Chapter 3

Phonological Phrasing

3.1 Introduction

Given the research objective to design, build, and test a linguistically-based SLG system for generating the most acceptable pitch accents and intonational boundaries, a preliminary grammar for pitch accent placement was presented in Section 2.5.4. Part of this grammar is an accent deletion rule called the *Rhythm Rule* (RR).

(1) **Rhythm Rule** (RR)

$$* \;\rightarrow\; \emptyset \;/\; [_{\varphi} \;\; * \;\; \_ \;\; * \;\;]$$

This rule deletes an accent occurring between two other accents if all of them are within the same phonological phrase (φ) (Section 2.2). It accounts for the observation that accentuation is influenced by rhythmic considerations, that is, some accents are deleted to obtain a more rhythmical pattern of accented and unaccented syllables. We will follow the strong hypothesis that the rule is applied both obligatory and exhaustively.

RR crucially depends on starting from the right φ's. Consider the example in (2). We assume that the whole phrase is focused, which will be the default assumption for all examples in this chapter. As a result, all content words are initially accented. Given this accent distribution, RR correctly deletes the accent on *afkomstige* (as indicated by ø) under the assumption that the whole phrase is a φ (2-a). However, if we were to assume that *afkomstige* is followed by a φ-boundary, then RR would have been blocked, and would have failed to delete the accent (2-b). This type of failure will be called an *undershoot error*.

(2) a. [$_\varphi$ een uit N$\overset{*}{\text{e}}$derland afk$\overset{ø}{\text{o}}$mstige pl$\overset{*}{\text{a}}$nt]
a from Netherlands originating plant
'a plant originating from the Netherlands'

b. ?[$_\varphi$ een uit N$\overset{*}{\text{e}}$derland afk$\overset{*}{\text{o}}$mstige] [$_\varphi$ pl$\overset{*}{\text{a}}$nt]

Analogously, the accent on *Nederlandse* is correctly preserved in (3-a) as long as a φ-boundary after *zeldzame* is assumed. Without this boundary, RR would have deleted the accent (3-b), and *Nederlandse* could only have been interpreted as unfocused, which contradicts our default assumption that the whole phrase is focused. This will be called an *overshoot error*.

(3) a. [$_\varphi$ een z$\overset{*}{\text{e}}$ldzame] [$_\varphi$ N$\overset{*}{\text{e}}$derlandse pl$\overset{*}{\text{a}}$nt]
a rare Dutch plant

b. ?[$_\varphi$ een z$\overset{*}{\text{e}}$ldzame N$\overset{ø}{\text{e}}$derlandse pl$\overset{*}{\text{a}}$nt]

Evidently, our algorithm for pitch accent placement using RR requires the φ-boundaries to be correctly determined. In this chapter, therefore, we will investigate phonological phrasing in Dutch. We will start in Section 3.2 with reviewing some of the better-known proposals for φ-phrasing, arguing that these are inadequate for Dutch. On the basis of a careful examination of examples of φ-phrasing in a number of syntactic structures, we will propose a new φ-phrasing algorithm (for Dutch) that is observationally adequate in Section 3.3. Next, we will report on a corpus-based evaluation of this algorithm in Section 3.4. We will conclude with a summary in Section 3.5.

3.2 A review of phonological phrase definitions

There are quite a few proposals for the derivation of φ's. Some are intended to be language-specific, e.g. Ferreira (1993), Gee and Grosjean (1983), Hayes (1986), Selkirk (1984, 1996), van Wijk (1987a), Zec and Inkelas (1990) for English and Neijt (1984) for Dutch, whereas others are claimed to be universal (Nespor and Vogel 1986).[1] We will first discuss some preliminary considerations with respect to the syntactic input and the evidence for φ-phrasing. Next, we will review two representative proposals, by Nespor and Vogel (1986) and by Hayes (1986), and show that these are inadequate in the case of Dutch.

3.2.1 Preliminaries

Recall from Section 2.2 that prosodic constituents are derived from the morpho-syntactic and semantic structure of an utterance by means of mapping rules. Nearly all φ-phrasing algorithms require as input a syntactic tree based on some version of $\overline{X}$-theory (Jackendoff 1977); semantic information appears to

[1] See Inkelas and Zec (1990) for phonological phrasing in some other languages.

be unnecessary for φ-phrasing. Although different authors may assume slightly different syntactic structures for certain inputs, most algorithms are insensitive to this variation in analysis, since they refer only to noncontroversial categories, like lexical categories (X), phrases (XP), and their dominance relations. The syntactic structures assumed in the current work (Section 2.6) provide sufficient information for all algorithms discussed here.

There are two types of evidence for prosodic constituents. First, there are phonological rules that are sensitive to the edges of a certain prosodic category (Section 2.2). Nespor and Vogel (1986) mention a number of segmental rules – for different languages – that apply within a φ, but not across a φ-boundary. For Dutch, however, no segmental rules are known that apply *exclusively* within φ's. To cite (Booij 1995, p. 146) in his survey of Dutch phonology: "For Dutch, the issue of the prosodic domains of rules above the level of the prosodic word is an underresearched area. [... T]he application of P-rules [EM phonological rules conditioned by phonological conditions only] above the word level decreases as the relevant prosodic domain of application becomes larger. In other words, I do not assign a particular prosodic domain to P-rules [...]". Hence, we will not use this type of evidence.[2]

Second, prosodic phonology predicts that the higher the rank of a prosodic category, the larger the amount of preboundary lengthening and the likelihood of a pause. Phonetic evidence for φ's in Dutch in the form of preboundary lengthening can be found in Hofhuis, Gussenhoven, and Rietveld (1995), and Gussenhoven and Rietveld (1992a), although this work addresses only a few (uncontroversial) φ-boundaries. Phonetic evidence may become stronger with the advent of Dutch speech corpora.

In the absence of reliable segmental evidence or easily obtainable phonetic evidence, the only clear evidence we have is the applicability of RR. This then, brings us to some remarks about using RR as a diagnostic tool for φ-phrasing. First, its application depends to a certain degree on speaking style and speaking rate. In slow and careful speech, it may apply optionally, while in fast and sloppy speech, it may at times even cross φ-boundaries (Hayes 1986, p. 215). Second, application may be to be sensitive to the spacing between accented syllables. Hayes (1984) claims the existence of *eurhythmy*, the tendency to produce equal spacings of accents. Gussenhoven (1991a) acknowledges the existence of eurhythmy, but argues that its role is rather limited. Bearing the above in mind, we will assume that all examples are spoken at an average speaking rate, abstracting away from the possible effect of very fast or slow speaking. Furthermore, we will as much as possible compare alternative φ-phrasings and the resulting accentuations, using the symbol '?' to the mark the one that is comparatively less acceptable. In other words, the claim is not that the questionable examples are wrong, but rather that they are less acceptable compared with the preferred ones. Finally, unless mentioned otherwise, all examples are assumed to have broad focus.

[2] However, see Loots (1983) and Menert (1994).

3.2.2 Phonological phrasing according to Nespor & Vogel

Nespor and Vogel (1986) propose the following algorithm to derive φ's.

(4) **Phonological phrase formation**

a. *φ-domain*
The domain of φ consists of a C which contains a lexical head (X) and all C's on its nonrecursive side up to the C that contains another head outside of the maximal projection of X.

b. *φ-construction*
Join into an n-ary branching φ all C's included in a string delimited by the definition of φ.

The letter 'C' stands for *Clitic Group*, which is a content word with optionally attached clitic words. It is irrelevant for the current discussion[3], so it can be interpreted as 'word'.

When we try to apply this algorithm to Dutch, we encounter a number of problems. The first problem is also recognised by Nespor and Vogel (1986, p. 186, footnote 2): Dutch syntax shows both left-recursive and right-recursive structures, so it is not clear what the nonrecursive side should be. Similar to an English NP, a Dutch NP can have an arbitrary number of embedded PP's and relative clauses to the right of its head, just as in English. Assuming that Dutch is left-recursive predicts that all these PP's and S's belong to one and the same φ, which is of course an unintended result. We are therefore forced to assume that the left side is the nonrecursive one.

Another minor issue involves the definition of lexical head. According to Nespor & Vogel, only V, N and A are lexical heads. Even if we are prepared to categorise adverbs as A, there remain other other content words that trigger a φ-boundary. Example (5) contains two numerals, both of which are capable of carrying two pitch accents. Without counting numerals as lexical heads, we get the huge φ in (5-b), in which RR can apply three times, resulting in a totally unacceptable accent distribution.[4] In contrast, counting numerals as lexical heads results in an acceptable accent distribution (5-c).

(5) a. van de ($_{Num}$ vijfennegentig) hebben we er ($_{Num}$
of the five-and-ninety have we there
tweeenveertig) afgewezen
two-and-forty rejected
'out of the 95, we rejected 42'

b. *[$_{\varphi}$ van de vijfennegentig hebben we er tweeenveertig afgewezen]

[3]However, the relation between clitics and φ's will be discussed in Section 3.2.3.

[4]It is possible to claim that both numerals modify a nominal head (N) that is empty, and that empty heads still count as lexical heads for the purpose of phonological phrasing. However, the assumption that empty elements (traces) affect the phonological structure is explicitly rejected by Nespor and Vogel (1986, ch. 2).

c. [$_\varphi$ van de v$\overset{*}{\mathrm{i}}$jfenn$\overset{*}{\mathrm{e}}$gentig] [$_\varphi$ hebben we er tw$\overset{*}{\mathrm{e}}$eenv$\overset{*}{\mathrm{e}}$ertig]
[$_\varphi$ $\overset{*}{\mathrm{a}}$fgewezen]

Yet another problem involves words that are not included in a φ domain and consequently are not dominated by a φ node at all. In Dutch – an SOV language – auxiliary verbs can occur in the sentence-final verb cluster; see *zou*, *moeten* en *hebben* in (6). Since auxiliaries do not count as lexical heads, they do not trigger a φ-domain of their own. Neither are they included in the φ-domain of the main verb *gezien*, because they reside at its recursive side. As a consequence, the auxiliaries remain unparsed (6-b). This then, violates the exhaustive parsing condition on prosodic constituent structures (Nespor and Vogel 1986, p. 7, Principle 2).

