

A sentence accentuation algorithm for a Dutch text-to-speech system

1. Introduction

In recent years, considerable progress has been made in the construction of linguistic components of text-to-speech conversion systems for Dutch, especially in the domains of grapheme-to-phoneme conversion, both rule-based and lexicon-based (Kerkhoff, Wester, and Boves 1984; Berendsen, Langeweg, and Van Leeuwen 1986; Daelemans 1987; Lammens 1987; Baart and Heemskerk 1988). Furthermore, F_0 -contours can be automatically generated on the basis of the grammar of Dutch intonation ('t Hart and Collier 1975; Van Wijk and Kempen 1985), where sentence accents and prosodic boundaries are assumed to be marked. Little attention has been paid, however, to the automatic location of sentence accents and prosodic boundaries in text. Clearly, sentence accents and prosodic phrases depend to a high degree on information which is difficult to obtain from raw text: pragmatic context and thematic structure, as well as detailed syntactic structure.

The sentence accentuation algorithm PROS, which we will describe here, is an attempt to arrive at a fairly acceptable sentence accentuation output at low cost, by means of avoiding the complications inherent in exhaustive syntactic and pragmatic text analysis. It contains probabilistic accentuation rules based on large quantities of text material. In addition, it has profited much from recent theoretical descriptions of Dutch sentence accentuation (Gussenhoven 1983, Kruyt 1985, Terken 1985, Baart 1987). In auditory perceptual evaluation, PROS has been found to produce a satisfactory output (see Van Bezooijen 1989, this volume).

2. Sentence accentuation from a theoretical perspective

A perfect sentence accentuation algorithm should have access to at least the following types of information. First, discourse context, including topic-comment structure and the distinction between **given** and **new** information. See (1), where accentuated words are in capitals:

- (1) Question: Houdt u van SPINAZIE?
 "Do you like SPINACH?"
Answer: Nee, ik HAAT spinazie!
 No, I HATE spinach!

In the answer the word *spinazie* will remain unaccentuated since it refers to information that was **given** in the question. In Kruyt (1985), Terken (1985), and Nooteboom and Kruyt (1987), the role of given and new information in accentuation has been confirmed experimentally.

Second, thematic structure should be taken into account, as shown in the examples below, each of which can be used as answers to the question "Wat heeft Mieke gedaan?" ("What did Mieke do?"):

- (2) a. Ze heeft het GAZON geruïneerd
 She has the LAWN ruined
 ("She has ruined the lawn")
b. Ze heeft het GAZON binnen een MINUUT GERUINEERD
 She has the LAWN within a MINUTE RUINED
 ("She has ruined the lawn within a minute")
c. Ze heeft het gazon van de BUREN geruïneerd
 She has the lawn of the NEIGHBOURS ruined
 ("She has ruined the neighbours' lawn")

Gussenhoven (1983) shows that in predicate-argument structures, the argument is accentuated (2a), unless a non-argument adverbial expression intervenes (2b), in which cases the verb is accentuated as well. The relevance of the independent non-argument status of the intervenient can be inferred from (2c), where *van de buren* is a non-argument as *binnen een minuut* is in (2b), but syntactically depends on *het gazon* (within a complex argument *het gazon van de buren*).

Third, rhythmic factors influence accentuation. This can be illustrated with the sentence pair below, where the verb in (3a), *geprotesteerd*, is accentuated following an adverbial *heftig*, whereas in (3b) the verb *protesterende* is not.

- (3) a. De BUREN hebben HEFTIG GEPROTESTEERD
The NEIGHBOURS have FIERCELY PROTESTED
("The neighbours have protested fiercely")
b. de HEFTIG protesterende BUREN
the FIERCELY protesting NEIGHBOURS

The difference stems from the fact that the verb is not followed by an accentuated word in the former case, whereas it is in the latter. See Kager and Visch (1987) for a metrical analysis of phrasal rhythmic accentuation in Dutch.

These three observations on sentence accents lead Gussenhoven (1983) and Baart (1987) to (implicitly) propose the model of sentence accentuation represented below:

- (4) Ideal Sentence Accentuation Algorithm
a. Determine [+focus]/[-focus] of constituents on the basis of **given** and **new** information, **predictability** within a discourse context, etc.
b. Determine **thematic structure** in terms of predicates and arguments (Gussenhoven 1983), or heads, complements, and specifiers (Baart 1987).
c. Form **focus-domains**, and assign sentence accents.
d. Idealise by means of rhythmic adjustments.

