterances according to context, and producing the corresponding realizations, are thus important steps in generating natural system output.

Outline We show how we improve the spoken output of an information-state based dialogue system by controlling intonation using IS.

§2 summarizes related work on controlling intonation in context. §3 gives background on IS,

[1]We print words bearing a pitch accent in SMALL CAPITALS and use the ToBI ("Tones, Breaks and Indices") notation for intonation, cf. *http://www.ling.ohio-state.edu/~tobi/*

§4 on the information-state approach to dialogue. §5 presents rules for determining IS from the information state. §6 describes the generation of spoken output with contextually varied intonation in our system using two off-the-shelf speech synthesis systems. §7 presents evaluation results. We close with a summary and outlook in §8.

2 Related Work

Early work on controlling intonation of synthesized speech in context concerned mainly accenting open-class items on first mention, and deaccenting previously mentioned or otherwise "given" items (Hirschberg, 1993; Monaghan, 1994). But algorithms based on givennenss fail to account for certain accentuation patterns, such as marking explicit contrast among salient items. Givenness alone also does not seem sufficient to motivate accent type variation.

(Prevost, 1995) models contrastive accent patterns and some accent type variation using Steedman's approach to IS in English (§3). In one application he handles question-answer pairs where the question intonation analysis in IS terms is used to motivate the IS of the corresponding answer, realized through intonation. Another application concerns intonation in generation of short descriptions of objects, where Theme/Rheme partitioning is motivated on text progression grounds, and Background/Focus partitioning distinguishes between alternatives in context.

Our approach to assigning IS is similar to Prevost's in assigning IS according to the preceding context, both in terms of what question is being answered and what alternatives are salient. In our dialogue system, context is represented in the information state, which evolves dynamically as the dialogue progresses. In addition, we also determine IS using domain knowledge.

3 Information Structure

Information structure (IS) is an inherent aspect of meaning; it is important for establishing coherence and getting the intended message across. IS partitioning refers to the organization speakers impose on utterances to reflect the context (what they believe is shared between them and the hearer(s)) and the intended context change. Despite conceptual similarities, various terminologies exist to describe IS and its semantics (Steedman and Kruijff-Korbayová, 2003).

We follow (Steedman, 2000), because of the insights he incorporates and the degree of their explicit formalization. In a number of respects Steedman offers a synthesis of earlier proposals. His main point is to provide a compositional analysis of English intonation in IS terms. Of importance to our current enterprise are (i) the discourse-semantic interpretation of IS and (ii) the concrete correlations between IS and intonation. Finally, Steedman's approach to IS in English has been used earlier to control the intonation of synthesized speech in context (cf. §2).

3.1 IS Partitioning

Steedman recognizes two dimensions of IS: a *Theme/Rheme* partitioning at the utterance-level, and a further *Background/Focus* partitioning of both Theme and Rheme. For example:

Theme/Rheme partitioning reflects an *aboutness* relation: the Rheme is semantically predicated over the Theme. In terms of a question test, Theme corresponds to what the question sets up, and Rheme is what answers it.

The Background/Focus partitioning reflects contrast between alternatives, against which the actual Theme and Rheme are cast.

3.2 IS Semantics

Elaborating on (Rooth, 1992; Büring, 1997), (Steedman, 2000) defines this semantics for IS:

Rheme presupposes a *Rheme-alternative set.* (ρ-*AS*). Rheme-Focus selects one element from ρ-AS. **Theme** presupposes a *Theme-alternative set* (θ-*AS*). Without Focus in Theme, θ-AS is a singleton set. Otherwise, θ-AS has more elements, and the Theme-Focus selects one of them.

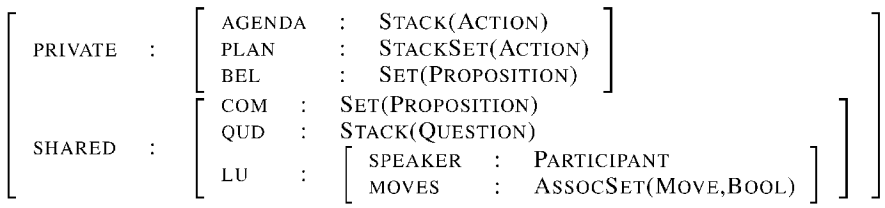


Figure 1: The Information State in GoDIS

3.3 IS and Intonation

IS can be realized by various means such as intonation, word order, grammatical structure or morphological marking. Here we concentrate on IS realization by intonation. Steedman has argued extensively that in English IS is homomorphic to intonation structure. In a nutshell:

