# Chapter 2

# Theoretical Framework

## 2.1 Introduction

The previous chapter introduced the two main issues addressed in this thesis: pitch accent placement and intonational phrasing in SLG. This chapter presents the theoretical framework for our work. It provides a more precise interpretation of these issues in terms of phonological theories, and a first suggestion as to how we may account for them. This will then be worked out in the following chapters.

Section 2.2 defines prosody, and summarises prosodic phonology, the study of prosody from a phonological point of view. Section 2.3.1 introduces intonational phonology, and describes the intonational phonology of Dutch. It should be noted, however, that material in this chapter is not intended as review of prosodic phonology or intonational phonology, but focuses on those aspects that are relevant to our research aims.[1] This theoretical framework enables a more precise definition of the challenges involved in generating intonation in Section 2.4. As a first step towards a solution, Section 2.5 proposes a linguistic account of pitch accent placement in the form of a generative grammar. As most of the work reported in this thesis is in one way or another dependent on syntax, Section 2.6 briefly describes the syntactic analysis of Dutch used throughout our work. The chapter ends with a summary in Section 2.7.

## 2.2 Prosodic phonology

### 2.2.1 Prosody

Speech is more than a succession of vowels and consonants. These speech segments are organised into higher units: segments are structured into syllables,

[1] General introductions to prosodic phonology can be found in Booij (1983), Nespor and Vogel (1986), Inkelas and Zec (1995), and Shattuck-Hufnagel and Turk (1996); the best source for intonational phonology is Ladd (1992, 1996).

syllables into feet, etc. In addition, there appear to be patterns of relative prominence among these higher units. For example, the most prominent foot in a word is called the *main-stressed* foot, which contains the main stressed syllable. These intuitions are supported by phonetic measurements which show that these perceived higher units, and their relative prominence relations, are correlated with acoustic parameters such as pitch, duration, and segmental quality. Accordingly, syllables with main-stress tend to be longer and more carefully pronounced than syllables with secondary or no stress (see for example Sluijter (1995) and Rietveld, Kerkhoff, and Gussenhoven (1999) for Dutch).

The hierarchical organisation of segments into higher units, the patterns of relative prominence among these units, and their phonetic reflection in terms of phonetic parameters are collectively referred to as *prosody*. Although other definitions of prosody have been proposed, the above definition is the one that emerges from recent work that aims at integrating insights from phonological and phonetic research. Accordingly, a recent tutorial paper on prosody defines it as:

> "(1) acoustic patterns of $F_0$, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and (2) the higher-level structures that best account for these patterns" (Shattuck-Hufnagel and Turk 1996, p. 196).

### 2.2.2 A phonological account of prosody

Initially, phonological accounts of prosody attempted to identify the hierarchical organisation of an utterance with its syntactic structure. However, it became clear that some important aspects of the syntactic structure are not reflected in the prosody, and conversely, that some prosodic aspects cannot be predicted from the syntactic structure.

> "It has become a majority view among researchers [...] that syntax does not provide domains for phonological rules in a direct fashion. Both the impoverished amount of syntactic information needed by the phonological module and the variety of mismatches between phonological rule domains and syntactic constituency argue for positing another level of representation" (Inkelas and Zec 1995, p. 537).

In reaction to this discrepancy between syntax and prosody, a number of phonologists, most prominently Selkirk (1984, 1996), Nespor and Vogel (1986), and Hayes (1986), developed a theory called *prosodic phonology*. It assumes the existence of a *prosodic structure*, an independent structure that consists of *prosodic constituents* labelled according to their *prosodic category*. An example of a prosodic structure is given in Figure 2.1.

Following to Nespor and Vogel (1986) and Hayes (1986), we distinguish seven prosodic categories, which are strictly ranked on a scale called the *proso-*

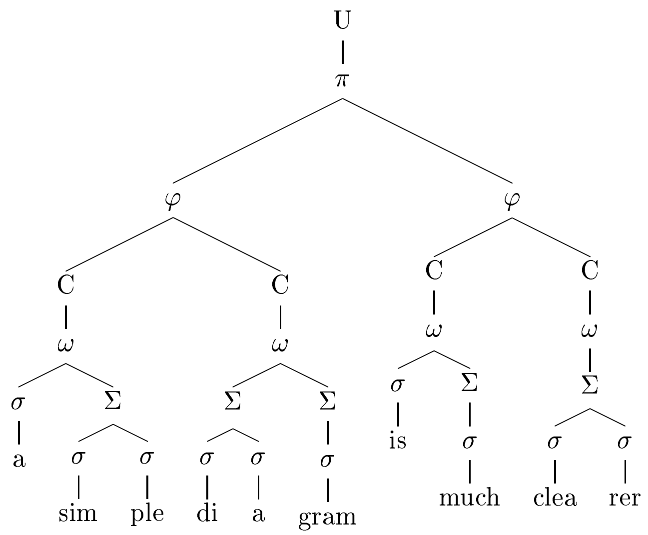



*Figure 2.1: Example of a prosodic structure.*

*dic hierarchy*; see Table 2.1. It is assumed that every utterance must be *fully* parsed in terms of *all* prosodic categories. Moreover, the resulting structure must adhere to the prosodic hierarchy. That is, a prosodic constituent of a certain rank cannot be dominated by a prosodic constituent of lower rank. A stronger version of this assumption claims that a prosodic constituent of a certain rank can only dominate a constituent of the directly subordinated rank, thereby excluding the possibility of recursion. This is known as the *Strict Layer Hypothesis*.

(1) **Strict Layer Hypothesis** (SLH)

The categories of the Prosodic Hierarchy may be ranked in a sequence $C_1, C_2, \ldots, C_n$ such that

a. all segmental material is directly dominated by the category $C_n$, and

b. for all categories $C_i$, $i \neq n$, $C_i$ directly dominates all and only constituents of category $C_{i+1}$.

(Hayes 1986)

By implication, prosodic trees must have n-ary branching nodes. Notice, however, that the prosodic structure in Figure 2.1 partly violates the SLH: the determiner *a* is directly dominated by a $\omega$ rather than a $\Sigma$. The reason is that a schwa is unstressed, and cannot constitute a $\Sigma$ or $\omega$ of its own. Some limited violation of the SLH is assumed to be necessary for prosodic constituents below

*Table 2.1: Hierarchy of prosodic categories.*

| Rank: | Category: | Symbol: |
|---|---|---|
| 1. | Utterance | U |
| 2. | Intonational Phrase | $\pi$ |
| 3. | Phonological Phrase | $\varphi$ |
| 4. | Clitic Group | C |
| 5. | Prosodic Word | $\omega$ |
| 6. | Foot | $\Sigma$ |
| 7. | Syllable | $\sigma$ |

the $\varphi$ (Booij 1995, p.144-145). Yet, in the current work, we will adhere to the SLH for $\varphi$'s and higher categories.

The motivation for distinguishing these particular prosodic categories is based on two types of arguments. First, prosodic constituents may be the application domain for certain segmental rules. These *P-rules* are shown to apply within (or at) the boundaries of a prosodic constituent of a certain rank, but not across its boundaries. An example is *Degemination* in Dutch (Booij 1995, p. 151): in a sequence of two identical consonants, one of them may be deleted. In (2-a), which constitutes a single $\pi$, Degemination is responsible for deleting one of the /n/'s. In contrast, it cannot apply in (2-b), because of the intervening $\pi$-boundary. As we will see later, the reason for this boundary is that tag questions such as *niet* obligatory form an $\pi$ of their own (Section 4.4.1).

(2) a. [$_\pi$ Is banaa/n/ /n/iet gezond dan? ]
Is banana not sound then

b. [$_\pi$ Apen houden van banaa/n/ ] [$_\pi$ /n/iet? ]
Monkeys like of banana not
'Monkeys like banana, don't they?'

However, undisputed evidence of this type is rare, especially for the prosodic categories between $\omega$ and $\pi$. The application of P-rules is thought to be dependent on speaking rate and speaking style as well. In all fairness, the generalisation appears to be that most P-rules apply more frequently in smaller domains than in larger domains (Booij 1995, p. 146).

The second type of argument in support of a prosodic constituent is based on the assumption that prosodic constituents account for the prosodic boundaries observed in an utterance. The higher the rank of a prosodic category, the more prominently its boundaries are marked by phonetic cues such as pause and lengthening of the preboundary syllable (Ferreira 1993; Menert 1994; Hofhuis, Gussenhoven, and Rietveld 1995; Rietveld, Kerkhoff, and Gussenhoven 1999; Cambier-Langeveld 2000). For instance, Gussenhoven and Rietveld (1992b) present evidence from a perception experiment that English listeners expect longer preboundary syllable durations as the rank of the prosodic boundary is higher.