The notion of **focus-domain** mentioned in (4c) deserves some further clarification. Its functional definition is as follows:

- (5) Focus-domain:
A constituent which can be interpreted as being integrally highlighted, i.e. [+focus] by means of one sentence accent (the **integrative** accent).

Returning to (2) above, the sequence *het gazon geruïneerd* in (2a) is a focus-domain, since the integrative accent on *gazon* places the entire sequence in focus. But in (2b), the sequence *het gazon binnen een minuut geruïneerd* is not a focus-domain, as suppressing any of the three accents will lead to a [-focus] interpretation of the relevant part. Gussenhoven (1983) and Baart (1987) seem to differ as to the question of whether a one-to-one correspondence exists between focus-domains and accents, but this matter is not highly relevant for our exposition.

For ease of reference, we will restrict ourselves here to one proposal for domain formation and accentuation, the one by Gussenhoven (1983). In (6), **P** stands for 'predicate', and **A** for 'argument'. Underlining represents a terms being [+focus]:

- (6) Sentence Accent Assignment Rule:
a. Domain Assignment: $\underline{P} (X) \underline{A} \rightarrow [P(X)]$
(ordered) $\underline{A} (X) \underline{P} \rightarrow [A(X)P]$
 $\underline{Y} \rightarrow [Y]$
b. Accent Assignment: In [Y], accent Y.
In [AP] or [PA], accent A.

The SAAR pairs sequences of thematic constituents (predicate-argument, argument-predicate) into focus-domains, where only non-focussed terms (X) may intervene.¹ As an example of (6), see (7) -- an elaborated version of (2):

- (10) De ZOEKACTIE op TERSCHELLING naar de met OLIE BESMEURDE VOGELS is dit WEEKEINDE ERNSTIG BEMOEILIJKT door de ZWARE SNEEUWVAL ZATERDAGNACHT en ZONDAG. De SNEEUW STAPELDE zich op tot een HOOGTE van ONGEVEER VEERTIG CENTIMETER. BOVENDIEN werd met NAME op TERSCHELLING VEEL HINDER ONDERVONDEN van IJSVORMING. Hierdoor konden de MEDEWERKERS van de VOGELWACHT VEEL PLAATSEN waar zich OLIESLACHTOFFERS BEVONDEN NIET BEREIKEN. Ondanks de BARRE WEERSOMSTANDIGHEDEN SLAAGDE men er in DERTIG EIDEREENDEN te VANGEN.
 "The SEARCH on TERSCHELLING for the with OIL POLLUTED BIRDS was this WEEKEND SERIOUSLY HAMPERED by the HEAVY SNOWFALL SATURDAY-NIGHT and SUNDAY. The SNOW PILED itself up to a LEVEL of SOME FORTY CENTIMETERS. ALSO was in PARTICULAR on TERSCHELLING MUCH HINDRANCE EXPERIENCED from ICE-FORMATION. By this could the CO-OPERATORS of the BIRD-WATCH MANY SITES where themselves OIL-VICTIMS FOUND NOT REACH. Notwithstanding the SEVERE WEATHER-CONDITIONS SUCCEEDED one in THIRTY EIDERDUCKS to CATCH." 2

As a first improvement on such a simple algorithm, one might take into account statistical accentuation properties of lexical categories and individual words. In the MITalk system (Allen, Hunnicutt, and Klatt 1987), categorial membership is used to determine relative prominence (height of F_0 -excursions), according to a scale proposed in O'Shaughnessy (1976). For Dutch, a similar type of scale has been established by Kruyt (1985):

- (11) Categorial Accentuation Scale
 NUM - proper name - NOUN - ADJ - VERB - ADV - etc.

This scale may be augmented by a list of words with specific accentual properties. Words whose meaning involves **negation** are strong candidates for being specified as inherently accentuated. Together these improvements lead to the algorithm below:

- (13) Improved CW-FW algorithm
Differentiate the CW-FW bipartition into a scale:
 a. Label lexical categories, and accentuate according to categorial likeliness as established in sufficient material.
 b. Include statistically (un)accentuated words in an additional lexicon, and make these specifications overrule categorial accentuation specifications.