(6) a. … dat ik dat ($_N$ probleem) ($_V$ gezien) zou moeten
that I that problem seen should must
hebben
have
'… that I should have seen that problem'

b. … dat ik dat probleem] [$_\varphi$ gezien] zou moeten hebben

The most serious problem, however, is one of indeterminacy. Consider (7), consisting of a lexical head V, followed by an NP containing a lexical head A and a lexical head N. Deriving the φ-domain for *rode* results in (7-b), while deriving the φ-domain for *bloem* gives (7-c). Clearly, these φ-domains are incompatible with the Strict Layer Hypothesis (Section 2.2), which prevents recursive embedding of prosodic constituents of the same category.

(7) a. de plant ($_V$ draagt) ($_{NP}$ ($_{Det}$ een) ($_{AP}$ ($_A$ rode)) ($_N$
the plant carries a red
bloem))
flower

b. de plant draagt [$_\varphi$ een rode] bloem

c. de plant draagt [$_\varphi$ een rode bloem]

Now assume, for the sake of argument, that we always give priority the *largest* φ-domain. So in (7), we take (7-c) rather than (7-b). This implies, however, that the φ-domain for a lexical N includes *all* of its premodifiers. For example, the NP in (8) would be one big φ. Moreover, as the number of premodifiers for a noun is in theory unlimited, it follows that the length of a φ is in theory unlimited as well. This conclusion is absurd, as it predicts that accents on *voltallige* and *voorzitter* would suffice to focus the whole NP. A similar argument can be made for the premodifiers of V.

(8) *[$_\varphi$ de door de voltallige raad van bestuur met
the by the entire board of management with
meerderheid van stemmen gekozen nieuwe voorzitter]
majority of votes elected new chairman

To sum up so far, the φ-algorithm by Nespor & Vogel is inadequate for Dutch because it is not evident what the nonrecursive side should be, the class of lexical heads is to narrow, and sentence final auxiliaries are left unparsed. Moreover, the algorithm is non-deterministic and cannot be easily repaired.

In addition to algorithm (4), Nespor & Vogel postulate an optional restructuring rule.

(9) **φ-restructuring (optional)**

A nonbranching φ which is the first complement of X on its recursive side is joined into the φ that contains X.

Restructuring accounts for observations of stress shift (Nespor and Vogel 1986, p. 178). Stress shift is assumed to result from a φ-bound rule that moves the main stress position of a word backwards in case of a *stress clash*, i.e. when the main stress of a word is too close to the main stress of the next word. In (10), for example, *quickly* forms a nonbranching φ that is the first complement of *reproduce* on its recursive side, and is therefore joined in a single φ with *reproduce*. Moreover, the normal main stress of *reproduce* clashes with the main stress of *quickly*. As a consequence of these two circumstances, stress shifts from the penultimate to the initial syllable of *reproduce*.[5]

(10) a. Given the chance, rabbits
($_{VP}$ ($_V$ reproduce) ($_{AdvP}$ ($_{Adv}$ quickly)))

b. Given the chance, rabbits [$_\varphi$ réproduce quíckly]

In contrast, the complement *very quickly* in (11) is branching, which prevents restructuring, and therefore stress shift is blocked.

(11) a. Given the chance, rabbits
($_{VP}$ ($_V$ reproduce) ($_{AdvP}$ ($_{Adv}$ very) ($_{Adv}$ quickly)))

b. Given the chance, rabbits [$_\varphi$ reprodúce] [$_\varphi$ véry quickly]

A similar pair of examples, involving a branching coordination of two adverbs, is given in (12) and (13).

(12) a. John ($_V$ perseveres) ($_{AdvP}$ ($_{Adv}$ gladly))

b. John [$_\varphi$ pérseveres gládly]

[5] Nespor & Vogel use an accentmark (´) to indicate primary stress.

(13) a. John ($_V$ perseveres) ($_{AdvP}$ ($_{Adv}$ gladly) and ($_{Adv}$ diligently))

b. John [$_\varphi$ persevéres] [$_\varphi$ gládly] [$_\varphi$ and diligently]

However, the restructuring option gives the wrong predictions for Dutch. First, observe that an adjective like *razendsnel* can carry two accents.

(14) die auto is r$\overset{*}{\text{a}}$zendsn$\overset{*}{\text{e}}$l
that car is super-fast

Second, if *razendsnel* directly precedes the accented verb *gereden*, its second accent may be deleted (15). Our explanation is that both words belong to the same φ, and that the second accent of *razendsnel* is deleted by RR.

(15) a. we hebben ($_A$ razendsnel) ($_V$ gereden)
we have super-fast driven

b. [$_\varphi$ we hebben r$\overset{*}{\text{a}}$zendsn$\overset{\varnothing}{\text{e}}$l ger$\overset{*}{\text{e}}$den]

Now consider (16), where the verb precedes *razendsnel* (cf. *reproduce quickly*). According to (9), *razendsnel* is a nonbranching φ which is the first complement of the verb on its recursive (right) side, and is therefore joined into the φ that contains the verb. This allows RR to delete the first accent of *razendsnel*. However, (16-b) is far less acceptable than (16-c), if not unacceptable. Apparently, the φ-boundary between between the verb and the adjective is obligatory, and cannot be subject to φ-restructuring.

(16) a. we ($_V$ rijden) ($_A$ razendsnel)
we drive super-fast

b. ?[$_\varphi$ we r$\overset{*}{\text{i}}$jden r$\overset{\varnothing}{\text{a}}$zendsn$\overset{*}{\text{e}}$l]

c. [$_\varphi$ we r$\overset{*}{\text{i}}$jden] [$_\varphi$ r$\overset{*}{\text{a}}$zendsn$\overset{*}{\text{e}}$l]

In conclusion, both the initial algorithm for φ-phrasing and the φ-restructuring rule proposed by Nespor & Vogel show serious defects when applied to Dutch.

3.2.3 Phonological phrasing according to Hayes

A somewhat different definition of the phonological phrase for English is proposed by Hayes (1986), where the formation of Clitic Groups (17) precedes the construction of Phonological Phrases (18).

(17) **Clitic Group Formation**

a. Every content word (lexical category) belongs to a separate Clitic Group.
b. Definition: The *host* of a Clitic Group is the content word it contains.
c. Definition: X and Y *share category membership* in C if C dominates both X and Y.
d. Rule: Clitic words are incorporated leftward or rightward into an adjacent Clitic Group. The group selected is the one in which the clitic shares more category membership with the host.

(18) **Phonological Phrase Construction**[6] (English)

In the configuration $(_{XP}$ ...X YP ...)
a. The sequence (...X) obligatory occupies the same φ,
b. YP may optionally adjoin to the φ of X if it contains only one Clitic Group, and
c. all Clitic Groups unaffected by rules (a) and (b) form φ's.

Applying these definitions to the syntactic structure in Figure 3.1 gives the C's in (19-a) and the φ's in (19-b).

(19) a. $[_{\mathrm{C}}$ he kept it] $[_{\mathrm{C}}$ in a large] $[_{\mathrm{C}}$ jar]

b. $[_{\varphi}$ he kept it] $[_{\varphi}$ in a large jar]

Hayes' definition of clitics is much broader than that of Nespor & Vogel: every function word counts as a clitic word. As a result, the problem with the unparsed auxiliaries in (6) does not occur, because they are incorporated in the Clitic Group of the content verb. At the same time, however, the notion of shared category membership introduces new problems with regard to the direction of incorporation. Consider the sentence in (20-a). It is important to notice that the syntactic structure above the PP is irrelevant with respect to Clitic Group Formation: whatever the VP and S structure may be, the preposition *in* will *always* share more category membership with the host *regen* than with *fietste*. Therefore, *fietste* and *in* must belong to different C's (20-b). However, this is contradicted by phonological evidence. First, the schwa of *fietste* can be deleted by the rule *Prevocalic Schwa Deletion* (20-c), which cannot apply across prosodic word (ω) boundaries (Booij 1995, p. 49), and by implication, not across a C boundary either. Second, resyllabification is likely, which can only occur within a ω.

[6]The rendering is slightly different from the original statement: φ instead of *P-phrase*, parenthesis instead of square brackets, XP and YP instead on X'' and Y'', and X instead of X^0. These notational variations do not alter its meaning.

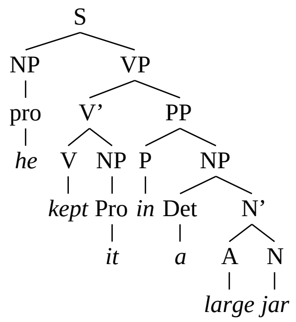

Figure 3.1: Syntactic tree for examples in (19).

