1 Introduction

1.1 DEFINING INTONATION AND PROSODIC STRUCTURE

In intonation patterns, not all parts of the melody are equally relevant: there are targets or turning points, which are highly significant for phonological modelling, and there are transitional parts, which are less crucial and simply relate two targets to each other. Languages vary systematically as to which parts of the contours are crucial and which are not. Let us begin with a comparison between sets of possible intonational renderings of short expressions in English, an intonation language, and in Beijing Mandarin Chinese, a tone language. Compare first the English expression many cars in Figure 1.1, provided with different melodies and embedded in different contexts. An intonation language like English assigns melodies to words and sentences, and these melodies have several sources. The notation used in Figure 1.1 represents melodic turning points with dots of different sizes. The first dot, the larger one in the contours, stands for the tonal target on the accented syllable in the expression, i.e. the pitch accent, which is on cars or on many. The syllable carrying the pitch accent is written in capital letters. The second dot, the smaller one, stands for the last tonal target, the boundary tone. It is realized at the end of the expression, here always on cars, but after the pitch accent on the same word in Figure 1.1a and b. The graphs in Figure 1.1 use a stylized notation, with lines indicating the melodic variation of the voice, rising, level or falling being the three possible contours. These can be combined to deliver more complex contours, see Figure 1.2 that shows contours with two pitch accents, one on many and the other on cars. As before, the dots indicate the tonal targets, in the case of English, pitch accents and boundary tones. This terminology is introduced in detail later.

A falling contour on cars, as in Figure 1.1a, stands for a neutral declarative intonation, for instance, as an answer to a question like What did you see? The words many and cars differ in their
prominence: *many* is not accented and is uttered at the mid-level of the speaker’s range, while *cars* is the accented word of the sentence. It starts at a higher level than *many* and it has a falling contour. The end of the expression is low. The word *cars* can also be realized with a rising contour, as in Figure 1.1b, in which case the rising contour confers an interrogative meaning. Again *many* is realized at mid-level, but now *cars* starts lower and ends higher than the first word. The difference in prominence between the two words as shown in Figure 1.1a and b illustrates the default realization of such an expression: the noun *cars* is more prominent than the quantifier *many*; *cars* carries the most perceptible melodic change of the two. The primarily accented word of a sentence bears what is called the ‘nuclear accent’ (Newman 1946; Bolinger 1958; Chomsky & Halle 1968) or just the ‘nucleus’ of the sentence. In a Germanic language like English, this default prominence pattern can be changed if needed by the context. This is illustrated in Figure 1.1c and d, where the falling or rising tone is now realized on *many*. In the melodic pattern illustrated in Figure 1.1c, the first syllable of *many* is accented and carries the falling contour that was illustrated in Figure 1.1a. The word *cars* is now postnuclear, deaccented, and it has a low and flat melody. In Figure 1.1d, *many* has a rising intonation, comparable to Figure 1.1b, and *cars* is again postnuclear. In this case, the first syllable of *many* is low, and starts a rising pattern that continues into the second syllable and optionally into the following word.

Further contours are also possible, as shown in Figure 1.2: both words may carry a melody of their own, both falling (a) or both rising (b), or the two words may have opposite contours, one falling and one rising (c and d). Furthermore, special contours, like a vocative ‘stylized’ contour with a lengthened syllable and a mid-final value for the last tone (e), are also an option.

The crucial observation is that *many cars* can have different contours without affecting the lexical meaning of the words: *cars*...
1.1 Defining Intonation and Prosodic Structure

always means ‘vehicles’, and many is always a quantifier meaning numerous, regardless of the tonal pattern of the expression many cars. The difference in tonal contours just illustrated is used for the expression of pragmatic and modal roles, but not to distinguish lexical meaning.

