Chapter 1

Introduction

When two theories purport to explain the same set of data, we will choose the one that is simpler, or if they seem much the same on that score, the one that additionally explains new data. In intonation research, where variant theories commonly exist, evaluation is in practice hampered by the circumstance that the new data required for testing predictions of these theories do not exist. The data that *are* available – examples used in the literature – are limited in number, and possibly selective in various ways. One way of overcoming this problem is to implement theories in a computer program, get them to *generate* spoken language, and see whether their predictions with respect to the intonation make sense. The question then is whether the generative capacity of the theory is similar – by some criterion – to that of humans. This is the question that the investigation reported here tries to answer.

To pursue our aim, we would need to build a *spoken language generation system*, whose output can be observed as an acoustic signal. Given such a system, however, we would still be faced with the task of evaluating the output. As will be explained below, comparing outputs by just listening to them will not lead to decisive answers as to which theory is to be preferred. In order to make convincing comparisons, we will need to isolate just those aspects which the theories that are tested actually describe, and generate, and see how these match up with what humans do. In the case of intonation, there is the additional problem of variability, in the sense that different speakers may choose different intonation contours for the same text, as well as inappropriateness, in the sense that many speakers are not capable of reading texts without disfluencies, or even obviously incoherent placements of pitch accents or boundaries. As we will explain below, obtaining a responsible data set that can serve as a reference in this respect is a problem that we have solved by eliciting the best estimate of the most acceptable realization of a text.

This chapter is organised as follows. In Section 1.1, we introduce three aspects of intonation: pitch accent, intonational phrase, and tune. We discuss linguistic intuitions about intonation in the form of acceptability judgements, how to cope with variance in acceptability judgements, and how to properly

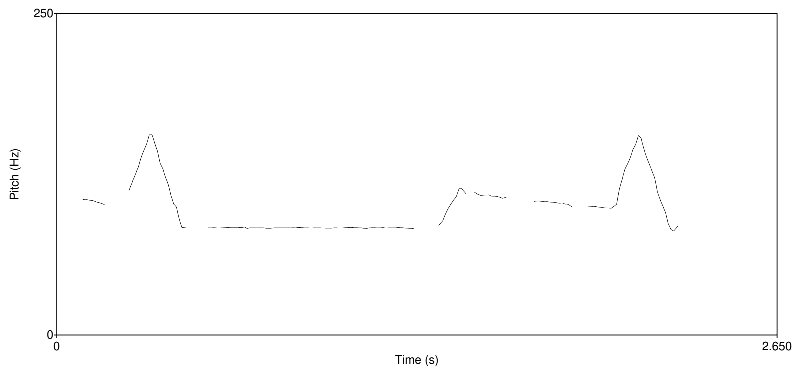

Figure 1.1: Pitch contour of the (synthesised) utterance [De sn*eltrein naar Nijmegen] [is al gew*eest] *'The fast-train to Nijmegen has already left'.*

assess these judgements. In Section 1.2, we introduce spoken language generation, and the way it is used in the current work. In Section 1.3, we describe our main research objective, identifying both a number of restrictions on its scope, and several more specific research questions. An outline of the remainder of the thesis concludes this chapter.

1.1 Intonation

1.1.1 Pitch accent, intonational phrase, and tune

Languages contain sounds that are *voiced* (all vowels and some of the consonants). The source of voiced sounds is a vibration of the speaker's vocal cords. The inverse time of one cycle of opening and closure of the vocal cords is called *fundamental frequency*, (F_0) or *pitch*.[1] Pitch can be measured during the production of a voiced sound, and is traditionally expressed in hertz (Hz), i.e. the number of cycles per second. If we measure the pitch of an utterance at fixed intervals of time, we obtain a sequence of pitch values through which can we draw a *pitch contour* (Figure 1.1).