The prosodic structure can be seen as the interface that connects morphology, syntax and semantics to phonology. It is obtained by filtering out all linguistic aspects that are not relevant to phonology. For instance, the complex, recursive syntactic structure of an utterance is reduced to a strictly-layered prosodic structure, which has only a few fixed categories and no recursion. This reduction is accomplished by means of so-called *mapping rules*, which map segmental, morpho-syntactic and semantic information onto prosodic constituents. Various mapping rules have been proposed, with slightly different predictions. Since only $\varphi$'s and $\pi$'s are directly relevant for accent placement and intonational phrasing, we will concentrate on the mapping rules for these constituents in Chapter 3 and Chapter 4, respectively. Some remarks regarding the lower prosodic constituents and their mapping rules can be found in Section 5.6.

## 2.3 Intonational Phonology of Dutch

### 2.3.1 Intonational phonology

Segmental phonology is concerned with describing the sound of a language at the phonemic level. As is well-known, the definition of phonemes is based on their capability to distinguish one word from another. For instance, Dutch has a phonemic distinction between the vowels /e/ and /u/, as the two words of the minimal pair /pen/ (*peen*, 'carrot') versus /pun/ (*poen*, slang for 'money') have different meanings. Other differences between speech sounds are non-distinctive, and gradual in nature. For example, although the vowel /e/ is 'coloured' by the following /r/ in /per/ (*peer* 'pear'), it is not distinct from the /e/ in /pen/ in the sense that it can differentiate Dutch words, and is therefore not a different phoneme. The distinction between phonemes is categorical, whereas the difference between allophones (non-distinctive variants of a phoneme) may be gradient in nature.

Analogously, *intonational phonology* assumes that intonation is phonologically structured, and that intonation contours can be analysed in terms of a limited inventory of categorical, contrastive intonational patterns. Additional gradient, non-distinctive variation in pitch is assumed to be a matter of phonetics. Accordingly, intonational phonology is concerned with the phonological aspects of pitch, representing pitch at an abstract, symbolic level. It describes intonation in the form of a grammar, that is, a combinatorial system consisting of a limited inventory of atomic elements, and rules to combine these elements into more complex structures. The complement of intonational phonology is *phonetic implementation*, which concerns the actual realization of these phonological abstractions as concrete pitch contours.

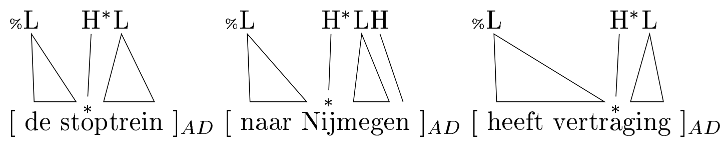



*Figure 2.2: Example of Gussenhoven's description of Dutch intonation. The utterance means 'The slow-train to Nijmegen has delay'. The top line is the tone string consisting of boundary tones and pitch accents. The bottom line is the segmental string, where square brackets indicate association domains, and stars indicate accented syllables. In between are the tiers that align the two level of description.*

The first model of intonation that introduced an explicit distinction between phonological description and phonetic implementation has been attributed to Pierrehumbert (1980). Her description of American-English intonation is called *autosegmental*, because it describes the intonation not in terms of features of phonemes, but as a separate string of intonational events, which are synchronised with the segmental string by means of *tiers*. The tonal events on the tone string are either *pitch accents* or *boundary tones*, both of which are composed of one or more *tone segments* (or *tones* for short). There are only two of these, *H* and *L*, which represent phonological abstractions of a high and a low pitch level, respectively.[2]

Following the work of Pierrehumbert (1980) for American-English, intonational phonologies of other languages have appeared. Gussenhoven (1984, 1988, 1991b, 1992b) developed an autosegmental description of English and Dutch intonation. It is this model that forms the theoretical basis of the work on intonation in SLG presented in this thesis. The relevant aspects of Gussenhoven's model will be presented in the remainder of this section.

### 2.3.2 Pitch accents

Earlier on, a *pitch accent* was defined as a perceptually salient pitch event, which results from a relatively sudden change in pitch level. A phonological description must specify more precisely which pitch patterns count as a pitch accent. According to Gussenhoven's description, Dutch has three types of pitch accents: H*L, L*H, and H*LH. The diacritic * indicates that the preceding tone segment, the *starred tone*, must be linked to a syllable in the segmental string. For instance, the first H* in Figure 2.2 is linked to the syllable *stop*, which is

[2]Notice that these are pitch *levels*, not pitch *movements*; in other words, points, not trajectories. This is one of the differences with primarily phonetic descriptions of intonation like the IPO-description of Dutch intonation ('t Hart, Collier, and Cohen 1990), which describe an intonation contour as a concatenation of pitch movements.

the main stressed syllable of the word *stoptrein*. Anchoring the starred tones of the tone string to stressed syllables in the segmental string is one of the ways in which the two descriptions are aligned.

It is important to keep apart the notions of (secondary and primary) *stress* and *pitch accent*. Although *stress* has been used as a term to cover prominence both at the lexical level (i.e. relative prominence patterns among syllables) and at the phrasal level (i.e. prominence patterns among words and phrases) current convention reserves *stress* for prominence at the lexical level and *pitch accent* for prominence at the phrasal level. This difference in prominence categories is supported by phonetic evidence (Beckman and Edwards 1994): the primary empirical cues for stress are segmental quality and duration, whereas the main cue for a pitch accent is pitch.

### 2.3.3 Boundary tones

An *intonational phrase* was defined earlier as a stretch of speech covered by a perceptually coherent intonation contour. This implies that the boundaries of an intonational phrase are marked by some discontinuity in the intonation contour. There appear to be at least two ways in which an intonational phrase boundary can be signalled: a change in pitch level, described by the occurrence of a boundary tone, or a change in register, described by the interruption of downstep. It is sometimes assumed they can be also be signalled by a pause of sufficient length. Although we agree that a tonally marked boundary is often supported by a pause, we reject the view that a pause is a sufficient cue for an intonational phrase boundary, because this gives rise to the inconsistencies (to be discussed in Section 2.3.7).

The change in pitch level near an intonational boundary is analysed in autosegmental approaches as a *boundary tone*, either H or L. Just as the starred tone segment of a pitch accent, it must be aligned with the segmental string. In Gussenhoven's model, it is linked to an edge of *association domain* (AD). For the time being, we can think of an AD as equivalent to an intonational phrase; we will discuss the differences in Section 2.3.7. As an example, the utterance in Figure 2.2 has three AD's. The edge the boundary tone is linked to is indicated by a prefixed (left edge) or suffixed (right edge) %-symbol. This amounts to four possibilities: a high initial boundary tone (%H), a low initial boundary tone (%L), a high final boundary tone (H%), and a low final boundary tone (L%). For example, the first %L in Figure 2.2 is an initial boundary tone that is linked to the beginning of the first AD. We postpone a discussion of boundary marking by a change in register to Section 2.3.6.

### 2.3.4 Intonation contours

So far, we have introduced the segmental string and the tonal string, which are synchronised by association lines which link every starred tone to a stressed syllable, and every boundary tone to an AD-boundary. Not every stressed syllable or AD-boundary needs to be linked. In Figure 2.2, for example, the syllables *de*, *naar*, and *heeft* are not linked to a starred tone; in other words, they are not accented. Similarly, the first AD in Figure 2.2 is not linked to a final boundary tone. The tone string in Figure 2.2 (%L H*L %L H*LH %L H*L) forms a complete description of the intonation, but before the representation is sent to the phonetic implementation rules for conversion to an acoustic signal, two things will happen.

First, tonal rules may modify the representation as a result of *AD-restructuring*, as described in Section 2.3.5. Second, default boundary tones are inserted. For one thing, the final H of a H*LH pitch accent in AD-final position functions as an AD-final boundary tone, H%. Second, an Utterance-final H*L is provided with a final L%, as shown in Figure 2.3, while an Utterance-final L*H is provided with a H%. Since we want to be able to discuss different intonational patterns without going into all the details, we will make use of an abstract, graphical representation of the pitch contour, one that dispenses with the default boundary tones and thus shows only AD-structure, the AD-initial boundary tone, and the pitch accents.

