A generation perspective on speech production in dialogue*

Eli Hagen, Elke Teich, Brigitte Grote, Adelheit Stein

Abstract. The present paper discusses speech production from the perspective of NL (natural language) generation in the context of dialogue, focusing on the selection of appropriate intonation. Bringing forward arguments from functional linguistics, we present a model of *communicative-context-to-speech* as an alternative to *text-to-speech* and *concept-to-speech* approaches.

1 Introduction

Both speech recognition and speech synthesis take an analysis perspective on speech; speech recognition deals with incomplete, incorrect input and repairs, while the major issue in speech synthesis is how to best render *text to speech*. In terms of applications of speech recognition and speech synthesis, what appears currently in the focus of attention is automatic interpretation, i.e., human-human dialogue in multiple languages translated by machine, and phone-in information systems, which also brings with it an analysis focus on speech processing with very specific requirements, some of which are given in the call to this workshop. In this paper we discuss an alternative view on speech *production* that bears the promise of overcoming some of the problems of text-to-speech by taking a natural language generation view.

While clearly, the processing of speech involves some speech-specific problems, there is a large degree of overlap between text and speech processing. We argue here that therefore, it is worthwhile considering an integration of speech and natural language processing more seriously (cf. [Black and Campbell, 1995, Hirschberg *et al.*, 1995]).

We introduce an approach that moves away from *text-to-speech* and *concept-to-speech* generation towards *communicative-context-to-speech* generation. It has been developed in the framework of SPEAK!, a joint collaborative project between the Technical Universities of Budapest and Darmstadt ([Grote, 1995, Hagen and Stein, 1996, Olaszy *et al.*, 1992]). The main goal of the project is the development of a generic dialogue-based interface for information retrieval, which generates German speech output for conveying *meta-communication* that complements the visual output and direct manipulation input. We propose to employ an explicit representation of the communicative context of dialogue contributions to select appropriate intonation patterns and thus improve the quality of the spoken output. Combining a state-of-the-art dialogue model with a state-of-the-art NL text generator, we determine the kinds of pragmatic and linguistic knowledge needed to constrain intonational choices.

As an introduction, we discuss the state-of-the-art in speech production (Section 2) and present the context of human-machine dialogue in information seeking processes (Section 3). Then we briefly present the NL generation perspective, focusing on the assignment/selection of intonation (Section 4). Section 4 also presents a functionally and contextually motivated model of speech production in dialogue that unites NL generation, speech production, and dialogue modelling. Section 5 concludes with a summary of the arguments brought forward in the course of this paper and points out a number of open issues.

2 Speech production

In spoken information seeking dialogues, intonation is often the only means to distinguish between different dialogue acts, thus making the selection of the appropriate intonation crucial to the success of the dialogue (see e.g., [Prevost and Steedman, 1994] for English). To illustrate this point, imagine an information dialogue where the user wants to know a specific train connection. At some point in the interaction, the system produces a sentence like *Sie fahren um drei Uhr von Darmstadt nach Heidelberg* ("You travel at three o'clock from

*This work was partially funded by the EU Programme Copernicus, Project No. 10393 (SPEAK!).—Eli Hagen and Adelheit Stein are employed at GMD-IPSI, Dolivostr. 15, 64293 Darmstadt, Germany; Elke Teich is employed at the University of the Saarland, Department of Applied Linguistics, Translation and Interpretation, 66041 Saarbrücken, Germany; and Brigitte Grote is employed at FAW Ulm, 89018 Ulm, Germany.

Darmstadt to Heidelberg."). Depending on the intonation this utterance can be interpreted differently; the most obvious interpretation being that the system presents some kind of information to the reader, but it could also be part of a clarification dialogue, where the system wants to reassure itself that it got the user's request right.

