Prosodic sentence analysis without exhaustive parsing

Hugo Quené—René Kager

Abstract

Suprasegmental phenomena in synthetic speech should reflect the linguistic structure of the input text. An algorithm is described which establishes the prosodic sentence structure (PSS). This can be achieved without exhaustive syntactic parsing, using a dictionary of 550 function words. Subsequently, phrase and accent locations are derived from the PSS; accentuation is also affected by some semantic and contextual information. Some exemplary rules for labeling, prosodic domain construction and accentuation are discussed. Comparison of the resulting sentence prosody with natural speech suggests that more detailed syntactic analysis may be necessary. Most of the accentuation errors are caused by semantic, pragmatic and contextual factors. These can only partly be imitated, since the relations between linguistic representations and real-world knowledge are not yet fully understood.

1. Introduction

Our ultimate aim, when listening to speech (be it synthetic or natural), is to understand the message it conveys. Usually, this boils down to comprehension of the semantic content of an utterance. This task is considerably easier, if the speech utterance contains adequate prosodic cues with regard to its semantic and pragmatic content (e.g. Collier—'t Hart, 1975; Wingfield 1975; Nooteboom—Brokx—de Rooij 1978; Cutler 1982; Cutler—Clifton 1984; Nooteboom 1985; Nooteboom—Kruyt 1987). Using suprasegmental cues, the listener can extract the linguistic structure of the message. Pauses, for example, divide the continuous speech signal into (linguistically) coherent word groups or chunks of acoustic-phonetic information. Hence, adequate pauses increase the intelligibility of the speech signal (Scharpff—van Heuven 1988). Likewise, *accentuation* guides a listener's attention to the words which are considered important by the speaker, and which are acoustically most reliable.

These perceptual functions of sentence prosody become even more important when the speech signal is less redundant, providing fewer segmental cues to the intended speech sounds, words, and meaning. Since the output of a text-to-speech conversion system lacks the normal degree of acoustic-phonetic redundancy, suprasegmental cues may significantly improve its perception and comprehension. Text-to-speech systems should aim at optimizing sentence prosody, in order to compensate for their reduced overall speech quality.

In the production of natural speech, various suprasegmental phenomena depend on the abstract linguistic notions *phrasing* and *accent*: F_0 movements (Collier—'t Hart 1975; Cooper—Sorensen 1977), location and duration of silent intervals (Goldman-Eisler 1972), segmental durations (Klatt 1975, 1976), intensity pattern (Lehiste 1970), coarticulation and sandhi (Cooper—Paccia-Cooper 1980) and vowel reduction (Koopmans-van Beinum 1980). All these suprasegmental phenomena can (in theory) be derived from the "abstract prosody". Considering the perceptual interest of adequate prosody, a high-quality text-to-speech system should attempt to establish both *phrasing* and *accentuation* and to convert these into adequate suprasegmental properties of the synthetic speech signal.

In natural speech, the (abstract) phrasing and accentuation are assumed to be related to the linguistic structure of an utterance. According to nonlinear phonological theory, abstract sentence prosody does not depend directly on the syntactic surface structure, but rather on the related *prosodic sentence structure* (Nespor—Vogel 1982, 1986; Gee—Grosjean 1983; Selkirk 1984, 1986). In addition, abstract accentuation is also affected by (a) the *thematic* structure, which specifies the semantic constituents functioning as Predicate, Argument and Modifier (Gussenhoven 1984), as well as by (b) the *focus* structure, which relates the syntactic constituents to be emphasized with specific accents (Ladd 1987; Baart 1987).

Extending this line of thought, we assume that accentuation and phrasing can both be derived from the prosodic sentence structure, provided that thematic structure and focus information are taken into account. Tô illustrate matters, our view of the various levels of sentence prosody is represented in Figure 1.

linguistic:	prosodic structure, thematic structure, focus structure
abstract prosody:	phrasing, accentuation
phonetic prosody:	segmental durations, pausing, F_0 movements, intensity, vowel reduction, coarticulation, sandhi, etc.