In order to derive the intonation contour, it is assumed that every tone yields a *target*. For instance, the first H* in Figure 2.3 yields a high target, indicated by a bullet. The dashed arrow indicates the tone of which each target is the phonetic implementation. Some tone segments, such as the initial boundary tones in Figure 2.3, produce two targets. The first target is located at the boundary, whereas the second target is located at some fixed distance from some other target. This is called *spreading*, and it is indicated in the figure by a triangle between the tone and the text showing the extent of the spreading. In addition to initial boundary tones, the *second* tone of the *last* pitch accent in an AD spreads up to the next tone in the AD, or up to the end of the AD if it is the last tone. For example, the L of the first H*L in Figure 2.3 spreads up to the end of the AD; the L of the H*LH spreads up to the following H, which, as a boundary tone, is in turn located at the end of the AD. When all targets are determined, connecting the consecutive targets yields the intonation contour. The computation of the actual pitch contour by the phonetic implementation module will be described more fully in Section 5.6.2.2.

### 2.3.5 Tonal rules

In addition to being autosegmental, Gussenhoven's description is also transformational. Analogously to segmental phonology, it assumes an underlying representation of the intonation that is rewritten into a surface representation by a number of optional *tonal rules*. The underlying representation requires that the utterance is exhaustively parsed into one or more AD's, and that each of these contains exactly one pitch accent.[3]

One of the tonal rules that may modify the underlying representation is called *AD-restructuring*. It merges two AD's into a single one by deleting their

[3]There is one exception: AD's without a pitch accent may occur for reporting clauses, where a rule called *Tone Copy* provides a pitch contour to the AD without an accent (Gussenhoven 1990; Gussenhoven 1992a).

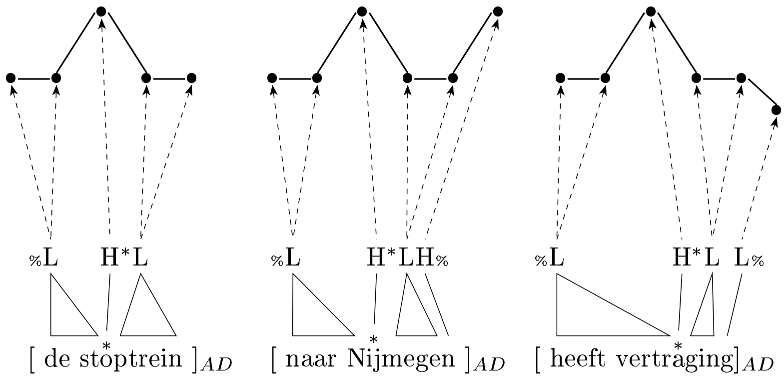



*Figure 2.3: Abstract representation of the intonation contour corresponding to the description of Figure 2.2.*

adjacent boundaries. As explained above, AD's serve as a reference for linking the non-starred tone segments of a pitch accent. Therefore, AD-restructuring will alter the alignment between the segmental string and the tone string, and consequently the shape of the intonation contour. An example of restructuring is given in Figure 2.4. Notice that the %L that was linked to the beginning of the second AD in Figure 2.3 is no longer present, since there is no legitimate position available to link it to. Furthermore, the first H*L accent is no longer the last accent in an AD. Recall that spreading only applies to the second tone segment of the *last* pitch accent in an AD. Therefore, the L tone segment of the first H*L in Figure 2.3 does not spread, but moves up to the following H*LH. The result is known as a *partly linked* contour.

Another tonal rule is called *Tone Linking.* It deletes the non-starred tone segments of the penultimate pitch accent in an AD. Figure 2.5 shows the effect of applying Tone Linking to the first AD in Figure 2.4. The result is an intonation contour that is known as a *fully linked* or *flat hat* contour.

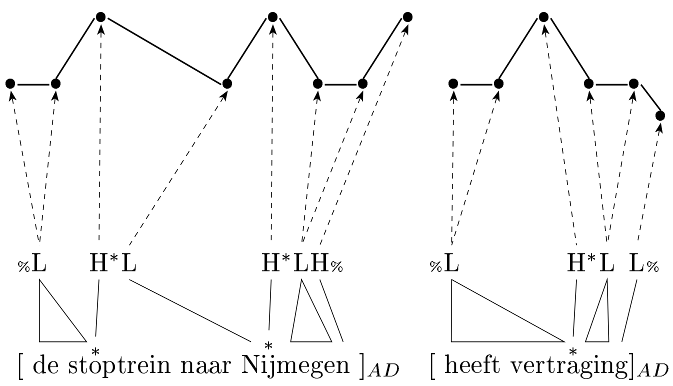



*Figure 2.4: Example of AD-restructuring, resulting in a partly linked contour.*

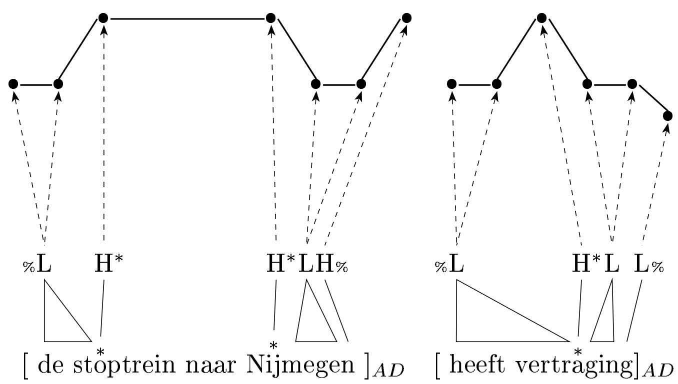



*Figure 2.5: Example of Tone Linking, resulting in a fully linked contour.*

### 2.3.6 Downstep

As mentioned earlier, an intonational phrase boundary can be signalled by a boundary tone or by a register change. The register corresponds roughly to the pitch span between H and L tones.[4] It can be compressed or expanded, and shifted upwards or downwards. One of its uses is to set off the boundaries of appositions, such as parenthetical remarks and non-restrictive relative clauses, where the register is compressed and shifted downwards at the start of the apposition, and restored to its former size and level at the end of the apposition.

Related to this is the phenomenon of gradual lowering of the pitch in the course of an intonational phrase, sometimes referred to as *declination* ('t Hart, Collier, and Cohen 1990). Declination has not been implemented in our intonation model, since it is believed that *accentual downstep* and *phrasal downstep* cause the intonation contour to have the required effect. The former involves stepwise lowering of the H* targets in an AD during phonetic implementation, the latter involves stepwise lowering of the AD's in an utterance (see Section 5.6.2.2). In Gussenhoven's model, these are described by an optional *tone morpheme* ('!') that is attached to an AD or an utterance, respectively.

The domain of accentual downstep is the AD. This means that it can be used as a cue to signal intonational phrasing: as long as there is downstep, there can be no intervening intonational boundary, whereas the register reset following downstep indicates the presence of a boundary.[5]

### 2.3.7 Association Domain

Section 2.2 defined prosody as the hierarchical organisation of segments into higher units, the patterns of relative prominence among these units, and their reflection in terms of phonetic parameters such as $F_0$ and duration. Clearly, the intonational phrase meets the criteria for a prosodic unit, and has therefore been incorporated into prosodic phonology as well. Most descriptions of intonation define an intonational phrase as a stretch of speech covered by a coherent, continuous pitch contour, and delimited by audible breaks. This definition of intonational phrase was also adopted by prosodic phonology: "The formulation of the basic I[ntontonational phrase] formation rule is based on the notions that the intonational phrase is the domain of an intonation contour and that the end of intonational phrases coincide with the positions in which pauses may be introduced in a sentence." (Nespor and Vogel 1986, p. 188). The intonational phrase was incorporated as a prosodic constituent ($\pi$), slotted in between the Utterance (U) and the Phonological Phrase ($\varphi$). In addition, mapping rules were formulated to derive $\pi$'s from the semantic and syntactic structure of an utterance.

Incorporating the intonational phrase as a prosodic constituent ($\pi$) implies that it becomes subject to the general constraints on prosodic structure imposed

[4] See van den Berg, Gussenhoven, and Rietveld (1992) for a more careful definition.

[5] This not always true, since there is a variant of downstep in which the last accent of a downstepped AD is unaffected.

by the theory of prosodic phonology. For instance, the Strict Layer Hypothesis (Section 2.2) requires that every $\pi$ is directly dominated by an U, and that every $\pi$ directly dominates one or more $\varphi$'s. In addition, the empirical predictions regarding prosodic constituents must hold for $\pi$ as well. One of these is that the prominence of the edge of a prosodic constituent is proportional to its rank in the prosodic hierarchy. In other words, the higher the constituent, the more its boundaries will be emphasised by prosodic means such as preboundary lengthening and pausing. This explains the remark about pause in the citation above: since the $\pi$ is the last but one highest in rank, its boundaries are likely to coincide with pauses. Consequently, whenever a clear silent[6] pause occurs, this must be considered strong evidence for the presence of a $\pi$-boundary. Another prediction is that the application of segmental rules is likely to be blocked by the intervention of an $\pi$-boundary.