Intonation can now be defined as:

> "[..] all aspects of the perceived pitch pattern that the speaker intends for the hearer to use in understanding the utterance, or that the hearer does use whether intentionally controlled by the speaker or not." (Beckman 1995)

[1] Strictly *pitch* refers to the auditory sensation of tone height, but the term will be used here to mean fundamental frequency.

Although not all theories of intonation are equally specific on this point, most of them distinguish at least three relevant aspects of the pitch pattern: the distribution of pitch accents, the structuring of speech by intonational phrases, and the specification of tune. A *pitch accent* can be defined as a perceptually salient pitch event, which usually results from a relatively sudden change in pitch level marking a particular syllable. An *intonational phrase* is a stretch of speech covered by a perceptually coherent intonation contour. Intonational phrase boundaries are also marked by a change in pitch level. *Tune* (or *melody*) refers to the particular shape of the intonation contour, which may vary independently of the locations of accents and boundaries. For instance, a statement may be turned into a question by changing only the tune, leaving the place of pitch accents and intonational boundaries unchanged. More precise definitions of these terms will be given in Chapter 2.

As an example, consider the question-answer pair in (1), and imagine the most natural intonation for the answer. The pitch accents in the answer are indicated by a star (*): the first is aligned with the stressed syllable of the word *sneltrein*; the second with the stressed syllable of *geweest*. These correspond to the pitch peaks in Figure 1.1, also referred to as *pointed hats* ('t Hart, Collier, and Cohen 1990). Furthermore, there are two intonational phrases, as indicated by by the square brackets. The boundary between them corresponds to the pitch discontinuity in Figure 1.1, i.e. a rise to a high pitch level at the end of *Nijmegen*, and a subsequent reset to the start pitch level beginning on *is*. Again, we leave a precise account of the relation between this symbolic representation of the intonation and the actual pitch track to Chapter 2.

(1) Question: Waarom nemen we de stoptrein naar Nijmegen?
why take we the slow-train to Nijmegen?
'why are we taking the slow train to Nijmegen?'

Answer: [$_{AD}$ De sn$\overset{*}{e}$ltrein naar Nijmegen] [$_{AD}$ is al
The fast-train to Nijmegen has already
gew$\overset{*}{e}$est]
left

In the current work, we are primarily concerned with pitch accents and intonational boundaries, largely ignoring the aspect of tune.

1.1.2 Acceptability of intonation

Native speakers have linguistic intuitions: they can usually tell if some expression is acceptable in their language. A native speaker of Dutch, for example, would agree that (2-a) is an acceptable expression, whereas (2-b) is unacceptable.[2]

[2]Following common practice, unacceptable expressions are preceded by a star (*).

(2) a. de trein vertrekt naar Amsterdam
the train leaves to Amsterdam

b. *de trein vertrekken naar Amsterdam
the train leaves to Amsterdam

From a syntactic point of view, the challenge is to account for these acceptability judgements. This requires analysing an expression's syntactic structure in abstract terms. For instance, we can say that (2-b) is ill-formed because the *subject* (*de trein*) is not in agreement with the *verb* (*vertrekken*) with respect to the feature *number*. In the tradition of generative linguistics, the syntactic analysis is formalised as a grammar that should generate all and only well-formed expressions.

Similarly, native speakers have linguistic intuitions about the intonation of an expression. Listening to the utterances[3] in (3), Dutch speakers would agree that the intonation of the first is perfectly acceptable, while that of the second is not.

(3) a. $[_{\mathrm{AD}}$ de tr$\overset{*}{\mathrm{e}}$in vertrekt naar Amsterd$\overset{*}{\mathrm{a}}$m]
the train leaves to Amsterdam

b. *$[_{\mathrm{AD}}$ de trein] $[_{\mathrm{AD}}$ v$\overset{*}{\mathrm{e}}$rtrekt naar Amsterdam]
the train leaves to Amsterdam

Again, accounting for these judgements requires analysing an expression's intonational structure. We can attribute the unacceptability of (3-b) to the fact that: (1) the last intonational phrase does not contain any accents (and would presumably have low pitch), and (2) the accent falls on the wrong syllable of *vertrekt*. In the next chapter, we will present a more elaborate analysis, and a grammar of Dutch intonation that is intended to generate all and only well-formed intonational structures.

