2 The Processing of Information Structure

CARSTEN GÜNTHER, CLAUDIA MAIENBORN,
AND ANDREA SCHOPP

Abstract

This chapter sketches an integrated view on processing information structure in a cognitively motivated computational linguistic model of language production. The approach extends from computing the information structure of an utterance out of a conceptual structure and a relevant contextual embedding to its corresponding prosodic realization. Apparently conflicting requirements of structurally oriented theories of information structure, on the one hand, and incrementality as one major property of human language processing, on the other hand, are shown to be reconcilable within the language production system advocated here.

1 Introduction

Within theoretical linguistics, information structure is generally seen as a phenomenon that concerns propositional units. Current theories of information structure consider whole sentences as the relevant level at which relational notions such as focus/background structure and topic/comment structure are determined (cf., e.g., Jacobs 1992, this volume). From the viewpoint of language processing, however, the sentence level is surely ruled out as a primary processing unit. Research in psycholinguistics as well as in computational linguistics has shown that incrementality is an essential property of efficient language processing (cf., e.g., Kempen and Hoenkamp 1987, Levelt 1989). This means that the components of a processing system for language production or understanding are enabled to process fragmentary input (so-called increments) rather than processing only complete input structures. While increments pass sequentially through succeeding processing components, each component operates in parallel on a distinct fragment of the input structure.[1]

Thus, the question arises of how the declarative view on information structure taken within theoretical linguistics can be brought into accord with a procedural model of language processing pursued within psycholinguistics. That is, how is information structuring performed under the circumstances of incremental language

The work reported on here was carried out in a research project funded by the German Science Foundation (DFG) within the research program of Cognitive Linguistics under grant no. Ha 1237/4. We would like to thank Bernd Abb, Uta Arnold, Ingo Schröder, and Soenke Ziesche for their substantial contributions to the design and implementation of the SYNPHONICS system.

processing? An answer to this question might allow us to draw some conclusions about the validity of both theoretical linguistic models of information structure as well as psycholinguistic models of language processing.

In this chapter, we concentrate on focus/background structure, that is, the dimension of information structure that divides a proposition into an informationally more relevant (new) part, that is, the *focus*, and a less relevant part that is mutually known by the speaker and the hearer, that is, the *background*. The goal of the chapter is to provide a procedural account of focus/background structure within language production.[2] This comprises the issue of generating a focus/background structure in the course of building up the propositional content of an utterance and its subsequent grammatical realization. In a language like German, which we have chosen as output language, grammatical realization of focus/background structure is accomplished primarily by prosodic means. As concerns the semantic foundation of focus/background structure, emphasis will be placed on reconstructing the formal notion of *alternative sets* referred to in theoretical linguistic approaches to the semantics of focus since the work of Rooth (1985) in cognitive linguistic terms that are suitable for a psycholinguistically biased model of language production.

The computational linguistic model of language production proposed here is being developed within the SYNPHONICS project (*Syn*tactic and *Phon*ological realization of *I*ncrementally Generated *C*onceptual *S*tructures), which aims at linking psycholinguistic insights with well established assumptions in theoretical linguistics. In Section 2, we shall give a brief overview of the SYNPHONICS system with special emphasis on the basic foundations that are relevant for our present topic. Section 3 elaborates our account of incrementally computing focus/background structure. It will be shown how a global focus/background structure for a propositional unit is built up under recourse to local focus information of single propositional fragments. Section 4 deals with the incremental realization of information structuring in terms of accent assignment. Both processing steps, semantic computation as well as phonological realization of focus/background structure, will turn out to benefit crucially from independently motivated properties of the SYNPHONICS system.

For expository reasons, we will use rather simple examples in order to illustrate the procedural approach to the different configurations of information structure that are of interest here (viz. wide focus, narrow focus, contrastive focus, and double focus). The general account claims to have a broader coverage, though. It has proved to provide an adequate basis for dealing with more intricate cases of focus/background structure, which can be found, for instance, in the domain of adjuncts too (cf. Maienborn 1994).

2 Overview of the Language-Production System SYNPHONICS

The SYNPHONICS project 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 concerning the representation of semantic, syntactic, phonological, and phonetic knowledge. The aim of the project consists in developing a system that covers the incremental generation of utterances from prelinguistic conceptualization to the formation of phonological structures, which are in turn interpreted phonetically, yielding an articulatorically specified input to a speech synthesis module. Among the linguistic phenomena that are analyzed within the SYNPHONICS framework, emphasis is placed on investigations concerning the syntactic and prosodic realization of information structures that vary in accordance with conceptual and contextual variations. We argue in particular that certain meaning distinctions triggered by changes in information structure are reflected by prosodic means without any additional support from syntax (cf. Günther et al. 1993). Therefore, SYNPHONICS conjectures a direct semantics/phonology interface in addition to the commonly assumed syntax/semantics and syntax/phonology interfaces. This enables the phonological component directly to access semantic information.[3]

The grammar formalism used for encoding declarative linguistic knowledge is HPSG.[4] Its multidimensional constraint-based architecture provides suitable means for the representation of the dependencies among semantic, syntactic, and phonological structure that are crucial for an account of information structure. Conceptual and linguistic objects are represented formally as typed feature structures in ALE (Attribute Logic Engine, cf. Carpenter 1992).[5] Figure 2.1 gives an overview of the architecture of the SYNPHONICS system with its three central processing units: the Conceptualizer, which plans the conceptual representation of an intended utterance as preverbal message; the Formulator, which encodes the preverbal message in terms of grammatical structure; and the Articulator, which finally generates a speech signal. For a detailed description of the SYNPHONICS system compare Herweg (1992), Schopp (1993, 1994), Günther (1994a), and Abb et al. (1995).

The Conceptualizer operates on a conceptual knowledge base (CKB), which comprises facts and rules representing so-called world knowledge as well as episodic knowledge corresponding to a particular scene representation. One of the peculiarities of the SYNPHONICS system as opposed to other language production systems (cf., e.g., Levelt 1989) is that it acknowledges the influence that the context has on an utterance by distinguishing analytically two parts of the preverbal message: The Conceptualizer creates a conceptual structure (CS) that comprises the propositional content of the planned utterance and a contextual structure (CT) that contains the currently relevant parts of the contextual environment. Our central claim with respect to the notion of context is that context should not be viewed as just a collection of discourse information, monotonically increasing while discourse develops but rather as a result of an active construction process that selects only the relevant pieces of information according to the intended utterance (cf. Herweg and Maienborn 1992, Günther et al. 1993 for a discussion of this topic). We therefore favor a dynamic and selective view of context instead of a uniform allocation of the whole discourse information. The SYNPHONICS architecture reflects this view by assuming a bipartite output stream of the conceptualization process, CS and CT.[6]

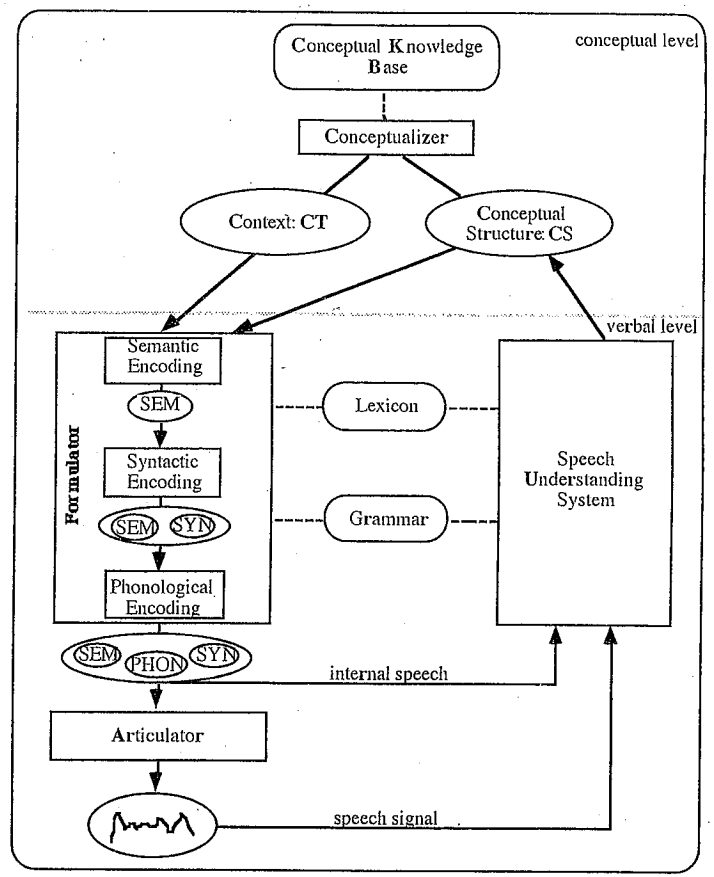

Fig. 2.1. The SYNPHONICS architecture

It is worthwhile to add that CS and CT comprise conceptual representations of the same formal type. Therefore, each conceptual entity might be assigned dynamically either to CS or to CT in the course of conceptual planning. Figure 2.2 shows a (slightly simplified) schematic representation of a corresponding conceptual entity.