However, this combination of theoretical constraints and empirical predictions, which result from incorporating the intonational phrase in the framework of prosodic phonology, leads to some serious problems. The core of these problems is a discrepancy between the two criteria for an intonational phrase, i.e. a coherent intonation contour and an audible pause.

Gussenhoven and Rietveld (1992a) gives the following pair of utterances.

(3) a. $[_\pi$ But we're n$\overset{*}{\text{o}}$t telling John ]

b. $[_\pi$ But we're n$\overset{*}{\text{o}}$t going ] $[_\pi$ John ]

While (3-a) has a single intonation contour corresponding to a single $\pi$, (3-b) has the same contour, but two $\pi$'s. This is because John is a indirect object in (3-a), but a vocative in (3-b). The latter constitutes an obligatory $\pi$ according to the rule for $\pi$-formation (see Section 4.4.1), but is unaccented, like all final vocatives. On the one hand, the intonational boundary in (3-b) is correct, since *John* is preceded by a prosodic break, which is clearly missing in (3-a). On the other hand, the boundary is incorrect, since not every $\pi$ in (3-b) corresponds to a coherent intonation contour, which violates the definition of an intonational phrase.

Another example of the same phenomenon is (4).

(4) a. $[_\pi$ To b$\overset{*}{\text{o}}$rrow the book may be che$\overset{*}{\text{a}}$per ]

b. $[_\pi$ Tom$\overset{*}{\text{o}}$rrow ] $[_\pi$ the book may be che$\overset{*}{\text{a}}$per ]

The preposed adverb *tomorrow* can occur in exactly the same intonation contour as the verb *to borrow*. That is, from an intonational point of view, both utterances ought to be analysed as single $\pi$'s. Yet, from a rhythmical point of view, there is an audible break after *tomorrow*, which is absent after *to borrow*. For additional evidence involving the application of segmental rules, see Gussenhoven and Rietveld (1992a)

[6] That is, not a filled pause, hesitation, etc.

Whereas these examples concern instances where a single intonation contour spans two $\pi$'s, (5) involves a single $\pi$ that contains at least two intonation contours. Gussenhoven points out that in the following pair of utterances the contours on *grey* and *bread* (5-b) can be identical to the contours on *Grey* and *bread* in (5-a), that is, a sequence of rise-falls, with an intonational break before *bread*.

(5) a. I brought [$_\varphi$ Gr$\overset{*}{\text{e}}$y ] [$_\varphi$ br$\overset{*}{\text{e}}$ad ]

b. I brought [$_\varphi$ gr$\overset{*}{\text{e}}$y br$\overset{*}{\text{e}}$ad ]

While (5-a) may be accounted for by claiming that the $\varphi$-boundary after *Grey* coincides with a $\pi$-boundary, this solution is unavailable in (5-b), since the principles of prosodic phonology forbid a $\pi$ to be dominated by a $\varphi$.

The crucial point is to show that *grey bread* in (5-b) really constitutes a single $\varphi$. One argument is that a pause may be inserted after *Grey* in (5-a), which is far less acceptable after *grey* in (5-b). A better argument is found in the facts of 'stress shift'. It is generally accepted that the $\varphi$ forms the domain for 'stress shift' phenomena (see Section 2.5.3). A word like *unbaked* normally has its main stress on the second syllable (cf. (6-a)).[7] However, if it is immediately followed by another stressed word within the *same* $\varphi$, its main stress retracts to the first syllable in order to avoid a stress clash with the next word. As this is exactly what happens when we substitute *unbaked* for *grey* (cf. (6-a) versus (6-b)), we can infer that *grey bread* is indeed a single $\varphi$. In contrast, substituting *Marie-Lou* for *Grey* does not trigger stress shift (cf. (6-c) versus (6-d)), indicating that *Mary-Lou bread* is not a single $\varphi$, and by implication, cannot contain a $\pi$-boundary either without violating the principles of prosodic structure.[8]

(6) a. I brought bread that is ùnbáked.

b. I brought únbàked bréad

c. I brought Màrie-Lóu bréad

d. *I brought Márie-Lòu bréad

To sum up, the incorporation of the intonational phrase as a prosodic constituent ($\pi$) in the framework of prosodic phonology gives rise to contradictions. Gussenhoven and Rietveld (1992a, p. 286) conclude: "[...] in general terms, the problem is that the stretch of speech covered by an intonation contour cannot consistently be identified with a particular prosodic constituent." The core of the problem is a mismatch between the two criteria for an intonational phrase,

[7] Main stress is indicated by ´, secondary stress by `.

[8] Yet another problem is that a $\varphi$, by definition, must contain at least one lexical head, i.e. a noun, verb, or adjective that is a content word. However, a tag question as in *You're not going, are you?* does not contain a lexical head; just an auxiliary and a pronoun. Therefore, it cannot be a $\varphi$. However, tag questions are considered to constitute separate $\pi$'s, for both intonational and durational reasons (cf. Section 4.4.1). As a result, we have a $\pi$ that does not dominate a $\varphi$, which violates the principles of prosodic structure.

i.e. a coherent intonation contour and an audible pause, which is also discussed in Ladd (1986).

The introduction of AD's can be seen as a solution to this problem: "[...] there is an independent constituent over which the tone segments of a pitch accent are distributed: the association domain (AD). The AD is determined by the structure provided by the conventional prosodic constituents, but is not itself part of the prosodic hierarchy. On this view the status of the conventional I[$\pi$, in our notation] is similar to that of the other prosodic constituents, like Ph[$\varphi$, in our notation]. It will be pause-defined, and may play a role in the structural description of phonological rules, but it would not be definable on the basis of the intonation contour." (Gussenhoven and Rietveld 1992a, p. 287).

Unfortunately, the resulting terminology is a potential source of confusion, so let us state clearly what we mean by the terms *intonational phrase*, $\pi$, and AD. We will keep using *intonational phrase* as a general, theory-neutral term that refers to a stretch of speech with a perceptually coherent intonation contour. The $\pi$ and AD are theory-specific interpretation of this: $\pi$ is the intonational phrase as a prosodic constituent in prosodic phonology; AD is the intonational phrase in Gussenhoven's intonational phonology. We will use *$\pi$-phrasing* for the determination of $\pi$'s, and *AD-derivation* for the derivation of AD's. However, determining intonational phrases is largely independent of the question about the correct theoretical status of the intonational phrase. That is, whether we prefer to analyse intonational phrases as $\pi$'s, AD's, or some other theoretical construct, the problem of how to derive them for a particular utterance in a particular context remains essentially the same. Therefore, we will also continue to use the generally accepted term *intonational phrasing* whenever the difference between $\pi$'s and AD's is irrelevant.

### 2.3.8 Summary

To sum up, we have assumed an intonational model with the following properties.

- Following the principles of intonational phonology, there are two levels of analysis: phonological description and phonetic implementation.

- The description of the intonation of an utterance is autosegmental, consisting of a tone string of tonal events that is aligned with a segmental string by means of tiers. The tonal events are either pitch accents (H*L, L*H, or H*LH) or boundary tones (%H, %L, H%, or L%), which are both composed of the primitive tones H and L. Every starred tone of a pitch accent must be linked to a stressed syllable, and every boundary tone must be linked to an AD-boundary; the remaining tones either spread, or move up to the next tone or AD-boundary.

- The point of departure is an underlying representation in which every utterance is exhaustively parsed into AD's, and every AD contains exactly one pitch accent. AD-restructuring allows for merging of AD's. Tonal rules like Tone Linking may transform the underlying representation to a surface representation. This will then be mapped onto an actual pitch contour by phonetic implementation rules, taking into account tone morphemes like Downstep.

- The intonational phrase cannot be incorporated into the prosodic hierarchy ($\pi$), but must be analysed as a independent domain called the association domain (AD).

## 2.4 Generating intonation

The intonational phonology of Dutch described in the previous section aims to define all well-formed intonational patterns, or more precisely, all well-formed combinations of tone strings and segmental strings. It allows for variation along three dimensions.

1. An utterance may be divided into AD's in many different ways. Taking Figure 2.6 as a (arbitrary) reference, Figure 2.7 shows an alternative division into AD's, in which the first two AD's have been restructured to a single one. We will refer to the choice of AD's as *AD-derivation*, or generally as *intonational phrasing* if there is no chance of confusion with $\pi$-phrasing.

2. The number of pitch accents, or their locations, may be different. Example Figure 2.8 shows a variant in which the word *stoptrein* is unaccented.[9] The choice of the number of accents and their locations in an utterance will be referred to as *pitch accent placement*.

