A Dialogue System with Contextually Appropriate Spoken Output Intonation

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

Abstract

We demonstrate the production of spoken output with contextually appropriate intonation in the information-state based dialogue system GoDiS. We exploit the context representation in the information state to determine the information structure of system utterances, which we use to control the intonation of synthesized spoken output.

1 Introduction

Producing spoken output with contextually appropriate intonation is one of the challenges for flexible dialogue systems with dynamically constructed output and synthesized speech. It is a well known fact that intonation reflects the relation of an utterance to the context, and that contextually inappropriate intonation may have negative effect on intelligibility or lead to confusion.

We demonstrate improvements of contextual appropriateness of English and German intonation in the GoDiS system. Intonation is controlled by information structure, which is determined from the context representation in the information state of the system using the information-state update approach to dialogue.

This note is structured as follows. In §2 we give an overview of GoDiS and its information-state update approach. In §3 we introduce the information structure partitioning we employ, and the rules we use to determine it from the information state. In §4 we describe the generation of spoken output with contextually varied intonation in GoDiS using the FESTIVAL and MARY text-to-speech synthesis systems. In §5 we summarize and indicate our further research plans.

2 GoDiS

GoDiS (Gothenburg Dialogue System) is an experimental dialogue system implemented using the TrindiKit, a toolkit for implementing dialogue move engines and dialogue systems based on the information-state update approach (TRINDI, 2001; Larsson and Traum, to appear).

One of the goals of the information-state update approach is to encourage modularity, reusability and plug-and-play; to demonstrate this, GoDiS has been adapted to several different dialogue types (information-seeking, action-oriented), domains (travel agency, autoroute, mobile phone, VCR) and languages (English, Swedish, German) (Larsson, 2002).Speech input and output are also supported in GoDiS.

The GoDiS architecture is shown in Fig. 1. It is an instantiation of the general TrindiKit architecture (Larsson and Ericsson, 2002).

The information state in GoDiS represented as a record (Fig. 2) is a modified version of the dialogue game board (Ginzburg, 1996). The main division is between information which is PRIVATE to an agent and that which is SHARED between agents. In PRIVATE, the PLAN field contains a list of long-term goals; AGENDA contains more immediate dialogue actions; BEL is a set of assumed propositions; TMP keeps track of information that

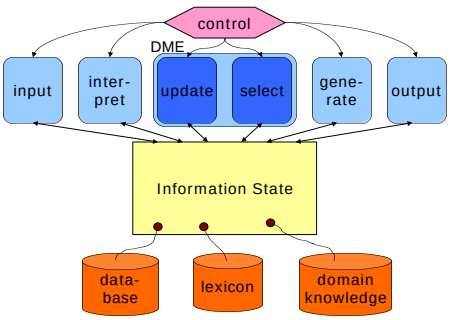


Figure 1: GoDiS architecture

has not yet been grounded. The SHARED part contains information about the latest utterance, a set of established shared commitments and a stack of questions raised in the dialogue that are currently under discussion.

What we concentrate on in this demonstration is our extension of GoDiS, enabling it to dynamically produce contextually appropriate intonation by assigning the system utterances information structure partitioning according to the information state, and controlling the output intonation accordingly.

3 Information Structure

Information structure (IS) refers to the organization speakers impose on their utterances to relate them to the context (what they believe is shared) and the intended context change (corresponding to their communicative intentions).

The approach to IS we employ follows (Steedman, 2000). This choice is motivated by the insights that Steedman incorporates and the degree of their explicit formalization. We thus use two dimensions of IS: (i) a partitioning into *Theme* and *Rheme*, corresponding to a semantic aboutness relation; and (ii) a further partitioning of both Theme and Rheme into *Background* and *Focus*, reflecting a contrast between alternatives in the context against which the actual Theme and Rheme are cast. E.g., the IS-partitioning suitable in the context of *The heater in the hall is out. But what is the status of the light in the hall?*:[1]