To illustrate the performance of an algorithm such as (13), consider (14). VERBs and some specific CWs (*ongeveer*) have been left unaccentuated, whereas specific FWs (*ondanks*) are accentuated.

- (14) De ZOEKACTIE op TERSCHELLING naar de met OLIE BESMEURDE VOGELS is dit WEEKEINDE ERNSTIG bemoeilijkt door de ZWARE SNEEUWVAL ZATERDAGNACHT en ZONDAG. De SNEEUW stapelde zich op tot een HOOGTE van ongeveer VEERTIG CENTIMETER. BOVENDIEN werd met NAME op TERSCHELLING VEEL HINDER ondervonden van IJSVORMING. Hierdoor konden de MEDEWERKERS van de VOGELWACHT VEEL PLAATSEN waar zich OLIESLACHTOFFERS bevonden NIET bereiken. ONDANKS de BARRE WEERSOMSTANDIGHEDEN slaagde men er in DERTIG EIDEREENDEN te vangen.

Notice that (13) induces a new type of error in (14), viz. verbs incorrectly left unaccentuated (such as *bereiken* and *slaagde*).

4. PROS

4.1 Introduction

We will now discuss another series of improvements on simple CW-FW algorithms, which have been implemented in PROS. In addition to accentuation which is based on category membership and individual words, PROS contains three types of rules to improve accentuation. The set-up of PROS is as below:

(15) PROS

Accentuate CWs except for (i) verbs and (ii) words specified in the lexicon as [-accent].
Adjust accentuation by means of three types of rules:

- a. Rhythmic **de**accentuation.
- b. **De**accentuation of words in contexts signalling **given** information.
- c. Default accentuation of VERBs in specific contexts.

In addition to such accentuation rules, PROS contains rules for word class assignment and prosodic phrase construction. These modules will not be discussed here, but we will assume their output to be accessible to the accentuation module. See for more specific information on labelling and phrase formation Kager and Quené (1987). In the following sections we will discuss (15a-c).

4.2 Rhythmic deaccentuation

If a word occurs between two accented words (of specific categories), then this word is deaccentuated. All three words must belong to the same prosodic phrase. The common general context of rhythmic deaccentuation rules is in (16), which also shows two instantiations of this context.

(16) Rhythmic-DEaccentuation (general context)

[+ACC] → [-ACC] / [+ACC] ___ [+ACC]

- | | | | | |
|----|------------|-------------------|--------------|---------------------------|
| a. | Q-TERM | CW | CW | |
| | drie | duizend | jaar | "three thousand years" |
| | half | miljoen | gulden | "half million guilders" |
| | zeer | lage | temperaturen | "very low temperatures" |
| b. | ADV | PART ³ | CW | |
| | droevig | kijkende | buren | "sad looking neighbours" |
| | zeer | verbaasde | toeschouwers | "very amazed spectators" |
| | evenwijdig | lopende | spoorlijnen | "parallel running tracks" |

Rhythmic deaccentuation is restricted to contexts where it is certain that it should apply, in order to avoid incorrectly deaccentuating the middle term in e.g. an ADV-ADJ-NOUN sequence such as *ongelooflijk simpele middelen* ("incredibly simple means"), or *typisch tijdelijke arbeid* ("typically temporary employment").

4.3 Deaccentuation in contexts signalling given information

The second type of rule in (15) is deaccentuation of words in contexts signalling **given** information. Qualifiers such as *dit*, *deze* ("this"), *dergelijke*, *zo'n*, *zulke* ("such") imply that the referent of the qualified term has been introduced earlier in the text. Therefore, these words signal **given** information, or [-focus]. Two separate classes of deaccentuating qualifiers are distinguished, for reasons related to default verb accentuation.

(17) Deaccentuation of words in contexts signalling given information.