Compare next how the words of Beijing Mandarin Chinese acquire their contours. The words themselves are associated with melodies, and these melodies are intrinsic and thus largely invariant (with the exception of assimilation and sandhi phenomena that are addressed in Chapter 7). The melodic pattern of a monosyllabic word such as ma can change in the same way as that of cars in English does, but the changes in melodic pattern go along with changes in lexical meaning. Mandarin assigns lexical tones to nearly all syllables of all words. It has four full tones and one neutral tone. The neutral tone acquires its shape according to its environment, but the four full tones are largely invariant. The widely cited ma example is reproduced in Table 1.1. This syllable can be assigned all four full tones, and in each case, it is a different lexical item. The widely cited ma example is reproduced in Table 1.1. This syllable can be assigned all four full tones, and in each case, it is a different lexical item. The notations in Table 1.1 are variants of each other. The first column gives a common notation, in which the relative height or movement of the tones is described. The corresponding tone sequence notation, with H standing for high and L for low, is indicated in parentheses. The second column approximates the contour of the tones with numbers; 5 is the highest one and 1 is the lowest one so that the pitch value 51 signifies a contour falling from the top to the bottom of a speaker’s range. The third column shows the most commonly used notation convention, with diacritic features on the vowels.

The four tones are illustrated schematically in Figure 1.3.

In Figures 1.1 and 1.2, a rudimentary and abstract annotation has been used to illustrate melodic patterns. Readers probably have had no difficulty in reconstructing the melodic patterns from the lines and dots. In fact, a stylized annotation is often easier to read than real pitch tracks; see, for instance, Figure 2.8 for comparison.

Speakers of different languages use the same articulatory organs and it comes as no surprise that a statement often made for
segmental phonology, namely that all languages select their contrastive features from the same phonetic and phonological universal set, also holds for intonation. We all use the same variations of fundamental frequency ($F_0$). This is true for English and Mandarin alike, as for all languages. When we speak, our voice goes up and down at a specific temporal pace and by specific height intervals. No language systematically sustains one certain level of $F_0$ throughout entire sentences, for instance. The falling and rising contours that we produce occur mostly at intervals of the size of one or two syllables. They are thus short in time. In most cases, $F_0$ changes follow each other in rapid succession. Even a short sentence usually contains several melodic changes of different magnitudes. Moreover, the intervals in $F_0$ are similar in all languages: we usually use only a portion of our voice range for speech, generally the one we are most comfortable with. The universal set of melodic tunes relates to the range of our voices, which is itself dependent on the size of our larynx and our vocal folds. Anatomical diversity gives rise to individual differences concerning voice range. Listeners adapt easily to the different ranges and voice qualities of the speakers around them.

However, despite evident similarities, the melodic patterns are the consequence of different grammars in different types of languages, as has been demonstrated in the preceding text. The same kind of melodic rise or fall can be the result of different grammatical features or properties, assigned in different ways to prosodic constituents.
1.1 Defining Intonation and Prosodic Structure

This book examines intonation and prosody from a grammatical perspective. Its approach is decidedly phonological. In the phonological view of intonation, tones and resulting tonal contours are grammatical objects, sometimes compared to morphemes (see Pierrehumbert 1980), which interact with other components of grammar, such as syntax and semantics. Typological comparison will play an important part. The typology used in this book not only considers lexical differences, but also looks at the level of the phrase and the clause, which roughly correspond to phonological or intonational phrases in prosodic structure, as shown in detail in Chapter 4.

The comparison between English and Mandarin illustrates that tones are assigned to the segmental material (the text) in a systematic way in both languages, although the tonal system underlying tone assignment differs. The contour illustrated in Figure 1.1a, a falling pattern on cars in English, is similar to T4 in Figure 1.3, the falling tone on ma in Mandarin, but the role of the falling contour is different in both languages. In Mandarin, tones are part of the lexical meaning of words, and thus changing the tonal pattern is used to change the meaning of a word. Tones have no or only little influence on the prominence relationships at the level of the sentence. The contour of the words does not change as a function of the modality of the sentence or its accent pattern. In English, by contrast, accented syllables indicate prominence at the level of the words, and they are assigned different kinds of tones according to their position in the sentence and the information structure of the sentence (see Chapter 6). In this language, sentence melody is generally falling for declarative sentences and it may be rising for prototypical questions (see the examples in Figures 1.1 and 1.2), and the tones assigned to the words not only respect this pattern, but also contribute to it. The tones have no influence on the meaning of the words in English. Lexical meaning is largely determined by the segmental content, and in some cases by stress. We will consider the differences in tone assignment among languages as fundamental to understanding intonation and show that languages differ not only in lexically assigned tones, but also in the way melodies are mapped to sentences.