$$\begin{bmatrix} \text{concpred: } \{\text{pred1}\} \\ \text{r\_pointer: r1} \\ \text{rel\_set: } \{\} \\ \textit{object\_refo} \end{bmatrix}$$

Fig. 2.2. Schematic representation of a referential object

Relevant information about conceptual entities is represented in terms of referential objects (*refo*), which are subclassified further into the subtypes *sit-refo* and *object-refo*, corresponding to the ontological distinction of situations and objects. Conceptual entities are characterized by referential, sortal, and relational information. The attribute *r-pointer* fixes the referential status of a refo, *concpred* (*conceptual predicate*) supplies sortal information, and *rel-set* specifies relational information about the refo in terms of, for example, thematic linkings to other *refos* relevant for the current utterance. CS and CT each are built up by a collection of *refos*.

Within the Formulator, a processing unit called "Semantic Encoder" generates a genuine linguistic meaning representation, namely, the semantic structure SEM (cf. Fig. 2.3), from the extralinguistic input structures CS and CT.[7] Within SYNPHONICS, SEM is divided into three major parts: Referential information is collected at a *ref_info* attribute that is mapped onto the partition of the lexicon that contains functional elements (determiner, complementizer, etc.) in the course of lexical access. The descriptive content of a *refo* is accounted for by *the core_info* attribute, which triggers the selection of suitable lemmata. And finally, information about the thematic embedding of a *refo* with regard to the actual CS configuration is collected in an *embed_info* attribute. The embedding information triggers in turn the selection of semantic/syntactic schemata (head-complement schema, head-adjunct schema) that build up the structural environment for the linguistic expression that is associated to a *refo*. In the course of semantic encoding, sortal information of CS is always mapped onto the *core_info* part of SEM, whereas relational information of CS may become part of *core_info* or of *embed_info*, depending on whether the relation expressed turns out to be an integral part of the *refo*'s semantic representation or not. That is, inherently relational expressions, as for instance verbs and relational nouns, assign thematic roles to their environment. The corresponding relational information is therefore mapped onto the *core_info* part. (We use the notion of *x-giver* and its ramifications *agent-giver*, *theme-giver*, and so on, to express this relational dependency.) Nonrelational nouns, on the other hand, have no relational content on their own but are assigned thematic roles (cf. the notion of *x-taker*). The corresponding relational information is thus mapped onto the *embed_info* part.

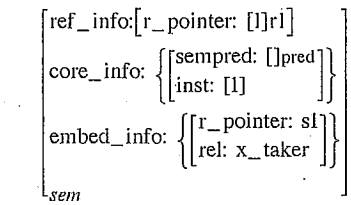

Fig. 2.3. Schematic representation of a semantic structure

SEM is integrated into a complex HPSG sign that is successively augmented by the syntactic structure SYN and the phonological structure PHON, while passing subsequently the corresponding processing units of the Formulator. Finally, the Articulator interprets PHON phonetically and yields an articulatorically specified input to a speech synthesis module.

As we saw in Section 1, incremental language production presupposes that the system components are enabled to process fragmentary input. As soon as a particular component has passed on the results to its successor component, it is ready to process the next incoming increment. This processing mode allows information fragments to be minimal, always provided that the maintenance of the connections between increments is somehow ensured. In SYNPHONICS, this task of ensuring coherence is mainly accomplished by the conception of embedding information as illustrated for the case of thematic relations. Embedding information is – besides referential specification – the major means to guarantee that a coherent linking between refos and, thus, the global shape of a propositional unit are recoverable under the circumstances of processing incrementally fragmentary information. This overall system design has already been shown to cope with a multitude of linguistic phenomena (cf., e.g., Herweg 1992, Schopp 1993, 1994). In the following, it will be applied to the issue of incrementally computing focus/background structure.

3 Incremental Computation of Information Structure

Focus/background structure is a purely linguistic means of signaling the organization of information within a propositional unit. Yet, it originates from genuinely conceptual configurations. Thus, the computation of information structure in terms of focus/background structure must take place at the interface between language independent and language specific processing units. Therefore, this task is performed at the level of semantic encoding, in SYNPHONICS.

There are two subtasks that have to be carried out in order to establish the focus/background structure of an incrementally generated utterance. First, we have to determine the informational status of the increment that is currently being processed. Second, we need some hints about the informational status of the increment's environment, that is, information on how the focus/background structure of the increment fits into the focus/background structure of the whole utterance. As mentioned in Section 1, we are not entitled to assume a complete picture of the planned utterance, according to psycholinguistic evidence. This gives rise to the following questions: What kind of embedding information concerning focus/background structure turns out to be minimally necessary, and what is a psycholinguistically legitimate source for this type of information?

The approach taken in SYNPHONICS assumes that already a minimal amount of information about the environment of an increment suffices to handle focus/

background structure under the circumstances of incremental language production. We argue, in particular, that the only information needed, besides information about the focus/background structure of the increment itself, is information about whether a focused increment is part of a larger focus domain or not. In this section, we will show how this minimal amount of embedding information concerning focus/background structure is computed at the level of semantic encoding in the course of processing every single increment. In the next section, we will demonstrate that exactly this kind of information suffices for incremental realization of focus/background structure at the level of phonological encoding.

A solution for the issue of determining an increment's own informational status (viz. local focus information) as well as its focus/background embedding (viz. global focus information) can be provided by the notion of context established in SYNPHONICS for independent reasons (cf. Section 2). The context representation can be seen as expressing the informational demand that the speaker wants to fulfill with his utterance. This informational demand might originate from the previous discourse, from a question posed explicitly by the hearer, for instance, or it might result from other types of contextual influences, such as perceptions that the speaker wants to communicate to the hearer.[8] In any case, all contextual parameters that are relevant for the actual utterance are collected into the context representation CT. In order to be contextually adequate, an utterance has to meet the contextual requirements expressed by CT. Therefore, the structure of CT may give us some hints about the global shape of the utterance; it may facilitate inferences about the propositional content to be uttered that go beyond the fragmentary information supplied by an increment in isolation. Having identified the primary source for the construction of focus/background structure, let us proceed to the discussion of the concrete algorithm for computing focus/background structure employed in SYNPHONICS.

During semantic encoding, each increment is checked as to whether its information fulfills the informational demand expressed by CT – in this case it belongs to the focus of the utterance – or whether it pertains to the part of CT mutually known by speaker and hearer, that is, the background. In the focus case, we need to determine furthermore whether the currently processed increment fulfills the informational demand of the utterance 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: that is, we are dealing with wide focus. Finally, contrastive focus is given in case CT does not express an informational demand but rather a claim, which might have been established, for instance, by a previous assertion of the hearer. Here, the speaker's utterance is linked to CT by correcting the claim expressed in CT according to the speaker's own beliefs. The result of the computing algorithm consists of the classification of the increment's semantic representation as *widely focused*, *narrowly focused*, *contrastively focused*, or *non-focused* (i.e., *background*). In SYNPHONICS, this differentiation is expressed by a corresponding type of distinction of the semantic restriction

elements. Notice that focus/background structure is thereby accounted for at the level of semantic representation.