[+ACC] → [-ACC] / KNOWN-QUALIFIER ___

- | | | |
|----|--|-----|
| a. | {dit, deze, dergelijk(e), zo'n, zulke, ...} | ___ |
| b. | {ander(e), volgend(e), vorig(e), (de)zelfde} | ___ |
| | "different" "next" "previous" "same" | |

Obviously, the basis of this type of rule is probabilistic. In a minority of relevant cases, the qualified term is [+focus], but the examples of (18) certainly represent the majority:

- (18) a. Het GEVOLG van zulke temperaturen is VRESELIJK
The CONSEQUENCE of such temperatures is AWFUL

- b. Ze heeft de VOLGENDE trein genomen
 She has the NEXT train caught
 ("She has caught the next train")

4.4 Default accentuation of verbs in specific contexts

According to rule (15c), verbs are accentuated in specific contexts. As will be clear from the discussion of thematic structure and accentuation above, a verb (being a predicate) will be accentuated if it is in a focus-domain on its own. This will happen in two cases. First, if a [+focus] adverbial term intervenes between an argument and the predicate (cf. 7c). Second, if no [+focus] argument is present to form a domain with the predicate, for instance in case the argument is either [-focus] or simply absent. Let us now discuss these cases one by one.

4.4.1 Adverbial constituents

In order to mimick the accentuation effects of blocked focus-domain assignment after adverbials, both thematic and focus information is needed. Again, we take advantage of heuristics. Adverbial terms can consist either of single adverbs or of adverbial phrases, which are often PP's. In the former case, an adverb will trigger accentuation of a following verb:

- (19) Default accentuation of VERBs following ADVerbs
 VERB → [+ACC] / ADV ___ INT

Note that there is no direct or indirect check on the ADV being [+focus]. The reason for this is probabilistic again, as we have found that even particle-like ADVs such as *al* are usually accompanied by stressed verbs. The rule refers to a following INT (intonational phrase boundary), which is inserted in the prosodic phrase structure module. It serves to guarantee that the verb to be accentuated is syntactically clause-final and not clause-medial, where it would typically be followed by an argument. Dutch has a basic SOV word order, but the finite verb is moved to second position in the matrix clause. See the following example (cf. 3a):

- (20) De BUREN hebben HEFTIGGEPROTESTEERD
 ADV VERB INT

PROS can detect adverbial phrases consisting of PP's through their initial preposition. That is, prepositions such as *ondanks*, *wegens*, *sinds* introduce constituents which take the place of ADV in (19). To check the constituency of a sequence of words starting with these prepositions, PROS contains heuristics for detecting (simple) constituents, which will not be discussed here. An example of a sentence containing an adverbial PP is below:

- (21) De BUREN hebben ONDANKS onze HULP GEPROTESTEERD
 [ADV] VERB INT

The simple constituent *ondanks onze hulp* is detected as an adverbial phrase, triggering default accentuation of the verb by rule (19).

4.4.1 Deaccentuated constituents

We will now turn to the interesting effect of default accentuation of verbs accompanied by [-focus] arguments. As (17) gives us a probabilistic indication of [-focus] terms, we can employ these cues to arrive at a more adequate accentuation of CW verbs. In the rule below, KQ(17a) is the **Known-Qualifier**, and CW a deaccentuated term following it:

- (22) Default accentuation of VERBs following Known-Qualifiers
 VERB → [+ACC] / KQ(17a) CW ___ INT

BEMOEILIJKT, stapelde, ondervonden, bevonden, BEREIKEN, vangen), and the verbal particle *op*. But certainly the output of PROS cannot be considered ideal, as will be clear from (27).

Deaccentuation of 'known qualifiers' over-applies in *dit weekeinde*, where the anaphor *dit* does not introduce a **known** or **given** referent. As we have remarked above, however, this use of anaphoric qualifier represents a minority of the actual cases attested in written text. Conversely, the **known**, or [-focus] status of *De sneeuw* is not signalled by an explicit 'known qualifier'⁵, and therefore *SNEEUW* is incorrectly accentuated. The noun *plaatsen* is preferably unaccentuated, probably since it is semantically fairly empty. However, accentuating it does not produce a severe error. Finally, PROS leaves the verb *slaagde* unaccentuated, as it does not detect the prosodic boundary after *er in*. Had this boundary been detected, then *slaagde* would have been the only CW in a prosodic domain, and it would have been accentuated (by a rule that has not been discussed here).

5. Conclusions

PROS represents an attempt to arrive at fairly acceptable sentence accentuation of raw text. In this respect, it performs considerably better than algorithms using only lexical information, viz. FW-CW distinctions and the accentuation properties of categories and individual words. In addition to lexical information, PROS takes into account accentuation properties of sequences of words. This is achieved by means of probabilistic rules; the principal ideas behind these rules, however, are well-grounded in theoretical descriptions of sentence accentuation and prosodic phrase structure.