Figure 1. Three levels of sentence prosody

The aim of the present project is to investigate whether "abstract prosody" can be derived automatically for text-to-speech conversion. To this end, we have developed an experimental algorithm (Pros). First, a hybrid prosodic sentence structure is established (hence PSS). Subsequently, accentuation and phrasing are derived from this PSS, while contextual, thematic and rhythmic information is also taken into account. The output of the PROS component is converted to suprasegmental parameters by other components: silence segments are inserted in the phoneme string, segmental durations are adjusted, and an F_0 contour is calculated (te Lindert—Doedens—van Leeuwen 1989; Terken—Collier 1989; van Leeuwen—te Lindert, this volume).

The algorithm presented in this chapter is the latest member of a steadily growing family of similar algorithms (Kulas—Rühl 1985; Allen—Hunnicutt—Klatt 1987; Ladd 1987; Monaghan 1990a, 1990b, 1990c; Quazza—Varese—Vivalda 1989; O'Shaughnessy 1989; Carlson—Granström—Hunnicutt 1989; Bailly 1989; Hirschberg 1990; Bachenko—Fitzpatrick 1990). All these algorithms acknowledge that suprasegmental phenomena cannot be derived from syntactic information alone; non-syntactic information also plays an important role in determining sentence prosody (e.g. phrase length, reference to "given" vs. "new" concepts). Since syntactic factors can be "overruled" by non-syntactic factors, an exhaustive syntactic analysis is generally superfluous. In spite of these similarities, however, our method differs in several respects from those mentioned above (apart from the language for which it was developed). These differences mostly stem from our distinction between three separate tasks in generating sentence prosody, viz. (a) derivation of sentence structure, (b) derivation of abstract phrasing and accentuation, (c) generation of suprasegmental parameters.

This three-step approach offers many advantages over other methods, where two or more of these tasks are combined. Firstly, it seems to be a better approximation of human speech production. The suprasegmental phenomena associated with phrasing and accentuation (pausing, intonation, segment duration, etc.) are obviously intertwined. Hence, both phrasing and accentuation are likely to be derived from the same sentence representation. This representation should combine syntactic, semantic and rhythmic relations; the adjusted PSS matches these requirements—as opposed to syntactic surface structure (e.g. Quazza et al. 1989), Tone Group structure (e.g. Ladd 1987) or phrase structure. Secondly, the PSS may also be used to control other suprasegmental phenomena in the resulting speech signal, such as those indicated in Figure 1. Thirdly, there will be no conflicting suprasegmental cues in the resulting synthetic speech: all cues indicate the same abstract prosody and linguistic structure. This makes the sentence structure relatively easy to retrieve.

2. Prosodic sentence structure

From many languages, it is known that sandhi phenomena have their own specific domains of application, not necessarily identical to syntactic constituents. Among others, Nespor—Vogel (1982, 1986) and Selkirk (1984, 1986) have made proposals as to the mapping between syntactic constituents and *prosodic domains*. Two prosodic domains employed in this sense are the phonological phrase (Phi) and the intonational phrase (Int). These domains are part of a hierarchical tree structure, viz. the prosodic sentence structure. In their definition, the distinction between content words and function words plays a crucial role. The main semantic load of a sentence is carried by its *content words* (hence CWs: nouns, verbs, adjectives and adverbs; note that these are extendable word classes). *Function words* (FWs) express the relations between the content words, while they have hardly any independent meaning (prepositions, conjunctions, complementizers, copula, etc; fixed word classes).

In languages such as English and Dutch, the *Phi* domain (phonological phrase) is built around a lexical head, i.e., the content word which is the head of a syntactic constituent. This *prosodic head* plays an important role in accentuation. The Phi domain includes this head, its preceding specifiers, as well as all preceding FWs.