This book is about intonation and prosodic structure, which are concerned with the melodic part of spoken language. Here I adopt the clear distinction between intonation and prosodic structure that has emerged in the past decades. In the following chapters, ‘intonation’ (tonal patterns) is a cover term for all tonal phenomena. A definition follows in (1); see Chapter 2 for the phonetics of intonation.
INTRODUCTION

(1) Intonation

Intonation is the tonal structure of speech expressed by the melody produced by our larynx. It has a phonetic aspect, the fundamental frequency ($F_0$), and a grammatical (phonological) aspect.

As for the term ‘prosodic structure’ (a definition follows), it will chiefly refer to the phonological domains arising from syntactic structure, information structure and pragmatic roles.

(2) Prosodic Structure

Prosodic structure refers to the parsing of continuous speech in hierarchically organized in prosodic domains – mora, syllable, foot, prosodic phrase and intonation phrase.

Even though the terms ‘intonation’ and ‘prosodic structure’ are straightforward, this is not true for the term ‘prosody’; see Ladd (2014) for a useful discussion of the history of this term. The term has a Greek origin and until the twentieth century was used to denote the musical accompaniment of the segmental parts of the words and sentences (or verses in poetry). It is only with Trubetsksyo (1939) that the term entered linguistics and denoted other properties besides word-level accentual features. According to Trubetsksyo, prosodic features refer to all rhythmic-melodic aspects of speech. Consequently, the term has been used as a synonym for suprasegmental phonology for several decades, until Bruce (1977), Liberman (1976) and Pierrehumbert (1980) separated tonal structure from metrical structure, that is, the accentual properties of speech. The next step leading to the definition of ‘prosodic structure’ as used in this book was the development of the prosodic hierarchy by Liberman (1975), Selkirk (1984) and Nespor and Vogel (1986); see Chapters 3 and 4. In sum, prosody is characterized by intonational features, and intonation is realized within prosodic domains. For this reason, since the overlapping of intonation and prosody is intrinsic to the organization of speech, an overlapping of the terminology is unavoidable. As a result, the terms ‘intonation’ and ‘prosody’ are often used interchangeably. The shared meaning of intonation and prosody includes what is sometimes referred to as ‘suprasegmental phonology’, and thus tonal structure, pitch accents, phonological boundaries, duration and intensity. The terminological confusion is acknowledged in this book, and no attempt to escape it or correct it is provided, except for the definitions in (1) and (2).
1.1 Defining Intonation and Prosodic Structure

In the past, intonation had been considered to be ‘around the edge of language’, to use Bolinger’s (1964) well-known expression. It had been assumed that intonation differs from other parts of language in being divided into a grammatical and a nongrammatical component, the latter called ‘paralinguistic’ (Ladd 1980), ‘half-tamed savage’ (Bolinger 1978) or ‘biological’ (Gussenhoven 2004). Some authors assert the importance of the ‘emotional’ or ‘expressive’ side of intonation. Until the 1970s, the predominant view was that intonation was superimposed on the grammatical structure of a sentence. For Pike (1945), for example, intonation has a special role, namely to express the speaker’s attitude, the content of the message being communicated by the text. In his conception, intonation is dissociated from grammar and should be investigated separately. As intonation was not perceived as being part of grammar proper, it was also not investigated as such. It was – and still is – often assumed that intonation has its own meaning, which can be superimposed on the lexical and syntactic meaning, contributing to pragmatic nuances or to attitude. A large number of experiments have been conducted by just changing the intonation of a sentence and looking at the perceived changes in attitudes elicited by the tonal changes (Lieberman and Michaels 1962; Osser 1964; Uldall 1964; see Ladd 1980:129ff for a survey and criticisms). It must be noted that the voice of an individual can reflect emotional attitudes: the F0 range becomes smaller or larger, intensity decreases or increases, and the speech tempo becomes faster or slower, as the entire body contracts or relaxes. In other words, the same phonetic means are used to express the grammatical aspects and the ‘paralinguistic’ ones. But crucially, the grammatical aspects of intonation do not change as a result of emotion: there is no change in lexical tone, or in the place of a lexical stress, and a prosodic boundary is never intentionally inserted at the wrong place because of happiness, anger or fear.

