INCREMENTAL FOCUS-ACCENT-REALISATION IN SYNPHONICS

Carsten Günther, Claudia Maienborn, Andrea Schopp, Soenke Ziesche
University of Hamburg, FB Informatik, AB WSV, D-22527 Hamburg, Germany

ABSTRACT

The SYNPHONICS system represents a new, incrementally based approach towards the prosodic focus-accent realisation within the framework of a cognitively motivated concept-to-speech architecture. We want to provide an answer to the question how focus/ background calculation determines an appropriate metrical and tonal planning under recourse to semantic focus information of fragmentary increments.

1. SURVEY OF THE CSS SYNPHONICS

The Concept-to-Speech system SYNPHONICS[1] adopts a cognitive approach to a computational linguistic model of language production that combines results from psycholinguistic research about the time course of human language production with recent developments in theoretical linguistics and phonetics concerning the representation of semantic, syntactic, phonological, and phonetic-articulatory knowledge. The aim of the project consists in developing a system that covers the incremental generation of utterances from pre-linguistic conceptualisations to the formation of phonological structures, which are in turn interpreted phonetically, yielding an articulatorically specified input to a speech synthesis module. The SYNPHONICS system consists of three central processing units: a Conceptualizer, a Formulator (grammatical and phonological encoding), and an Articulator (Figure 1). Linguistic objects and rules are represented as typed feature structures in a formal specification language (ALE, Attribute Logic Engine [2]).

[1] SYNPHONICS is an acronym for Syntactic and Phonological Realization of Incrementally Generated Conceptual Structures, for a detailed description of implementational issues cf. [1].

2. INCREMENTAL COMPUTATION OF INFORMATION STRUCTURE

Among the linguistic phenomena which are analysed within the SYNPHONICS framework, emphasis is placed on investigations concerning the syntactic and prosodic realization of different information structures (e.g. focus/background structure) in accordance with conceptual and contextual variations. We argue that certain meaning distinctions triggered by changes in information structure are reflected by prosodic means without any additional support from syntax [3]. Therefore, within SYNPHONICS, a direct semantics/phonology interface is conjectured in addition to the commonly assumed syntax/semantics and syntax/phonology interfaces. This enables the phonological component to access semantic information directly. We want to provide an answer to the question how focus/ background calculation determines an appropriate metrical and tonal planning under recourse to semantic focus information of fragmentary increments.

Generally, theories of focus/ background structure and their accentual realisation consider whole sentences as the relevant domain of application. From the viewpoint of language processing, however, the sentence level is surely ruled out as primary processing unit. Incremental language production implies that the components of the language production system are enabled to process fragmentary input (so-called increments). Increments pass sequentially through succeeding components, so that each component operates in parallel on a distinct fragment of the input structure.

Under recourse to a conceptual knowledge base, the Conceptualizer of the SYNPHONICS system creates a conceptual structure CS comprising the propositional content of the planned utterance and a contextual structure CT containing the currently