The next higher prosodic constituent is the *Int* domain (intonational phrase), constructed by grouping adjacent Phi domains. Hence, a whole Phi domain is always contained within a single Int domain. In addition, however, important syntactic breaks are also respected; each syntactic constituent which is attached to any S-node (in the syntactic surface structure) establishes a separate Int domain. Consequently, (a) displaced syntactic constituents, (b) most subordinate clauses, and (c) parentheticals, are all separate Int domains.

In the following example, the Phi and Int domains are illustrated in a flat representation (where '##' indicates an Int-boundary, and '#' a Phi-boundary). Note that prosodic domains do not necessarily correspond to syntactic constituents.

(1) ## Kasyapa's great war elephant # turned aside
## to avoid # a patch # of marshy ground ##

(2) ## De computer # spreekt # tot de bemanning ## op de betweterige # en begrijpende toon ## die we kennen # uit de zachte sector ##
'## the computer # speaks # to the crew ## in the pedantic # and understanding tone ## which we know # from the soft sector ##'

Prosodic domains tend to be of equal length as much as possible, and their length increases in faster speech. To account for these effects, separate rules restructure the prosodic domains (by merging and dividing them).

3. Automatic prosodic analysis

According to the linguistic theory described above, the prosodic sentence structure is initially derived from the syntactic (surface) sentence structure. Hence, syntactic parsing is necessary for any text-to-speech system performing prosodic analysis. For two reasons, however, this linguistically motivated method is not applicable to *automatic* prosodic sentence analysis. Firstly, there is no algorithm for syntactic analysis available with satisfactory performance (speed, accuracy, hardware requirements, Dutch language). In fact, it may be doubted whether any parser could perform satisfactorily, when confronted with ambiguous sentences like the following:

(3) a. I (have mown) (the lawn with the flowers).
b. *I (have mown) (the lawn) (with the flowers).

(4) a. Het was (ondanks de luchtverversing) (door de tv-lampen
it was in-spite-of the air-conditioning because-of the TV-lights
snikheet).
suffocatingly-hot
b. *Het was (ondanks de luchtverversing door de tv-lampen)
it was in-spite-of the air-conditioning by-means-of the TV-lights
(snikheet)
suffocatingly-hot

Solving this type of thematic ambiguity (with solutions yielding different prosody) requires a semantic and pragmatic analysis; the parser must "know" that one cannot use flowers to mow a lawn, and that TV lights produce heat rather than fresh air. Again, no system exists for this type of sentence analysis.

Secondly, syntactic information is insufficient for establishing sentence prosody. Syntactic factors may be overruled by non-syntactic factors (e.g. phrase length, reference to "given" or "new" concepts): some results of the syntactic analysis are subsequently discarded. Hence, an exhaustive syntactic analysis seems to be superfluous.

For these reasons, we have chosen to derive a correct PSS directly, using the minimum amount of syntactic information as input. Shortcomings of the correct *syntactic* sentence structure, or even its absence, need not detract from the adequacy of the resulting prosody, as long as a correct *prosodic* sentence structure underlies the latter. This follows from our assumption that prosody can be derived correctly from the PSS (as explained in section 1). The present project aims at (a) establishing the PSS from incomplete syntactic information; (b) deriving abstract sentence prosody (phrasing and accentuation) from the resulting PSS. This approach is illustrated in Figure 2 below.

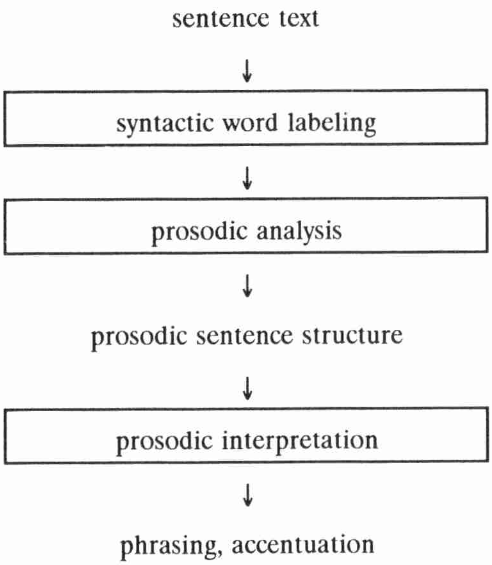

Figure 2. Alternative method for deriving the prosodic sentence structure (PSS) directly from an input sentence