(20) a. ze ($_V$ fietste) ($_{PP}$ ($_P$ in) ($_{NP}$ de regen))
she cycled in the rain

b. [$_C$ ze fietste] [$_C$ in de regen]

c. fietst in
/fitstə in/ ⟶ /fit-stin/

Analogously, the predicted C's for sentence (21-a) are given in (21-b), but are contradicted by the fact that resyllabification is certainly allowed here (21-c).[7]

(21) a. ik ($_V$ zie) ($_{NP}$ ($_{Det}$ het) ($_N$ water)) verdampen
I see the water evaporate

b. [$_C$ ik zie] [$_C$ het verdampen [

c. zie het
/si ət/ ⟶ /si-jət/

In conclusion, Hayes' algorithm for Clitic Group Formation based on maximal shared membership leads to false predictions for Dutch. Since Phonological Phrase Construction crucially depends on Clitic Group Formation, it follows that Phonological Phrase Construction makes false predictions as well.[8]

In support of his φ-construction algorithm, Hayes gives the following examples of stress shift.

(22) a. ($_{NP}$ ($_{AP}$ Jápanese) ($_N$ connéctions))

[7] The /j/ is inserted by the rule *Homorganic Glide Insertion*, which is normally applied within ω's, but appears to be across constituents higher as well in casual speech (Booij 1995, p. 150).

[8] This, however, is difficult to prove by means of RR, because clitics are virtually always unaccented.

b. ($_{NP}$ ($_{NP}$ Tóscanini's) ($_{N}$ íce cream))

c. ($_{AP}$ ($_{Adv}$ évidently) ($_{A}$ trúe))

d. he'll ($_{VP}$ ($_{Adv}$ ábsolutely) ($_{V}$ flíp))

Clause (18-a) is very close, but not identical to Nespor & Vogel's algorithm. The difference is that Hayes' formulation does not suffer from indeterminacy. Consider example (23): substituting *true* for X in (18), we can deduce that the sequence *evidently true* must occupy the same φ; substituting *statement* instead, entails that the sequence *an evidently true statement* must occupy that same φ. Both results are compatible without the need for embedded φ's (23-b).

(23) a. ($_{NP}$ an ($_{AP}$ ($_{Adv}$ evidently) ($_{A}$ true)) ($_{N}$ statement))

b. [$_{\varphi}$ an évidently true státement]

As far as Dutch is concerned, however, (18-a) is still inadequate, because it predicts that all modifiers preceding the head of an NP occupy the same φ. As mentioned earlier in relation to the Nespor & Vogel algorithm, this results in improbably long φ's (cf. example (8)) and overapplication of RR.[9] The same problem occurs with respect to a VP: according to (18-a), the sequence ($_{VP}$ ... V) in (24) obligatory occupies the same φ.

(24) a. ($_{S}$ dat hij ($_{VP}$ ($_{PP}$ door de voltallige raad van
that he by the entire board of
bestuur) ($_{PP}$ met meerderheid van stemmen) ($_{PP}$ tot
management with majority of votes to
nieuwe voorzitter) is ($_{V}$ gekozen))$_{VP}$)$_{S}$
new chairman is elected

b. *[$_{\varphi}$ dat hij door de voltallige raad van bestuur met meerderheid van stemmen tot nieuwe voorzitter is gekozen]

Clause (18-b) is in effect identical to Nespor & Vogel's optional restructuring rule (9).[10] Hayes gives the English examples in (25), which are similar to (10) and (12).

(25) a. ($_{VP}$ ($_{V}$ cómprehending) ($_{NP}$ éverything))

b. it'll ($_{V}$ ínterfere) ($_{PP}$ with télevision)

c. to ($_{V}$ íntervene) ($_{AdvP}$ intélligently)

[9] In Hayes' view, the spacing between two accented syllables is also relevant for the application of RR. Thus, a large φ does not necessarily imply overapplication of RR. We will show in the next section, however, that even in a short sequence of premodifiers, RR is often not applicable.

[10] Notice, however, that Hayes' formulation is more precise, as it avoids the notion *non-branching*. In (25-b), for instance, the PP *with television* is – strictly speaking – binary branching.

d. ($_{NP}$ the ($_N$ jápanese) ($_{PP}$ of Hónshu))

We have already argued that the Dutch equivalent of (25-c) is not a single φ (cf. *rijden razendsnel* in (10)). In addition, the Dutch equivalent of (25-d) is almost never a single φ either. Consider (26), for example, where the accent on *porselein* cannot be deleted without altering the focus interpretation.

(26) ($_{NP}$ orgine*el ($_N$ porsele*in) ($_{PP}$ uit De*lft))
original porcelain from Delft

This appears to hold for nearly all postmodifying PP's in Dutch, with the exception of a some (idiomatic) expressions like (27), where *lagen* is deaccented.

(27) ($_{PP}$ in ($_{NP}$ bre*de ($_N$ la∅gen) ($_{PP}$ van de bevo*lking)))
in broad layers of the population
'in large sections of the population'

In conclusion, the φ-algorithm by Hayes is not adequate for Dutch either. The problems appear to be caused by the syntactic differences between English and Dutch. First, Dutch offers more complex possibilities for premodifying a noun, spanning more than one φ, and has a right-headed VP structure. Second, nonbranching – or single C – postmodifiers are almost never merged with the preceding φ. Third, C-formation based on shared membership contradicts phonological evidence.

3.3 The phonological phrase in Dutch

In this section, a better φ-algorithm for Dutch will be proposed. We will first present a preliminary version of the algorithm. Next, we will show that it accounts well for φ-phrasing in many different syntactic configurations, but is not yet entirely adequate. This motivates the subsequent revision of the algorithm.

3.3.1 Preliminary algorithm

We start with a preliminary formulation of φ-phrasing.

(28) **Phonological phrase construction for Dutch** (1st version)

A φ-boundary occurs after every content word that is not an adjacent modifier.

(29) **Adjacent modification** (1st version)

X is an *adjacent modifier* if and only if there is a head such that both

a. X is left-adjacent to Y, and
b. X is dominated by YP.

The class of content words excludes function words belonging to a closed class, such as determiners, complementizers, prepositions, auxiliaries, pronouns, conjunctions, etc. The notion of adjacent modification can be explained by means of example (30). In (30-a), the adjective *gemetselde* is both adjacent to the noun *schoorsteen* and dominated by the maximal projection of *schoorsteen* (NP), and is therefore an adjacent modifier. In contrast, the adjective *onhandige* is adjacent to the adjective *gemetselde*, but not dominated by the maximal projection of *gemetselde* (AP), and therefore *not* an adjacent modifier. We can contrast this with (30-b), where *gemetselde* is again an adjacent modifier of *schoorsteen*, but in addition the adjective *onhandig* is an adjacent modifier of *gemetselde*.[11]

(30) a. ($_{NP}$ ($_{Det}$ een) ($_{AP}$ ($_{A}$ onhandige)) ($_{AP}$ ($_{A}$ gemetselde))
a awkward bricked
($_{N}$ schoorsteen))
chimney
'a chimney that is awkward and made out of bricks'

b. ($_{NP}$ ($_{Det}$ een) ($_{AP}$ ($_{A}$ onhandig) ($_{A}$ gemetselde)) ($_{N}$
a clumsy bricked
schoorsteen))
chimney
'a chimney that has been clumsily bricked'

The relation of adjacent modification is related to the inflection of adjectives. In Dutch, attributive adjectives are inflected with the suffix *-e*, except when modifying an indefinite neuter singular noun.[12] In (30-a), both *onhandige* and *gemetselde* are inflected. In (30-b), only *gemetselde* is inflected, as *onhandig* is adverbially used, in which case it is never inflected. In the particular case of a noun that is indefinite, neuter and singular, however, a modifying adjective is not inflected either and an NP can become syntactically ambiguous as in (31).

(31) a. ($_{NP}$ ($_{Det}$ een) ($_{AP}$ ($_{A}$ onhandig)) ($_{AP}$ ($_{A}$ gemetseld)) ($_{N}$
a awkward brick
schoorsteentje))
chimney + DIM
'a small chimney that is awkward and made out of bricks'

[11] Most adjectives in Dutch can also be used adverbially to modify a verb or another adjective.

[12] Although this is certainly the most frequent pattern, it is an oversimplification. See Haeseryn et al. (1997, ch. 6.4) for exceptions.

b. $(_{NP}\ (_{Det}$ een) $(_{AP}\ (_{A}$ onhandig) $(_{A}$ gemetseld)) $(_{N}$
a clumsily bricked
schoorsteentje))
chimney + DIM
'a small chimney that has been clumsily bricked'

Interestingly, the two readings correspond to different phonological phrasings, resulting in different accent distributions. Although the adjective *onhandig* is a content word, it is not an adjacent modifier in (31-a), and is followed by a φ-boundary according to definition (28). Therefore, its accent cannot be deleted by RR (32-a). In contrast, *onhandig* is both a content word and an adjacent modifier in (31-b), thus no φ-boundary is predicted. Hence, RR deletes the accent on *gemetseld* (32-b). In short, the two sentences in (32) form a minimal pair, where the relation between accentuation and interpretation can be explained in terms of φ-phrasing and RR.

(32) a. $[_{\varphi}$ een onh$\overset{*}{\text{a}}$ndig] $[_{\varphi}$ gem$\overset{*}{\text{e}}$tseld scho$\overset{*}{\text{o}}$rsteentje]

b. $[_{\varphi}$ een onh$\overset{*}{\text{a}}$ndig gem$\overset{\text{ø}}{\text{e}}$tseld scho$\overset{*}{\text{o}}$rsteentje]

This contrast cannot be explained by the algorithms reviewed earlier. Additionally, we will to show that our proposal can cope with the types of modification that the reviewed algorithms could not. In order to demonstrate that it yields all and only correct φ's, we need to do two things. First, we have to investigate all syntactic structures with an adjacent modification relation, i.e. of the type $(_{YP} \ldots (_{X}\ )\ (_{Y}\ ) \ldots)$, and show that the RR can apply freely, i.e. across X and Y. Second, we have to show that in all syntactic structures where a content word is *not* an adjacent modifier, application of RR is blocked. Of course, investigating all relevant syntactic structures is practically impossible, for one because the number of structures that are distinguished is to some extent dependent on the particular syntactic theory adopted. Nevertheless, we will show in the next two sections that our preliminary algorithm based on the notion of adjacent modification makes the correct predictions in many different syntactic structures.