However, in addition to being structurally well-formed, an expression must also fit into the context. Compare the two pieces of text in (4) and (5), which differ only with respect to their second sentence. Most native speakers of Dutch would probably agree that (4) is better than (5). Although (5-b) is perfectly acceptable in isolation, it is less appropriate in the given context when compared with the alternative in (4-b).[4]

(4) a. U neemt de trein naar Amsterdam.
You take the train to Amsterdam.

b. Die vertrekt om acht uur tien.
Which leaves at eight hour ten.

[3] In general, we will use the term *utterance* to refer to the audible, phonological form, and the term *sentence* to refer to the readable, written form.

[4] While comparing expressions, the less acceptable expression is preceded by a question mark.

(5) a. U neemt de trein naar Amsterdam.
You take the train to Amsterdam.

b. ?Om acht uur tien vertrekt de trein.
At eight hour ten leaves a train.

Accounting for this type of observation requires not only an analysis of the syntactic structure of an expression, but also an analysis of its context. Although it is not the topic of this thesis, we will elaborate on the relation between the syntactic structure of an expression and its context in Chapter 2 and Chapter 6. The point here is that acceptability judgements may be *context-dependent*. Another way to look upon this is to say that the set of possible syntactic forms to express a piece of information is reduced when it is placed in a more specific context.

The same goes for acceptability judgements about intonation. Accentuation is highly context-dependent. For example, most speakers would agree that the accentuation in (6-b) is inappropriate as a reply to question (6-a).

(6) a. Waarom de stoptrein naar Nijmegen?
why the slow-train to Nijmegen?

b. ?De sneltrein naar Nijmegen is al geweest
The fast-train to Nijmegen has already past

By contrast, the same accentuation is perfectly acceptable as a reply to question (7-b)

(7) a. Waarom de sneltrein naar Arnhem?
why the slow-train to Arnhem?

b. De sneltrein naar Nijmegen is al geweest
The fast-train to Nijmegen has already past

It is generally assumed that the reason for this is that pitch accents signal the information structure of an utterance.[5] In general, words that convey important or new information will be accented, whereas those conveying less important, old, or predictable information will remain unaccented. Hence, *sneltrein* can remain unaccented in answer (7-b), as it was already mentioned in the preceding question. whereas *Nijmegen* and *geweest* are both accented, because they contribute important new information. The fact that (6-b) is unacceptable can be explained in the same way. Although this relation between the pitch accents of an utterance and its information structure is backed up by substantial experimental evidence (Birch and Clifton 1995; Brown 1983; van Donzel 1999; Fowler and Housum 1987; Horne 1991; Kruyt 1985; Nooteboom

[5] We will elaborate on information structure, including the notion of *focus*, in Section 2.5.1.

and Terken 1982; Nooteboom and Kruyt 1987), the correspondence is definitely not as simple as these examples might suggest. Not only because an adequate definition of 'new' or 'important' has proved to be difficult, but also because of interactions with other aspects of the semantic, morpho-syntactic, and phonological structure of an utterance, preventing a one-to-one relation between 'new' and 'accented'.

In sum, native speakers can judge the acceptability of an expression's intonation. In order to be acceptable, the intonation has to be structurally well-formed and contextually appropriate. From a linguistic point of view, the challenge is to account for these judgements by means of a grammar.

1.1.3 Variance in acceptability

Despite the clear-cut examples above, it is not uncommon to find variation in acceptability judgements. There appear to at least three sources of variation:

1. acceptability judgements may be gradual rather than categorical, ranging from entirely unacceptable to highly acceptable
2. alternative intonational forms may be judged *equally* acceptable
3. acceptability judgements may vary from person to person

Any linguistic theory or computational system that tries to account for the acceptability of intonational forms has to deal with this variation. In the current work, the following assumptions are made.