Current German speech synthesizers can support sophisticated variation of intonation, but no existing system provides the semantic nor pragmatic guidance necessary for selecting intonation appropriately. Traditional *text-to-speech* systems (e.g., [Hemert *et al.*, 1987, Huber *et al.*, 1987, Olaszy *et al.*, 1992]) perform a linguistic analysis on the written text and use the resulting syntactic structure to control prosodic features. *Concept-to-speech* systems (see e.g., [Dorffner *et al.*, 1990]) synthesize from a pre-linguistic conceptual structure but only use this structure to avoid the problem that syntactic analysis produces incomplete and multiple structures. *Discourse model based systems* (see e.g., [Hirschberg, 1992, Prevost and Steedman, 1994]) also start from a syntactic structure and hence the discourse information is restricted to those kinds that can be reconstructed from that structure.

Moreover, the above mentioned systems assume a one-to-one mapping between syntactic structure and intonational features, therefore they cannot account for phenomena where the same syntactic structure can be realized with differing intonation. Assuming that intonation is more than the mere reflection of the surface linguistic form (see [Halliday, 1967, Pheby, 1969]), and further, that intonation is selected to express particular communicative goals and intentions, we argue that an effective control of intonation requires synthesizing from *meanings* rather than word sequences. This is acknowledged in the SYNPHONICS system (see [Abb *et al.*, 1993]), which allows prosodic features to be controlled by various factors other than syntax, e.g., by the information structure (focus-background or topic-comment structure). The function of intonation is, however, still restricted to grammatical function, more specifically *textual* function, without considering *interpersonal aspects*, i.e., communicative goals and speaker attitudes (see [Halliday, 1967]). Yet, in the context of generating speech in information seeking dialogues where intonational features are often the only means to signal a type of dialogue act, these aspects have to be taken into account.

Furthermore, in a dialogue situation it is not sufficient to look at isolated sentences; instead one has to consider an utterance as part of a larger interaction. Intonation is not only used to mark sentence-internal information structures, but additionally it is employed in the management of the communicative demands of interaction partners. Therefore, we also have to consider the function of intonation with respect to the context, taking into account the discourse (dialogue) history (see also [Fawcett *et al.*, 1988]). Intonation as realization of interactional features thus draws on dialogue and user model as sources of constraints.

3 Human-machine dialogue

Although production of linguistic output plays a decisive role in the SPEAK! project, we employ a dialogue model that was originally developed in support of *multimodal* human-machine information seeking dialogue. Still, it was a speech act oriented approach to modelling dialogue.

Its central part, the COR (Conversational Roles) model (cf. [Sitter and Stein, 1992]), regards information seeking dialogue as collaborative action whose basic units are linguistic and non-linguistic communicative acts. COR is a task-independent dynamic model of dialogue and has been combined with a model of global dialogue structure represented by *scripts* for information seeking strategies [Belkin *et al.*, 1995]. We use both scripts and a modified version of the COR model [Hagen and Stein, 1996].

We do not claim that COR is a generic model for all types of genre, however, it does cover all kinds of information seeking dialogues. Focussing on *information* does not exclude *action*. In information retrieval systems, for example, the interactants do not merely exchange information but may also negotiate for actions or goods and services (e.g., searching/filtering, adaptation of presentation forms, editing/storing data). Since they are closely coupled, the COR model does not distinguish between moves for action and moves for information.

Comparable task-independent models of exchange have been developed, for example, by [Bilange, 1991, Fawcett *et al.*, 1988, O'Donnell, 1990, Traum and Hinkelman, 1992]. They are, however, mainly concerned with NL dialogue, and most of them seem to address specific kinds of information dialogue, i.e., exchange of factual information within well defined task-settings (tasks and goals are usually not that well defined in information retrieval interactions).