Theme/Rheme partitioning determines what accents are used: L+H*, L*+H in Theme and H*, L*, H*+L, H+L* in Rheme. Focus/Background partitioning determines the placement of pitch accents: they are assigned to words realizing the Focus elements. Words realizing Background elements do not carry an accent. A Rheme always contains a Focus, while Themes are marked (with Focus) or unmarked (without Focus).

An L or H boundary marks the end of an intermediate phrase, and an L% or H% boundary tone the end of an intonational phrase. Themes or Rhemes can constitute intonational phrases.

Accents, appropriate boundaries and boundary tones create tunes. Steedman argues that in English L+H*LH% is a marked-Theme tune, and H*LL% is one of the Rheme tunes in assertions. (Uhmann, 1991) suggests similar default tunes for German. We deviated from that by preserving the basic tunes of the German speech synthesis system we used, namely L+H*H-% for marked Theme and H+L*LL% for Rheme.

4 Information State Based Approach

We implemented the generation of contextually varied intonation in GoDIS, an experimental system within the Information State framework, built using TrindiKit[2]. GoDIS handles information exchange dialogue in travel agency and autoroute domains, and action-oriented dialogue at the interface to a mobile phone, VCR and some other home devices (Larsson, 2002).

The Information State approach to dialogue modeling views dialogue as moves made by the participants. Their content updates the information state in various ways. The type of record assumed for the GoDIS information state is a version of the *dialogue gameboard* (Ginzburg, 1996) (Fig. 1). It is divided into a PRIVATE and a SHARED part, the latter containing information that the agent assumes to be shared by the participants in the dialogue. Besides information about the latest utterance (speaker and move(s)), the SHARED part contains shared commitments (a set of propositions) and QUD (a stack of questions under discussion). When a question is asked, it is pushed onto the QUD, and is popped off when it is answered. In the PRIVATE part, the plan contains the system's long-term goals, while the agenda contains more immediate actions.

A user utterance like "I'd like to go to London" is recognized as a move giving a destination and its content is represented in the shared commitments as the proposition $dest(london)$. The corresponding question, where does the user want to go, is represented on the QUD as $?\lambda x.dest(x)$.

GoDIS also contains modules for input interpretation, updating the information state, selection of next system move, and output generation, as well as resources such as lexicon and domain knowledge. The domain knowledge includes, e.g., dialogue plans and semantic sorts.

5 Information Structure Determination

We present our approach to determining IS from the information state here, and the corresponding generation of varied intonation in §6.

[2] *http://www.ling.gu.se/projekt/trindi/trindikit/*

5.1 IS Determination Rules

(Ginzburg, 1996) describes the felicity of an IS partitioning as requiring that a certain question is topmost on QUD. Based on that, we formulate the QUD-based Theme/Rheme determination (**QudTR rule**): If there is a question q topmost on QUD, and an utterance u with content c is to be uttered, where q is obtained by λ-abstracting over c, then that part of c which corresponds to q belongs to the Theme of u, and the other part of c is the informative part which constitutes the Rheme of u. For example, if the question under discussion is $?\lambda x \lambda y.price(x, y)$, then the propositional content $price(200, euro)$ of an answer can be partitioned as $\langle_T\ \lambda x.price(x)\ _T\rangle$ $\langle_R$ 200,Euro $_R\rangle$, where the Rheme corresponds to the value of the price parameter.

The Focus/Background determination within Theme and Rheme is done using (semantic) parallelism, which we define as follows (an information unit is a basic term, a Theme or a Rheme, or a proposition without Theme-Rheme partitioning): Two information units, $a = a1 \circ a2$ and $b = b1 \circ b2$ ($\circ$ means composition), are parallel when $a1$ is parallel with $b1$ and $a2$ is parallel with $b2$. Two basic terms are parallel when they are either identical or alternatives (belonging to the same sort but non-identical). For example, $class(business)$ and $class(economy)$ are parallel since the two instances of $class$ are identical, and $business$ and $economy$ are alternative.