3. The kind of pitch accent may vary. In Figure 2.9, both accents of the type L*H; the H tone at the end of the final AD will cause this utterance to be interpreted as a question. In addition, tonal rules like Tone Linking and tone morphemes like Downstep may alter the shape of the contour. The choice of the type of pitch accents and boundary tones, and the application of tonal rules and tone morphemes in an utterance will be called *tune choice*.

The intonational grammar describes the variation along these three dimensions from a purely structural point of view, analogous to the way an autonomous syntactic grammar attempts to define all grammatical strings of a language without taking their meaning into account. However, many, if not most, of the intonational variants that are structurally well-formed will not be appropriate for a particular utterance in a particular context. It is generally accepted that this is because the intonational form is constrained by

[9] As a consequence, the utterance can have at most two AD's, since AD's without an accent are ill-formed. This shows that the three dimensions of variation are not independent of each other.

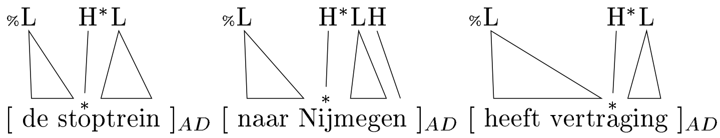



*Figure 2.6: An example of an intonational form for the utterance 'The slow-train to Nijmegen has delay'.*

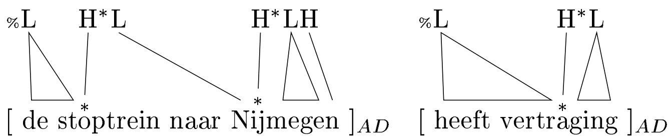



*Figure 2.7: The same utterance with a different division into AD's.*

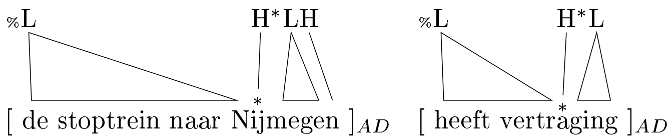



*Figure 2.8: The same utterance with a different number of pitch accents as well.*

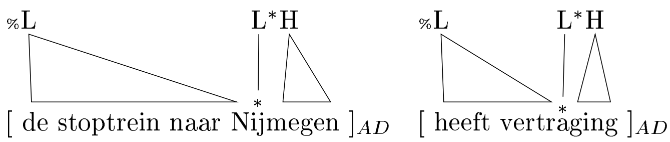



*Figure 2.9: The same utterance with a different type of pitch accent as well.*

the semantic, morpho-syntactic and phonological structure of the utterance, in combination with contextual and pragmatic factors. This has proved to be a complex relation, and has given rise to a large body of research in theoretical and computational linguistics, psycholinguistics, phonetics and speech technology. In the current work, we will address this issue from the point of view of generating intonational forms in an SLG system. The challenge of *generating intonation* is thus to perform intonational phrasing, pitch accents placement and tune choice in accordance with the linguistic structure and context of an utterance.

The complement of generating intonation may be called *interpreting intonation*, and concerns the way the intonation of an utterance constrains its structure and meaning, and that of the discourse it is part of. Whereas the former is typically found in TTS and SLG, the latter may occur in automatic speech recognition (ASR) and spoken dialogue systems. Although the present work is primarily concerned with generating intonation, a considerable part of the underlying theory actually concerns the general relation between intonation on the one hand and syntax, semantics, phonology and discourse on the other hand, and may therefore be relevant to the problem of intonation interpretation as well.

Pitch accent placement, as was already demonstrated in Section 1.1.2, is sensitive to the distinction between 'new' and 'given' information. In the next section, we will present a provisional algorithm for pitch accent placement that requires other input as well. It will be argued that other relevant factors include the predicate-argument structure, adjacency relations in the surface string, and the prosodic structure (phonological phrases). The main purpose of presenting this grammar is to explain why we will be concerned with phonological phrasing in Chapter 3, designing an appropriate SLG environment that can supply the required information in Chapter 5, and generating syntactic structures in Chapter 6. A more elaborate algorithm for pitch accent placement will then be presented in Chapter 7.

Intonational phrasing appears to be mainly dependent on the syntactic structure. In Chapter 4, several algorithms for intonational phrasing proposed in the literature will be systematically compared. This comparison will reveal a number of flaws in these algorithms. On the basis of this, we will propose and test a new algorithm.

Tune choice appears to be closer to pitch accent placement in the sense that it is affected by many factors, among others, the sentence mood (e.g. statement versus question) and the discourse structure. However, tune choice is beyond the scope of our research objective, and will only be superficially addressed in the current work.

The theoretical framework described so far contributes in at least two ways to our research goals. First, it allows us to focus on generating intonational forms at a phonological level, abstracting away from the details of phonetic implementation. Second, it defines well-formed intonational forms, ruling out those forms that are unacceptable on purely phonological grounds. In addition, it provides a precise definition of tune choice, from which we intend to abstract away, versus pitch accent placement and intonational phrasing.

## 2.5 Pitch accent placement

Pitch accent placement, or accentuation for short, refers to predicting the distribution of pitch accents in an utterance, i.e. the locations of pitch accents. At the level of intonational phonology, a pitch accent is linked to a stressed syllable, whereas at the level of the phonetic implementation, it is even more precisely located, when its targets are timed relative to the segments that make up the syllable (Section 5.6.2.2). For current purposes, however, it is convenient to specify the locations of pitch accents at a higher level of abstraction: we will assume that pitch accents are attached to words. Our goal is thus to explain and predict which words of an utterance are accented.

Distributing pitch accents over words abstracts away from the fact that some words can carry more than one pitch accent. For example, the word *roodbruin* 'redbrown', which is a compound adjective, may in some contexts carry pitch accents on both constituents. This implies that, in principle, we have to go down to the morphological level of non-compound words to provide a pitch accent with a unique location. This issue will be further discussed in Section 7.3.1.1. In most cases, however, it will suffice to locate each pitch accent by anchoring it to a word. We will continue to use a star (*) above the stressed syllable(s) of a word to indicate that it is pitch accented.

In accordance with our linguistic motivation, our goal is a *grammar for pitch accent placement*, that is, a generative grammar that takes into account all relevant factors in order to produce an acceptable accent distribution. A number of such grammars have been proposed, for example by Selkirk (1984, 1995) for English, or Baart (1987), Dirksen (1992, Dirksen and Quené (1993), and van Wijk (1987b) for Dutch. The grammar used in the present work is due to Gussenhoven (1984, 1991a, 1992b, 1999). Below, we will review the relevant factors and consider how they can be modelled the in grammar. The general strategy is one of deletion. That is, we start from the assumption that every word is is accented in the underlying representation, and our ambition is to predict which accents survive in the surface representation.

### 2.5.1 Focus

Language is typically used to exchange information by means of utterances that are not just sets of primitive propositions, but complex structured messages, which convey more aspects of meaning than just the propositional content. One of these aspects concerns the relative importance of the parts of information that constitute the message, which has been called the *information structure*.

> "Information structure is the encoding of the relative salience of the constituents of a clause, especially nominals, and is realized as

> choices among alternative syntactic arrangements. The information structure of a clause is determined by the larger sentence or discourse of which it is part (i.e. its context). The communicative effect of the information structure is to foreground certain aspects of the message of the clause, but to background others. The need to encode information structure is a language universal, but the formal means to do so vary widely across the languages of the world" (Foley 1994)

Almost every theory about grammar encompasses some notion of information structure, although the differences in terminology and definitions is bewildering, making it one of the most confusing issues in linguistics (Grosz, Joshi, and Weinstein 1995; van Kuppevelt 1991; Rooth 1995; Bosch and van der Sandt 1999). Nevertheless, some notion of information structure is essential for the purpose of pitch accent placement. In addition, it is important for the NLG component, because it affects syntactic variations in word choice and word order. We therefore will attempt to explicate our assumptions about information structure.

Foley's definition of information structure has three important aspects:

1. some information is more salient than others;

2. the distinction is formally (syntactically) encoded;

3. and depends on the context.

With respect to the first point, we will assume a binary distinction between *focus* and *background*. The partition of an utterance into focused and unfocused parts will be called its *focus distribution*. The focus distribution is not necessarily equivalent to the information structure; rather, the former constitutes a part of the latter, as the information structure is often considered to include additional distinctions. For instance, that between *topic* and *comment*, where the former represents what the sentence is about, and the latter what the sentence has to say about this: "The topic is presupposed information that the remainder of the sentence comments upon." (Foley 1994). In functional approaches to grammar (Halliday 1994), this is referred to as *theme* versus *rheme*. We will ignore other aspects of the information structure here, and concentrate on the focus distribution.