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

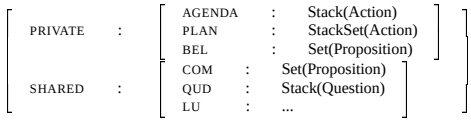
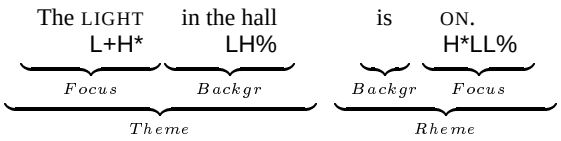


Figure 2: GoDiS Information State Record Type



3.1 Information structure and Intonation

Intonation is one of the means by which IS can be realized. For English, Steedman has argued that IS is homomorphic to intonation structure:

The Theme/Rheme partitioning determines the overall intonation pattern: different accents are used within the Theme (L+H*, L*+H) and within the Rheme (H*, L*, H*+L, H+L*).

The Focus/Background partitioning determines the placement of pitch accents: they are assigned to the words realizing the Focus within Theme and Rheme. A Rheme must always contain a Focus, while Themes can be unmarked (without Focus) or marked (with Focus).

Tunes are obtained by combining accents with appropriate boundaries and boundary tones. Steedman has argued that the contour H*LL% is one of the "rheme tunes" in assertions in English, and L+H*LH% is a (marked) Theme-tune. For German, we adopt H+L*LL% as a default rheme tune, and L+H*H-% for a marked theme (these accents are the ones implemented as defaults in the Mary system we use to synthesise German; cf. (Kruijff-Korbayová et al., 2003) for more discussion).[2]

3.2 Information Structure Determination

We have implemented IS-assignment to system moves in GoDiS as a module invoked from the selection algorithm (cf. Fig. 1). The module takes as input the propositional content of a dialogue move, and returns this content IS-partitioned. The process of IS assignment has several phases shown schematically in Fig. 3.

[2]For German ToBI cf. (Grice et al., to appear).

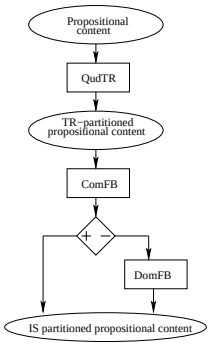


Figure 3: IS-Assignment in GoDiS

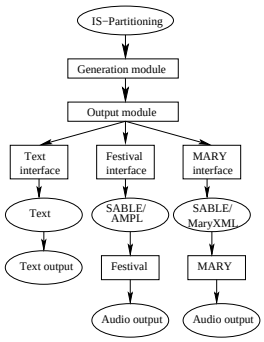


Figure 4: GoDiS-TTS Interfaces

First, the QudTR rule partitions the content into Rheme and Theme, according to the question topmost on QUD. Next, the determination of the Background/Focus partitioning within each Theme and Rheme is done using a notion of (semantic) parallelism, by two complementary rules which differ in what the source of alternatives is taken to be: The ComFB rule tries to assign Focus on the basis of the previous dialogue context, looking for alternatives in the SHARED.COM field of the information state. If this fails to assign any Focus, the rule DomFB assigns Focus by looking for alternatives in the domain representation. (See (Prevost, 1995) for a similar algorithm.)

The IS partitioning of a dialogue move content is encoded by the operators *rh* for Rheme, *foc_rh* for Rheme-Focus and *foc_th* for Theme-Focus.

Finally, the IS-partitioned content is sent to the generation module, which produces a string of words with an annotation of the IS partitioning using an internal set of labels `<RH>`, `<F_RH>` and `<F_TH>`, respectively.

4 Producing Speech Output with Intonation Variation

In order to produce contextually varied synthesized speech output we use the FESTIVAL TTS for English and the MARY TTS for German which are publicly available. We chose these systems because they support not only the SABLE intonation mark up standard,[3] but also a more abstract ToBI-based intonation annotation.

FESTIVAL is a multi-lingual TTS system developed at CSTR, University of Edinburgh. We use an experimental set of patches (APML) developed by Robert Clark at the University of Edinburgh, that allows to annotate FESTIVAL input with higher levels of information including speech-act type and turn-talking information, as well as a ToBI-based intonation markup.

MARY (Schröder and Trouvain, 2001) is a TTS System developed by the DFKI language technology lab and the Institute of Phonetics at Saarland University. MARY supports the full inventory of tones defined in GToBI, and allows partial annotation at any level in its input.