With respect to the gradual nature of acceptability, the aim is to predict the *most acceptable* intonation of an utterance, alternatively referred to as the *preferred intonation*. For example, it is claimed in the current work that the accentuation in (8-a) is more acceptable than that in (8-b), and that the intonational phrasing of (9-a) is more acceptable than that of (9-b). This is not to say that the b-examples are unacceptable, but rather that a-examples are preferred in unemotional utterances in informationally neutral contexts.[6]

(8) a. honderd v$\overset{*}{\text{i}}$eren d$\overset{*}{\text{e}}$rtig
hundred four-and thirty
'134'

b. ?h$\overset{*}{\text{o}}$nderd v$\overset{*}{\text{i}}$eren d$\overset{*}{\text{e}}$rtig

(9) a. [Die vertr$\overset{*}{\text{e}}$kt van perron $\overset{*}{\text{e}}$lf] [om $\overset{*}{\text{a}}$cht uur ti$\overset{*}{\text{e}}$n]
Which leaves from platform eleven at eight hour ten

b. ?[Die vertr$\overset{*}{\text{e}}$kt van perron $\overset{*}{\text{e}}$lf om $\overset{*}{\text{a}}$cht uur ti$\overset{*}{\text{e}}$n]

[6] Kruyt (1985, p. 115) proposed a similar distinction between *acceptable accentuation* and *optimal accentuation* in order to explain the variation between subjects in a number of perception and production experiments on accentuation in Dutch.

If alternative intonational forms are equally acceptable, we are satisfied with predicting at least one of these. In general, we will ignore equally acceptable alternatives, because they are of no value in distinguishing one theory from another. After all, if an utterance *with* a certain accent (boundary) is equally acceptable *without* it, the theory that predicts this accent (boundary) is just as good as the one that does not.

Finally, the possibility of intersubjective differences in acceptability judgements is addressed by relying on the judgements of *experts*. That is, the most acceptable intonation is determined by combining the judgements of a number of subjects who are experienced in analysing and transcribing Dutch intonation. The point is that experts are better judges of what is normal than laymen, because they have been exposed more to linguistic data and have developed a greater awareness of them. This situation is parallel to one in which a person who has learnt to write alphabetically is better able to identify the segments in a CVC syllable than an illiterate (Morais, Cary, Alegria, and Bertelson 1979).

In sum, the challenge addressed in the current study is to account for the intonation that is *most acceptable* (preferred) according to the judgements of experts, ignoring differences between equally acceptable forms.

1.1.4 Acoustic versus symbolic assessment of acceptability

The scientific paradigm of the current study is that of generative linguistics. Accordingly, the facts on which the theories are built are the intuitions of native speakers about the acceptability of intonational forms (competence), rather than intonational forms that are observed in actual speech (performance). Whereas intuitions about syntactic form are relatively easy to obtain, intuitions about intonational form are less easily accessed, and bring along a number of methodological considerations.

In order to tap the intuitions of *naive* subjects, the data, i.e. carrier utterances with the relevant patterns of accents and intonational phrases, will have to be presented acoustically. That is, rather than reading them, subjects must listen to the utterances, because naive subjects cannot be expected to judge a symbolic representation of the intonation. To make the data audible, we need either a human speaker, some form of speech manipulation, or speech synthesis. Whichever way this is done, the danger will be that acceptability judgements are influenced by external factors, such as segmental quality or speaker characteristics. Although listeners may be explicitly instructed to ignore these aspects, and to pay attention to the intonation only, it is doubtful whether naive listeners are capable of this. The only reliable method to prevent segmental quality and speaker characteristics from affecting a subject's judgement appears to be to present pairs of utterances that differ only with respect to their intonation.