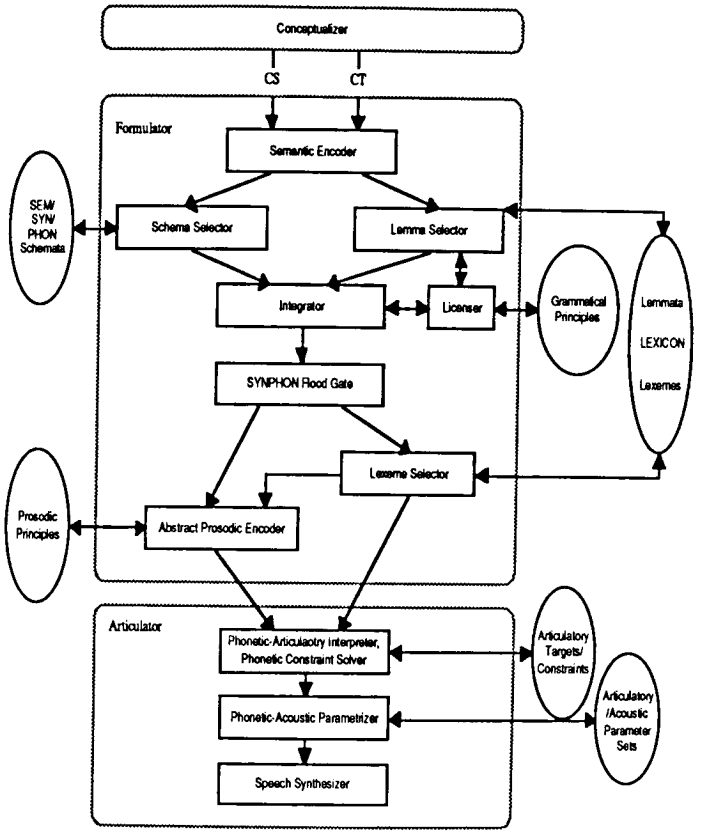

Figure 1. The SYNPHONICS architecture

relevant parts of the contextual environment. Relevant information about conceptual entities is represented in terms of referential objects (*refo*). Within the Formulator, a processing unit called Semantic Encoder generates a genuine linguistic meaning representation SEM from these input structures CS and CT.

The task of computing information structure in terms of focus/background structure is performed at the level of semantic encoding by evaluating how the informational status of the current increment fits into the focus/background structure of the whole utterance. We argue that the only information needed, besides information about the focus/background structure of the increment itself, is information whether a focused increment is part of a larger focus domain or not. A solution for the issue of this determining can be provided by the notion of context established in SYNPHONICS. The context representation can be seen as expressing the informational demand, the speaker wants to fulfil with his utterance. All contextual parameters relevant for the actual utterance are collected into the context representation CT. An adequate utterance has to meet these contextual requirements.

During semantic encoding, each increment is checked whether the information supplied fulfils the informational demand expressed by CT – in this case it belongs to the focus of the utterance – or wether it pertains to the

part of CT mutually known by speaker and hearer, i.e. the background. In the focus case, we further need to determine whether the currently processed increment fulfils the utterance's informational demand exhaustively or only partially. In the former case, we are dealing with narrow focus, in the latter the information expressed by the increment is part of a larger focus domain, i.e. we are dealing with wide focus. In case, CS directly contradicts the actual context CT, contrastive focus is assigned. The result of the computing algorithm consists in the classification of the corresponding restriction elements of the increment's semantic representation as *widely_focused, narrowly_focused, contrastively_focused,* or *non_focused* (i.e. *background*).

3. INCREMENTAL PROSODIC REALISATION OF FOCUS STRUCTURE

Differences in focus/background partitioning of semantic representation trigger different phonetic realisations by prosodic means. In German, focus type information is prosodically marked essentially by a F0-movement, the pitch accent, but also by lengthening and an intensity peak. The different focus structures cause different accent patterns. An abstract prosodic planning process interprets focus type information into an abstract prosodic feature representation (in terms of metrical pattern and accent tones) which is transformed into concrete tonal, durational and intensity parameters. Dealing with incrementallity at the processing stage of focus realisation, the complete Focus Domain (even in the case of wide focus) has not necessarily to be exhaustively specified and only partial syntactic tree structures are accessible for accentual planning.

In the following, focus realisation rules are presented that cover the determination of prominence degrees of constituents in the case of different Focus Domain sizes. These rules are variants of one general focus realisation rule and refer to different structural conditions. Due to lack of space, we neglect the prosodic interpretation of narrow and contrast focus and sketch solely the accent realisation of a wide Focus Domain. The realisation of wide focus turns out to be a more intricate problem since global structural knowledge has to be taken into account at large. The complete Focus Domain is expressed by one nuclear accent, but the exact accent placement depends on semantic and syntactic conditions. Generally speaking, in case of wide focus on VP level (focusing of the verb and its complements and adjuncts), focus is realised on the verb adjacent complement or, in case of the occurrence of a verb adjacent adjunct, on the verb itself. Preceding constituents (either complements or adjuncts) of the Focus Domain carry phrasal accent (secondary stress). Thus, a focus-accent mapping that proceeds incrementally has to check whether the current increment is situated in a verb adjacent position or not. Therefore, in (1), the non-verb-adjacent argument *Peter* is assigned phrasal accent.

(1) Maria hat [PETER das BUCH gegeben]$_F$.

(Mary has given the book to Peter.)