In order to achieve the *first* aim, two types of syntactic information are indispensable. Firstly, lexical heads must be identified, since prosodic domains are constructed around these. To this end, word classification as CW or FW is sufficient. In theory, each last CW preceding an FW (or sentence boundary) constitutes a lexical/prosodic head. The CW-or-FW status of a word (hence "prosodic label") is determined by means of lexical lookup, in a dictionary containing all Dutch FWs (554 types). In addition, the dictionary contains another 300 CWs which behave anomalously for one reason or another. For all words, the syntactic word class is also specified. This syntactic labeling serves two purposes: it is used for (a) determining the syntactic labels of the words not found in the dictionary, and (b) the construction of prosodic domains, as explained below. (More details on this dictionary are given by Quené—Kager 1990, 1992).

Secondly, syntactic phrasing must be known, since it is relevant for prosodic domains. For example, the main verb (or verb group) in a sentence must be identified (cf. O'Shaughnessy 1989). This word (group) establishes a predicate constituent, corresponding to a separate Phi domain (Gee—Grosjean 1983). In Dutch, as in English, this Phi domain may separate the subject and object arguments of the predicate. Likewise, subordinate clauses must be identified, because they usually establish separate Int domains (see section 2). However, an exhaustive syntactic analysis appears to be superfluous for establishing prosodic domains: it is not necessary to determine the structural relations between the words in a sentence. Returning to (1), for example, it is not necessary to decide which is the internal structure of the subject NP:

(1) a. (Kasyapa's (great (war elephant)))
b. (Kasyapa's ((great war) elephant))
c. ((Kasyapa's (great war)) elephant)

Instead, it suffices to determine that these four CWs together establish a single constituent, which in turn establishes a separate Phi domain. This information can be derived from the sequence of syntactic word classes (the colon indicates a sub-classification):

(1)	Kasyapa's	great	war	elephant	#	turned	aside ...
	N	A	N	N	#	V:Infl	A ...

The words *elephant* and *turned* must belong to separate syntactic constituents, as a noun cannot be a specifier to an inflected verb. Since both words are lexical heads of syntactic constituents, they cannot both belong to the same Phi domain. Consequently, a Phi boundary must separate these two words. In more general terms, syntactic constituency is derived from restrictions on possible sequences of syntactic labels within a syntactic constituent. If constraints on the possible label sequences are violated, then we may assume a syntactic boundary. Not all syntactic boundaries, however, are equally important for the PSS. For the purposes of sentence prosody, it generally suffices to demarcate those syntactic constituents which are respected by prosodic domains, while ignoring other syntactic structures. It should be noted, however, that this resulting PSS is an *approximation* of the theoretical PSS, since not all relevant syntactic information is available for the prosodic analysis.

Because the algorithm demarcates prosodic domains on the basis of both the CW/FW distinction and syntactic word labels, it can deal with a large variety of input text types. For its development, a large corpus of newspaper articles was used (approx. 450,000 word tokens). Its CW:FW ratio is roughly 1:1 (46:54). Although the analysis rules (described in more detail below) were primarily based on observations regarding this corpus, they also apply to other types of input text. If an input text would contain predominantly FWs, then the rules based on syntactic label sequences ensure correct demarcation of prosodic domains (5a); if no syntactic labels could be determined, then prosodic domains are derived from the CW/FW distinction (5b). Only in this latter case are problems to be expected, if subsequent unlabeled words are all classified as CW (5c).