COR uses recursive state-transition networks as representation formalism, i.e., a *dialogue* network and several *move* networks (for details see [Hagen and Stein, 1996]).[1] Transitions in the dialogue net are moves, e.g., REQUEST, OFFER, REJECT, WITHDRAW, ACCEPT, INFORM, EVALUATE (cf. Table 1). Transitions in a move net may include atomic *acts* (of the same speech act type as the

[1] The representation formalism for the linguistic resources (semantics, grammar) are *system networks*. Since the COR model was developed independently of the KOMET-PENMAN generator, representation not being uniform has simply historical reasons.

superordinate move), complex *moves* (to supply additional information related to the atomic act), and optional sub-*dialogues* for clarification (to seek additional information).

Table 1 includes the definition of types of dialogue moves, and thus the corresponding acts. It also describes the differentiation between initial vs. responding and expected vs. unexpected moves, which are a reflection of speaker–hearer role relations and expectations. The possible sequences of turns and moves can be followed moving through the table across and down.

4 What a generation perspective can offer to speech production in dialogue

In Section 2 we noted that most current speech production systems do not have enough knowledge about the resources constraining intonation.

Language can be described through three sources of meaning: *propositional, textual*, and *interpersonal* (cf. [Halliday, 1985]. The most sophisticated speech generation systems of today use propositional and possibly textual meaning as their only constraint on intonation (cf. Section 2). While this is an improvement of the purely syntactic approaches, they do not suffice; it remains difficult to derive meaning from strings, and since the interpersonal aspects are left out, the derived meanings are functionally under-specified. The interpersonal source is, however, the major constraint on intonation, since it brings speech function, speakers' attitudes, hearers' expectations, and speaker–hearer role relations in the dialogue into consideration.

When generating written natural language, it is essential to have knowledge about the *type of discourse or genre* that is to be generated and the *communicative goals* the text shall fulfill. Without this knowledge linguistic-semantic choice and grammatical and lexical expression will be underconstrained and a generation grammar will fail to produce any sensible output. In order to generate spoken language with appropriate intonation, we need access to the same types of knowledge and in particular, we need a source of interpersonal knowledge, otherwise the intonation control is under-specified.

We argue that our dialogue module can provide the missing interpersonal information: The dialogue model itself defines our genre; the moves describe the stages in information seeking dialogues and the ordering of these stages is inherent to the model (see [Hasan, 1978] for a general discussion on genre). The definition of the dialogue moves (see Table 1) together with the dialogue history provide the communicative goals.

In the following sections we discuss in more detail how we employ the dialogue module to constrain semantic and grammatical choice including the selection of intonation in the German version of a Systemic Functional Linguistic (SFL) based generation system, the KOMET-PENMAN text generator [Bateman and Teich, 1995, Penman Project, 1989].

4.1 Intonation as integral part of generation resources

From a *systemic-functional linguistics* viewpoint intonation is just one means among others—such as syntax and lexis—to realize grammatical selection (see [Fawcett *et al.*, 1988, Halliday, 1985, Matthiessen, 1995]). This implies that choices underlying the realization of intonation may be organized in exactly the same way as other choices in the grammar (see [Fawcett *et al.*, 1988, Halliday, 1967]). Hence, the intonational control required for speech generation in a dialogue system has been built into the existing generation grammar for German ([Grote, 1995, Teich, 1992]). The added discriminations—i.e., more delicate choices in those areas of the grammar where intonational distinctions exist—are constraints on the specification of an appropriate intonation rather than constraints on syntactic structure.

There are three distinct kinds of phonological categories, i.e., *tone group, tonic syllable* and *tone* (see e.g., [Bierwisch, 1973, Pheby, 1969]), the determination of which requires a representation of:

- **Tonality:** The division of a text into a certain number of tone groups
- **Tonicity:** The placing of the tonic within the tone group
- **Tone:** The choice of a tone for each tone group

Choices in **tonality** and **tonicity** lead to an information constituent structure independent of the grammatical constituency, whereas choices in **tone** result in the assignment of a tone contour for each identified tone group in an utterance. The choices and hence the assignment of intonation features are constrained by knowledge of the intended textual and interpersonal meaning of the text.