We now define two complementary rules for determining Focus/Background based on parallelism. The difference between them lies in what the source of alternatives (and identicals) is. Focus is assigned to any element in an informativity unit having an alternative:

(i) In the shared commitments (**ComFB rule**): If $price(1000, euro)$ is in the shared commitments, and $price(500, euro)$ is to be uttered, Focus will be assigned to 500, because that is what distinguishes the price alternatives.

(ii) In the domain (**DomFB rule**): Given $business$ and $economy$ are alternatives in the domain, DomFB assigns Focus to $economy$ in $class(economy)$.

5.2 Implementation

In our experimental implementation in GoDIS, the selection algorithm evokes for each system move the module for IS assignment. It takes as input the propositional content of the move, and returns it partitioned. IS assignment has several phases (Fig. 2). First, the QudTR rule partitions the semantic form into Theme and Rheme. Then, the ComFB rule fires. If it fails to assign any Focus, the DomFB rule fires.

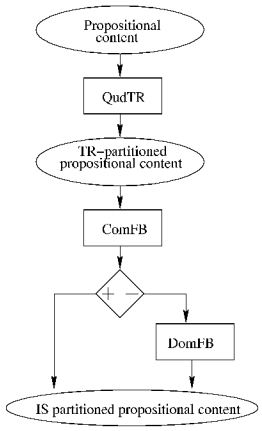
The IS assigned to the content of a move is encoded by the operators rh for Rheme, foc_rh for Rheme-Focus and foc_th for Theme-Focus. The IS-partitioned content is sent to the generation module, which produces a string of words with an IS annotation using an internal set of labels: `<RH>`, `<F_RH>` and `<F_TH>` respectively. For instance, a fully partitioned proposition is *class(foc_th(business)), price(rh(foc_rh(1000)))*. The generated utterance labeled with information structure is: `<F_TH>` Business `</F_TH>` class costs `<RH>` `<F_RH>` 1000 `</F_RH>` Euro`</RH>`.

The generation of the corresponding contexually varied spoken output in GoDIS is described in §6. The sections below detail out how we get the IS-partitioning of the propositions.

5.2.1 QudTR

The QudTR rule is implemented as four disjunctive selection rules which fire depending on the semantics of the move to be generated and the content of QUD. For example, the rule below

Figure 2: Information Structure Assignment



is applied if there is a question topmost on QUD which the proposition of the next move resolves.[3]

RULE: **qudTR**

CLASS: select_tr

PRE: { fst($CONTENT_OF_NEXT_MOVES, answer(A))
or fst($CONTENT_OF_NEXT_MOVES, inform(A))
fst($QUD, $?A.B$)
$DOMAIN :: resolves($B$, C)

EFF: { rheme1(CONTENT_OF_NEXT_MOVES)

The assignment of Rheme, Rheme-Focus and Theme-Focus is done by a number of operators. For example, *rheme1* defined for an answer move assigns Rheme to the argument of a proposition.

```
operation(rheme1,oqueue([Move|T]),[],
                  oqueue([Move1|T])):-
        Move = answer(A),
        A =..[Functor,Argument],
        A2 =.. [Functor, rh(Argument)],
        Move1 = answer(A2).
```

Each information unit corresponding to a Theme or a Rheme is further processed by the rules assigning the Focus/Background partitioning using parallelism, which we turn to below.

5.2.2 ComFB

The ComFB rule currently applies to *inform* or *answer* move. The *in_set* operator checks if SHARED/COM contains a parallel proposition.

RULE: **comFB**

CLASS: select_fb

PRE: { fst($CONTENT_OF_NEXT_MOVES,
inform(rh([A | _]))) or
fst($CONTENT_OF_NEXT_MOVES,
answer(A))
in_set($/SHARED/COM, A)

EFF: { focus_arg(CONTENT_OF_NEXT_MOVES)

5.2.3 DomFB

The DomFB rule is implemented in three separate selection rules. The general case is covered by the rule below that takes an *answer* or *inform* move and tests for an alternative in the domain.

RULE: **domFB**

CLASS: select_fb

PRE: { fst($CONTENT_OF_NEXT_MOVES, answer(A)) or
fst($CONTENT_OF_NEXT_MOVES, inform([A]))
$DOMAIN :: proposition($A$)

EFF: { focus_arg(CONTENT_OF_NEXT_MOVES)

[3] PRE and EFF abbreviate the precondition(s) and effect(s) of a rule, respectively.