The integration of FESTIVAL and MARY into GoDiS is shown in Fig. 4. The interface between GoDiS and MARY/FESTIVAL works as follows: The output module of GoDiS takes a string annotated with IS partitioning and calls a Linux/Unix shell. A program written in PERL converts the string into the corresponding SABLE/MaryXML/APML tags. The result is saved into a SABLE/MaryXML/APML output file. The mapping of tags for German using MaryXML is shown in Table 1, for English using APML in FESTIVAL in Table 2.

Both MARY and FESTIVAL can be run locally or as servers. The output module of GoDiS calls a Linux/Unix shell and sends the SABLE/MaryXML/APML file to MARY/FESTIVAL.

More detailed information about the system can be found in (Kruijff-Korbayová et al., 2002).

5 Summary and Future Work

Our goal is to explore the use of the information state in GoDiS to control the intonation of system

[3] http://www1.bell-labs.com/project/tts/sable.html

IS-partitioning	GToBI
Focus within Theme	L+H*
Focus within Rheme	H+L*
Unmarked-Theme boundary (before Rheme)	none
Marked-Theme boundary (before Rheme)	H-
Rheme boundary (before Theme)	none

Table 1: Mapping of IS partitioning tags to MaryXML intonation annotation for German

IS-partition	APML-label
Begin Rheme	<rheme>
End Rheme	</rheme>
Theme Focus	<emphasis x-pitchaccent="Hstar">
Rheme Focus	<emphasis x-pitchaccent="LplusHstar">

Table 2: Mapping of IS-partitioning tags into APML-labels in FESTIVAL for English

output. We demonstrate an experimental implementation using the FESTIVAL and MARY TTS systems which support the SABLE standard as well as a ToBI-based intonation markup.

Our implementation allows us to test hypotheses concerning contextually appropriate intonation in dialogue. A pilot evaluation of the German output produced with MARY yielded encouraging results suggesting that in general users find the controlled contextually appropriate intonation better (Kruijff-Korbayová et al., 2003).

Although we have so far only exploited intonation, one goal for the future is to let various information structure realization means interplay.[4]

References

[Ginzburg1996] Jonathan Ginzburg. 1996. Interrogatives: Questions, Facts and Dialogue. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*.Blackwell Publishers.

[Grice et al. to appear] 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.

[Kruijff-Korbayová et al.2002] Ivana Kruijff-Korbayová, Stina Ericsson, Carlos Garcia, Rebecca Jonson, Elena Karagjosova, Pilar Manchón, Kepa J. Rodriguez, and José Quesada. 2002. Improving System Output Using the Information State. Deliverable D5.1, SIRIDUS.

[Kruijff-Korbayová et al.2003] Ivana Kruijff-Korbayová, Stina Ericsson, Kepa Joseba Rodriguez, and Elena Karagjosova. 2003. Producing Contextually Appropriate Intonation in an Information-State Based Dialogue System. In *Proceedings of the 10th Conference of the European Chapter of the ACL*. forthcoming.

[Larsson and Ericsson2002] Staffan Larsson and Stina Ericsson. 2002. GoDiS - Issue-Based Dialogue Management in a Multi-Domain, Multi-Language Dialogue System. Demo-abstract. The 40th Annual Meeting of the ACL, University of Pennsylvania, Philadelphia.

[Larsson and Traum to appear] Staffan Larsson and R. Traum, David. to appear. Information State and Dialogue Management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering*.

[Larsson2002] Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University.

[Prevost1995] Scott Prevost. 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. dissertation, University of Pennsylvania, Philadelphia.

[Schröder and Trouvain2001] 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.*

[Steedman2000] Mark Steedman. 2000. Information Structure and The Syntax-Phonology Interface. *Linguistic Inquiry*, 31(4):649–689.

[TRINDI2001] TRINDI. 2001. The TRINDI Book: Task Oriented Instructional Dialogue. Technical Report LE4-8314, Gothenburg University, Sweden. http://www.ling.gu.se/projekt/trindi/book.ps.

[4]This work was supported by the EU project SIRIDUS (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems, IST-1999-10516). We are grateful to Robin Cooper, Geert-Jan Kruijff and Staffan Larsson for discussions and comments, as well as to the 42 subjects.