Dialogue Move			Follow-up Move		
name & state transition		definition	expected	unexpected	
request (S,K)	1 $\rightarrow$ 2	S wants (K does A)	promise (K,S) + inform (K,S)	reject-request (K,S) withdraw-request (S,K) sub-dialogue (K,S)	2 $\rightarrow$ 1/7 2 $\rightarrow$ 1/7
offer (K,S)	1 $\rightarrow$ 2'	K intends (K does A)	accept (S,K)	reject-offer (S,K) withdraw-offer (K,S) sub-dialogue (S,K)	2' $\rightarrow$ 1/7' 2' $\rightarrow$ 1/7'
promise (K,S)	2 $\rightarrow$ 3	K intends (K does A*)	inform (K,S)	withdraw-promise (K,S)	3 $\rightarrow$ 1/8
accept (S,K)	2' $\rightarrow$ 3	S wants (K does A*)	inform (K,S)	withdraw-accept (S,K) sub-dialogue (K,S)	3 $\rightarrow$ 1/8
inform (K,S)	3 $\rightarrow$ 4	K believes (P)	evaluate (S,K)	evaluate + quit (S,K) sub-dialogue (S,K)	4 $\rightarrow$ 5
evaluate (S,K)	4 $\rightarrow$ 1	S believes (P)	request (S,K) offer (K,S)	sub-dialogue (K,S)	
reject-request (K,S)	2 $\rightarrow$ 1	K intends [not (K does A*)]	request (S,K) offer (K,S)	meta-dialogue (–,–) withdraw (–,–)	1 $\rightarrow$ 1 1 $\rightarrow$ 6
reject-offer (S,K)	2' $\rightarrow$ 1	S wants [not (K does A*)]	request (S,K) offer (K,S)	meta-dialogue (–,–) withdraw (–,–)	1 $\rightarrow$ 1 1 $\rightarrow$ 6
reject-request (K,S)	2 $\rightarrow$ 7	K intends [not (K does A*)]	— *end* —	sub-dialogue (S,K)	
......	......	......	......	......	

Table 1: COR dialogue moves and role assignments. Notation: Order of parameters: first = speaker; second = addressee; S = information Seeker; K = information Knower; – = S or K; P= Proposition; A = Action (as defined here); $A*$ = A adopted from preceeding move; numbers indicate dialogue states: states 1–4 build a dialogue cycle; states 5-8 are terminal states.

While tonality and tonicity encode textual meaning, it is mostly tone that carries interpersonal meaning. Following [Pheby, 1969], we assume five *primary tones*, plus a number of *secondary tones* that are necessary for the description of German intonation contours. These tones are: *fall* (tone1), *rise* (tone2), *progredient* (tone3), *fall-rise* (tone4), *rise-fall* (tone5), where the first four can be further differentiated into secondary *a* and *b* tones.[2]

The primary tones are the undifferentiated variants, whereas the secondary tones are interpreted as realizing additional meaning. The (primary) tone selection in a tone group serves to realize a number of speech functional distinctions. For instance, depending on the tone contour selected, the system output *//sie wollen um fünfzehn uhr fahren//* ("You want to leave at 3 pm.") can be either interpreted as a question (tone **2a**) or a statement (tone **1a**). The (secondary) tone is conditioned by attitudinal options such as the speaker's attitude towards the proposition being expressed (surprise, reservation ...), what answer is being expected, emphasis on the proposition etc., referred to as *key*. Consider the following example taken from an information seeking dialogue: The computer has retrieved an answer to a query, and this answer is presented graphically to the user. As a default, the system would generate a neutral statement choosing tone **1a** to accompany the presentation, as in //**1a** *die ergebnisse sind unten DARgestellt//*[3] ("The results are given below"). If, however, the results had so far been presented at a different position on the screen, the system would generate tone **1b** in order to place special emphasis on the statement: // **1b** *die ergebnisse sind UNTEN dargestellt//*.