Let us go now into some more detail, having a look at some illustrative examples. As we have seen, the Semantic Encoder determines the focus/background structure of an increment by matching it against CT. Let us assume that Figure 2.4 comprises an actual context selection that has to be taken into account by a speaker who plans an utterance. Figure 2.4 expresses that an object referent r1, *Peter*; an object referent r2, of type 'book'; and a situational referent 1, of type 'give' are given in the context. Furthermore, the situational referent establishes certain thematic relations to other referential entities, namely, *agent*, *experiencer*, and *theme*. Agent and theme role, are assigned to r1 and r2, respectively, whereas the experiencer role remains underspecified (*r_var* stands for an object variable).

$$
\text{CT:}\left\{\begin{bmatrix}\text{concpred: \{peter\}}\\ \text{r\_pointer: r1}\\ \text{rel\_set: }\left\{\begin{bmatrix}\text{r\_pointer: s1}\\ \text{rel: agent\_taker}\end{bmatrix}\right\}\\ \textit{obj\_refo}\end{bmatrix},\begin{bmatrix}\text{concpred: \{book\}}\\ \text{r\_pointer: r2}\\ \text{rel\_set: }\left\{\begin{bmatrix}\text{r\_pointer: s1}\\ \text{rel: theme\_taker}\end{bmatrix}\right\}\\ \textit{obj\_refo}\end{bmatrix},\right.
$$

$$
\left.\begin{bmatrix}\text{concpred: \{give\}}\\ \text{r\_pointer: s1}\\ \text{rel\_set: }\left\{\begin{bmatrix}\text{r\_pointer: r1}\\ \text{rel: agent\_giver}\end{bmatrix},\begin{bmatrix}\text{r\_pointer: r2}\\ \text{rel: theme\_giver}\end{bmatrix},\begin{bmatrix}\text{r\_pointer: r\_var}\\ \text{rel: experiencer\_giver}\end{bmatrix}\right\}\\ \textit{sit\_refo}\end{bmatrix}\right\}
$$

Fig. 2.4. Sample context representation

Every context restricts possible utterances to a certain extent. The context representation assumed in Figure 2.4 turns out to be very restrictive, because it delimits precisely the informational demand that the utterance is supposed to fulfill, namely, information about the referent that is assigned the experiencer role. The context representation in Figure 2.4 may thus be seen as expressing the implicit question *To whom did Peter give the book?*[9] A plausible utterance that takes into account these contextual requirements could be an utterance with a propositional content corresponding to *Peter gave Mary the book.* Now, what happens if the first increment reaches the Semantic Encoder? Let us assume, for the time being, that the first increment comprises conceptual information about the topic of the utterance, namely, the referential object r1. Its conceptual structure is given in Figure 2.5.

In the course of semantic processing, the CS fragment is compared with CT. Since it matches exactly one of the CT elements, the conclusion can be drawn that the increment currently processed does not contribute any new information in the

actual context setting but rather supplies background information. The increment's semantic representation is therefore assigned the informational status *non_focused* (cf. Fig. 2.6).

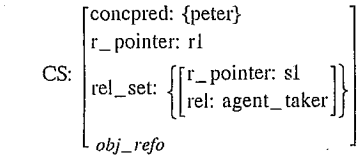

Fig. 2.5. CS increment 'Peter'

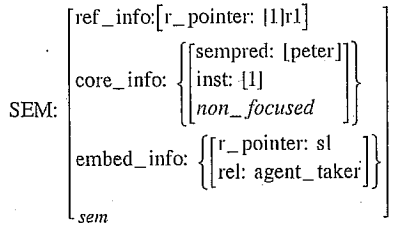

Fig. 2.6. Nonfocused semantic representation

What about the increment comprising conceptual information about the referential object *Mary* (cf. Fig. 2.7)? There is no element in CT that matches this CS fragment. On the contrary, the information supplied by CS fits exactly into one clearly shaped gap within the conceptual specification of s1, namely, the gap corresponding to the lack of information about the bearer of the experiencer role in s1. With respect to the context setting in Figure 2.4, the CS fragment in Figure 2.7 supplies new information that is suitable to fulfill exhaustively a precisely delimited informational demand in CT. Therefore, the increment's semantic representation is assigned the informational status *narrowly focused* (cf. Fig. 2.8).

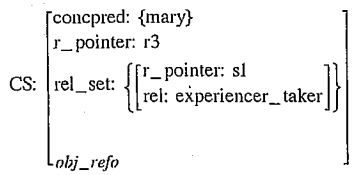

Fig. 2.7. CS increment 'Mary'

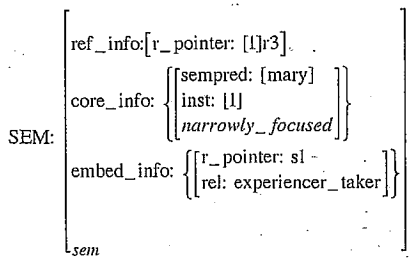

Fig. 2.8. Narrowly focused semantic representation

If the increment supplies more complex information about the bearer of the experiencer role, as in Figure 2.7′ each of the corresponding semantic restriction elements is assigned the informational status *narrowly focused* (cf. Fig. 2.8′).

In general terms, the structural configuration that is decisive for the assignment of narrow focus can be characterized as follows: CT contains a variable that is related to the contextual environment by a thematic link, and CS provides an instantiation of the variable equipped with a thematic linking that is compatible with the relation established in CT. The information that fills up the gap constitutes one single focus domain and thus qualifies as narrow focus. Under this view, multiple focus is

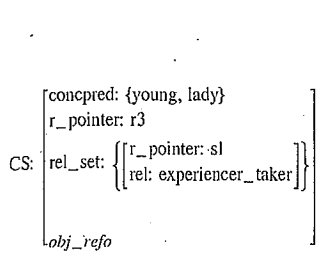

Fig. 2.7′. CS increment 'young lady'

Fig. 2.8′. Narrowly focused semantic representation

simply a phenomenon that originates from a contextual setting with two or more clearly shaped but distinct informational gaps. Each of the increments that supply information about one of the gaps qualifies as *narrowly focused.*

Having discussed so far the issues of background and narrow focus, let us turn now to the apparently more intricate problem of determining incrementally wide focus. In contrast to narrow focus, wide focus originates from contextual configurations that are less restrictive with regard to the propositional content of adequate utterances. If CT raises only few conditions that have to be met by the actual utterance, that is, the informational demand established by CT is more or less underspecified, there is a broader range of possible utterances and each utterance conveys a greater extent of new information, depending on the determinacy of CT. Take, for instance, the context representation in Figure 2.9, which is a slightly changed version of Figure 2.4.

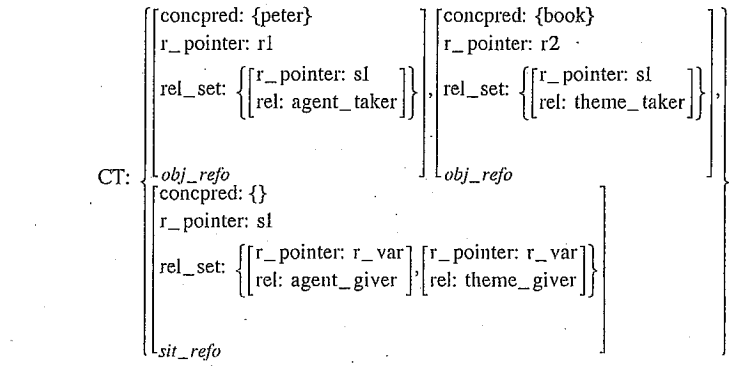

Fig. 2.9. Sample context representation

Again, r1, r2, and s1 are selected as contextually relevant discourse referents. But in this case, there is less information available about s1. The only thing we know about s1 is that it refers to a situation with at least one agent and one theme. The sortal type of the situation and possibly further participants remain,

however, unspecified. That is, Figure 2.9 may be understood as expressing the implicit question *What has Peter done with the book?* The informational demand of this kind of context is not as sharply delimited as in Figure 2.4, because any kind of information may be suitable, provided it meets the minimal contextual requirements concerning referential identity and thematic linking.[10] Let us assume that, within the contextual setting of Figure 2.9, the speaker wants to say that Peter gave the book to Mary, and let us turn immediately to the semantic processing of the increment comprising conceptual information about Mary (cf. Fig. 2.7). This time, the only hint to an adequate integration of CS that is provided by CT concerns referential identity of the situation referents in CT and CS. We are therefore entitled to assume that the fragment supplied by CS constitutes one piece of information that conveys further information about the situation referent and thus is part of the focus of the utterance. In particular, no evidence is given whether the increment's information is the only new information communicated by the speaker, or whether it might be augmented by subsequent increments. Therefore, the corresponding semantic restriction element is assigned the type *widely focused* (Fig. 2.10).