3.3.2 Phonological phrasing in NP, AP and AdvP

3.3.2.1 Adjacent modifiers preventing a φ-boundary

We will start with adjacent modifiers before a noun. Example (33) shows an adjective premodifying a noun: (33-a) shows the syntactic structure; (33-b) shows the predicted φ-phrasing and the accent distribution after RR has applied; (33-c) shows that the compound adjective *ijskoud* can indeed carry two accents in an appropriate context. Example (34) does the same for a noun premodifying another noun.[13]

[13] The structure in which an adverb premodifies a noun, i.e. $(_{NP} \ldots (_{Adv}\ )\ (_{N}\ ) \ldots)$ appears to be non-existent in Dutch. A phrase like *jouw keuze is een altijd gelukkige* 'your

(33) $({}_{NP} \ldots ({}_{A}\ )\ ({}_{N}\ )\ \ldots)$

a. $({}_{NP}\ ({}_{AP}\ ({}_{A}$ ijskoud)) $({}_{N}$ bier))
ice-cold beer

b. $[{}_{\varphi}$ $\overset{*}{\text{i}}$jsk$\overset{\varnothing}{\text{o}}$ud b$\overset{*}{\text{i}}$er]

c. het b$\overset{*}{\text{i}}$er is $\overset{*}{\text{i}}$jsk$\overset{*}{\text{o}}$ud
the beer is ice-cold

(34) $({}_{NP} \ldots ({}_{N}\ )\ ({}_{N}\ \ldots\ )$

a. $({}_{NP}\ ({}_{NP}$ een halve $({}_{N}$ kilo)) $({}_{N}$ kaas))
a half kilo cheese

b. $[{}_{\varphi}$ een h$\overset{*}{\text{a}}$lve k$\overset{\varnothing}{\text{i}}$lo ka$\overset{*}{\text{a}}$s]

c. graag een h$\overset{*}{\text{a}}$lve k$\overset{*}{\text{i}}$lo
please a half kilo

The next three examples explore the predictions with respect to an adjacent modifier before an adjective.

(35) $({}_{AP} \ldots ({}_{A}\ )\ ({}_{A}\ )\ \ldots)$

a. $({}_{NP}$ een $({}_{AP}\ ({}_{AP}\ ({}_{A}$ lichamelijk)) $({}_{A}$ gehandicapt)) $({}_{N}$
a physically handicapped
kind))
child

b. $[{}_{\varphi}$ een lich$\overset{*}{\text{a}}$melijk geh$\overset{\varnothing}{\text{a}}$ndicapt k$\overset{*}{\text{i}}$nd]

c. het k$\overset{*}{\text{i}}$nd is lich$\overset{*}{\text{a}}$melijk geh$\overset{*}{\text{a}}$ndicapt
the child is physically handicapped

(36) $({}_{AP} \ldots ({}_{Adv}\ )\ ({}_{A}\ )\ \ldots)$

a. $({}_{NP}$ een $({}_{AP}\ ({}_{Adv}$ erg) $({}_{A}$ mooie)) $({}_{N}$ tuin))
a very beautiful garden

b. $[{}_{\varphi}$ een $\overset{*}{\text{e}}$rg mo$\overset{\varnothing}{\text{o}}$ie t$\overset{*}{\text{u}}$in]

c. hun tuin is $\overset{*}{\text{e}}$rg mo$\overset{*}{\text{o}}$i
their garden is very beautiful

choice is an always lucky one' might appear to fit this pattern. However, we would argue that *gelukkkige* is an adjective, as indicated by its inflection, modifying an empty N.

(37) $(_{AP} \ldots (_{N}\ )\ (_{A}\ )\ \ldots)$

a. $(_{NP}$ de $(_{AP}$ $(_{NP}$ $(_{N}$ maandag)) $(_{A}$ aangekondigde)) $(_{N}$ maatregelen))
the Monday announced measures

b. $[_{\varphi}$ de m$\overset{*}{a}$andag aangek$\overset{ø}{o}$ndigde m$\overset{*}{a}$atregelen]

c. m$\overset{*}{a}$atregelen werden m$\overset{*}{a}$andag aangek$\overset{*}{o}$ndigd
measures were Monday announced
'measures were announced on Monday'

What remains is the case of an adjacent modifier before an adverb. Again, one of the logical possibilities, the structure $(_{AdvP}\ (_{A}\ )\ (_{Adv}\ )\ \ldots\ )$, would appear to be lacking in Dutch.

(38) $(_{AdvP} \ldots (_{Adv}\ )\ (_{Adv}\ )\ \ldots\ )$

a. $(_{AP}$ $(_{AdvP}$ $(_{AdvP}$ $(_{Adv}$ erg)) $(_{Adv}$ vaak)) $(_{A}$ ziek))
very often ill

b. $[_{\varphi}$ $\overset{*}{e}$rg v$\overset{ø}{a}$ak zi$\overset{*}{e}$k]

c. hij sliep $\overset{*}{e}$rg v$\overset{*}{a}$ak
he slept very often

(39) $(_{AdvP} \ldots (_{N}\ )\ (_{Adv}\ )\ \ldots\ )$

a. $(_{AdvP}$ $(_{NP}$ een $(_{A}$ heel) $(_{N}$ eind)) $(_{Adv}$ verder))
a whole piece further
'a whole lot further'

b. $[_{\varphi}$ een h$\overset{*}{e}$el $\overset{ø}{e}$ind v$\overset{*}{e}$rder]

c. naar de top bleek een h$\overset{*}{e}$el $\overset{*}{e}$ind
to the top proved a whole piece
'to the top turned out to be a long way'

So far, we have looked at relatively simple syntactic structures. More complex examples are given in (40) and (41). Notice that in the latter RR applies twice.

(40) a. $(_{NP}$ de $(_{AP}$ $(_{PP}$ van $(_{NP}$ $(_{N}$ water))) $(_{A}$ afhankelijke)) $(_{N}$ eenden))
the on water dependent ducks
'the ducks which depend on water'

b. [$_{\varphi}$ de van w$\overset{*}{\text{a}}$ter afh$\overset{\text{ø}}{\text{a}}$nkelijke $\overset{*}{\text{e}}$enden]

(41) a. ($_{NP}$ de ($_{AP}$ ($_{PP}$ op ($_{NP}$ ($_{AP}$ ($_{A}$ goed)) ($_{N}$ geluk)))) ($_{A}$
the at good luck
gekozen)) ($_{N}$ deelnemers))
chosen participants
'the participants (who were) chosen hoping for the best'

b. [$_{\varphi}$ de op go$\overset{*}{\text{e}}$d gel$\overset{\text{ø}}{\text{u}}$k gek$\overset{\text{ø}}{\text{o}}$zen d$\overset{*}{\text{e}}$elnemers]

3.3.2.2 Non-adjacent modifiers triggering a φ-boundary

We now turn to syntactic configurations where a content word is *not* an adjacent modifier and therefore triggers the insertion of a φ-boundary. As far as NP's are concerned, these involve coordinations and postmodifiers. Example (42) has two coordinated AP's, showing that deletion of the second accent is incompatible with a full focus interpretation (42-c). Example (43) has two coordinated AP's of which the first contains the compound adjective *peperduur*, which can carry two accents. Deletion of its second accent is less unacceptable (43-c).

(42) a. ($_{NP}$ ($_{AP}$ ($_{A}$ eindeloze)) ($_{AP}$ ($_{A}$ saaie)) ($_{N}$ vlaktes))
endless boring plains

b. [$_{\varphi}$ $\overset{*}{\text{e}}$indeloze] [$_{\varphi}$ sa$\overset{*}{\text{a}}$ie vl$\overset{*}{\text{a}}$ktes]

c. ?[$_{\varphi}$ $\overset{*}{\text{e}}$indeloze sa$\overset{\text{ø}}{\text{a}}$ie vl$\overset{*}{\text{a}}$ktes]

(43) a. ($_{NP}$ een ($_{AP}$ ($_{A}$ peperduur)) ($_{AP}$ ($_{A}$ nieuw)) ($_{N}$
a very-expensive new
speeltje))
toy

b. [$_{\varphi}$ een p$\overset{*}{\text{e}}$perdu$\overset{*}{\text{u}}$r] [$_{\varphi}$ ni$\overset{*}{\text{e}}$uw spe$\overset{*}{\text{e}}$ltje]

c. ?[$_{\varphi}$ een p$\overset{*}{\text{e}}$perdu$\overset{\text{ø}}{\text{u}}$r ni$\overset{\text{ø}}{\text{e}}$uw spe$\overset{*}{\text{e}}$ltje]

A case of two coordinated NP's is given in (44). While the noun *maandag* is an adjacent modifier of the adjective *gesloten*, the noun *zondag* is not, and is therefore followed by a φ-boundary. This correctly predicts that the accent on *gesloten* can be deleted, whereas that on *maandag* must be preserved.