The interpersonal meaning encoded by tone is speech function (statement, question, command, offer) in the first instance and speaker's attitude in the second instance. In the grammar, these are realized in *mood* and *key*. In SFL the *key* options are refinements of *mood* and have accordingly been implemented as an integral part of the grammar.

[2]The criteria for the distinction of primary tones is the *type* of the tone movement, for instance rising or falling tone contour, whereas the *degree* of the movement, i.e., whether it is strong or weak in expression, is considered to be a variation within a given tone contour.

[3]In this paper, the following notational conventions hold: // marks tone group boundaries, CAPITAL LETTERS are used to mark the tonic element of a tone group. Numbers following the // at the beginning of a tone group indicate the type of tone contour. The MULTIVOX text-to-speech system has been extended to interpret these notational conventions [Olaszy *et al.*, 1995].

4.2 Integrating NL generation, dialogue and speech

With the intonational resources in place as described in the preceding section, we now proceed to our proposal of integrating generation, dialogue, and speech.

It is generally acknowledged that intonation is in the first instance the realization of interpersonal-semantic *speech function* (command, statement, question, offer). In a generation perspective, speech function meanings are accomodated on the semantic stratum of the generation resources and related to the grammar by a realization relation. Speech function is typically recoded grammatically in sentence mood (declarative, interrogative, imperative). However, the relation between speech function and mood is not one-to-one, and the speech function 'offer' has no prototypical grammaticalization at all. Imperative, interrogative, and declarative mood may encode a command, for example, *Schließ die Tür!* ("Close the door!"), *Würdest Du die Tür schließen, bitte?* ("Would you close the door, please?"), *Die Tür soll geschlossen werden.* ("The door should be closed."). This means that the mapping between speech function and mood must be constrained by some other, additional factors—and so must be the mapping to the *key* options opened up by mood selection which are realized in intonation (tone).

When we look at dialogue — rather than monologue — these other factors that come into play are the *type of dialogue move* to be verbalized and the *history of the dialogue* taking place. For instance, the genre of information-seeking, human-machine dialogues is characterized by certain genre-specific types of moves (see Section 3). A typical move in this genre is the REQUEST move. In terms of speech function, a REQUEST is often a question. The REQUEST–question correlation in the kind of dialogue we are dealing with here constrains the choice of mood to *interrogative* or *declarative*, e.g., (1) *Wohin möchten Sie fahren* ("Where do you want to go") (interrogative)— (2) *Sie wollen um drei Uhr fahren* ("You want to go at three o'clock") (declarative). So, in information-seeking dialogues, the type of move largely constrains the selection of speech function, but it only partially constrains the mapping of speech function and mood.

Deciding between declarative and interrogative as linguistic realization of a move REQUEST requires information about the immediate *context* of the utterance, i.e., about the dialogue history. It is in the dialogue history that speaker's attitudes and intentions and hearer's expectations are implicitly encoded. The area in the grammar recoding this kind of interpersonal information is *key*. The *key* systems are subsystems of the basic *mood* options (see Section 4.1) and realized by tone. Consider the contexts in which (1) or (2) would be appropriate: (1) would typically be used as an *initiating* move of an exchange, where there is no immediately preceding context—the speaker's attitude is essentially neutral and tone 1 is appropriate. (2) would typically be used in an exchange as the realization of a *responding* move; in terms of the dialogue model, (2) would be a possible realization of a REQUEST in the context of an INFORM or REQUEST—the speaker wants to make sure she has understood correctly and tone 2 is the appropriate intonation.