(5) a. van al wie er toen was ## is hij het geweest ## die ...
V\FW ## V\FW V\FW ## Comp\FW
of all who there then was ## is he (it) been ## who ...
b. de aan mijn broer verkochte ## zeer gevaarlijke auto
?\CW ## ?\FW
the to my brother sold ## very dangerous car
c. de onlangs dubbel verkochte (##) uiterst gevaarlijke auto
?\CW ?\CW ?\CW ?\CW ?\CW
the recently doubly sold (##) extremely dangerous car

In order to achieve its second aim, viz. derivation of phrasing and accentuation from the PSS, additional information is again indispensable. As regards phrasing, boundaries of Int domains are simply converted into pause locations. As regards accentuation, however, matters are considerably more complex. This prosodic phenomenon is known to be influenced by three types of additional information, which our algorithm should also take into account. The first factor is the discourse context of a sentence, specifically the distinction between "given" and "new" concepts (Fuchs 1984). A phrase is unaccented if the speaker assumes its content already known by the listener ("given"). Usually, only domains introducing new information are accented, as demonstrated below (accent indicated by capitals):

(6) a. he came # by CAR
b. his car # was BLUE

Secondly, thematic relations between prosodic domains also play an important role (Gussenhoven 1984; Baart 1987). Predicates are usually left unaccented; they receive accent only if their argument(s) are unaccented or if they are not adjacent to their argument(s). Thirdly, rhythmic factors affect accentuation: speakers prefer some alternation between accented and unaccented words. If multiple adjacent accents should occur, then one of the accents is removed or shifted to a different word (Kager—Visch 1988; Visch 1990).

In summary, then, the algorithm starts by providing the words constituting the sentence with a syntactic label. Subsequently, the PSS is derived from both the orthographic input sentence (text string), and from the syntactic labeling of its constituent words. Finally, (abstract) phrase boundaries and accents are derived from the PSS, as well as from additional information.

4. Some typical examples

4.1. Syntactic disambiguation

In the algorithm's representation, words can have multiple syntactic labels. These can be copied from the dictionary, output by the morphological analysis (Heemskerk—van Heuven, this volume), or generated by rule (on the basis of string features, e.g. affixes). Prior to the demarcation of prosodic domains, these syntactic label ambiguities must be resolved. As explained in section 3 above, this is done on the basis of restrictions on label sequences: linguistically motivated disambiguation can only be achieved when taking account of the context.

This selection of the appropriate syntactic label is often based on linguistic constraints on word sequences: for example, an inflected verb may not be preceded by an article (since the article label of the first word comes from the dictionary, the verb label of the second word should be discarded). In addition, several rules employ statistical constraints: for example, prepositions are usually followed by nouns, rather than by verbs. Such probabilistic observations are based on our analyses of the newspaper text corpus. Although these rules filter out many incorrect labels (7a), they may occasionally introduce errors (7b):

(7) a. Vogels broeden **in nesten.**
birds breed in nests
b. Hij wil zich daar **in mengen**.
he wants (to) himself there in mingle

In addition, orthographic conventions guide the selection of syntactic labels: strings containing a hyphen are compound nouns; strings containing digits are numerals; words starting with a capital (not sentence-initial) are proper names; long words (over 13 characters) are probably nouns (e.g. *wapenhandelaren* 'weapon-traders') rather than verbs (although the latter are possible, e.g. *herprogrammeren* 'reprogram').

4.2. Separate Int's for sub-clauses

As explained in section 2 above, each sub-clause constitutes a separate Int domain. To produce such domains, several rules in the algorithm insert an Int boundary between two adjacent words which cannot both belong to the same clause (as may be deduced from their syntactic and prosodic labels). Note that Dutch is an SOV language, where the inflected verb takes the final position in subordinate clauses, and the second position in main clauses. Some generalizations in our rules for Dutch do not apply to English, since that is an SVO language (cf. glosses (8-10)).

Rule (8) demarcates the end of a subordinate clause (with final verb). Rules demarcating the beginning of a subordinate clause are exemplified by (9). Two juxtaposed main clauses are demarcated by rule (10); it is determined from the right-hand context of the conjunctive whether it links two clauses (and not two other constituents).