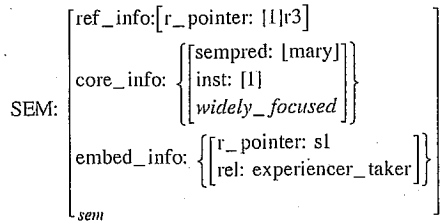

Fig. 2.10. Widely focused semantic representation

If, for instance, a subsequent increment conveys information about the location where the situation takes place, or any other kind of additional information about the situation referent, the same conditions concerning its contextual embedding may be applied. Therefore, the semantic structure is qualified as *widely_focused*, too. The same holds true for an increment conveying sortal information about the situation referent. Thus, to sum up so far: If the informational content of a CS increment does not belong to the background specified in CT and it does not fit an exactly delimited CT gap either, the Semantic Encoder computes a semantic structure of type *widely focused*.[11] In the course of processing one increment after the other, the collection of restriction elements that were assigned wide focus builds up step by step a complete focus domain.

Finally, the structural configurations that trigger the assignment of contrastive focus remain to be accounted for. As we have seen, in the case of contrastive focus there is no informational demand expressed in CT, but rather a claim that the speaker wants to contradict. The refusal of CT may refer to any part of the conceptual information, that is, the referential (*this book vs. that book*), sortal (*book vs. newspaper* or *borrow vs. buy*), or relational part (*Peter gave the book to Mary vs. Mary gave the book to Peter*), whereas the remaining parts of the CS increment are presupposed to match with the corresponding CT opponent. Given for instance the CT in Figure 2.11, which expresses the claim that Peter borrowed the book from Mary, the speaker may want to correct this claim by uttering that it was the case that Peter bought the book from Mary. That is, the propositional content of CS differs from CT only with respect to the sortal information about the situational referent. Consequently, the corresponding semantic restriction element is marked as *contrastively focused* (cf. Figs. 2.12 and 2.13).

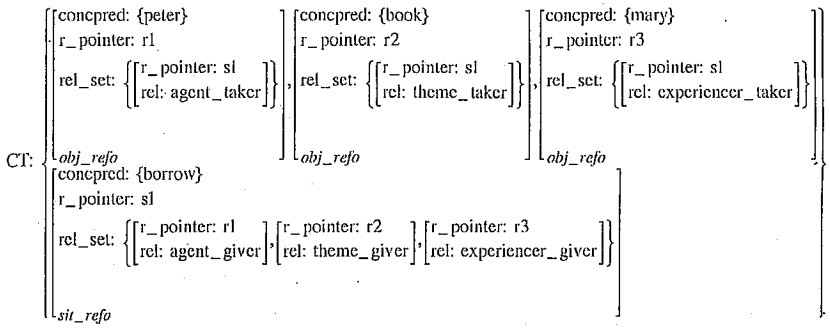

Fig. 2.11. Sample context representation

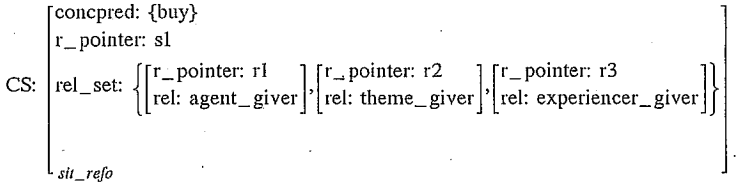

Fig. 2.12. CS increment 'buy'

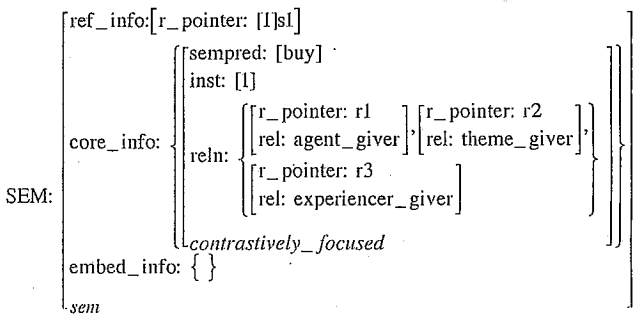

Fig. 2.13. Contrastively focused semantic representation

This completes the survey of the SYNPHONICS approach to the incremental computation of focus/background structure.[12] Local as well as global aspects of focus/background structure are determined by making intensive use of the CS-CT distinction developed in SYNPHONICS for independent reasons. We will now proceed with a sketch of the incremental realization of information structure.

4 Incremental Realization of Information Structure

Differences in focus/background partitioning of a semantic representation trigger different phonetic realizations by prosodic means. Generally speaking, focused constituents are realized more prominently than nonfocused constituents. The concrete phonetic realization of prominence is language and context dependent. In German, for instance, focus is prosodically marked essentially by an outstanding movement of fundamental frequency (F0), that is, the pitch accent, but also by lengthening and an intensity peak. Yet, it is not possible to map semantic focus features directly onto phonetic parameters. An abstract prosodic rule inventory is required, instead, that interprets focus-type information into an abstract prosodic feature representation in terms of accent pattern and accent tones. These abstract prosodic feature values are converted into concrete tonal, durational, and intensity parameters afterwards.

The accent placement and the corresponding tone contour on focused constituents depend on the assigned focus-type information and the argument/modifier status of the verb-adjacent constituent. In the case of narrow focus, the prosodic focus realization is restricted to the focused constituent. Figure 2.14 shows the analysis of the speech signal[13] of the utterance *Er hat den Mimen die MAHNUNG geschickt* (*He has sent the mimes the warning*)[14] as a possible answer to the question *What has the manager sent to the mimes?* With respect to this context setting, the constituent *die Mahnung* is marked as *narrowly_focused*, according to the computation of information structure explained earlier. Thus, it becomes the most prominent constituent of the utterance. Within a narrowly focused constituent, the metrically strongest word of the constituent (*Mahnung*) carries the nuclear accent, which is tonally realized by an initial dip and a following rise of the fundamental frequency (transcribed as a complex accent tone L + H*) on the word accent bearing syllable. The nuclear accent is finally bound by a drop of the fundamental frequency onto the speaker's baseline (transcribed as a phrasal tone L−). If an utterance comprises multiple narrow foci, each focused constituent is assigned a nuclear accent, which will be realized in turn as an L + H* L− tone contour.

In the case of wide focus, on the other hand, the prosodic realization of the information structure yields a more complex accent distribution and tone contour. Figure 2.15 shows the prosodic realization of the same utterance but with a different, less restricted context setting. This utterance is a possible answer to the question *What has the manager done?*, which causes a wide focusing of the complex constituent *[$_F$ den Mimen die Mahnung geschickt]* (*sent the mimes the warning*). The f0-contour in Figure 2.15 shows two pitch accents on the arguments

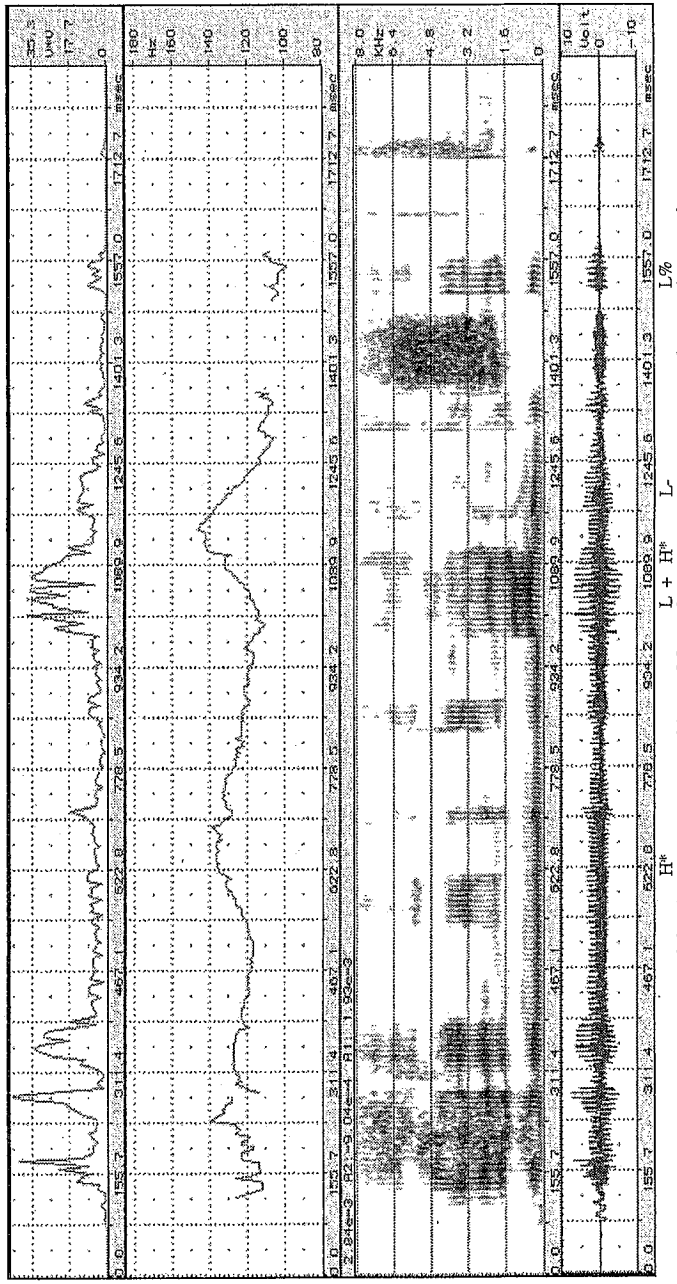

Fig. 2.14. Prosodic narrow focus realization of the utterance *Er hat den Mimen [$_F$ die MAHNUNG] geschickt*