For the representation of constraints between *dialogue move* and *speech function* on the side of interpersonal semantics and *mood* and *key* on the part of the grammar, the type of dialogue move *in context* (the dialogue history) suggests itself as the ultimate constraint on tone selection. Given that all of the parameters mentioned (dialogue move type, dialogue history, speech function, mood/key) are logically independent and that different combinations of them go together with different selections of tone, an organization of these parameters in terms of stratification as providing the required flexibility in mapping the different categories appears appropriate. Such an organization is for instance proposed in systemic functional work on interaction and dialogue, but without considering intonation [Berry, 1981, Martin, 1992, Ventola, 1987].

In the systemic functional model, the strata assumed are context (extra-linguistic), semantics and grammar (linguistic). On the semantic stratum, general knowledge about interactions is located, described in terms of the *negotiation* network (cf. [Martin, 1992]). A pass (or passes, since *negotiation* is recursive) through the network results in a syntagmatic structure of an interaction called *exchange structure*. An exchange structure consists of *moves* which are the units for which the *speech function* network holds. *Negotiation* and *speech function* are the two ranks of the stratum of interpersonal semantics. The *mood* and *key* systems represent the grammatical realization of a move (given that a move is realized as a clause).[4] The ultimate constraint on the selection of features in the interpersonal semantics and grammar is the information located at the stratum of context. This is knowledge about the type of discourse or *genre*. In the present scenario, this contextual knowledge is provided by the dialogue model, reflecting the genre of *information-seeking, human-machine dialogue*. Since the stratum of context is extra-linguistic, locating the dialogue model—which has originally not been designed to be a model of *linguistic* dialogue, but of retrieval dialogue in general—here is a straightforward step.

[4]A similar approach as developed here for the scenario of human-machine dialogue integrating speech has been suggested by [Fawcett, 1990].

For a graphical overview of the stratified architecture we just described briefly see Figure 1.

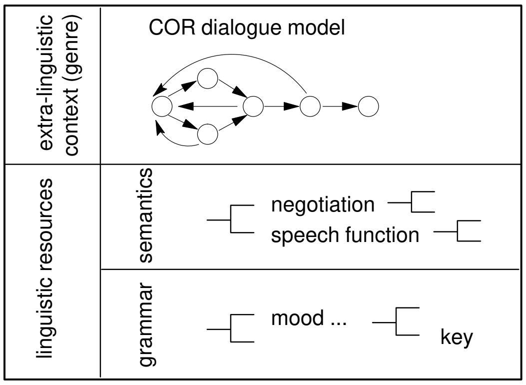

Figure 1: The stratified model.

The general relation that holds between the strata is that of *realization*, which is implemented in KOMET-PENMAN-style generation as *inter-stratal preselection* [Bateman and Teich, 1995] and as *chooser-inquiry interface* [Matthiessen, 1988]. The interstratal mappings can then be flexibly formulated depending on the dialogue history and the classification of dialogue moves (see Table 1), hence choice of intonation is sufficiently constrained.

5 Conclusions

In this paper we have argued that one of the shortcomings of state-of-the-art speech production is the selection of appropriate intonation contours. We have pointed out that at the source of the problem is the analysis orientation of most speech production research. We suggested that it is worthwhile to consider the problem of intonation selection from the perspective of natural language generation and more seriously attempt an integration of natural language and speech processing. Taking such a perspective, we have presented a speech production approach that is embedded in a systemic-functionally based generation system used for the generation of spoken output in a human-machine dialogue scenario.

There is a number of open issues that we could not touch upon here which we are currently investigating. These include:

- accounting also for the *textual* meanings of intonation, such as information structure and thematic progression, which are reflected in the selection of the tonic element;
- carrying out more extensive analyses of human-machine dialogue in order to validate the model;
- trying to adapt some of the text planning mechanisms we are currently working on for

monologic text in the KOMET-PENMAN system for the mapping between the contextual and the semantic level in our dialogue system;

- empirically backing up the intonational classification ([Pheby, 1969]) by acoustic analyses and auditive experiments, and, if necessary, revising it.