(44) a. ($_{NP}$ een ($_{AP}$ ($_{NP}$ ($_{NP}$ ($_{N}$ zondag)) ($_{Conj}$ en) ($_{NP}$ ($_{N}$
a Sunday and
maandag))) ($_{A}$ gesloten)) ($_{N}$ winkel))
Monday closed shop
'a shop (which is) closed on Sundays and Mondays'

b. [$_{\varphi}$ een z$\overset{*}{\text{o}}$ndag] [$_{\varphi}$ en m$\overset{*}{\text{a}}$andag gesl$\overset{\emptyset}{\text{o}}$ten w$\overset{*}{\text{i}}$nkel]

c. ?[$_{\varphi}$ een z$\overset{*}{\text{o}}$ndag en m$\overset{\emptyset}{\text{a}}$andag gesl$\overset{\emptyset}{\text{o}}$ten w$\overset{*}{\text{i}}$nkel]

The data in (45) concerns two coordinated AP's of which the first is complex in the sense that it contains a PP modifier. Since *onsnapte* is not an adjacent modifier and must thus be followed by a φ-boundary, none of the accents in (45) can be deleted by RR.

(45) a. ($_{NP}$ de ($_{AP}$ ($_{PP}$ op ($_{NP}$ ($_{N}$ maandag))) ($_{A}$ ontsnapte))
the on Monday escaped
($_{AP}$ ($_{A}$ gevaarlijke)) ($_{N}$ criminelen))
dangerous criminals
'the dangerous criminals (who) escaped on Monday'

b. [$_{\varphi}$ de op m$\overset{*}{\text{a}}$andag ontsn$\overset{*}{\text{a}}$pte] [$_{\varphi}$ gev$\overset{*}{\text{a}}$arlijke crimin$\overset{*}{\text{e}}$len]

c. ?[$_{\varphi}$ de op m$\overset{*}{\text{a}}$andag ontsn$\overset{\emptyset}{\text{a}}$pte gev$\overset{\emptyset}{\text{a}}$arlijke crimin$\overset{*}{\text{e}}$len]

The last example involving coordination is a rather complex structure (46). The noun *huis* is adjacently modified by *verborgen*, which is itself adjacently modified by the PP *achter ... bomen*. Within this PP, the noun *bomen* is modified by the two coordinated adjectives *hoge* and *groene*. Since only *groene* is an adjacent modifier, we predict a φ-boundary after *hoge*. As result, only the accents on *bomen* and *verborgen* are in the scope of RR.

(46) a. ($_{NP}$ een ($_{AP}$ ($_{PP}$ achter ($_{NP}$ ($_{AP}$ ($_{A}$ hoge)) ($_{AP}$ ($_{A}$
a behind high
groene)) ($_{N}$ bomen))) ($_{A}$ verborgen)) ($_{N}$ huis))
green trees hidden huis
'a house (that was) hidden behind high green trees'

b. [$_{\varphi}$ een achter h$\overset{*}{\text{o}}$ge] [$_{\varphi}$ gro$\overset{*}{\text{e}}$ne b$\overset{\emptyset}{\text{o}}$men verb$\overset{\emptyset}{\text{o}}$rgen hu$\overset{*}{\text{i}}$s]

c. ?[$_{\varphi}$ een achter h$\overset{*}{\text{o}}$ge gro$\overset{\emptyset}{\text{e}}$ne b$\overset{\emptyset}{\text{o}}$men verb$\overset{\emptyset}{\text{o}}$rgen hu$\overset{*}{\text{i}}$s]

Postmodifiers of N's most frequently occur in the form of PP's and relative S's. These are shown to trigger a φ-boundary in (47) and (48) respectively.

(47) a. ($_{NP}$ een ($_{N}$ tent) ($_{PP}$ met lichtgewicht stokken))
a tent with lightweight poles

b. [$_\varphi$ een t$\overset{*}{\text{e}}$nt] [$_\varphi$ met l$\overset{*}{\text{i}}$chtgewicht st$\overset{*}{\text{o}}$kken]

c. ?[$_\varphi$ een t$\overset{*}{\text{e}}$nt met l$\overset{\o}{\text{i}}$chtgewicht st$\overset{*}{\text{o}}$kken]

(48) a. ($_{NP}$ een ($_N$ tent)) ($_S$ die ($_{NP}$ lichtgewicht stokken)
a tent that lightweight poles
heeft))
has

b. [$_\varphi$ een t$\overset{*}{\text{e}}$nt] [$_\varphi$ die l$\overset{*}{\text{i}}$chtgewicht st$\overset{*}{\text{o}}$kken heeft]

c. ?[$_\varphi$ een t$\overset{*}{\text{e}}$nt die l$\overset{\o}{\text{i}}$chtgewicht st$\overset{*}{\text{o}}$kken heeft]

3.3.3 Phonological phrasing in VP and S

It was already shown in example (15) that φ-phrasing and RR also occur at the level of VP's and clauses. Another example is given in (49). The compound adjective *ijskoud* can carry two accents (cf. example (33)). However, if it directly precedes a verb, its second accent can be deleted. This is in line with our expectations, as *ijskoud* adjacently modifies *drinken*, and does therefore not trigger a φ-boundary.

(49) a. bier moet men ($_{VP}$ ($_A$ ijskoud) ($_V$ drinken))
beer should one ice-cold drink
'one should drink beer cold as ice'

b. bier [$_\varphi$ moet men $\overset{*}{\text{i}}$jsk$\overset{\o}{\text{o}}$ud dr$\overset{*}{\text{i}}$nken]

c. bier [$_\varphi$ moet men $\overset{*}{\text{i}}$jsk$\overset{*}{\text{o}}$ud] [$_\varphi$ dr$\overset{*}{\text{i}}$nken]

Analogously to φ-phrasing in NP's, two subsequent modifiers are separated by a φ-boundary when both modify the verb (50), but are merged into the same φ when the first is an adjacent modifier of the second (51).

(50) a. koffie kan men ($_{VP}$ ($_A$ koud) ($_A$ onmogelijk) ($_V$
coffee can one cold impossibly
drinken))
drink
'it is impossible to drink coffee cold'

b. koffie [$_\varphi$ kan men k$\overset{*}{\text{o}}$ud] [$_\varphi$ onm$\overset{*}{\text{o}}$gelijk dr$\overset{*}{\text{i}}$nken]

c. ?koffie [$_\varphi$ kan men k$\overset{*}{\text{o}}$ud onm$\overset{\o}{\text{o}}$gelijk dr$\overset{*}{\text{i}}$nken]

(51) a. koffie moet men ($_{VP}$ ($_{AP}$ ($_{Adv}$ zeer) ($_{A}$ heet)) ($_{V}$
coffee must one very hot
drinken))
drink
'one should drink coffee very hot'

b. koffie [$_\varphi$ moet men z$\overset{*}{\text{e}}$er h$\overset{ø}{\text{e}}$et dr$\overset{*}{\text{i}}$nken]

c. ?koffie [$_\varphi$ moet men z$\overset{*}{\text{e}}$er h$\overset{*}{\text{e}}$et] [$_\varphi$ dr$\overset{*}{\text{i}}$nken]

Another example of this contrast is given in (52).

(52) a. ?onze software kunt u sn$\overset{*}{\text{e}}$l doeltr$\overset{ø}{\text{e}}$ffend gebr$\overset{*}{\text{u}}$iken
our software can you quickly effectively use
'you will quickly be able to use our software effectively'

b. onze software laat zich z$\overset{*}{\text{e}}$er doeltr$\overset{ø}{\text{e}}$ffend gebr$\overset{*}{\text{u}}$iken
our software lets itself very effectively use
'our software can be used very effectively'

These examples suggest that φ-phrasing in clauses is governed by the same principles as φ-phrasing in NP's. Nevertheless, there are some exceptions. To begin with, we can insert an auxiliary between an adjective and a main verb without affecting the φ-phrasing (53). This might be explained by assuming that the adjective is an adjacent modifier of the auxiliary, but since this is not the only problematic example, we will not further explore this option.

(53) a. bier moet ($_{VP}$ ($_{A}$ ijskoud) ($_{Aux}$ worden) ($_{V}$
beer should ice-cold be
gedronken))
drunk
'serve beer cold as ice'

b. bier [$_\varphi$ moet $\overset{*}{\text{i}}$jsk$\overset{ø}{\text{o}}$ud worden gedr$\overset{*}{\text{o}}$nken]

In addition, a noun that adjacently modifies a verb appears to resist incorporation in the same φ, even though its nominal head is an adjacent modifier of the verb; see the NP in (54) [14] and the PP in (55) respectively.

(54) a. goede voorbereiding kan ($_{VP}$ ($_{NP}$ nare verrassingen) ($_{V}$
good preparation can dismal surprises
voorkomen))
prevent
'good preparation can prevent nasty surprises'

[14] One might expect the predicate *voorkomen* to be unaccented as result of applying SAAR. However, even though it has an adjacent focused argument, the utterance is contingent rather than eventive, and therefore an exception to SAAR (Gussenhoven 1992b, Appendix).

b. goede voorbereiding [$_\varphi$ kan n$\overset{*}{\text{a}}$re verr$\overset{*}{\text{a}}$ssingen] [$_\varphi$ voork$\overset{*}{\text{o}}$men]

c. ?goede voorbereiding [$_\varphi$ kan n$\overset{*}{\text{a}}$re verr$\overset{\o}{\text{a}}$ssingen voork$\overset{*}{\text{o}}$men]

(55) a. Frankrijk wordt ($_{VP}$ ($_{PP}$ door Nederlandse toeristen) ($_V$
France is by Dutch tourists
gemeden))
avoided
'France is avoided by Dutch tourists'

b. Frankrijk [$_\varphi$ wordt door N$\overset{*}{\text{e}}$derlandse toer$\overset{*}{\text{i}}$sten] [$_\varphi$ gem$\overset{*}{\text{e}}$den]

c. ?Frankrijk [$_\varphi$ wordt door N$\overset{*}{\text{e}}$derlandse t$\overset{\o}{\text{o}}$eristen gem$\overset{*}{\text{e}}$den]

While all preceding examples had the finite verb in sentence-final position, its *base* or *deep-structure* position in the terminology of transformational grammar, it can also occur in the (syntactically) second position of a declarative main clause. The question arises if the changed position affects the φ-phrasing behaviour of the verb. The following sentences show the verb in second position preceded by an AP (56), an NP (57), and a PP (58) respectively. The resulting accent patterns suggest that the changed position is irrelevant to φ-phrasing: only an A merges with a following V into a single φ.