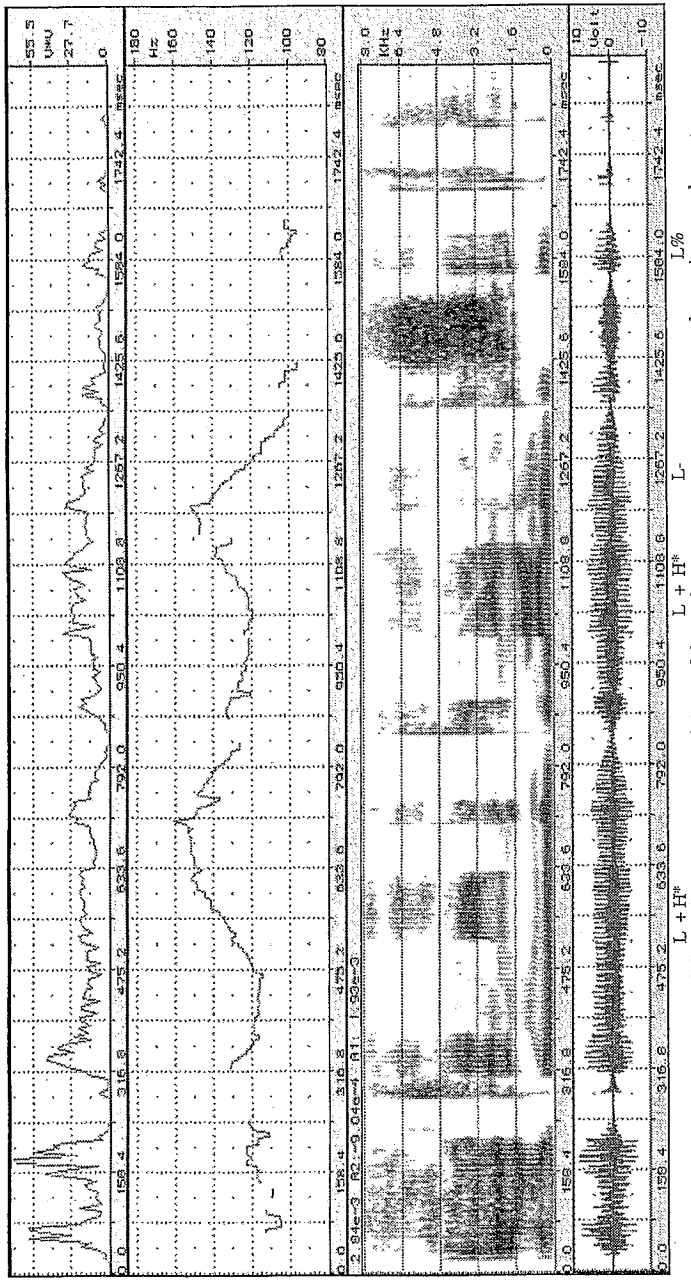

Fig. 2.15. Prosodic wide focus realization of the utterance *Er hat* [$_F$ *den MIMEN die MAHNUNG geschickt*]

of the verb. The second accent is perceived as the stronger because of its late position in the utterance and the immediately following strong fall. Because of this prominence distinction, the last accent will be called *nuclear accent* and the preceding ones *prenuclear accents*. Both accent types are realized tonally by a rising pitch accent L + H*, but the nuclear accent comprises additionally a low phrasal tone L−. Notice that the exact boundaries of the focus domain are not prosodically marked: The verb is part of the focus domain, yet, it does not carry a pitch accent due to its sentence-final position. The underdetermined marking of focus domains is the reason for the phenomenon of so-called focus ambiguities (cf. Jacobs 1991).

The nonaccenting of widely focused verbs in sentence-final position is restricted to argument-verb constellations. Syntactic constructions of verbs with a verb-adjacent adjunct exhibit a different accent pattern (cf., e.g., Jacobs 1991, Maienborn 1994). Figure 2.16 shows the accent distribution in the utterance *Sie hat auf der Wiese geschlafen.* (*She has slept in the meadow.*) Contrary to the accent pattern in Fig. 2.15, the V^0-constituent *geschlafen* carries the nuclear accent (L + H* L−) and the verb-adjacent adjunct *auf der Wiese* the prenuclear accent (L + H*), which is in turn realized on the noun in accordance with general metrical principles.

Generally speaking, in the case of wide focus on VP level (focusing of the verb and its arguments and adjuncts), the nuclear accent is assigned to the verb-adjacent arguments or, in the case of the occurrence of a verb-adjacent adjunct, to the verb itself.[15] Table 2.1 gives a summary of the prosodic realization of different focus-type information. Besides narrow and wide focus, contrastive focus is characterized by its outstanding tonal realization and the f_0-dip immediately on the accent-bearing syllable (transcribed as L* + H).

Table 2.1. *Prototypical prosodic realization of focus in German*

Semantic Focus Type	Accent Type	Pitch Accent
narrow focus	nuclear accent	accent tone: L+H* phrasal tone: L-
wide focus	prenuclear accent	accent tone: L+H*
	nuclear accent (assigned either to the verb-adjacent argument or to the verb itself)	accent tone: L+H* phrasal tone: L-
contrastive focus	contrastive accent	accent tone: L*+H phrasal tone: L-

In the following, we present the incremental derivation of different accent types, based on the semantic focus computation developed in Section 3. Dealing with incrementality as a characteristic processing property with regard to focus realization,

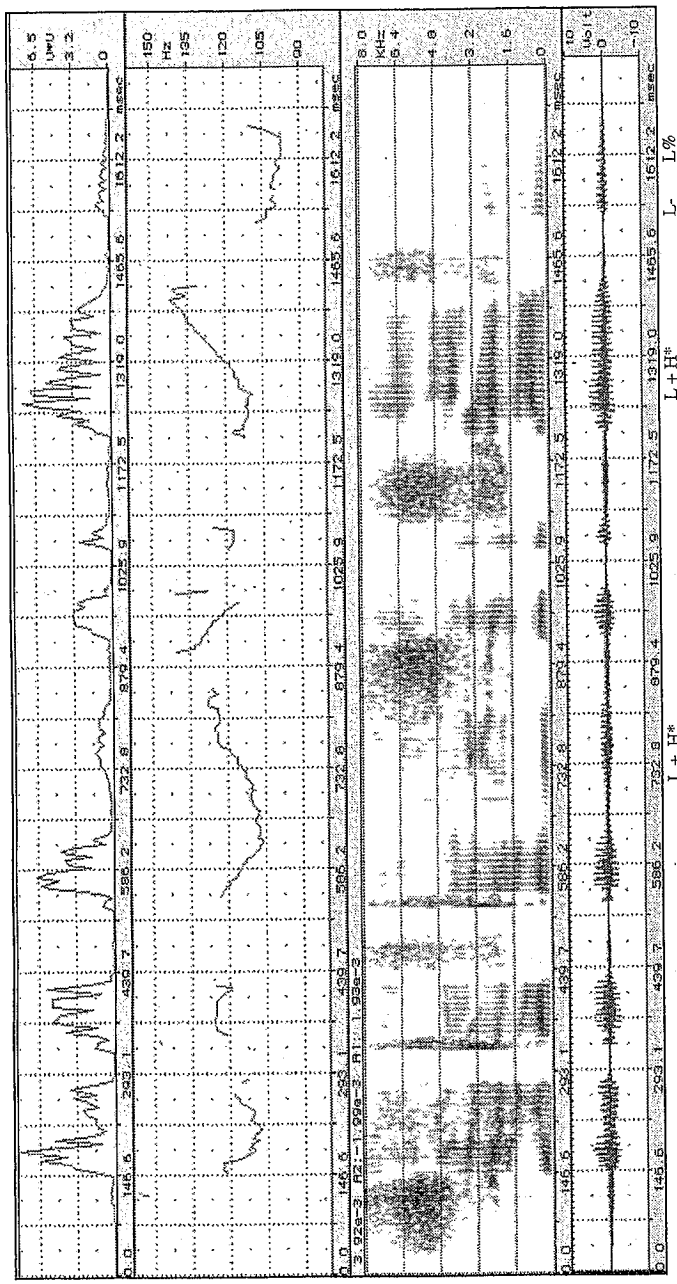

Fig. 2.16. Prosodic wide focus realization of the utterance *Sie hat* [$_F$ *auf der WIESE GESCHLAFEN*]

we have to bear in mind that the focus domain is not necessarily completely specified when single increments are processed.