References

[Abb *et al.*, 1993] B. Abb, C. Günther, M. Herweg, C. Maienborn, and A. Schopp. Incremental syntactic and phonological encoding – an outline of the synphonics formulator. In *Proceedings of the Fourth European Workshop on Natural Language Generation*, pages 19–29, 1993.

[Bateman and Teich, 1995] J. A. Bateman and E. Teich. Selective information presentation in an integrated publication system: an application of genre-driven text generation. *Information Processing and Management: an international journal; Special Issue on Summarizing Text*, 31(5):753–768, September 1995.

[Belkin *et al.*, 1995] N. Belkin, C. Cool, A. Stein, and U. Thiel. Cases, scripts, and information seeking strategies: On the design of interactive information retrieval systems. *Expert Systems and Applications*, 9(3):379–395, 1995. Also available as: Arbeitspapiere der GMD, No. 875, Sankt Augustin, Germany, 1994.

[Berry, 1981] M. Berry. Systemic linguistics and discourse analysis: a multi-layered approach to exchange structure. In Malcolm Coulthard and Michael Montgomery, editors, *Studies in Discourse Analysis.* Routledge and Kegan Paul, London, 1981.

[Bierwisch, 1973] M. Bierwisch. Regeln für die Intonation deutscher Sätze. In *Studia Grammatica VII: Untersuchungen über Akzent und Intonation im Deutschen*, pages 99–201. Berlin: Akademie Verlag, 1973.

[Bilange, 1991] E. Bilange. A task independent oral dialogue model. In *Proceedings of the European Chapter of the ACL 1991*, pages 83–87, 1991.

[Black and Campbell, 1995] A. W. Black and N. Campbell. Predicting the intonation of discourse segments from examples in dialogue speech. In P. Dalsgaard, L.B. Larsen, L. Boves, and I. Thomsen, editors, *Proceedings of the ESCA Workshop on Spoken Dialogue Systems; Theories and Applications*, pages 197–200, 1995.

[Dorffner *et al.*, 1990] G. Dorffner, E. Buchberger, and M. Kommenda. Integrating stress and intonation into a concept-to-speech system. In *Proc. of the 14th Intl. Conf. on Computational Linguistics (COLING'90)*, pages 89–94, 1990.

[Fawcett *et al.*, 1988] R.P. Fawcett, A. van der Mije, and C. van Wissen. Towards a systemic flowchart model for discourse. In R.P. Fawcett and D. Young, editors, *New Developments in Systemic Linguistics*, volume 2, pages 116–143. Pinter, London, 1988.

[Fawcett, 1990] Robin P. Fawcett. The computer generation of speech with discoursally and semantically motivated intonation. In *5th International Workshop on Natural Language Generation, 3-6 June 1990*, Pittsburgh, PA., 1990.

[Grote, 1995] B. Grote. Specifications of grammar/semantic extensions for inclusion of intonation within the KOMET grammar of German. COPERNICUS '93 Project No. 10393 SPEAK!, Deliverable R2.1.1, 1995.

[Hagen and Stein, 1996] E. Hagen and A. Stein. Automatic generation of a complex dialogue history. In *Proc. 11th Canadian Conference on Artificial Intelligence (AI96).* Canadian Society for Computational Studies of Intelligence, 1996.

[Halliday, 1967] M.A.K. Halliday. *Intonation and Grammar in British English.* The Hague: Mouto, 1967.

[Halliday, 1985] M.A.K. Halliday. *An Introduction to Functional Grammar.* Edward Arnold, London, 1985.

[Hasan, 1978] R. Hasan. Text in the systemic-functional model. In W. Dressler, editor, *Current Trends in Text Linguistics*, pages 228–246. de Gruyter, Berlin, 1978.

[Hemert *et al.*, 1987] J. P. Hemert, U. Adriaens-Porzig, and L. M. Adriaens. Speech synthesis in the SPICOS project. In H.G. Tillmann and G. Willee, editors, *Analyse und Synthese gesprochener Sprache. Jahrestagung der GLDV*, pages 34–39. Hildesheim: Georg Olms, 1987.