(56) a. ($_{AP}$ jarenlang) ($_V$ meed) hij haar
years-long avoided he her
'for years, he avoided her'

b. [$_\varphi$ j$\overset{*}{\text{a}}$renl$\overset{\o}{\text{a}}$ng m$\overset{*}{\text{e}}$ed hij haar]

(57) a. ($_{NP}$ vele jaren) ($_V$ sleet) hij daar
many years spent he there
'many a year he spent there'

b. ?[$_\varphi$ v$\overset{*}{\text{e}}$le j$\overset{\o}{\text{a}}$ren sl$\overset{*}{\text{e}}$et hij daar]

(58) a. ($_{PP}$ na vele jaren) ($_V$ schreef) hij haar
after many years wrote he her
'after many years, he wrote her'

b. ?[$_\varphi$ na v$\overset{*}{\text{e}}$le j$\overset{\o}{\text{a}}$ren schr$\overset{*}{\text{e}}$ef hij haar]

There is an additional complication here, however. Since the moved V is no longer the *head* of a VP, the preceding A is not dominated by the VP in (56), and therefore does not match the definition of an adjacent modifier.

Besides movement of the verb, Dutch syntax allows extraposition – rightward movement across the verb cluster – of S's, PP's and heavy NP's. Without exploring all combinations in depth, the examples in (59) and (60) show that our φ-algorithm correctly predicts that a φ-boundary is obligatory between a content verb and an extraposed constituent.

(59) a. voor een stoplicht moet men ($_{VP}$ ($_V$ wachten)) ($_{PP}$ op groen licht)
before a traffic-light must one wait for green light
'before a traffic light, one has to wait for the green light'

b. voor een stoplicht [$_\varphi$ moet men w$\overset{*}{a}$chten] [$_\varphi$ op gr$\overset{*}{o}$en l$\overset{*}{i}$cht]

c. ?voor een stoplicht [$_\varphi$ moet men w$\overset{*}{a}$chten op gr$\overset{ø}{o}$en l$\overset{*}{i}$cht]

(60) a. ze zijn ($_{VP}$ ($_V$ vertrokken)) ($_{PP}$ met al hun bezittingen)
they are left with all their possessions
'they left with all their possessions'

b. [$_\varphi$ ze zijn vertr$\overset{*}{o}$kken] [$_\varphi$ met $\overset{*}{a}$l hun bez$\overset{*}{i}$ttingen]

c. ?[$_\varphi$ ze zijn vertr$\overset{*}{o}$kken met $\overset{ø}{a}$l hun bez$\overset{*}{i}$ttingen]

We finish this discussion of φ-phrasing in S's and VP's by signalling a few problems that also plagued the other algorithms. First, our algorithm suffers from the problem of unparsed auxiliaries (cf. example (6)), because it fails to predict a sentence-final φ-boundary when the sentence ends with a function word. Moreover, without the notion of a Clitic Group, a similar problem occurs with clitics (61).

(61) a. ik ($_{VP}$ ($_V$ geef) 't aan 'm)
I give it to him

b. [$_\varphi$ ik geef] 't aan 'm

In general, as observed in (20) and (21), clitics may end up in the wrong φ because their host triggers a φ-boundary.

3.3.3.1 Revised algorithm

To sum up so far, we have shown that our preliminary proposal for φ-phrasing in Dutch based on the notion of adjacent modification works well in NP's, AP's and AdvP's. However, it proved to be not entirely adequate in VP's and S's, because it falsely predicts that an NP/PP and a following verb are merged into

Table 3.1: Syntactic structures in which no φ-boundary is inserted after a content word.

Left context:	*Word:*	*Right context:*	
$(_{NP} \ldots$	*A*	*N*	$\ldots)$
$(_{NP} \ldots$	*N*	*N*	$\ldots)$
$(_{AP} \ldots$	*A*	*A*	$\ldots)$
$(_{AP} \ldots$	*Adv*	*A*	$\ldots)$
$(_{AP} \ldots$	*N*	*A*	$\ldots)$
$(_{AdvP} \ldots$	*Adv*	*Adv*	$\ldots)$
$(_{AdvP} \ldots$	*N*	*Adv*	$\ldots)$
	A	Aux^* V	
	Adv	Aux^* V	
	V	Aux^+	
	V	$Clitic^+$	

one φ. In addition, it suffers from the unparsed auxiliaries problem, and from clitics that are either unparsed or in the wrong φ. Table 3.1 summarises our current state of knowledge.[15]

Our preliminary proposal needs to be revised. The problem with the unparsed auxiliaries and clitics can be solved by incorporating them into the φ of the preceding V, as will be proposed in Section 7.2.1. The observation that a N adjacently modifying a V still triggers a φ-boundary can be accounted for by assuming that a V cannot be adjacently modified at all, that is, by removing the option Y=V from the definition of adjacent modification (62). Interestingly, the observation that V differs from other lexical heads with respect to φ-phrasing is not unique to Dutch (Hayes and Lahiri 1991).

(62) **Adjacent modification** (final version)

For every $X, Y \in \{N, A, Adv\}$:
X is an *adjacent modifier* of Y if and only if

a. X is left-adjacent to Y, and
b. X is dominated by YP.

As a result, the fact that an A/Adv before a V does not trigger a boundary can no longer be accounted for in terms of adjacent modification, but must be enforced explicitly (63-b). However, this stipulation is needed anyway for an A/Adv preceding either an auxiliary (cf. example (53)) or a fronted verb (cf. example (56)).

[15] Using regular expressions, the superscript star (*) means zero or more occurrences of the superscripted category, and superscript plus (+) means one or more occurrences.

(63) **Phonological phrase construction for Dutch** (final version)

A φ-boundary occurs after every content word W unless

a. W is an adjacent modifier, or

b. W is an A or Adv directly preceding a verb.

Although there is room for improvement, (63) provides an observationally adequate definition of φ-construction for Dutch.

A few issues have remained unexplored. In Section 7.3.5, we will show that our φ-phrasing algorithm handles numeral expressions as well. We leave φ-phrasing with respect to quantifiers and idiomatic expressions to future research.

3.4 Evaluation

In this section, we will report on an evaluation of the proposed φ-algorithm. Recall that the reason for investigating phonological phrasing is that the grammar for pitch accent placement using RR crucially depends on the φ-boundaries being correctly determined. It is from this perspective that the algorithm was evaluated. That is, rather than looking for phonemic or phonetic evidence (Section 3.2.1) for the φ-boundaries, we checked for a substantial number of utterances whether applying RR within the predicted φ's led to the desired accentuation. Below, we will describe the method, material, and results of this evaluation.

3.4.1 Method

As explained in Section 1.1.3, acceptability judgements regarding the accentuation may vary, because (1) they may be gradual rather than categorical, (2) some variants may be equally acceptable, and (3) intersubjective differences may occur. To cope with this variation, we have chosen to aim for the intonation that is most acceptable according to the judgements of intonational experts, ignoring differences between equally acceptable forms. In addition, we argued in Section 1.1.4 that the disadvantage of a perceptual evaluation is that other aspects of the wave form, such as segmental errors, inadequate phonetic realization, or speaker characteristics, may interfere with judgement of the intonation proper. We therefore decided to rely on a symbolic assessment of the acceptability judgements of the experts, asking them to prescribe the preferred intonational form for a given texts, and comparing this with the prediction by our algorithms. The method that we used for the evaluation of our phonological phrasing algorithm is a practical implementation of these methodological choices. It encompasses the following steps.

1. **Construction of the corpus**
   Assemble a corpus of sentences. For each sentence, add its syntactic structure and its preferred distribution of pitch accents.

2. **Implementation of accentuation algorithm**
   Implement the accentuation algorithm that includes RR.

3. **Evaluation of accentuation algorithm**

   (a) *Automatic comparison*
   Apply the algorithm to the sentences in the corpus. For each sentence, automatically compare the predicted accentuation with the preferred accentuation. Report all differences and generate a confusion matrix.

   (b) *Error analysis*
   On the basis of the results of the automatic comparison, determine which accentuation errors are due to φ-phrasing errors.

Notice that a complete accentuation algorithm is used rather than only RR. The reasons for this will be explained further on (Section 3.4.3). The above steps will described in detail in the next three sections.

3.4.2 Corpus

3.4.2.1 Text

It is evident that evaluating φ-phrasing on text material in which RR rarely or never applies makes no sense. The text material has to be relevant in the sense that it must contain sentences in which words are frequently deaccented for rhythmic reasons. We found botanical texts to be an appropriate challenge, since most of their sentences are relatively long and have a rather complex syntactic structure, giving rise to a large number of accents and many opportunities to apply RR. Our corpus consisted of 10 descriptions of flowers, which were taken from a botanical handbook about wild flowers in Europe (Forey 1992). Each description consisted of 4 to 8 sentences. The total number of sentences was 52, the total number of orthographic words 907, an average of 17.4 words per sentence. Some words, like the noun *Zuid-Europa* 'Southern-Europe' or the compound adjective *eirond* 'egg-round', can receive two accents. Counting these as two words, which reflects the actual number of potential accent locations, the total number of words is 935. An example sentence is given in (64); a number of full texts are in given in Appendix A and Appendix B.