Within the SYNPHONICS System, abstract prosodic planning of focus realization is part of the phonological encoding, whereas the acoustic parametrization takes place within the phonetic-articulatory planning module after a phonetic interpretation of sub- and suprasegmental structures (cf. Section 2). There is a current controversy about which aspects determine prosodic planning units and how the Phonology-Syntax interface should be designed (indirect or direct reference approach, cf. Inkelas & Zec 1990). The debate centers around the problem of how to ensure a uniform notion of the prosodic structure units either in syntactic terms (maximal phrasal projections according to the definition of Phonological Phrases, cf. Nespor & Vogel 1986) or in mere semantic terms (argument, predicate, modifier structuring, cf. Gussenhoven 1992; sense units, cf. Selkirk 1984). These theories, which might be termed structural approaches, do not take into account procedural aspects of language processing.

According to our incremental approach, we advocate a dynamic view on the Syntax-Phonology interface where structure units and increment size are determined essentially by procedural aspects. At this formulator-internal interface, structure units (single lemmata or constituents) are taken out (from left to right) from the semantic and syntactic structure built up so far, provided that they are exhaustively morpho-syntactically specified (in terms of case, gender, number, person, and tense information) as well as standing in a left-to-right gap-free order. A complete, category-dependent morpho-syntactic specification is necessary for a correct morphophonological generation of word forms (lemmata) to take place. These are the only requirements increments have to fulfill in order to enter the Phonological Encoder. Thus, prosodic increment size turns up as a procedural result of the encoding processes performed so far. In fact, we argue that prosodic planning units are defined in procedural terms (reflected by morpho-syntactic completeness and linear order), rather than in terms of semantic or syntactic constituent structure. Thus we can dispense with an explicit transformation of semantic and syntactic structures into prosodic structures. Such a dynamic view on prosodic planning units reflects the overwhelming variance of speech chunks in natural spoken dialogues.

These rather parsimonious requirements on increment properties of prosodic planning units ensure that prosodic planning takes place without presupposing completeness of focus domains. Such a strict incremental proceeding enables a phonetic realization of fragments of a wide focus domain even in the case that succeeding parts of the utterance are not yet semantically processed. Thus, as in the case of semantic focus calculation, prosodic focus realization can be considered from a local as well as from a global point of view. The local domain covers the focus realization within the actual increment selected by the Syntax-Phonology interface. The global perspective takes into account focus-type information that goes beyond one single increment (narrow or wide focus) as well as the structural information conveyed by the syntactic environment and serves to license the application of focus projection rules.

In the following, we will present focus realization rules that cover the determination of the accent distribution on constituents for different focus domain sizes. These rules are instantiations of one general focus realization principle that take into account different structural conditions and focus-type information. In this sense, these rules manifest the direct link between Semantics and Phonology within SYNPHONICS. They are applied along the verbal projection line during the incremental structure-building process. Prosody internal metrical principles inherit the abstract accent information to the most prominent lexeme (DTE = designated terminal element) of a constituent where the accent is realized by a special tone contour. The different kinds of rules (focus-accent mapping, accent inheritance, accent-tone mapping) represent declarative grammatical constraints that relate different types of grammatical knowledge. They are applied at different time points during the course of language production.

The case of narrow focus assignment on a single constituent does not exhibit a severe problem for an incremental approach to focus-accent mapping because of the locality of rule restrictions and effects. In the example mentioned (see Figs. 2.4–2.8) *Peter hat* [$_F$ *MARIA*] *das Buch gegeben.* (*Peter has given Mary the book.*), the verbal argument *Maria* has been specified as *narrowly_focused*. The appropriate prosodic realization of this focus/background structure is guaranteed by the assignment of the nuclear accent to the narrowly focused constituent (cf. Table 2.1). This relation between semantic and prosodic structure is formulated according to the rule in Figure 2.17. The *Narrow-Focus Accent Rule*[16] checks the focus status of a constituent and assigns to the phonological feature PHON|ACC the value *nuclear_accent* (*nucl_acc*)[17] if this constituent is narrowly focused. The rule evaluates only local semantic-type information (*narrowly_focused*) and does not need any further semantic or syntactic structure information about preceding or following increments.

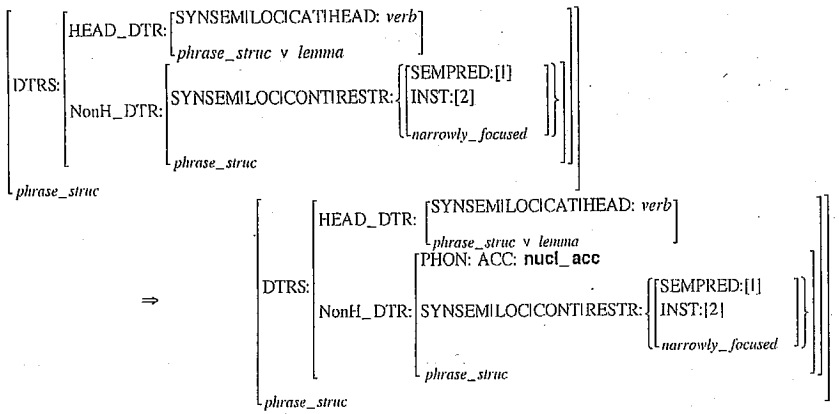

Fig. 2.17. Narrow-Focus Accent Rule

Likewise, the realization of contrastive focus is a local process, which leads, however, to a more prominent realization (contrastive accent, Table 2.1) during the subsequent acoustic realization than in the case of narrow focus (cf. Bartels and Kingston 1994).

The incremental realization of wide focus turns out to be a more intricate problem since global structural knowledge about the information structure of the whole utterance has to be taken into account at large. The complete focus domain is expressed by one nuclear accent and preceding prenuclear accents, but the exact accent placement depends on semantic and syntactic conditions. The nuclear accent will be realized either on the verb-adjacent complement or, in the case of the occurrence of a verb-adjacent adjunct, on the verb itself. Preceding constituents (whether complements or adjuncts) of the focus domain carry prenuclear accents. Thus, a focus-accent mapping that proceeds incrementally has to check first whether the currently processed focused constituent is in a verb-adjacent position. If this is not the case, the constituent will be marked for prenuclear accent. Therefore, in the example *Peter hat* [$_F$ *MARIA das BUCH gegeben]*, the non-verb-adjacent constituent *Maria* is assigned prenuclear accent. The rule in Figure 2.18 licenses assignment of prenuclear accent (*prenucl_acc*) to widely focused non-verb-adjacent constituents (NonH_DTR) by ensuring that the verbal head daughter (CAT:HEAD: *verb*) is not of category V^0.[18] Notice that this rule imposes minimal requirements on global information, ensuring incremental processing. For the prosodic processing of the structural increment *Maria*, it is only necessary to know that this increment is part of a larger focus domain and that the sister constituent is not the verbal head.