However, this method is unsuitable for comparing long utterances or paragraph-sized text, due to listeners' memory limitations. Moreover, it is difficult, if not impossible, for a human speaker to produce two versions of an utterance that differ only with respect to their intonation. Alternatively, a speech processing technique like LPC resynthesis or PSOLA may be used to manipulate the pitch and duration of the stimuli. The problem with this approach is that the quality of the phonetic realization of pitch accents and boundaries in terms of F_0 may be inadequate. For instance, the precise alignment of intonational peaks to the segmental content is a difficult problem that is not yet fully understood; see Wichmann, House, and Rietveld (2000) for an overview. This introduces the possibility that an accent or boundary is in principle correctly placed, but is judged unacceptable only because of an inadequate phonetic realization. In addition, it is known that a natural realization of pitch accents involves increasing duration and intensity, and altering spectral tilt (Sluijter 1995). Creating these effects by hand is a nontrivial task. The same problems occur when using a speech synthesiser, with the addition that errors in segmental synthesis may disturb a subject's judgement as well. In sum, acoustically-based approaches for assessing the intuitions of native speakers are faced with methodological problems, all of which are related to the fact that other aspects of the wave form may interfere with judgement of the intonation.

As an alternative, intuitions of native speakers may be assessed at the symbolic level. This means that, rather than acoustically, the data are presented orthographically as strings of words annotated with intonational symbols indicating pitch accents, intonational boundaries, and perhaps also tune. Subjects are then asked to judge the acceptability of the intonation on the basis of this symbolic representation. This type of *symbolic assessment* has been used for evaluating accent placement in TTS systems (Baart 1987; van Bezooijen 1989; Monaghan 1993), and is most extensively presented as a sound evaluation method in Monaghan and Ladd (1990).

Evidently, symbolic assessment precludes the use of naive subjects, as only experts can interpret the symbolic representation. However, the use of expert judgements was already proposed in the previous section in order to limit intersubjective variation. Moreover, it avoids the methodological problems associated with an acoustic presentation of the data. At the same time, results obtained with both types of presentation appear to be comparable: van Bezooijen (1989, p. 29) reports a significant, positive correlation between judgements about accent placement in a symbolic and acoustic evaluation of the output of a Dutch speech synthesiser.

In the current work, therefore, symbolic assessment was used to assess the intuitions of experts with respect to the preferred accent placement and intonational phrasing. The latter has, as far as we are aware, not been studied before using symbolic assessment. Furthermore, we follow van Bezooijen (1989) in asking subjects to *prescribe* the locations of pitch accents and intonational boundaries rather than to judge whether a certain distribution of accents and boundaries is acceptable. The result will be referred to as a *prescription* or *reference*. The advantage of this approach is that the resulting prescription remains a valuable piece of data, because it can be used time and again as a reference for verifying predictions.

1.2 Spoken language generation

Spoken language generation (SLG) is a form of *natural language generation* (NLG). According to two prominent researchers in the field, NLG may be defined in the following way:

> "The area of study called natural language generation is about twenty-five years old, having started as an offshoot of the discipline natural language processing (also called computational linguistics). It shares interests with a healthy mixture of linguistics, computer science (artificial intelligence), psycholinguistics, and occasionally, philosophy, sociology, and neurolinguistics. The central question studied is how one can program computers to produce high-quality text from some computer-internal representation of information. Motivation for this study ranges from entirely theoretic (linguistics, psycholinguistic) to the entirely practical (to produce output systems for computer programs)." (Bateman and Hovy 1991, p. 53)

NLG systems come in different sizes and flavours. Some of them are modest applications that use canned text to produce simple messages, while others are complex systems with ambitious goals such as generating weather forecasts in multiple languages, the technical documentation of an aircraft, social security forms, or personalised letters to convince a person to quit smoking. See Adorni and Zock (1993) and Paiva (1998) for a survey of NLG systems.