The focus-background distinction is closely related to the new-given distinction. Generally, new information, which is added to the discourse for the first time, will be focused. Conversely, old information, which has been recently mentioned in the discourse, will be left unfocused. The prototypical example of this involves a question-answer pair. Since a question is a request for a specific part of new information, it requires this piece of information to be focused in the reply. For example, *where* in (7-a) forces focus on the requested location, i.e. *a party*, in (7-b), where focus is indicated by underlining.

(7) a. Where did Sue go to?

b. She went to $\underline{\text{a p}\overset{*}{\text{a}}\text{rty}}$.

The question arises how a listener recognises the focus in (7). This brings us to the second point of the above definition of information structure: the distinction is formally encoded, by morphological, syntactic and phonological means. The reason for assuming this focus distribution is that the speaker of (7-b) used several linguistic means to signal it. To begin with, the speaker chose to refer to *Sue* by means of the pronoun *she*. A pronoun usually refers to an accessible antecedent, i.e. one that has been recently mentioned in the discourse or is contextually evoked, and is therefore unfocused. Second, *a party* is an indefinite noun phrase, which is normally used to add a new, and therefore focused, element to the discourse. Furthermore, the word order is such that the focused information is located at the end of the clause. This is in accordance with the canonical word order for a clause in English (and Dutch) in which unfocused information precedes focused information. Deviation of the normal word order often results in a distinct pragmatic effect (Rochemont and Cullicover 1990; Rats 1996); cf. *To a party she went*. Finally, and most importantly for the discussion at hand, the speaker chose to put a pitch accent on *party*. This illustrates the general tendency for focused information to be pitch accented, whereas unfocused information is not (Birch and Clifton 1995; Brown 1983; van Donzel 1999; Fowler and Housum 1987; Horne 1991; Kruyt 1985; Nooteboom and Terken 1982; Nooteboom and Kruyt 1987). This focus-accent correspondence forms the point of departure for our grammar for pitch accent placement.

If we alter the focus cues of an utterance, its focus distribution, and consequently its interpretation, will be affected to some extent. For instance, changing the pronoun *she* to the proper noun *Sue*, as in (8-a), renders the utterance somewhat formal. On the other hand, using a pronoun for the focused information, as in (8-b), makes it an incoherent answer, as its antecedent cannot be identified by the listener.[10] A similar effect arises from changing the pitch accent distribution in (7-b) to (8-c). Apparently, the pitch accent on *party* is essential for signalling the focus distribution required by preceding the wh-question. Putting a pitch accent on old information may result in pragmatic side-effects, as in (8-d), which appears to imply something about another person than Sue as well.

(8) a. Sue went to a party.

b. She went there.

c. S$\overset{*}{\text{u}}$e went to a party.

d. S$\overset{*}{\text{u}}$e went to a p$\overset{*}{\text{a}}$rty.

[10] Assuming there is no additional context from which the antecedent can be inferred. The antecedent of *there* would be clear if it were accompanied by a pointing gesture.

The listing of focus cues above is by no means exhaustive; different languages have shown to employ different strategies (Foley 1994). Moreover, there is no straightforward, fixed, or one-to-one correspondence between focus and its linguistic cues, because they may serve to signal other functions or may be bound by other grammatical constraints. For instance, an *indefinite* noun phrase may be *unfocused*, when its indefiniteness is meant to signal a generic reading instead of a specific one. For instance, *a printer* in (9) refers to a generic, or prototypical printer, and not to a specific individual we can point at.

(9) Printers are a good example of unreliable technology.
A printer will always break down when you need it the most.

Likewise, the preferred word order with unfocused preceding focused information may be disturbed due to purely syntactic constraints on word order (Section 6.3.7), or for stylistic purposes as in left-dislocation or similar marked word orders (Rochemont and Cullicover 1990).

(10) It is Sue who went to a party.

Often, there will be interaction among different cues. For instance, although a pronoun normally refers to unfocused information, it may be focused by giving the pronoun a pitch accent (Verhagen 1986). This often occurs in a context where previously introduced discourse entities are contrasted.

(11) She went to the party (while I had to stay home).

The linguistic realization of focus poses a challenge to linguistic theories, and is an important aspect of NLG and SLG. Although some aspects of the relation between focus and word order will be addressed in Chapter 6, our work is primarily concerned with the accent-focus relation. Important for the discussion at hand is that there is no one-to-one correspondence between the focus distribution of an utterance and its pitch accent distribution either. Consider example (12). Given its similarity to (7), it is reasonable to assume that the whole phrase *a very big party* is focused. Nevertheless, the word *big* remains unaccented. In fact, the determiner *a* also remains unaccented in both examples.

(12) a. Where did Sue go?

b. She went to a very big party.

A grammar of pitch accent placement has to account for these type of observations; they will be addressed in Section 2.5.3.

We continue with the third point of the definition of information structure: the it depends on the context. This has been demonstrated by the wh-question in (7), which determines the focus distribution in the answer. This raises the question where the focus distribution comes from. In most cases, the relation between the focus distribution and the context is less obvious. It was already mentioned that the focus-background distinction is closely related to the given-new distinction. However, the focus distribution cannot be equated to, or derived from, the given-new distinction. In (13), for instance, *Mike* and *Sue* are new information at the start of the utterance, and consequently focused, as indicated by the pitch accents. Even though the second instance of the word *Sue* can be assumed to convey old information, as it was recently mentioned, it remains focused and accented. A possible explanation is that the second time, it is focused to contrast it with the elements of a set that was introduced earlier (Theune 1997b; Theune 1997a; van Deemter and Odijk 1997). Thus contrast may be one of the reasons that old information is nevertheless focused.

(13) Both M$\overset{*}{\text{i}}$ke and S$\overset{*}{\text{u}}$e like h$\overset{*}{\text{i}}$king in the m$\overset{*}{\text{o}}$untains, although S$\overset{*}{\text{u}}$e only likes going d$\overset{*}{\text{o}}$wn.

Likewise, new information is not always focused. An example of this is (14). Although *the car* is a new piece of information, in the sense that the word has not been mentioned yet, it is unfocused. Apparently, the utterance, or rather the speaker, assumes that the listener knows that a Roll Royce is a kind of car, and can infer that *the car* is just another way of referring to a previously mentioned discourse entity, without adding any important information.

(14) The Rolls Royce was severely damaged.
It took s$\overset{*}{\text{i}}$x w$\overset{*}{\text{e}}$eks to rep$\overset{*}{\text{a}}$ir the car.

Examples like these indicate that the focus distribution is not equivalent to or derivable from the difference between given and new information in any obvious way. They have motivated some researchers to propose a classification of *givenness* on a more refined scale (instead of the binary distinction focus versus background) with classes like *brand-new*, *inferable*, *situationally invoked*, etc. (Prince 1981; van Deemter and Odijk 1997). These are interesting options to explore, but at our current state of knowledge it is hard to relate these categories to the accent distribution (see however van Donzel (1999)).

There may be many (pragmatic) reasons for a speaker to chose a particular distribution of focus. For current purposes, however, we will leave questions about the reasons for a particular focus distribution to semantic and pragmatic research, and concentrate on the question how focus is *realized*, in particular by means of pitch accent placement. In other words, we will assume that focus is a theoretical primitive, and abstract away from the underlying motivations that the speaker may have for a particular focus distribution. This entails that the underlying motivation for a particular focus distribution is not allowed to affect the distribution or the realization of accents. Evidence for this is provided by a study that addressed (among other thing) the relation between informational categories and their prosodic realization in Dutch (van Donzel 1999, p. 152):

"Measurements of the actual realization of the pitch accents in fun-

> damental frequency revealed that the type of information on which the pitch accent is realized has no effect. That is, pitch accents realized on new information are not realized differently from those realized on inferable information or modifiers, for instance."

Likewise, Kruyt (1985, ch. 6) reported on a perception experiment showing that neither the particular type of pitch accent, nor the presence or absence of a subsequent prosodic boundary has any effect on the interpretation as 'given' or 'new'.

### 2.5.2 Argument structure

In Gussenhoven (1992b), argument structure is defined as follows:

> "[T]hree types of constituent are assumed: predicates, arguments (also 'inherent arguments') and modifiers (also 'noninherent arguments'). Every predicate constitutes a clause, which will have from zero to three arguments, and any number of modifiers. The distinction between arguments and modifiers corresponds with the distinction between 'inherent' roles and 'noninherent' roles, that is, between roles that are specified in the lexical entry of a predicate and those that are not. Expressions that instantiate listed roles are inherent arguments, or simply arguments, while expressions filling other roles are noninherent arguments, or modifiers."