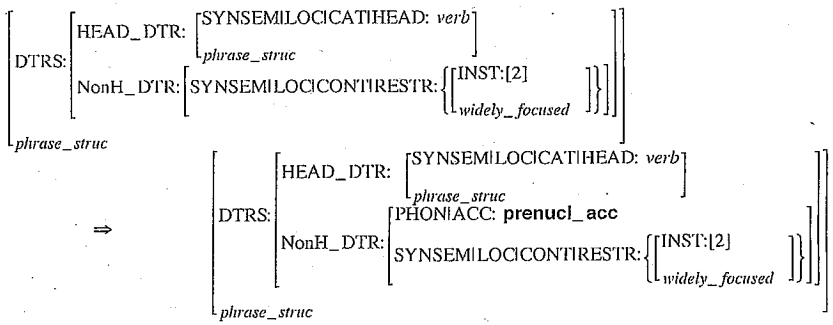

Fig. 2.18. Wide-Focus Accent Rule (for non-verb-adjacent constituents)

In the case that a verb-adjacent constituent is selected at the Syntax-Phonology interface, it must be checked whether this constituent is a complement or an adjunct. Figure 2.15 illustrates that verb-adjacent complements carry the nuclear accent. The rule in Figure 2.19 – usually termed the *Focus Projection Rule* – is applicable to a V^0-complement configuration[19] and assigns to the verb-adjacent focus exponent the nuclear accent (PHON|ACC: nucl_acc).

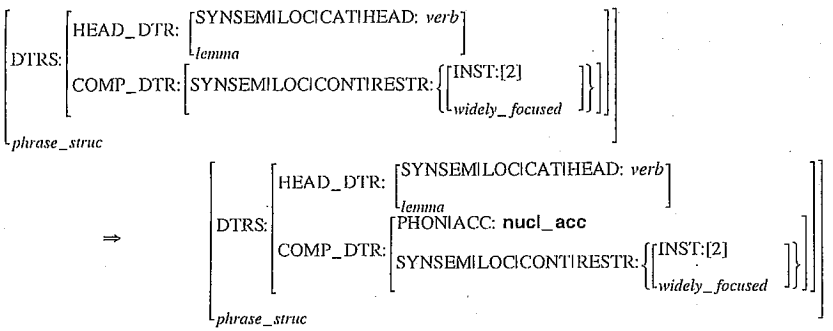

Fig. 2.19. Wide-Focus Accent Rule (for verb-adjacent complements)

In the case of widely focused verb-adjacent adjuncts, no focus projection takes place, and the nuclear accent is realized on the verbal head itself (cf. Fig. 2.16). This is ensured by the rule in Figure 2.20 that assigns a prenuclear accent to the adjunct and a nuclear accent to the verb, respectively.

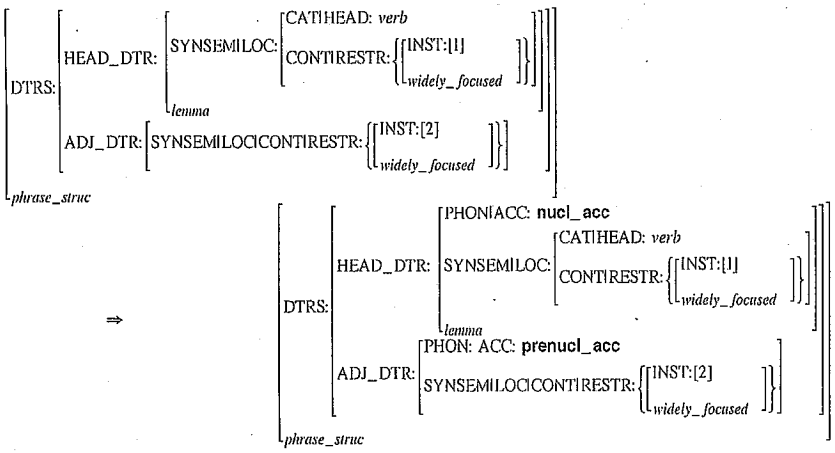

Fig. 2.20. Wide-Focus Accent Rule (for verb-adjacent adjuncts)

The subsequent prosodic encoding and its phonetic interpretation in terms of acoustic parameter settings ensure the appropriate acoustic realization of information structuring (cf. Günther 1994 for details). As presented in Table 2.1, the different accent types are associated with different tone contours. The accent tones are realized on the metrical strongest element of the focused constituents. The prosodic accent interpretation rules of Figures 2.21–2.23 evaluate the accent status (nuclear, prenuclear, and contrastive accent) of a given word (*lexeme*) and assign the appropriate tone information. The tone contour on a lexeme is postlexically specified

in terms of accent, phrasal, and boundary tones whereby tonal focus realization is restricted to accent tone (ACC_TONE) and phrasal tone (PHRAS_TONE).

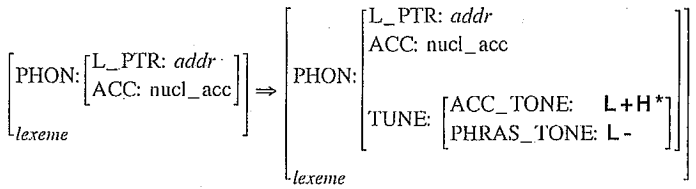

Fig. 2.21. Nuclear Accent Realization Rule

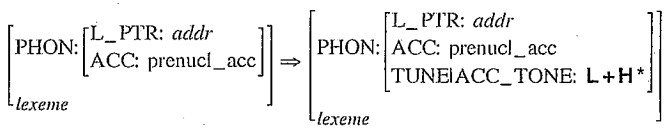

Fig. 2.22. Prenuclear Accent Realization Rule

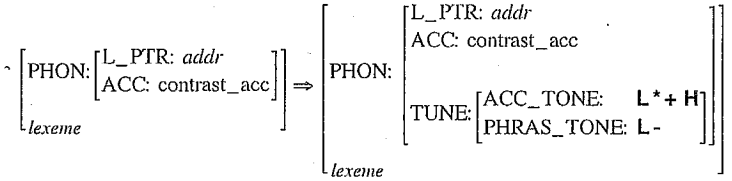

Fig. 2.23. Contrastive Accent Realization Rule

These abstract phonological tones are in turn interpreted phonetically, yielding concrete acoustic parameters as target values for the calculation and interpolation of the fundamental frequency contour.

To summarize the general approach to the processing of information structure within SYNPHONICS: Focus/background structure originates at the interface between the conceptual and the semantic levels as a result of comparing an actual conceptual fragment with its corresponding relevant context representation and is realized incrementally by prosodic means, thereby exploiting essential properties of the system, namely, the assumption of a conceptual representation, CS, and a relevant contextual environment, CT, as input for the linguistic components; the consideration of local information as well as global linkings of processing units by means of different types of embedding information; and, finally, the facility of a direct access of phonological processes to semantic representations by means of a direct semantics/phonology interface.