The more advanced NLG systems are firmly rooted in linguistic theory. For instance, discourse grammars are employed to generate coherent multi-sentence texts. Likewise, syntactic grammars are used to map abstract phrase specifications into grammatically correct expressions in the target language. These aspects, and NLG in general, will be discussed more thoroughly in Chapter 5. The key point here is that NLG systems incorporate a fair amount of linguistic knowledge in the form of grammars and intermediary levels of representation.

The topic of this thesis, *spoken language generation* (SLG), is an extension of natural language generation to spoken rather than written output. Analogously, the central question is how one can program computers to produce high-quality *speech* from some computer-internal representation of information. It is therefore best viewed as a combination of NLG and speech synthesis.

Our motivation for working on SLG is primarily linguistic, that is, our perspective is that of a theoretical/computational linguist working in the paradigm of generative linguistics. Our intention is to use SLG as a tool to test and improve linguistic theories that try to account for the intuitions of native speakers with respect to the intonation.

The advantages of using SLG to this end may need some clarification. As far as accent placement and intonational phrasing are concerned, theories are typically based on a small set of crucial examples. Since alternative theories account for the well-known examples equally well, it is difficult to compare them. These theories should be confronted with a wider range of material in a systematic way to see to what extent their empirical predictions hold good. The problem is that test material is not readily available. Testing theories about accent placement and intonational phrasing typically requires discourse, semantic and morpho-syntactic information. Although prosodically-annotated speech corpora have become available, these appear to lack the amount of structural analysis that is desirable for this purpose.

There is, however, a way of bypassing the problem of lack of adequate data. The NLG system that is a necessary part of any SLG system is in principle capable of generating all of the required discourse, semantic and syntactic information. It can be used to generate large quantities of reliable test material. An SLG system thus offers an excellent test bed for linguistic theories about intonation. In addition to verifying a theory, its implementation in an SLG system may reveal hidden assumptions or inconsistencies. In short, we think that both the observational and descriptive adequacy of a linguistic theory may benefit from its implementation in an SLG system.

It should be noted that we have no psycholinguistic ambitions in the sense that our SLG systems might model human language production. Neither do we have ambitions with respect to language engineering or language and speech technology in the strict sense of intending to solve some immediate practical problem by means of SLG technology. Nevertheless, theoretical research has often been relevant for applied research and vice versa, and some practical implications of our work will be discussed in Chapter 9.

According to the above definition of SLG, the goal is to produce *high-quality speech*. Evidently, a definition of the quality of the output, and of the quality of an SLG system in general, directly depends on the motivation for engaging in SLG. From a psycholinguistic point of view, the quality of the output is determined by the extent to which it resembles human speech performance, for example, with respect to timing, hesitations, speech errors and repairs (De Smedt 1996). From an applied perspective, however, it is defined in terms of its practical value, for instance, how actual users judge the intelligibility, naturalness, or pleasantness of the output, or how well it enables a user to accomplish a certain task. The motivation that underlies the current work is primarily linguistic: to account for the intonation that is most acceptable according to the judgements of experts. Hence, high-quality output is defined as speech with the pitch accents and intonational boundaries preferred by experts.

We think this goal is well suited to the research paradigm of NLG and SLG. With respect to syntax, for instance, few NLG systems aim at generating merely acceptable, or even all acceptable, syntactic forms. Rather, systems try to generate the *best* syntactic form, given the content and context of the message. In the same vein, an SLG system should try to generate the *best* intonational form.

1.3 Research objective

To sum up so far, our intention is to use SLG as a tool to test and improve linguistic theories about the relation between the intonation of an utterance and other aspects of its structure and context. We focus on the placement of pitch accents and intonational boundaries, largely ignoring matters of tune choice. Given our linguistic motivation, the SLG system should output the intonation that is preferred by experts. In short, the *research objective* of this study is:

> To design, build, and test a linguistically-based SLG system that generates speech with pitch accents and intonational boundaries that are most acceptable to experts.