The distinction between arguments and modifiers plays a role in morphology, syntax and phonology. For instance, it is important in NLG, because it constrains the word order in a clause (Section 6.3.7). Here, we will concentrate on its effect on accent placement. The following examples, which were taken from Gussenhoven (1992b), illustrate a systematic discrepancy between the focus distribution and the accent distribution, which occurs in English (15), as well as in Dutch (16). In (15-a), the verb *smoked* is focused and accented, whereas in (15-b), the verb *remained* is focused but *not* accented. The problem is to explain why in the (a) examples an accent on the verb is required to signal its focus, whereas in the (b) examples an accent on the verb can be omitted without affecting its focus.

(15) What did John do?

a. John smoked in the tent

b. John remained in de tent

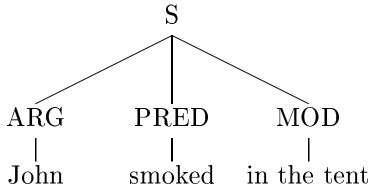
(16) Wat deed Jan?

a. Jan heeft in de tent gerookt

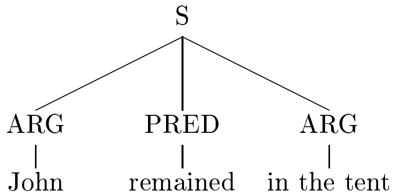
b. Jan is in de tent gebleven

Gussenhoven's solution is to take the argument structure of a clause into account (Gussenhoven 1984; Gussenhoven 1992b). The argument structures corresponding to (15-a) and (15-b) are given in (17-a) and (17-b), respectively.

(17) a.



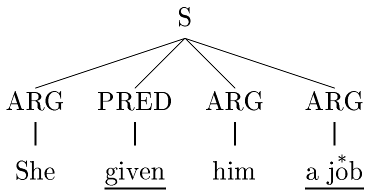

b.




The lack of an accent on the predicate is attributed to the fact that it is adjacent to a focused argument. Following Gussenhoven's terminology, we can say that the predicate *remained* and the argument *in the tent* merge into a single *focus domain*, which is satisfied with just one accent.

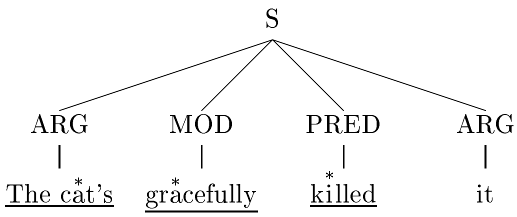
Actually, the situation is more complex, because unfocused elements are ignored. This is illustrated in (18), where the intervening unfocused argument does not prevent the predicate *given* from forming a single focus domain with its focused argument *a job*.

(18) Any news about Mary and John?




However, an intervening *focused* modifier does prevent a focused predicate from forming a single focus domain with its argument.

(19) Where's the canary?



Gussenhoven proposed to explain these patterns by the following generalisation.

(20) **Sentence Accent Assignment Rule** (SAAR)

If focused, every predicate, argument, and modifier must be accented, with the exception of a predicate that, discounting unfocused constituents, is adjacent to an argument

Experimental evidence in support of this generalisation is reported in Gussenhoven (1983b) and Birch and Clifton (1995). We will use SAAR as the basis for our accent placement algorithm. There are a number of exceptions to SAAR (Gussenhoven 1992b, Appendix), some of which will be addressed in Section 7.3.1.

### 2.5.3 Rhythm

Example (21) repeats another discrepancy between the focus distribution and the accent distribution.

(21) a. Where did Sue go?

b. She went to $\underline{\text{a v}\overset{*}{\text{e}}\text{ry big p}\overset{*}{\text{a}}\text{rty}}$

Although there is no reason to assume that *big* is not focused, it is unaccented. Again, the problem is to explain why the accent on *big* can be omitted without affecting the focus interpretation. Following Gussenhoven (1991a), we will assume that the accent on *big* may be deleted to obtain a more rhythmical distribution of pitch accents, i.e. a more regular pattern of accented and unaccented words. Responsible for this is an accent deletion rule called the *Rhythm Rule*.

(22) **Rhythm Rule** (RR)

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

This rule allows the deletion of an accent if it is surrounded by accents within the same phonological phrase, the prosodic constituent introduced in Section 2.2). Application of the rule is optional, and it can be applied more than once. The net effect of exhaustive application is that only the outermost accents within a phonological phrase are preserved. Empirical evidence for this rule in English is reported in Horne (1990); see also Grabe, Warren, and Nolan (1994).

The phonological phrases corresponding to (21-b) are as follows, where the former accent deleted by RR is indicated by ø.

(23) $[_{\varphi}$ She went $]$ $[_{\varphi}$ $\underline{\text{to a v}\overset{*}{\text{e}}\text{ry b}\overset{\text{ø}}{\text{i}}\text{g p}\overset{*}{\text{a}}\text{rty}}$ $]$

Notice that an alternative phonological phrasing, such as (24), would not lead to the desired result.

(24) [$_\varphi$ She went ] [$_\varphi$ to a vèry ] [$_\varphi$ bìg pàrty ]

Moreover, incorrect phonological phrasing may trigger an accent deletion that gives rise to an unintended focus distribution, as in (25).

(25) a. Where did Sue go?

b. ?[$_\varphi$ She went ] [$_\varphi$ to a bìg party in Àmsterdam ]

Thus, in order to prevent incorrect accent deletions, it is crucial that the boundaries of phonological phrases are correctly determined. Therefore, phonological phrasing in Dutch will be investigated in Chapter 3.[11]

### 2.5.4 A grammar for pitch accent placement

To sum up, our preliminary grammar for pitch accent placement requires four types of input:

- the focus distribution
- the argument structure (predicates, their arguments, and optional modifiers)
- the prosodic structure, minimally consisting of the phonological phrases
- the surface string

Given this input, the grammar consists of just two rules:

1. the Sentence Accent Assignment Rule (SAAR) to account for the focus-accent relation
2. the Rhythm Rule (RR) to account for rhythmical adjustments to the pitch accent distribution

An example of its application is given in (26).

This simple grammar forms the basis for the elaborate pitch accent placement algorithm in our SLG system, which will be described in Chapter 7.

[11]There is another type of stress shift, or pitch accent retraction, that appears to serve rhetorical purposes, and is called *rhetorical stress shift* in (Gussenhoven 1983a). For example, *ongelòoflijk* 'unbelievable' can become *òngelooflijk* in a phrase like *wat een òngelooflijke fòut!* 'What an unbelievable mistake!'. We abstracted away from this phenomenon, as it did not occur in our application domains.

(26) *Context:* What did you do last evening?

S

*Input:* ARG PRED ARG

[$_\varphi$ we <u>vistited</u> ] [$_\varphi$ <u>a very big party</u> ]

*SAAR:* we $\overset{\emptyset}{\text{visited}}$ a $\overset{*}{\text{very}}$ $\overset{*}{\text{big}}$ $\overset{*}{\text{party}}$

*RR:* we visited a $\overset{*}{\text{very}}$ $\overset{\emptyset}{\text{big}}$ $\overset{*}{\text{party}}$

*Output:* we visited a $\overset{*}{\text{very}}$ big $\overset{*}{\text{party}}$

## 2.6 Dutch syntax

There are at least three reasons why we need syntactic structures in the current work. First, correct application of RR depends on correct phonological phrases, which are derived from the syntactic structure of an utterance by means of mapping rules (Section 2.2). Second, syntactic information is essential for intonational phrasing. Third, syntactic knowledge is required for generating syntactically well-formed output in our SLG system (see Chapter 6). However, there are many different approaches to the syntactic analysis of Dutch, ranging from traditional structuralistic grammars (Haeseryn et al. 1997; Rijpma and Schuringa 1978; van den Toorn 1984) to modern transformational (Model 1991; Bennis and Hoekstra 1989; Zwart 1993), categorial (Bouma and van Noord 1995), and unification-based grammars (Bouma 1996). Rather than adopting a particular syntactic theory, we made some minimal assumptions about the phrasal structure that can be satisfied by most theories. This keeps our results compatible with different syntactic views and guarantees a wide range of application.

The following two sections illustrate our assumptions on the basis of a some examples. Admittedly, this is rather simplistic and ignores complications; a more elaborate account can be found in Chapter 6 about syntactic realization.

### 2.6.1 S/VP structure

We assume clauses (S) of two types: main clauses (27) and subordinate clauses (28). Both minimally contain a verb phrase (VP). Main clauses have a finite verb at the syntactically second position, directly dominated by the S-node. Declarative main clauses (and wh-questions) have a preposed constituent at the first position (27-a), whereas polar questions (and imperatives) do not allow this (27-b). Extraposed prepositional phrases (PP), clauses and heavy noun phrases (NP) may occur after the VP (27-c).