Example To show the assignment of both Theme-Focus and Rheme-Focus, consider (3):

(3) S1: Hello, how can I help you?
U1: What is the price of a flight from Paris to London on April fifth?
S2: What class did you have in mind?
U2: I don't know.
S3: BUSINESS class costs ONE THOUSAND euro. ECONOMY class costs FIVE HUNDRED euro.

The first utterance in (3S3) is an answer move already partitioned into Theme/Rheme: *price(rh(1000))*, *class(business)*. This Theme/Rheme partitioned move is the input of the Focus/Background rules. In this case, DomFB is applied since SHARED/COM contains no proposition parallel to the proposition in the answer *class(business)*. The operator *focus_arg* assigns Focus to the arguments of the Rheme and the Theme. The resulting partitioned propositions *price(rh(foc_rh(1000)))* and *class(foc_th(business))* serve as input to the generation of the surface realization.

6 Producing Speech Output with Intonation Variation

To produce contextualy varied spoken output in GoDIS, we use the Mary and Festival text-to-speech synthesis systems, and define mappings from our internal IS annotation to intonation annotation used by these systems. We chose these systems because they are both freely available, and they both support not only the SABLE intonation annotation standard[4] but also a more fine-grained ToBI-based intonation annotation.

The integration of Festival and Mary into GoDIS (Fig. 3) allows to experiment with: (i) Mary for German using SABLE or GToBI intonation annotation, (ii) Mary for English using SABLE intonation annotation and (iii) Festival for English using SABLE or the AMPL annotation (Kruijff-Korbayová et al., 2003).

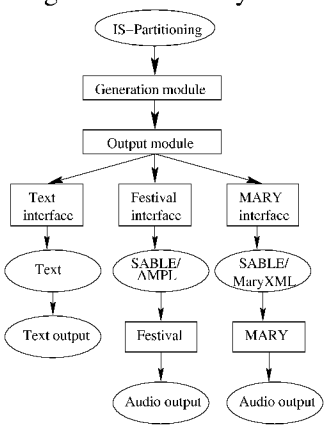
Here we concentrate on using Mary to generate German output, because that is the version for which we already have evaluation results. An evaluation of the English version is forthcoming.

[4] *http://www.bell-labs.com/project/tts/sable.html*

Table 1: Experimental mapping of IS-partitioning to intonation annotation for German in Mary

IS-partitioning	GToBI	SABLE
Focus within Theme	L+H*	`EMPH, PITCH BASE="+15%"`
Focus within Rheme	H+L*	`EMPH, PITCH BASE="+20%"`
Unmarked-Theme boundary (before Rheme)	none	-
Marked-Theme boundary (before Rheme)	H-% (break ind. 3)	-
Rheme boundary (before Theme)	none	-

Figure 3: Integration of TTS systems in GoDIS.



Mary is developed at DFKI and the Saarland University.[5] It is designed to be highly modular, focusing on transparency and accessibility of intermediate processing steps, which makes it a suitable research tool. Mary currently handles German and English. It supports SABLE for both. For German Mary also suports the full inventory of tones defined in the German ToBI (Grice et al., to appear), and a set of break indices that distinguish between a potential boundary location (which might be "stepped up" and thus realized by some phonological process later on), an intermediate phrase break and sentence-final and paragraph-final boundaries. To assign default accents, Mary treats clauses as intonation phrases, and phrases as intermediate phrases. Each intermediate phrase carries a pitch accent. The last pitch accent in an intonation phrase is H+L*, all others are L+H*. This default intonation structure corresponds roughly to an IS partitioning with a marked Theme before Rheme (cf. §3.3).

The GoDIS–Mary interface overrides these defaults. It converts the automatically assigned internal IS annotation tags into SABLE/GToBI (cf. the tag mapping in Tab. 1), and stores the result as SABLE- or Mary-XML, respectively.

7 Evaluation

To evaluate the impact of controlling intonation through IS on the acceptability of system turns, we conducted two experiments with the German output of Mary. We tested whether there are differences in acceptability between (i) default output and (ii) the controlled intontation, in general and for various IS patterns. First, we compared contextual appropriateness of output produced with (i) the default intonation and controlled intonation, using either (ii) GToBI or (iii) SABLE intonation markup. Second, we carried out a more detailed comparison of contextual appropriateness for (i) the default intonation and (ii) the controlled intonation using GToBI.