Notes

1 The ordering between the first two clauses of (6a) reflects the preference for analysing A-P-A into A-(PA) over (AP)-A. A sequence of Subject-VERB-Object will be organised into two focus domains: the Subject, and the predicate phrase.

2 The gloss reads: "The search on [the island of] Terschelling for the oil-polluted birds has been seriously hampered during the weekend by the heavy snowfall on saturday night and sunday. The snow piled up to a level of some 40 centimeters. Also, on Terschelling in particular, much hindrance was experienced from ice-formation. This kept bird watch co-operators from reaching the oil victim's sites. In spite of the severe weather conditions, one succeeded in catching 30 eiderducks."

3 The PARTICIPLES of (16b) are accentuated (in spite of their being VERBs) by rule (19).

4 We have not indicated phrase boundaries.

5 The definite article *De* is too weak an indicator of [-focus] status.

References

- ALLEN, J., M.S. HUNNICUTT, and D. KLATT
1987 *From Text to Speech: The MITalk system*, Cambridge, UK: Cambridge University Press
- BAART, J.L.G.
1987 *Focus, Syntax, and Accent Placement*, dissertation Rijksuniversiteit Leiden
- BAART, J.L.G. and J. HEEMSKERK
1988 The problem of ambiguity in morphological analysis for a Dutch text-to-speech system, *Proceedings SPEECH'88* (7th FASE symposium), Edinburgh, 959-965
- BERENDSEN, E., S. LANGEWEG, and H. VAN LEEUWEN
1986 Computational phonology: Merged, not mixed, *Proceedings COLING'86*, Bonn, 612-615

- BEZOOIJEN, R. van
 1989 Evaluation of an algorithm for the automatic assignment of sentence accents, in H. Bennis and A. van Kemenade (ed) *Linguistics in the Netherlands 1989*, Dordrecht: Foris (this volume)
- DAELEMANS, W.
 1987 *Studies in language technology: An object-oriented computer model of morpho-phonological aspects of Dutch*, Ph.D. thesis Leuven University
- GUSSENHOVEN, C.
 1983 *On the grammar and semantics of sentence accents*, Dordrecht: Foris
- 't HART, J. and R. COLLIER
 1975 Integrating different levels of intonation analysis, *Journal of Phonetics* 3:235-255
- KAGER, R. and H. QUENÉ
 1987 Deriving prosodic sentence structure without exhaustive syntactic analysis, in J. Laver and M.J. Jack (eds) *Proceedings European Conference on Speech Technology*, Edinburgh, 1:243-246
- KAGER, R. and E. VISCH
 1987 Metrical constituency and rhythmic adjustment, *Phonology* 5:21-71
- KERKHOFF, J., J. WESTER, and L. BOVES
 1984 A compiler for implementing the linguistic phase of a text-to-speech conversion system, in H. Bennis and W.U.S. van Lessen Kloeke (eds) *Linguistics in the Netherlands 1984*, Dordrecht: Foris, 111-117
- KRUYT, J.G.
 1985 *Accents from speakers to listeners. An experimental study of the production and perception of accent patterns in Dutch*, dissertation Rijkuniversiteit Leiden.
- LAMMENS, J.
 1987 A lexicon-based grapheme-to-phoneme conversion system, in J. Laver and M.A. Jack (eds) *Proceedings European Conference on Speech Technology*, Edinburgh, 1:281-284
- NOOTEBOOM, S.G., and J.G. KRUYT
 1987 Accents, focus distribution, and the perceived distribution of given and new information: An experiment, *Journal of the Acoustic Society of America* 82(5):1512-1524
- O'SHAUGHNESSY, D.
 1976 *Modelling fundamental frequency and its relationship to syntax, semantics, and phonetics*, Ph.D. thesis, M.I.T., Cambridge MA
- TERKEN, J.M.B.
 1985 *Use and function of intonation: Some experiments*, dissertation Rijksuniversiteit Leiden
- VAN WIJK, C., and G. KEMPEN
 1985 From sentence structure to intonation contour, in B.S. Muller (eds) *Sprachsynthese: Zur Synthese von natürlich gesprochener Sprache aus Texten und Konzepten*. Hildesheim: Georg Olms, 157-182