(8) ∅ ⇒ IntBound / <VERB:INFL> ___ <VERB:INFL>
— omdat het geen haast **had ## deed** ik het later
because it no hurry had ## did I it later
— de fresco's die hij **schilderde ## zien** er nog fris uit
the frescoes which he painted ## look (there) still fresh (out)

(9) ∅ ⇒ IntBound / <VERB:PART> ___ <NOT <VERB>>
— hij is erin **geslaagd ## om** dit goed weer te geven
he has (in-it) succeeded ## this correctly to represent
— hij heeft **beloofd ## morgen** te komen
he has promised ## tomorrow to come

(10) ∅ ⇒ IntBound / ___ <CONJ> {<VERB:INFL>}
{<FW>}
— hij raapte een steen op **## maar aarzelde** om die weg te gooien
he picked up a stone ## but hesitated it away to throw
— Thero excuseerde zichzelf **## en hij** verliet de kamer
Thero excused himself ## and he left the room

4.3. Phi demarcation

As explained in section 2 above, Phi domains are strictly enclosed within Int domains. Our algorithm produces this hierarchy by first demarcating Int domains; subsequently, Phi domains are demarcated within these Int domains. (This procedure deviates from the theoretical construction of prosodic domains, e.g. Nespor—Vogel 1986.) Analogously to Int domain demarcation, a Phi boundary is inserted between two adjacent words which cannot both belong to the same domain. The most important of these rules (11) uses the fact that the prosodic head of a Phi is typically its rightmost CW. FWs always belong to the prosodic head on their right-hand side, i.e. to the right-hand Phi domain. Hence, a Phi boundary is assumed between a CW (which is the prosodic head of the left-hand Phi domain) and a following FW (which is a specifier for the right-hand prosodic head):

(11) Ø ⇒ PhiBound / <CW> ___ <FW>
— she **prunes # the** red **roses # in** her garden

Of course, this requires additional rules for Phi domains whose prosodic head is not a CW, but rule (11) correctly identifies the majority of Phi domains.

4.4. Verb accentuation

Predicates are accented, if a sentence modifier intervenes between the verb (group) and its arguments, or if all of the arguments are unaccented (not in focus)—or simply absent (cf. section 3). The verb accentuation rules work on single CW verbs, as well as on longer sequences of verbs, of which the first CW is accented. In addition, they work on complex verbs, of which the stem and particle can occur separately. Stem and particle occur together only in participles (12a), where the two parts should be written as a single word. But in infinitive constructions, the infinitive marker *te* can be interposed (12b), whereas in inflected forms in Dutch, the inflected stem in second position in main clause may be far away from the particle in "original" clause-final position, as in (12c). Nevertheless, it is always the particle that is accented, even if the stem is far away (12b,c):

(12) a. Ik **heb** # mijn MOEDER # VANDAAG # **OPGEBELD.**
I've # my mother # today # phoned
b. Ik heb GEPROBEERD # mijn MOEDER # VANDAAG # **OP te bellen.**
I've tried # my mother # today # to phone
c. Ik **belde** # haar # VANDAAG # **OP.**
I phoned # her # today # Ø

However, particles are often identical to prepositions (e.g. *op* in (12)), which makes their detection difficult. Some can be identified because they occur immediately before the infinitive marker *te* (12b); these are treated as the prosodic head of the predicate Phi domain. Others may be identified because they occur in clause-final position (12c), where "real" prepositions cannot occur. The rules discussed below also work on these "stranded" clause-final particles.

A verb (group) is accented if it follows an adverb (13a) or adverbial phrase (13b). An adverbial PP is detected through its initial preposition, which is seldom used within an [NP PP]$_{NP}$ complex (argument) (e.g. *ondanks* 'notwithstanding', *sinds* 'since'). The verb (group) itself must be followed by an Int (or sentence) boundary, since a clause-internal predicate would typically be followed by an (adjacent) argument.