(64) De Reuzenbalsemien is uit de Himalaya naar Europa
The Impatiens-Glandulifera is from the Himalayas to Europe
gebracht en koloniseert hier waterkanten in vele gebieden ook
brought and colonises here watersides in many areas also
plaatselijk in Nederland en Belgie.
locally in the-Netherlands and Belgium

3.4.2.2 Accent distribution

A panel of three intonationally experienced, native speakers of Dutch was asked to prescribe the preferred accentuation for these texts. They worked on a version of the texts in which all abbreviations and numbers were expanded to their full lexical form. They were asked to specify the locations of the pitch accents to be used in a competent reading of the text. To accommodate for the possibility that alternative accentuations are equally acceptable (Section 1.1.3), a distinction was allowed between obligatory and optional accents. The panel discussed disagreements in their accentuations until a consensus was reached. In total, they assigned 435 obligatory accents and 10 optional accents. The prescription produced for (64) is given in (65); see Appendix A and Appendix B for other sentences.

(65) De R*euzenbalsemien is uit de Himal*aya naar Eur*opa gebracht en kolonis*eert hier w*aterkanten in v*ele gebi*eden ook pl*aatselijk in N*ederland en B*elgie.

3.4.2.3 Syntactic structure

The sentences were manually parsed by the author according to the syntactic analysis outlined in Section 2.6.[16] The syntactic structures provide all information that is required by the φ-phrasing algorithm, and by our accent placement algorithm in general (to be described in the next section). This amounts to an analysis in terms of major constituents (S, VP, PP, AP, AdvP), their heads (V, N, P, A, Adv) and several minor categories (Determiner, Quantifier, Complementizer, Conjunction). A number of word level features was used to mark pronouns, auxiliaries, copula, and the verbal part of compound verbs. Empty nodes or traces were not included. Figure 3.2 shows the syntactic structure corresponding to (64).

In addition to this purely syntactic information, the accentuation algorithm needs two other types of input. First, it requires the focus distribution. The default assumption is that all content word are focused (N, V, A, Adv, Q), whereas function words are not (determiners, prepositions, auxiliaries, copula, complementizers, dislocated verbal parts of compound verbs, conjunctions). Only non-default focus is explicitly indicated by means of focus and background features. This is rare, however, and only occurs on anaphoric expressions, i.e. some of the pronouns, definite noun phrases and pronominal adverbs. Second, the predicates and their corresponding arguments are labelled to enable application of SAAR.

[16] We experimented with syntactic parsers, but they either failed to handle the level of complexity of these sentences, or did not provide the required level of detail.

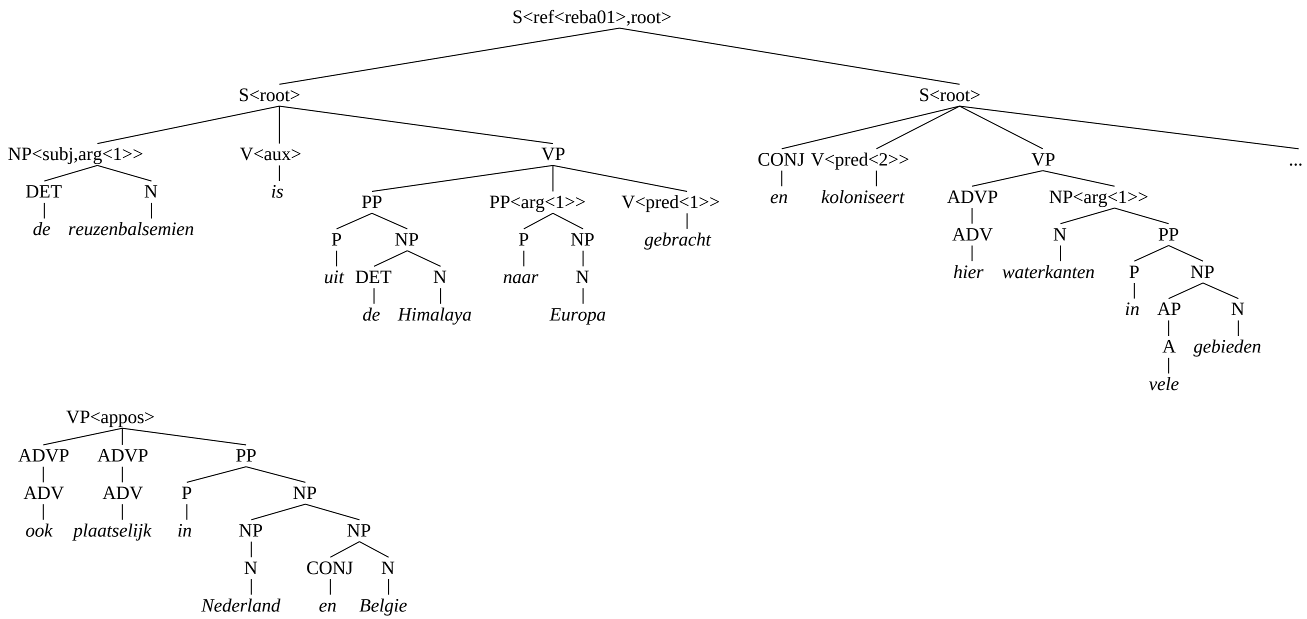

Figure 3.2: Syntactic tree for the sentence in (64).

3.4.3 Accentuation algorithm

In order to test RR, an initial accent distribution is needed. If the initial accentuation makes no sense, it is impossible to judge whether RR applied correctly. Therefore, the first step is accent assignment based on the focus distribution: every focused word is accented; compound words capable of bearing two accents receive two accents (e.g. *donkerbruin* 'dark brown').

However, this is not sufficient yet. Suppose we accent all focused content words in (66-a), and derive the φ's, then RR incorrectly deletes the accent on *later* (66-b). The error is not due to incorrect φ-phrasing, but to the fact that the predicate *verschijnt* should be unaccented. According to SAAR (Section 2.5.2), the predicate is adjacent to its focused argument *een witte bloem*, and must therefore be unaccented. That is, only by applying SAAR before RR can the correct accentuation in (66-c) be obtained .

(66) a. ($_{AdvP}$ ($_{NP}$ een maand) ($_{Adv}$ later)) ($_{V}$ verschijnt) ($_{NP}$
a month later occurs
een witte bloem)
a white flower

b. ?[$_{\varphi}$ een m$\overset{*}{\text{a}}$and l$\overset{\emptyset}{\text{a}}$ter versch$\overset{*}{\text{i}}$jnt] [$_{\varphi}$ een w$\overset{*}{\text{i}}$tte bl$\overset{*}{\text{o}}$em]

c. [$_{\varphi}$ een m$\overset{*}{\text{a}}$and l$\overset{*}{\text{a}}$ter versch$\overset{\emptyset}{\text{i}}$jnt] [$_{\varphi}$ een w$\overset{*}{\text{i}}$tte bl$\overset{*}{\text{o}}$em]

In conclusion, evaluating φ-phrasing requires an accentuation algorithm including initial accent assignment and SAAR in order to avoid unintended applications of RR. The algorithm in (67) is in fact identical to that proposed earlier in Section 2.5.4.[17]

(67) **Accentuation algorithm**

1. Assign initial accents
2. Apply SAAR
3. Derive φ's
4. Apply RR

3.4.4 Results and discussion

We distinguished two types of errors: *undershoot errors* are words[18] that are unaccented in the reference, but not deleted by the algorithm; *overshoot errors* are words are accented in the reference, but are deaccented by the algorithm. We assumed that the algorithm, which itself never predicts optional accents, can never make a wrong prediction with respects to an optional accent in the reference. That is, whether the algorithm predicts an accent that is optional

[17] It is implemented in GRAMTSY (Coppen 1990; Coppen 1991), a program to transform trees by means of transformational rules. See Section 6.3.1.

[18] This includes parts of a compound words

Table 3.2: Comparison of the predicted accentuation with the reference accentuation in the form of a confusion matrix.

		Predicted		
		Accented	*Unaccented*	*Total*
Reference	*Accented*	414 (44.76%)	13 (1.41%)	427
	Unaccented	42 (4.54%)	456 (49.30%)	498
	Total	456	469	925

Table 3.3: The number of correct and incorrect applications of the deaccentuation rules.

Rule:	*Correct:*	*Incorrect:*
SAAR	31	7
RR	39	5
Total:	70	12

according to the reference, or fails to predict this accent, either way is considered correct.

The results for one text as produced during automatic comparison are given in Appendix A. The overall results are summarised in Table 3.2 in the form of a confusion matrix. There were 42 undershoot errors and 13 overshoot errors. Since the ten words with optional accents were ignored, the total number of words was 925 (i.e. 935 minus 10). This means that the algorithm correctly predicted the accentuation of 94.05% of all words.

3.4.4.1 RR overshoot errors

Of course, not all overshoot errors are due to RR. Table 3.3 shows the contribution of SAAR and RR to the total number of overshoot errors.[19]

With only five incorrect applications against 39 correct applications, RR appears to perform reasonably well. Two of these errors are shown in (68). The adverb *ook* and the adjective *zeldzame* are both adjacent modifiers, while the pronoun *dit* and the copular verb *is* are function words, so the first φ-boundary

[19] One overshoot error is due to the fact that a word was not initially accented.

is as late as after the noun *soort*. After initial accent assignment[20], RR applies twice (68-b). Compared to the reference (68-c), we see one overshoot error (*ook*) and two undershoot errors (*dit* and *zeldzame*). The deaccentuation of *dit* would not have occurred without the overshoot error, i.e. the initial assignment of an accent to *ook*. However, the incorrect deaccentuation of *zeldzame* can only be attribted to the lack of a φ-boundary after *dit*. We may hypothesise that a focused function word gains the status equivalent to a content word, trigging a φ-boundary .