There has been a progressive change in perspective concerning the paralinguistic part of intonation. See Ladd (1980) for a detailed description of the evolution of attitude up through the late 1970s. Since then, the importance of the paralinguistic side of intonation has decreased over time as a result of the progress in analysis, especially in pragmatics. We will return to the notion of paralinguistics in Chapter 6. For now, it is sufficient to acknowledge that intonation may have an extralinguistic component, though this component is probably much smaller than often assumed. Depending on what pragmatics covers, large parts of intonation proper are language dependent, and can thus be considered as grammatical.
1.2 STRUCTURE OF THE BOOK

Chapter 2 introduces the phonetics of intonation. This short chapter provides basic knowledge about articulatory and acoustic aspects of intonation, and it also helps novice readers to create and read pitch tracks. Section 2.1 explains how the larynx functions, how the melody of speech arises by the action of the vocal tract, and how modal voicing and different types of phonation are produced. Important terms and terminological distinctions concerning the articulatory aspect of intonation are introduced here. Section 2.2 introduces the acoustic components of intonation: fundamental frequency (F0), intensity (decibel, dB), duration (millisecond, ms) and spectral analysis of vowels (formants). It describes sound waves and shows how F0 is calculated. Section 2.3 focuses on the generation of pitch tracks by computer algorithms. Some of the classic problems arising with computer algorithms are pointed out, as well as methods to circumvent them. Microprosodic effects are also discussed. The last part of the chapter gives some advice on how to design prosodic experiments.

Chapters 3 and 4 are dedicated to prosodic structure in the narrow definition of this term and thus to the organization of segmental sequences in phonological domains. An account of the grammatical role of intonation as intended in this book requires precise definitions not only of melodic excursions, but also of the prosodic constituents to which they are assigned. In other words, tonal cues need support to be realized, and this support is the text as organized in prosodic constituents. Chapter 3 introduces the ‘rhythmic’ prosodic constituents. After a general introduction to the indirect reference hypothesis, the prosodic hierarchy and the Strict Layer Hypothesis in Section 3.1, syllable, mora, foot and prosodic word are described, with an emphasis on their roles in providing support for tones. The syllable and the mora are the stress-bearing units (SBUs) and the tone-bearing units (TBUs) (Section 3.2). The mora plays an important role in quantity-sensitive languages, as it determines syllable weight: heavy syllables attract stress. The foot (Section 3.3) is a metrical unit, and as such, is important for stress assignment. Some elements of metrical phonology are introduced that are crucial to understanding intonation and prosody. These three prosodic constituents, mora, syllable and foot, are the rhythmic constituents of the prosodic hierarchy. The last section of Chapter 3 (Section 3.4) introduces the prosodic word (o-word), and at the same time, some more theoretical issues that will be important in the following chapters. One of these is the question of recursivity of prosodic structure. Ito
1.2 Structure of the Book

and Mester’s recursive ‘min–max’ model of prosodic constituency is introduced here and applied to English. The chapter ends with a summary of compounds in Japanese, illustrating how the prosodic hierarchy maps morphosyntactic constituents to prosodic ones.