For each experiment, we prepared 3-5 turn dialogues from the travel agency and home-device domains handled in GoDIS.[6] The last turn was the evaluated system utterance. The turns were constructed so that the context supported different IS patterns in the target: Marked or unmarked Theme before or after Rheme. Intonation annotation was assigned as described in §5 and §6.

The dialogues were presented on a web page[7] with the targets highlighted in bold. The subjects were asked to go through the dialogues one by one, read a dialogue, listen to the target audio, and judge the contextual appropriateness of its intonation on a scale from 1 (worst) to 5 (best). In the first experiment, 22 subjects judged 10 utterances in different intonation versions, in the

[5] *http://mary.dfki.de*; (Schröder and Trouvain, 2001)

[6] We had to use constructed fragments, because we do not have a corpus of GoDIS sessions.

[7] *http://www.coli.uni-sb.de/cl/projects/siridus/*

second one it was 20 subjects and 16 utterances.[8]

The default was generally judged worse than SABLE output and that was judged worse than GToBI output (Tab. 2, Ex.1). This was also the case for utterances with unmarked Theme, irrespective of Theme-Rheme order (Tab. 3-4, Ex.1). For marked Theme, SABLE shows a slight improvement over the default (Tab. 5-6, Ex.1). However, this is due to slight differences in pronunciation not due to the intonation annotation. More detailed analysis revealed a few exceptions when SABLE was judged better than GToBI. A possible source is that SABLE annotated input is additionally processed and possibly modified by applying Mary defaults in ways we cannot control. Thus, Mary may sometimes "improve" the SABLE intonation specification (towards default). With GToBI we give specific annotations that prevent the application of Mary defaults, but may sometimes result in less smooth output.

In the second experiment we restricted the comparison to the default and GToBI output, and we varied the IS patterns systematically.

The second experiment confirms the tendency that GToBI outputs get better acceptability judgements than Mary defaults both overall (Tab. 2, Ex. 2) and per IS pattern (Tab. 3-6, Ex.2).

Comparing average absolute judgements can be problematic, if the subjects place their judgements differently on the scale. However, the average differences between individual subjects' judgements of the GToBI and default version of each target were small and confirmed that GToBI output is judged better than the default (Table 7).

After the first experiment, we also realized that the setup we use cannot ensure the subjects actually take the context into account. We considered presenting the dialogues spoken (recorded human-user turns and synthesized system turns), but then the quality and intonation of turns other than the target could influence the judgements.

Instead, we included evaluation of the targets with default intonation in isolation before their evaluation in context in the second experiment.[9]

Table 2: Overal judgements

Exp.1 (10 sent.)	All	Default	GToBI	SABLE
mean/med.	3.35/3	3.23/3	3.62/4	3.19/3
stand.dev.	1.20	1.18	1.08	1.32
Exp.2 (16 sent.)	All	Default	GToBI	SABLE
mean/med.	3.59/4	3.47/4	3.71/4	-
stand.dev.	1.11	1.18	1.03	-

Table 3: Unmarked Theme, Theme before Rheme

Exp.1 (3 sent.)	All	Default	GToBI	SABLE
mean/med.	3.10/3	3.12/3	3.52/3	2.65/2
stand.dev.	1.32	1.42	1.09	1.36
Exp.2 (4 sent.)	All	Default	GToBI	SABLE
mean/med.	3.55/4	3.42/4	3.67/4	-
stand.dev.	1.11	1.13	1.08	-

Table 4: Unmarked Theme, Rheme before Theme

Exp.1 (2 sent.)	All	Default	GToBI	SABLE
meanmed.	3.67/4	3.68/4	3.70/4	3.60/3
stand.dev.	1.00	1.155	1.07	1.12
Exp.2 (4 sent.)	All	Default	GToBI	SABLE
mean/med.	3.63/4	3.56/4	3.70/4	-
stand.dev.	1.15	1.23	1.01	-