(13) a. Hij heeft # het GAZON # **VANDAAG**$_{\text{Adv}}$ # **GEMAAID.**
he has # the lawn # today # mown
b. Hij heeft # het GAZON # **ondanks** de REGEN # **GEMAAID.**
he has # the lawn # in-spite-of the rain # mown

If the predicate is accompanied by an un-accented argument, then the predicate (verb group) must be accented. Such arguments are indicated by pronominal FWs. In addition, deictic qualifiers (such as *dit, deze* 'this', *dergelijke, zulke* 'such') imply that the following term refers to "given" information. Any words following one of these cue words in the same Phi domain are de-accented, and an adjacent predicate receives integrative accent (14a). In cases with comparative adjectives, however, this qualifier takes the integrative accent, instead of the predicate (14b). Finally, the predicate is also accented if there is no argument at all.

(14) a. De ZEEHONDEN # hebben # **deze** lage temperaturen # **OVERLEEFD.**
the seals # have # these low temperatures # survived
b. Ze heeft # de **HOGERE** weg # **genomen**.
she has # the higher road # taken

5. Evaluation

5.1. Introduction

As explained in section 1, the algorithm described above aims at establishing the correct abstract sentence prosody (accentuation and phrasing). It is assumed that these aspects (or rather, their appropriate phonetic correlates) make synthetic speech more natural and intelligible. In order to evaluate whether our algorithm achieves this aim, two methods are possible.

First, its output can be evaluated from a perceptual viewpoint, viz. by investigating whether synthetic speech is sufficiently intelligible and adequate, if accents and pauses are derived automatically. Van Bezooijen (1989a) reported subjects' ratings on a 10-point adequacy scale of sentence prosody under four accentuation conditions: (a) derived automatically [mean 6.0], (b) same number of accents, distributed at random over CWs [mean 4.6], (c) subjects' preferred accentuation, as established earlier [mean 7.7], and (d) as produced by a human professional speaker [mean 7.4]. From these results, she concludes that the algorithm produces sufficiently adequate accentuation, although its output is still defective in several respects (see also van Bezooijen—Pols, this volume).

Secondly, we can compare the output accentuation and phrasing with those produced by a human speaker. Ideally, there should be no difference between the two. Even if we allow some variation, any differences found should not disturb the semantic and pragmatic equivalence between the natural and synthetic versions. This comparison is reported in the following section.

5.2. Comparison with natural prosody

The accentuation and phrasing produced by our algorithm was compared with natural speech on a word-by-word basis (a phrase-by-phrase comparison is reported by Quené—Dirksen 1990). A professional speaker, the designated talker of the ASSP program, read aloud several texts (10,766 words, of which 5,273 were CWs and 5,493 FWs, in 600 sentences, grouped into 43 texts). Recordings were transcribed with respect to accents and prosodic boundaries by the two authors (and occasionally, by a third transcriber). Subsequently, this transcription was compared with the phrasing and accentuation produced by our algorithm.

We must realize, however, that a difference in phrasing or accentuation between speaker and algorithm does not necessarily imply that the latter has been wrong. In fact, both versions may be equally acceptable, and roughly equivalent (15). All such discrepancies between equivalent versions were discarded.

(15) a. Een aantal ONDERZOEKERS meent overigens ## dat de VRAAG ## of passief meeroken SCHADELIJK is ## al LANG positief kan worden BEANTWOORD.

b. Een AANTAL onderzoekers MEENT overigens ## dat de vraag of PASSIEF MEEROKEN SCHADELIJK is ## al lang POSITIEF kan worden beantwoord.
a number (of) researchers thinks indeed ## that the question (##) whether passive smoking detrimental is ## already positively can be answered

The algorithm generated 86 percent of the naturally produced phrase boundaries (N=1,319), while 21 percent of its output *phrase boundaries* were incorrect. An error occurs in 5 percent of all word boundaries in the input sentences (N=10,167).

Of the naturally produced *accents* (N=4,173), 85 percent were also generated automatically, while 19 percent of the output accents were incorrect. It must be noted that the two error types are not independent here, since accents usually occur in a rhythmic pattern. Therefore, incorrect accentuation of one word corresponds to an incorrect non-accent on a neighboring word, as in the fragments *een aantal onderzoekers meent* and *al lang positief* in (15) above. The different accentuations of such fragments are not equivalent, since semantic prominence is guided towards different words.