Notes

1 Current topics of research in computational linguistics as well as in psycholinguistics concern, in particular, the sequence and size of increments, their internal structure, as well as their interrelations.
2 Compare Günther et al. (1993) for a procedural account of topic/comment structure in a language production system.
3 The assumption of a direct semantics/phonology interface is currently becoming more popular; compare, for example, Engdahl & Vallduví (1994).
4 While adopting HPSG as grammar formalism, we deviate substantially from parts of the standard HPSG grammar theory; compare, for example, Abb and Maienborn (1994), Abb et al. (1995).
5 ALE lends itself as a formal description language for a language processing system because it supports a uniform formal specification of structural as well as rule knowledge.
6 For present purposes, we put aside the issue of giving a psycholinguistically supported account of the corresponding CT and CS selection processes. Therefore, in the following we will just deal with the question of how linguistic processing proceeds given a certain CS-CT configuration.
7 Compare Bierwisch and Schreuder (1992), Herweg and Maienborn (1992) for a motivation of the distinction between a linguistic (viz semantic) versus an extralinguistic (viz. conceptual) level of meaning representation from the perspective of language production.
8 Compare the notion of implicit question or *quaestio* in Klein and von Stutterheim (1987).
9 In the present chapter, we neglect issues concerning the referential specification of discourse referents in terms of definiteness, tense, and so on.
10 In fact, maximal focus over the whole utterance originates from a minimally restrictive CT that just provides a "naked" situation referent without any sortal or relational information, thus corresponding to the implicit question *What's going on?*
11 Notice that, given a contextual setting with a precisely delimited informational demand, if the speaker decides to supply further information compatible with, but not strictly required by, CT, the resulting utterance turns out to be less acceptable, because it contravenes at least slightly conditions of contextual appropriateness. Compare, for instance, the utterance *At the bus stop, Peter gave the book to Mary*. which is not well suited in the contextual setting of Figure 2.4, which expresses the implicit question *To whom did Peter give the book?* Besides fulfilling the specific informational demand expressed in Figure 2.4, this utterance supplies further new information (in terms of a locative adverbial) that is not properly asked for in CT. When we continue the utterance with, for example, ... *and in the subway, he gave it to Paul*, the utterance becomes acceptable again. But in this case the speaker plans his utterance with regard to a more complex context representation that comprises pairs of situation referents and experiencers. For lack of space, we cannot go into more detail here.
12 Compare Ziesche (1994) for documentation of the implementation realized in SYNPHONICS.
13 The speech analysis figures show the following information from top to bottom: intensity, fundamental frequency in the range of 80–180 Hz, frequency spectrum in the range of 0–8 kHz, and the speech signal.
14 We illustrate our focus-accent rules with German example sentences.
15 Actually, this is a slight simplification. Compare Maienborn (1994) for a specific type of verb-adjacent adjunct that might also bear a sentence accent.
16 The focus-type information originates from the specification of the semantic structure SEM (cf. Figs. 2.3 and 2.8). During lexical access, SEM is integrated into a complex HPSG sign; the focus-type information then becomes part of the semantic restriction set (RESTR) of the CONTENT value of the corresponding HPSG sign. The hierarchical structure of a phrasal linguistic sign (*phrase_struc)* is coded as a binary branching tree (as the value of the daughters feature DTRS) whereby a phrasal sign branches into a head daughter (HEAD_DTR) and either a complement daugther (COMP_DTR) or an adjunct daughter (ADJ_DTR). In the case of narrow focus, the accent-type specification is independent of the argument/modifier status of a constituent. For the sake of a general presentation, the feature NonH_DTR is therefore an abbreviation for either a complement daughter or an adjunct daughter.
17 For the sake of illustration, only the focus-accent rules are presented. The determination of the metrical strongest constituent (DTE) within a complex PP or DP is not a topic of this chapter. This process is described in detail in Günther (1994).
18 In this case, the type of the head daughter must not be *lemma* but *phrase_struc*. (Terminal nodes of the syntactic tree structure are typed as *lemma*, whereas all hierarchically higher nodes are typed as *phrase_struc*.)
19 The type of the head daughter is *lemma*.

References

Abb, B., C. Günther, M. Herweg, K. Lebeth, C. Maienborn, and A. Schopp. 1995. Incremental syntactic and phonological encoding: An outline of the SYNPHONICS Formulator. In G. Adorni & M. Zock (eds.), *Trends in Natural Language Generation: An Artificial Intelligence Perspective*. Berlin: Springer.

Abb, B. and C. Maienborn. 1994. Adjuncts in HPSG. In H. Trost (ed.), *Conference Proceedings of KONVENS 94: Verarbeitung natürlicher Sprache. Wien, 28–30. September 1994*, Berlin: Springer, 13–22.

Bartels, C., and J. Kingston. 1994. Salient pitch cues in the perception of contrastive focus. In P. Bosch & R. van der Sandt (1994, eds.) Vol. 1, 1–10.

Bierwisch, M. and R. Schreuder. 1992. From concepts to lexical items. *Cognition* 42:23–60.

Bosch, P. and R. van der Sandt. 1994: Focus and Natural Language Processing. (Working Papers of the Institute for Logic and Linguistics) IBM Deutschland Heidelberg.

Carpenter, B. 1992. *The Logic of Typed Feature Structures*. Cambridge: Cambridge University Press.

Engdahl, E. and E. Vallduví. 1994. Information packaging and grammar architecture: A constraint-based approach. In: E. Engdahl (ed.), *Integrating Information Structure into Constraint-Based and Categorial Approaches* (DYANA-2 Report), ILLC: Amsterdam, 39–79.

Günther, C. 1994. *Planung und Repräsentation prosodischer und phonetisch-artikulatorischer Merkmale im Sprachproduktionssystem SYNPHONICS*. In: A. Schopp (ed.), 15–28.

Günther, C. ed. (1994a). *Hamburger Arbeitspapiere zur Sprachproduktion*. VI. GK-Kognitionswissenschaft, AP 20, Universität Hamburg.

Günther, C., C. Habel, C. Maienborn, and A. Schopp. 1993. *What's up with the printer? Context relative presentation of conceptual structure and its prosodic realization in a language production system*. In: A. Schopp (ed.), 5–16.

Gussenhoven, C. 1992. Sentence accents and argument structure. In I.M. Roca (ed.) *Thematic Structure: Its Role in Grammar*, Berlin, New York: Foris, 79–106.

Herweg, M., ed. 1992. *Hamburger Arbeitspapiere zur Sprachproduktion. I.* GK-Kognitionswissenschaft, AP 9, Universität Hamburg.

Herweg, M. and C. Maienborn. 1992. Konzept, Kontext, Bedeutung – Zur Rolle der Zwei-Ebenen-Semantik in einem Modell der Sprachproduktion. In: M. Herweg (ed.), 7–36.

Inkelas, S. and D. Zec. 1990. *The Phonology-Syntax Connection*. Chicago: University of Chicago Press.
Jacobs, J. 1991. Focus Ambiguities. *Journal of Semantics* 8, 1–36.
Jacobs, J., ed. 1992. *Informationsstruktur und Grammatik*. Opladen: Westdeutscher Verlag.
Kempen, G. and E. Hoenkamp. 1987. An Incremental Procedural Grammar for Sentence Formulation. *Cognitive Science* 11, 201–258.
Klein, W. and C. von Stutterheim. 1987: Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte* 109:163–183.
Levelt, W. J. 1989. *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.
Maienborn, C. 1994. Fokus und Adjunktion: Überlegungen zum Fokusprojektionsverhalten von Adjunkten am Beispiel der Lokaladverbiale. In: A. Schopp (ed.), 29–66.
Nespor, M. and I. Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
Rooth, Mats E. 1985. Association with focus. Dissertation, University of Massachusetts, Amherst.
Schopp, A., ed. 1993. *Hamburg Working Papers on Language Production. II*. GK-Kognitionswissenschaft, AP 13, Universität Hamburg.
Schopp, A., ed. 1994. *Hamburger Arbeitspapiere zur Sprachproduktion. III*. GK-Kognitionswissenschaft, AP 15, Universität Hamburg.
Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Mass.: MIT Press.
Ziesche, S. 1994. Der semantische Enkodierer im SYNPHONICS-Project. In: C. Günther (ed.), 53–112.

3 On the Limits of Focus Projection in English

CARLOS GUSSENHOVEN

Abstract

Focus projection, the ability of a pitch accent to mark a larger constituent than the word it is found on as focused, has been given different treatments in the recent literature. In one view, focus projection cannot extend beyond the confines of the word; in another it is restricted to a sequence of an argument and its predicate, while in a third it can extend all the way to the sentence (the latter two types of projection being subject to various conditions). This chapter argues on the basis of 'focus ambiguity' that the second view is to be preferred over the first as well as over the third. In particular, the third interpretation of focus projection, proposed by Selkirk, is shown to make the wrong predictions in a number of structures, while two arguments that have been put forward in support of the third view are shown to be flawed. By contrast, the second, more restricted view of focus projection is not only seen to make the correct predictions, in both simple sentences and sentences with embedded predicates, but is also conceptually simpler. It is argued that the cause of the failure of the sentence-wide view is the a priori assumption that the focus of the sentence must correspond to a single node in the syntactic tree. Unlike this view, Höhle's proposal that the focus may be discontinuous does appear to make the correct predictions.

1 Introduction

Like many languages, English expresses the focus of the sentence in the phonological structure. Most strikingly, pitch accents appear on focused constituents. The expectation, therefore, is that unaccented constituents are outside the focus. However, the situation is more complex than this. One of the most intriguing and widely debated issues in the expression of focus in English is what Chomsky (1972) called 'Focus Projection', the phenomenon that a pitch accent can serve to mark more than just the word it is placed on as focused (a phenomenon that is also to be found in the related languages German, Frisian, Dutch, and Afrikaans; see Schmerling 1976, Gussenhoven 1983, Ladd 1996). As a result, a sentence like (1) is ambiguous between a full-focus reading (as when it is an answer to "What

I thank Ad Foolen, Wus van Lessen Kloeke, and Erwin Marsi for useful discussion of the issues dealt with in this chapter. This is a revised version of the contribution to the Proceedings of the Focus Conference, which was entitled "Focus and Sentence Accents in English."