(68) a. ($_S$ ($_{NP}$ ($_{AdvP}$ ($_{Adv}$ ook)) ($_N$ dit)) ($_V$ is) ($_{VP}$ ($_{NP}$ een
also this is a
zeldzame soort)) ...)
rare species
'this is a rare species as well'

b. [$_\varphi$ $\overset{*}{\text{o}}$ok d$\overset{ø}{\text{i}}$t is een z$\overset{ø}{\text{e}}$ldzame so$\overset{*}{\text{o}}$rt]

c. ook d$\overset{*}{\text{i}}$t is een z$\overset{*}{\text{e}}$ldzame so$\overset{*}{\text{o}}$rt

The third incorrect application of the RR (69) appears to be caused again by initial accent assignment: the adverb *vaak* should be left unaccented.[21]

(69) a. Kruipend stalkruid heeft houtige vertakte ($_{NP}$ ($_{AP}$
Ononis-Repens has woody branching
($_{Adv}$ vaak) ($_A$ liggende)) ($_N$ stengels)) ...
often recumbent stems

b. [$_\varphi$ va$\overset{*}{\text{a}}$k l$\overset{ø}{\text{i}}$ggende st$\overset{*}{\text{e}}$ngels]

c. vaak l$\overset{*}{\text{i}}$ggende st$\overset{*}{\text{e}}$ngels

The fourth error is deletion of the accent on *stelen* in (70). We have no explanation here, except that the syntactic structure is rather complex, which may have caused the panel to add an extra accent. The same goes for the fifth and final erroneous application of the RR, not shown here.

(70) a. ... ($_{AP}$ ($_{PP}$ in ($_{NP}$ ($_N$ kluwens) ($_{PP}$ op ($_{NP}$ ($_A$
in tangles on
overhangende) ($_N$ stelen))))) ($_A$ bijeenstaand)) ...
overhanging stems gathered
'gathered in tangles on overhanging stems'

b. [$_\varphi$ in kl$\overset{*}{\text{u}}$wens] [$_\varphi$ op $\overset{*}{\text{o}}$verhangende st$\overset{ø}{\text{e}}$len bij$\overset{*}{\text{e}}$enstaand]

c. in kl$\overset{*}{\text{u}}$wens op $\overset{*}{\text{o}}$verhangende st$\overset{*}{\text{e}}$len bij$\overset{*}{\text{e}}$enstaand

[20] The pronoun *dit* is explicitly focused and therefore receives an accent by initial accent assignment.

[21] An alternative interpretation, in which *vaak* is a sentential adverb rather than a modifier of the following adjective, would at least have prevented the deletion of the accent on *liggende*.

3.4.4.2 RR undershoot errors

We encountered 42 undershoot errors. The question is to what extent we can blame the combination of φ-phrasing and RR for the failure to delete accents. This question is considerably harder to answer, because it means that for each error, we have to consider alternative φ-phrasings that would have preserved the accent.

To begin with, at least six errors involved accented content verbs. The accent on a verb can never be deleted by RR, because there is always a φ-boundary after a content verb or after a verb cluster, which means that it can never be followed by another accent in the same φ.

As many as 17 errors stem from accented adverbs. Here, the problem is the initial accent assignment: how to decide whether a given adverb can be accented at all. We have no solution to offer, but will further discuss this issue in Section 7.3.6.3.

The remaining 19 errors may indeed be attributed to φ-phrasing. For example, the fact that (71) has no accent on *soort* could be taken to indicate that *de bekendste soort van de familie* is in fact one φ.[22] Example (72) is structurally identical, and occurs five times in the corpus. These examples suggest, contrary to our position in Section 3.3.2, that a postmodifying PP does not always force a φ-boundary.[23]

(71) a. ... ($_{NP}$ de ($_{AP}$ bekendste) ($_{N}$ soort) ($_{PP}$ van de
the best-known species of the
familie))
family

b. [$_\varphi$ de bek*endste so*ort] [$_\varphi$ van de fam*ilie]

c. de bek*endste soort van de fam*ilie

(72) a. ... ($_{PP}$ in ($_{NP}$ ($_{AP}$ grote) ($_{N}$ delen) ($_{PP}$ van Europa)))
in large parts of Europe

b. [$_\varphi$ in gr*ote d*elen] [$_\varphi$ van Eur*opa]

c. in gr*ote delen van Eur*opa

Another class of undershoot errors consists of those in which the second accent of a compound adjective is not deleted. Example (73) contains the compound adjective *kalkhoudende*, which may carry two accents when it is in sentence-final position. The predicted φ-phrasing causes it to carry two accents (73-b), but the reference prescribes just one (73-c).

[22] Alternatively, we can assume that – for some reason – words like *soort* 'type', *deel* 'part', *grond* 'ground', *gebied* 'area' etc. are never accented when postmodified.

[23] As a matter of fact, we examined the effect of incorporating all postmodifying PP's in the φ of the preceding noun, but found that this results in far too many deaccentuation errors.

(73) a. … ($_{PP}$ op ($_{AP}$ droge) ($_{AP}$ kalkhoudende) of ($_{AP}$
on dry calcareous or
basische) ($_{N}$ bodems))
alkaline soils

b. [$_{\varphi}$ op dr$\overset{*}{\mathrm{o}}$ge] [$_{\varphi}$ k$\overset{*}{\mathrm{a}}$lkh$\overset{*}{\mathrm{o}}$udende]] of b$\overset{*}{\mathrm{a}}$sische b$\overset{*}{\mathrm{o}}$dems]

c. op dr$\overset{*}{\mathrm{o}}$ge k$\overset{*}{\mathrm{a}}$lkhoudende of b$\overset{*}{\mathrm{a}}$sische b$\overset{*}{\mathrm{o}}$dems

The same goes for all of the compound adjectives listed in (75).

(74) *losbloemig* loose-flowered
kalkhoudend calcareous
langwerpig elongated
eenjarig annual
kransstandig verticillate
veerlobbig feather-lobate
eironde egg-round
purperroze purple-pink

It should be noted that whether a compound adjective has one or two accents is not always clear. While many adjectives have clearly either one (75-a) or two potential accents (75-b), some appear to allow both possibilities, or at least show variation among speakers (cf. Gussenhoven (1983a, p. 314)).

(75) a. langdr$\overset{*}{\mathrm{a}}$dig
long-winded

b. l$\overset{*}{\mathrm{a}}$ngb$\overset{*}{\mathrm{e}}$nig
long-legged

c. l$\overset{*}{\mathrm{a}}$ngw$\overset{*}{\mathrm{e}}$rpig / langw$\overset{*}{\mathrm{e}}$rpig
long-shaped
'elongated'

Recall that the reference accentuation as produced by the panel did not specify if a compound can potentially have two accents.[24] Rather, this decision was made by the author while manually parsing the sentences. Thus, some of the deviations in (74) may be due to doubt or disagreement with respect to the accent potential of a compound adjective.

In conclusion, the general outcome of this evaluation is that pitch accent placement using a combination of SAAR, φ-phrasing, and RR correctly predicted 94.05% of all 925 accent decisions (ignoring optional accents). Considering the length and complexity of the sentences, this is a rather good result. There were more than three times as many undershoot errors as there were overshoot errors. In other words, we predicted too many rather than too few accents. This suggests that our grammar is too generous, and that additional rules for deaccentuation are needed.

[24] Unless, of course, two accents were prescribed.

With respect to RR, it was found that it applies correctly in 39 cases and incorrectly in only 5 cases. Two of these overshoot errors can be attributed to an incorrect initial accent assignment, whereas the remaining three may be real errors, requiring further research. Furthermore, 19 out of the 42 undershoot errors may in principle be blamed on either RR and/or on φ-phrasing. It would appear that a postmodifying PP is not always preceded by a φ-boundary, as predicted by our φ-phrasing algorithm, although the conditions for integration are unclear. It was suggested that undershoot errors in compound adjectives can be partly explained by variability in their accent potential (one or two accents). Yet, the undershoot errors do not provide us with compelling reasons to modify our φ-phrasing algorithm. At a minimum, we can conclude that the proposed φ-algorithm for Dutch is fairly accurate.

3.5 Summary

In this chapter, we investigated φ-phrasing in Dutch. The reason for this exercise is that our algorithm for pitch accent placement using RR requires that the φ-boundaries are correctly determined. We started by reviewing the φ-phrasing algorithms proposed by Nespor & Vogel and by Hayes, and argued that these are inadequate for Dutch. We then proposed a new φ-algorithm for Dutch based on the notion of adjacent modification. A careful examination of φ-phrasing in a number of syntactic structures revealed that it was not entirely adequate in the case of VP's and S's, which called for a minor revision of the definitions of φ-phrasing and adjacent modification. Although further generalisation is desirable, it provides us with an observationally adequate definition of the Dutch φ.

In the second part of this chapter, the proposed algorithm was evaluated by checking for a substantial number of utterances if applying RR within φ's leads to the preferred accentuation. Rather than a perceptual evaluation, we opted for an evaluation at the symbolic level, using a partly automatic, corpus-based method. The corpus consisted of flower descriptions containing 52 relatively long and complex sentences, with a total of 935 words. The text material was manually parsed and a preferred accentuation was prescribed by a panel of three experts. The predictions of the algorithm for pitch accent placement, which included φ-phrasing and RR application, were compared with the reference. Prediction and reference corresponded in 94.05% of all 925 accent placement decisions. A detailed analysis of the errors of RR showed no compelling reasons to revise the φ-phrasing algorithm. At a minimum, we can conclude that the proposed φ-algorithm for Dutch is fairly accurate. This result may contributes to a better description of the prosodic phonology of Dutch.