The rule in Figure 2 licenses the assignment of phrasal accent (*phras_acc*) to widely focused non-verb-adjacent complements. In case that a verb-adjacent constituent is selected at the Syntax-Phonology interface it has to be checked whether this constituent is a complement or an adjunct. Example (1) illustrates that verb-adjacent complements (*das Buch*) carry nuclear accent. The rule in Figure 3 (usually named *Focus Projection Rule*)

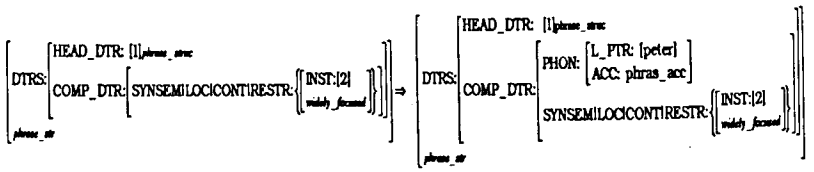

Figure 2. Phrasal-Accent Rule for non-verb-adjacent complements

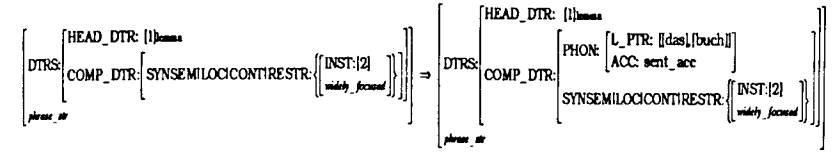

Figure 3. Sentence-Accent Rule for verb-adjacent complements

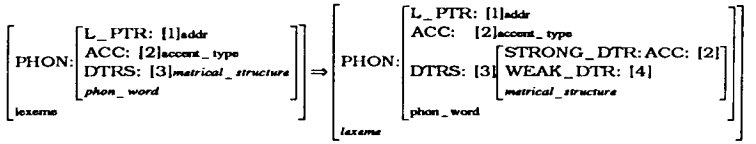

Figure 4. Accent Percolation Rule

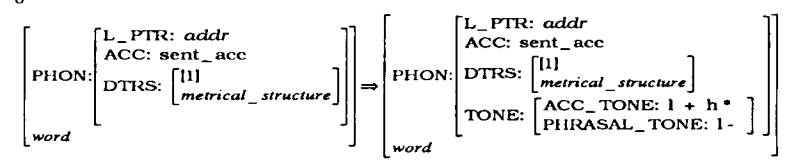

Figure 5. Tonal Sentence Accent Rule

determines the verb-adjacent focus exponent and assigns the sentence accent.

In case of widely focused verb-adjacent adjuncts (cf. 2), sentence accent is realised on the verbal head. This is ensured by a special rule set that assigns phrasal accent to the adjunct and sentence accent to the verb.

(2) Maria hat [nachts GESCHLAFEN]$_F$

(Mary has slept at night.)

The subsequent prosodic encoding and its phonetic interpretation (cf. Figure 1) in terms of articulatory and acoustic parameter settings ensures the appropriate acoustic realisation of information structuring. E.g., the application of the Accent Percolation Rule (Figure 4) licenses the realisation of the sentence or phrasal accent on the word accent bearing syllable, the designated terminal element (DTE). Prosodic rules operate on structures of the abstract data type *metrical tree*.

On the designated syllable, the prenuclear and nuclear accent is realised as bitonal (e.g. sentence accent: L+H*; contrastive accent: L*+H) (Figure 5) or monotonal (phrasal accent: H*) accent tones. The assignment of the phrasal tone L- reflects the end of a Focus Domain.

Phonetic-acoustic interpolation rules subsequently parametrise the F0-contour on the accent-tone adjacent syllables. For synthesizing speech, a Klatt-based *Formant Syntesizer* algorithm (TU Dresden) is applied.

ACKNOWLEDGEMENT

The research reported in this paper is carried out in a research project which is funded by the German Science Foundation (DFG) within the research program of Cognitive Linguistics under grant no. Ha 1237/4-3.

REFERENCES

[1] Günther, C. (ed.,1994), *Hamburger Arbeitspapiere zur Sprachproduktion VI*, University of Hamburg.

[2] Carpenter, B. (1992): *The Logic of Typed Feature Structures*. Cambridge: Cambridge Univ. Press.

[3] Günther, C., C. Maienborn, A. Schopp (1994): *The processing of information structure in SYNPHONICS*. In: P.Bosch & R. van der Sandt (eds.): *Proc. of the interdisciplinary conference of the 10th anniversary of the journal of semantics "Focus & natural language processing"*. Schloß Wolfsbrunnen.