Chapter 4 is dedicated to the prosodic phrase (Φ-phrase) and the intonation phrase (Ț-phrase), the most important prosodic constituents for intonation. Section 4.1 introduces higher-level constituents and describes their metrical organization. The chapter examines the relationship between syntax and intonation, the syntax–prosody interface, that is, how prosodic phrases and syntactic constituents are mapped to each other. Bruce (2005) defines two roles of intonation: the expression of grouping and the expression of prominence. The grouping of constituents is typically coupled to syntactic structure, although mismatches between syntax and prosodic structure are not infrequent, and prominence is rendered by metrical structure. Sections 4.2 to 4.5 give overviews of the most important models of the syntax–prosody interface. In Section 4.2, the relation-based models (Nespor & Vogel 1986) are described and illustrated, in Section 4.3 the edge-based models (M. Chen 1987; Selkirk 1986) and in Section 4.4 the alignment models (McCarthy & Prince 1993b; Truckenbrodt 1995). The common property of these models is that they assume strictly layered non-recursive constituents (Nespor & Vogel 1986; Selkirk 2004). Moreover, all traditional models define the prosodic constituents by delimiting them at one edge of the corresponding syntactic constituents only. In Section 4.5, evidence for the need for recursive prosodic models is presented. Ladd’s (1990) compound domains, Itô and Mester’s (2007, 2013) min–max model of prosodic constituency and Selkirk’s (2011) Match constraints are introduced and illustrated. It is shown in this chapter that the need to map syntactic constituents to prosodic constituents in a one-to-one-fashion is mitigated by so-called well-formedness constraints imposing formal and rhythmic constraints on prosodic constituents.

Chapter 5 introduces several models of intonation. After a general introduction to the chapter in Section 5.1, Section 5.2 presents the Tone-Sequencing (TS) model of Liberman (1975), Bruce (1977) and Pierrehumbert (1980), also called the Autosegmental-Metrical (AM) model by Ladd (1996). Section 5.3 covers the autosegmental aspect of the TS model with a special emphasis on African tone languages. The chapter also gives overviews of other models. First, ToBI, an annotation technique based on the TS model, is briefly addressed in Section 5.4. The optimality-theoretic model developed by Gussenhoven (2004) and others is summarized in Section 5.5. The British school model is
presented in Section 5.6. Finally, Xu’s (2005) Parallel ENcoding and Target Approximation (PENTA) model is shown in Section 5.7 as an example of a ‘functional’ phonetic system assuming a direct relationship between phonetics and semantics.

Chapter 6 contains an overview of the effect of pragmatics and semantics on the tonal structure. After an introduction to the chapter (Section 6.1), it is first shown how intonation is involved in the rendering of information structure. This is the topic of Section 6.2. In our English examples (Figures 1.1 and 1.2), cars is richly intonated as long as it is the most prominent word of the expression or sentence in which it appears. If many is focused, and thus more prominent, cars may be devoid of melody altogether. Compare the felicitous answer to the question in sentence 3a with the infelicitous one in 3b, where small capitals stand for pitch accent and lowercase for the lack thereof.

(3) {Who was elected president of Europe?}
   a. Jean-Claude was elected president.
   b. #Jean-Claude was elected president.

Sentence 3b is not ill formed in an absolute sense, but only in specific contexts. For instance, it is perfectly fine in a context asking for the function of Jean-Claude, whereas in such a case 3a is not felicitous. In the second part of the chapter, the assignment of precise pragmatic meanings to specific tonal contours is investigated in Section 6.3: Pierrehumbert and Hirschberg’s (1990) compositional proposal to relating tones and meanings in English is summed up in this section, as well as other proposals, some of which have been called ‘holistic’. In Figures 1.1 and 1.2, a relationship was established between falling contours and declarative sentences on the one hand, and between rising contours and questions on the other. This association is only a tendency and the choice between falling and rising melodies can serve other purposes. Moreover, more complex tone sequences may have more complex pragmatic roles. Pragmatic use of tones is largely language dependent, as the same contour can have different meanings or implications in another language. The last section of this chapter, 6.4, returns to the question of the paralinguistic use of intonation.

Chapters 7 and 8 propose a typology of intonational systems by comparing several types of languages with tones originating at different levels. Chapter 7 starts the typological overview by explaining