(27) a. ($_S$ ($_{NP}$ u ) ($_V$ kunt ) ($_{VP}$ ($_{PP}$ in Amsterdam ) ($_{NP}$ de<br>
you can in Amsterdam the<br>
stoptrein ) ($_V$ nemen ) ) )<br>
slow-train take<br>
'you can take the slow train in Amsterdam'

b. ($_S$ ($_V$ kan ) ($_{VP}$ ($_{NP}$ ik ) ($_{AdvP}$ daar ) ($_{NP}$ de stoptrein )<br>
can I there the slow-train<br>
($_V$ nemen ) ) )<br>
take<br>
'can I take the slow train there?'

c. ($_S$ u kunt ($_{VP}$ de stoptrein nemen ) ($_{PP}$ om twaalf uur<br>
you can the slow-train take at twelve hour<br>
vijf ) )<br>
five<br>
'you can take the slow-train of five past twelve'

Subordinated clauses have all verbs at the end of the VP. Bound subordinated clauses start with a complementizer (C) (28-a), and relative subordinated clauses with a relative expression (28-b).

(28) a. ($_S$ ($_C$ dat ) ($_{VP}$ u in Amsterdam de stoptrein kunt<br>
that you in Amsterdam the slow-train can<br>
nemen ) )<br>
take<br>
'that you can take the slow train in Amsterdam'

b. ($_{NP}$ de stoptrein ($_S$ ($_{NP}$ die ) ($_{VP}$ u in Amsterdam<br>
the slow-train that you in Amsterdam<br>
neemt ) ) )<br>
take<br>
'the slow train that you take in Amsterdam'

### 2.6.2 NP structure

Noun phrases are headed by (possibly empty) noun, may be preceded by a determiner (Det), a quantifier (Q), a numeral phrase (NumP)[12], and premodifying adjectival phrases (AP), a premodifying NP[13], and followed by postmodifying prepositional phrases (PP) or relative clauses (S).

12 See Section 6.2.3.4

13 See Section 7.3.2

(29) a. ($_{NP}$ ($_{Det}$ de ) ($_{AP}$ ($_{A}$ gele ) ) ($_{N}$ bloemen ) ($_{PP}$ ($_{P}$ in ) ($_{NP}$ het gras ) ) )
the yellow flowers in the grass

b. ($_{NP}$ ($_{Q}$ alle ) ($_{N}$ bloemen ) ($_{S}$ die geel zijn ) )
all flowers that yellow are

c. ($_{NP}$ ($_{NP}$ station ) ($_{N}$ Amsterdam-Centraal ) )
station Amsterdam-Central

### 2.6.3 Pronouns are determiners

In contrast to the above analyses, which are rather conventional, we assume that pronouns are actually determiners, following the analysis by Coppen (1991, p. 100-112), which is compatible with the DP analysis of Abney (1986). We repeat some of Coppen's arguments below.

First, a so-called demonstrative pronoun (Haeseryn et al. 1997, ch. 5) can precede a noun (30-a) in the same way a determiner can precede a noun (30-b), suggesting an identical syntactic structure. For independent demonstrative determiners, we assume the NP has an empty head (30-c).

(30) a. ($_{NP}$ ($_{Det}$ die ) ($_{N}$ boeken ) )
those books

b. ($_{NP}$ ($_{Det}$ de ) ($_{N}$ boeken ) )
the books

c. ($_{NP}$ ($_{Det}$ die ) ($_{N}$ $\emptyset$ ) )
those

Second, both so-called personal (31-b) and demonstratives pronouns (31-c) can be followed by a quantifier, and are similar to a determiner in this respect (31-a), whereas nominal heads do not have this possibility (31-d).[14]

(31) a. de twee boeken
the two books

b. wij twee(en) hebben dat gedaan
we two(+plur) have that done
'the two of us did that'

c. die twee hebben dat gedaan
those two have that done
'those two did that'

[14] Coppen remarks that this pattern occurs only with *plural* pronouns, and with numerals below ten.

d. *de kinderen twee hebben dat gedaan
the children two have that done

The third argument, which goes back to Postal (1969), is that pronouns can appear in the same position as determiners.[15]

(32) a. wij fietsenmakers
we bicycle-repairmen

b. jij kleine wijsneus
you little know-it-all

c. ik arme stumper
I poor wretch

See Coppen (1991) for additional arguments and a more extensive analysis.

## 2.7 Summary

This chapter presented the theoretical framework for our work on pitch accent placement and intonational phrasing in SLG, providing a more precise interpretation of these issues in terms of phonological theories, and a first suggestion as to how we may account for them.

In Section 2.2, prosody was defined as the hierarchical organisation of segments into higher units, the patterns of relative prominence among these units, and their phonetic reflection in terms of phonetic parameters. Prosodic phonology was presented as the study of prosody from a phonological point of view. We introduced its core concept: a prosodic structure consisting of seven prosodic categories derived by mapping rules, and serving as the interface from morphology, syntax and semantics to phonology. We discussed the well-formedness constraints imposed on prosodic structures by the prosodic hierarchy and the Strict Layer Hypothesis. We briefly reviewed two arguments for the existence of prosodic constituents. First, to explain segmental phenomena: certain segmental rules can only apply within – or at the boundary of – prosodic constituents of a certain rank. Second, to explain the distribution of prosodic boundaries in an utterance: the higher the rank of a predicted prosodic constituent, the more prominently its boundaries should be perceived as a result of phonetic cues such as pause and lengthening of the preboundary syllable (preboundary lengthening).

In Section 2.3, intonational phonology was introduced as the study of intonation from a phonological point of view, relying on a division of labour between phonological description and phonetic implementation. We summarised the intonational phonology of Dutch developed by Gussenhoven, consisting of a tone string of tonal events that is aligned with a segmental string by means of

[15] Coppen convincingly argues against the view that these constructions should be analysed as containing an appositional NP.

tiers. The tonal events are either pitch accents (H*L, L*H, or H*LH) or boundary tones (%H, %L, H%, or L%), both of which are composed of the primitive tones H and L. According to the association principles, every starred tone of a pitch accent is linked to a stressed syllable, and every boundary tone to an AD-boundary, while the remaining tones either spread, or move up to the next tone segment or AD-boundary.

The model is transformational in the sense that the underlying structure is rewritten to a surface structure. In the underlying representation, every utterance is exhaustively parsed into AD's, and every AD contains exactly one pitch accent. AD-restructuring allows for merging of AD's. Tonal rules like Tone Linking may modify the tone string, and tone morphemes like Downstep may be attached. The structure is mapped onto an actual pitch contour by the phonetic implementation module, taking into account tone morphemes like Downstep.

We also followed Gussenhoven in his view that the intonational phrase cannot be incorporated into the prosodic hierarchy, but must be analysed as a independent domain called the association domain (AD).

In Section 2.4, it was shown that the model provides a precise account for variation along three dimensions: (1) intonational phrasing (the partition of an utterance into one or more AD's); (2) pitch accent placement (the number of pitch accents and their association with stressed syllables in the utterance); (3) tune choice (the choice of the kind of pitch accents and boundary tones, and the application of tonal rules and tone morphemes). The challenge of *generating intonation* was defined as intonational phrasing, pitch accent placement and tune choice in such way that the resulting intonational form fits the linguistic structure and context of an utterance. The contribution of the intonational model in this respect is (1) that allows us to focus on generating intonational forms at a phonological level, abstracting away from the details of phonetic implementation, and (2) that it defines the well-formed intonational descriptions, ruling out those forms that are unacceptable on purely phonological grounds.

In Section 2.5, we reviewed three factors that affect pitch accent placement: focus, argument structure, and rhythm. We presented a preliminary grammar for pitch accent placement based on the work of Gussenhoven. It requires as input the argument structure (predicates, their arguments, and optional modifiers), the focus distribution, the prosodic structure (the phonological phrases), and the surface string. It consists of just two rules: (1) the Sentence Accent Assignment Rule (SAAR) to account for the focus-accent relation, in particular for focused yet unaccented predicates, and (2) the Rhythm Rule (RR) to account for rhythmical adjustments by deletion of pitch accents within phonological phrases. It was explained that in order to prevent incorrect accent deletions, it is crucial that the boundaries of phonological phrases are correctly determined. Finally, Section 2.6 described the syntactic analysis of Dutch used throughout our work, relying on minimal assumptions about the phrasal structure that are compatible with most syntactic theories.