Table 5: Marked Theme, Theme before Rheme

Exp.1 (4 sent.)	All	Default	GToBI	SABLE
mean/med.	3.22/3	3.08/3	3.43/4	3.15/3
stand.dev.	1.16	1.11	1.20	1.23
Exp.2 (4 sent.)	All	Default	GToBI	SABLE
mean/med.	3.62/4	3.54/4	3.69/4	-
stand.dev.	1.05	1.05	1.06	-

Table 6: Marked Theme, Rheme before Theme

Exp.1 (1 sent.)	All	Default	GToBI	SABLE
mean/med.	4.13/4	3.25/3	4.50/4	4.55/5
stand.dev.	1.05	1.056	1.10	1.23
Exp.2 (4 sent.)	All	Default	GToBI	SABLE
mean/med.	3.47/4	3.3/4	3.63/4	-
stand.dev.	1.15	1.27	1.00	-

Table 7: Default vs. GToBI judgment differences

Exp.1	Th before Rh	Rh before Th
+ Theme-Focus	0.44	1.53
- Theme-Focus	0.32	0.04
Exp.2	Th before Rh	Rh before Th
+ Theme-Focus	0.29	0.28
- Theme-Focus	0.3	0.26

[8]The subjects were mostly computational linguistics students. Some had previous knowledge of phonetics or experience with speech synthesis.

[9]We did not include the non-default versions in the eval-

Since the default intonation corresponds to the IS pattern with marked Theme before Rheme (what may differ is the location of pitch accents), we expected the judgements to remain about the same when the context supports an IS partitioning which results in the same intonation as the default. In other cases, we expected the judgements for GToBI output to be better than those for the default. These predictions were born out by the differences between judgments of individual sentences with the respective IS patterns.

8 Conclusions

Our goal was to explore the use of the information state to control the intonation of system output. We concentrated on intonation as the realization of IS. We defined a set of rules which derive IS from the information state in GoDIS.

The information state, together with the domain knowledge, has proven to accommodate the IS components and predications of (Steedman, 2000), which in turn is translatable into other approaches to IS. This in itself is a result and an indication of the viability of our approach.

We developed an experimental implementation that uses speech synthesis systems supporting SABLE and ToBI-based intonation markup. We presented the results of evaluating the implementation using Mary for generating contextually varied spoken output in German. The evaluation indicates that the contextual appropriateness of system output improves when intonation is assigned on the basis of IS.

A number of issues remain to be explored:

- Adjustments to the rules to properly cover specific cases of dialogue moves
- Accounting for the interplay between information in the information state and in the domain knowledge
- Taking more dialogue history into account
- Making more subtle decisions about the realization of information structure

The need for making more fine-grained semantic choices such as the IS partitioning also raises the issue of a suitable semantic representation. The semantics close to database contents we use now obscures many aspects of meaning important for more subtle dialoge modelling.

Acknowledgments This work has been supported by the EU project SIRIDUS (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems, IST-1999-10516). We thank Robin Cooper, Geert-Jan Kruijff, Staffan Larsson and David Milward for discussions. We are also grateful to the evaluation participants.

References

Daniel Büring. 1997. *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. Routledge.

Jonathan Ginzburg. 1996. Interrogatives: Questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.

Martine Grice, Stefan Baumann, and Ralf Benzmüller. to appear. German intonation in autosegmental-metrical phonology. In Jun Sun-Ah, editor, *Prosodic Typology*. Oxford University Press.

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

Ivana Kruijff-Korbayov´a, Elena Karagjosova, Kepa Joseba Rodr´ıguez, and Stina Ericsson. 2003. A dialogue system with contextually appropriate spoken output intonation. In *Proc. EACL'03*. Budapest, Hungary.

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University.

Alex Monaghan. 1994. Intonation accent placement in a concept-to-dialogue system. In *Proc. 2nd ESCA/IEEE Wsh on Speech Synthesis*, pp. 171–174. New Paltz, NY.

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.

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. In *Proc. 4th ISCA Wsh. on Speech Synthesis*. Blair Atholl, UK.

Mark Steedman and Ivana Kruijff-Korbayov´a. 2003 (to appear). Two dimensions of information structure in relation to discourse structure and discourse semantics. *Journal og Logic, Language and Information*.

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

Susanne Uhmann. 1991. *Fokusphonologie*. Niemeyer.

uation of targets in isolation, just to keep the down the number of judgements the subjects had to make.