There are a number of other restrictions on the scope of the current work. First, and rather trivial, the target language for our SLG system is *Dutch*. This means that not all of our conclusions are necessarily valid for other languages, but they will, at a minimum, be suggestive of how other languages could be dealt with. Second, it is concerned only with speech read aloud; the generation and synthesis of spontaneous speech is likely to bring along some different problems. Third, although the output is real speech, we focus on generating intonation at the symbolic, phonological level, and largely abstract away from matters of phonetic realization. Fourth, due to the limitations inherent in NLG, system output is limited to certain domains. In our case, these are botanical descriptions of wild flowers (the *Wild Flower Domain* or WFD), and instructions on travelling by means of the public transport system (the *Public Transport Domain* or PTD); the reasons for choosing these particular domains will be described in Chapter 5. Fifth, it is not practically feasible to study all of the competing theories about accent placement and intonational phrasing by means of SLG. Therefore, we commit ourselves to a number of specific linguistic theories, which will be introduced in Chapter 2.

Given the above objective and restrictions, the research questions addressed in this work belong to two categories: (1) questions regarding the linguistic theory, and (2) questions regarding the architecture of an SLG system.

As explained above, we use SLG as a tool for evaluating particular linguistic theories, more specifically, by asking the following questions:

1. How adequate is the theory?

2. How can it be improved?

Adequacy is primarily addressed in terms of observational adequacy, that is, correctness of predictions and coverage of the facts.

Although NLG has a longstanding tradition, the body of research on SLG is relatively small. There are still a number of moot questions regarding the architecture of an SLG system, including, but not limited to, the following:

1. What is the best way to combine NLG and speech synthesis?
2. What is the best place in the system to control the intonation?
3. What modifications to the 'standard' architecture of NLG and speech synthesis systems are required?
4. What additional information should be generated by the NLG system, and how can this be accomplished?

Admittedly, some of these questions may overlap, but given our current state of knowledge, more focused formulations cannot be provided yet. These questions underly the work presented in this thesis, the content of which will be outlined below.

1.4 Outline of thesis

This thesis is divided in four parts. The first part (Chapter 2) presents the theoretical framework, which forms the point of departure for all subsequent work on accent placement and intonational phrasing. The second part involves preliminary explorations into phonological phrasing (Chapter 3) and intonational phrasing (Chapter 4). This involves corpus-based work, rather than SLG proper, but provides a substantial part of the linguistic theory ultimately implemented in the SLG system. The third part (Chapters 5, 6, and 7) describes the SLG system we developed. The final part concerns an evaluation of its output (Chapter 8), followed by a general discussion (Chapter 9). The role and content of the individual chapters will be described below.

Chapter 2: Theoretical Framework
This chapter presents the linguistic framework for our work on accent placement and intonational phrasing in SLG.

First, we define the concept of *prosody*, and introduce *prosodic phonology*, which is a phonological theory about the hierarchical organisation of speech. It assumes the existence of a *prosodic structure* consisting of prosodic constituents (the phonological counterpart of a syntactic tree) which serves as the interface between semantics, syntax, and morphology on the one hand, and prosodic phenomena on the other. One of these prosodic constituents is the *phonological phrase*, which plays an important role in our approach to accent placement. It is commonly assumed that the *intonational phrase* is a prosodic constituent as well, but it is argued here that this view gives rise to contradictions, and an alternative is presented.

Second, we introduce *intonational phonology*, and a phonological description of Dutch intonation. This allows us to describe the intonation of utterances at a phonological level, and to abstract away from the details of phonetic implementation.[7] Furthermore, the research objective is interpreted as generating pitch accents and intonational boundaries that are not only well-formed according to the intonational phonology, but also fit the grammatical structure, meaning, and context of an utterance.

Third, we present a preliminary rule-based grammar for pitch accent placement, which forms the point of departure for accentuation in the SLG system. The role of focus and argument structure in accent placement is discussed. It is shown that phonological phrasing is crucial for pitch accent placement, as the phonological phrase forms the application domain for the *Rhythm Rule*. This is a phonological rule that deletes accents under certain conditions, which are also explained.