[Hirschberg *et al.*, 1995] J. Hirschberg, C.H. Nakatani, and B.J. Grosz. Conveying discourse structure through intonation variation. In P. Dalsgaard, L.B. Larsen, L. Boves, and I. Thomsen, editors, *Proceedings of the ESCA Workshop on Spoken Dialogue Systems; Theories and Applications*, pages 189–192, 1995.

[Hirschberg, 1992] J. Hirschberg. Using discourse context to guide pitch accent decisions in synthetic speech. In G. Bailly and C. Benoit, editors, *Talking machines: Theory, Models and Design*, pages 367–376. Amsterdam: North Holland, 1992.

[Huber *et al.*, 1987] K. Huber, H. Hunker, B. Pfister, T. Russi, and C. Traber. Sprachsynthese ab Text. In H.G. Tillmann and G. Willee, editors, *Analyse und Synthese gesprochener Sprache. Jahrestagung der GLDV, 1987*, pages 26–33. Hildesheim: Georg Olms, 1987.

[Martin, 1992] J.R. Martin. *English text; System and structure*, chapter 7, pages 493–573. John Benjamins Publishing Company, Philadelphia/Amsterdam, 1992.

[Matthiessen, 1988] C. M.I.M. Matthiessen. Semantics for a systemic grammar: the chooser and inquiry framework. In James D. Benson, Michael Cummings, and William S. Greaves, editors, *Linguistics in a Systemic Perspective.* Benjamins, Amsterdam, 1988. Also available as USC/Information Sciences Institute, Reprint Series Report ISI/RS-87-189, Marina del Rey, CA.

[Matthiessen, 1995] C.M.I.M. Matthiessen. *Lexicogrammatical cartography: English systems.* International Language Science Publishers, Tokyo, Taipei and Dallas, 1995.

[O'Donnell, 1990] M. O'Donnell. A dynamic model of exchange. *Word*, 41(3):293–327, 1990.

[Olaszy *et al.*, 1992] G. Olaszy, G. Gordos, and G. Nemeth. The MULTIVOX multilingual text-to-speech converter. In G. Bailly and C. Benoit, editors, *Talking machines: Theory, Models and Design*, pages 385–411. Amsterdam: North Holland, 1992.

[Olaszy *et al.*, 1995] G. Olaszy, G. Nemeth, A. Tihanyi, and G. Szentivanyi. Implementation of interface language in the SPEAK! dialogue system. Technical report, 1995. COPERNICUS '93 Project No. 10393 SPEAK!. Deliverable P2.3.1.

[Penman Project, 1989] Penman Project. PENMAN documentation: the Primer, the User Guide, the Reference Manual, and the Nigel manual. Technical report, USC/Information Sciences Institute, Marina del Rey, California, 1989.

[Pheby, 1969] J. Pheby. *Intonation und Grammatik im Deutschen.* Akademie-Verlag, Berlin, 1969. (2nd. edition, 1980).

[Prevost and Steedman, 1994]
S. Prevost and M. Steedman. Specifying intonation from context for speech synthesis. *Speech Communication*, 15(1–2):139–153, 1994. Also available as http://xxx.lanl.gov/cmp-lg/9407015.

[Sitter and Stein, 1992] S. Sitter and A. Stein. Modelling the illocutionary aspects of information-seeking dialogues. *Information Processing and Management*, 8(2):165–180, 1992.

[Teich, 1992] E. Teich. Komet: Grammar documentation. Technical Report, GMD/Institut für integrierte Publikations- und Informationssysteme, Darmstadt, 1992.

[Traum and Hinkelman, 1992]
D. Traum and E. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599, 1992.

[Ventola, 1987] E. Ventola. *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters.* Frances Pinter (publishers), London, 1987.