The "human" and automatic accentuations are in agreement in 87 percent of all words in the input sentences (N=10,766). Since FWs are seldom accented, the agreement in their (non-)accentuation is considerably higher (viz. 94 percent; this class includes all particles) as compared to CWs (77 percent).

5.3. Error analysis

In a post-hoc analysis of the discrepancies (Quené—Kager 1992), both "misses" and "false alarms" in the PROS output were classified according to their origin (e.g. labeling, syntax, rhythmic, "given"/"new", etc.). Results show that most *phrasing* errors (73 percent) are caused by incorrect mapping from syntactic constituents to prosodic domains. Most often, subordinate clauses are incorrectly demarcated, as in (16):

(16) De strepen worden doorgesneden ## langs de deuren # om het instappen # te vergemakkelijken.
the strips are cut ## along the doors # in-order-to the getting-in # to facilitate

These are serious errors, since they result in "garden path" sentences: the prosody signals an incorrect linguistic structure. In addition, they may also propagate into subsequent accentuation errors.

Most errors in *accentuation* (69 percent) are caused by semantic, pragmatic and contextual factors. Any accentuation algorithm should decide which words convey "given" information, and adjust the accentuation accordingly (see section 4.4). Moreover, it should also decide which words are important, and then assign a contrastive accent on these words (or an integrative accent on its prosodic head). Our algorithm performs poorly on both of these tasks. Regarding the first task, this is primarily due tot the fact that its scope is limited to the single sentence, yielding errors such as (17):

(17) a. De OLIEVLEK # zal aan DUIZENDEN VOGELS # het LEVEN kosten.
the oil-slick # will of thousands [of] birds # the life cost
b. Er zijn al # TIENTALLEN **VOGELS** # DOOD AANGETROFFEN.
there are already # tens [dozens of] birds # dead found

The obvious remedy might be to widen this scope, e.g. by maintaining a buffer of content words (Silverman 1987) or root morphemes (Hirschberg 1990) across sentences, and resetting this buffer at paragraph boundaries. This strategy would avoid some errors, but co-references involving synonyms and paraphrases are likely to go undetected — even though these are more frequent than the former type.

Moreover, it has been observed that there is no perfect correspondence between the distinctions "given"/"new" and ±accent. Words conveying "new" information can be unaccented (18a), and (implicitly) "given" words can be accented (18b):

(18) a. de HERENIGING # **van de twee Duitslanden**
the reunion # of the two Germanies
b. **ZIJN** salaris # is VEEL hoger # dan het UWE
his salary # is much higher # than yours

In the latter case, contrastive accents are selected solely on pragmatic and contextual grounds. Such accents can be derived partly from statistical regularities and other heuristics (Monaghan 1990c). Yet, the majority of such errors cannot be solved without a better understanding of the complex relation between accentuation and our knowledge of the real world.

6. Conclusion

In the present chapter, it was argued that suprasegmental phenomena in synthetic speech should be derived, via abstract accentuation and phrasing, from an underlying linguistic sentence representation. We therefore attempted to establish the prosodic sentence structure (PSS; Nespor—Vogel 1982, 1986). This structure is based on a well-established phonological theory (supported by independent evidence from various languages), and predicts many prosodic phenomena. Yet, it does not require exhaustive syntactic analysis of the input sentence, which (even if required at all) is currently unavailable. The algorithm described in this chapter approximates the PSS on the basis of syntactic word labeling. Subsequently, phrasing and accentuation are derived from the PSS, while for accentuation some semantic and contextual information is also taken into account.

A large proportion of the errors in the resulting phrase boundaries could be solved by more exhaustive and more advanced syntactic analysis. Accentuation errors are mainly caused by contextual and pragmatic factors. It is in the nature of things, however, that such errors cannot be avoided in a principled way. At present, linguistic generalizations cannot include the real-world knowledge shared by speakers and listeners, which seems to be responsible for most of the variation in our use of prosodic patterns.