Finally, this chapter outlines the syntactic analysis of Dutch assumed throughout our work.

Chapter 3: Phonological Phrasing
As already mentioned, correct phonological phrasing is considered crucial for accent placement. In this chapter, it is argued that the algorithms for phonological phrasing that have been proposed by other researchers are inadequate for Dutch. Therefore, a new algorithm for phonological phrasing in Dutch is proposed.

This algorithm is subsequently evaluated at the symbolic level: its consequences for accent placement are compared with a reference accentuation, resulting from a symbolic assessment of the preferences of several experts. The text material used in this evaluation consists of botanical descriptions of wild flowers (the wild flower domain or WFD). These involve rather long complex sentences, and are therefore highly challenging for a phonological phrasing algorithm. The results show that the newly proposed algorithm is fairly adequate.

Chapter 4: Intonational Phrasing
This chapter addresses intonational phrasing in Dutch in much the same way as phonological phrasing was addressed in Chapter 3. We start by describing several algorithms for intonational phrasing that have been proposed in the literature. Next, these algorithms are evaluated by comparing the intonational boundaries that they predict with those prescribed by a panel of experts. A detailed analysis of the results reveals that the algorithms have various flaws, and that none of them is entirely satisfactory. From these findings, a number of requirements for a better algorithm are derived.

The second part of this chapter presents a new algorithm for intonational phrasing. A similar evaluation of this algorithm shows that it performs better than the earlier algorithms.

[7] Although phonetic implementation will be addressed to some extent at the end of Chapter 5.

Chapter 5: Spoken Language Generation
This chapter primarily addresses the research questions regarding the architecture of an SLG system. It discusses alternative ways of combining NLG and speech synthesis. It summarises the 'standard' architecture for NLG systems and proposes an extension for SLG systems. In particular, it addresses the question where intonation should be generated in an SLG system.

Next, we introduce a reusable front-end for SLG systems called CONPAS ('CONtent to Prosodically Adequate Speech'). This component is shared in the two SLG systems developed in the current work. The first system is for generating flower descriptions (WFDGen), and is best considered as an attempt to reproduce the results of Chapter 3 and Chapter 4 in the context of an actual SLG system. The second system generates travel descriptions (PTDGen), and is intended as an independent domain to test the developed grammars for accent placement and intonational phrasing.

The remainder of this chapter provides an overview of the modules and intermediary representations in the SLG system. It describes the modules for text planning and speech synthesis, but leaves a description of the modules for syntactic and prosodic realization to the next two chapters.

Chapter 6: Syntactic Realization
Here, we describe a syntactic realization module for Dutch, which translates abstract phrase specifications into grammatically correct expressions and their syntactic structures. The main purpose of this chapter is to show how all of the syntactic information that is required for accentuation and intonational phrasing can indeed be generated in an SLG system.

Chapter 7: Prosodic Realization
This chapter brings together work of the preceding chapters by describing how the prosodic realization module handles the derivation of prosodic structure, pitch accents, and tune choice. The main contribution in this chapter is a further development of the preliminary algorithm for pitch accent placement presented in Chapter 2. One of the issues is the correct accentuation of numerals.

Chapter 8: Evaluation
This chapter reports on an evaluation of the SLG system. This is again carried out at the symbolic level by comparing the predicted pitch accents and intonational phrases with those prescribed by a panel of experts. However, the domain is different: travel descriptions generated by PTDGen rather than the flower descriptions that figured in the previous evaluations. Moreover, the prescription is provided by experts who did not participate in the previous evaluation. Roughly, this evaluation shows that the performance of our algorithms is comparable to, or even better

(depending on the evaluation metric) than that of other approaches, but still somewhat worse than that of humans.

Chapter 9: General Discussion
In this final chapter, we examine the current work in the light of the research objective, in particular with respect to the research question formulated above. In addition, we discuss some issues that were not explictly addressed, including the practical implications for speech technology.