Prosody and Intonation

Caroline Féry

2015

Content

Preface

Chapter 1 Introduction ……..……
- 1.1 Defining intonation and prosodic structure
- 1.2 Structure of the book

Chapter 2 Phonetic correlates of intonation……..…….....
- 2.1 Articulatory correlates of intonation
- 2.2 Acoustic correlates of intonation
- 2.3 Generating and reading pitch tracks
  - 2.3.1 Segmental effects (microprosody)
  - 2.3.2 Errors of the algorithm
  - 2.3.3 What to measure?

Chapter 3 Lower-level prosodic constituents……
- 3.1 Indirect reference hypothesis, prosodic hierarchy and Strict Layer Hypothesis
- 3.2 Syllables and moras
- 3.3 Feet
- 3.4 Prosodic words
  - 3.4.1 Min-max model and compounds in Japanese
  - 3.4.2 Prosodic word in Germanic languages

Chapter 4 Intonation and syntax: The higher-level prosodic constituents …….…...…
- 4.1 Introduction
- 4.2 Relation-based model of the syntax-prosody interface
- 4.3 Edge-based approach
- 4.4 Alignment approach
- 4.5 Recursivity in prosodic structure
  - 4.5.1 Evidence for recursive prosodic structure
  - 4.5.2 MATCH constraints and well-formedness constraints
  - 4.5.3 Comparing ALIGN and MATCH
- 4.6 Conclusion

Chapter 5 Models of intonation……..……
- 5.1 Introduction
- 5.2 Tone-sequence model of intonation
  - 5.2.1 Discrete tones
  - 5.2.2 Text-tone association
  - 5.2.3 Continuous phonetic melody
  - 5.2.4 Downtrends, reset and upstep
- 5.3 Tones in African tone languages
- 5.4 ToBI
- 5.5 Optimality-theoretic model of intonation
- 5.6 Nuclear tone tradition (British school)
- 5.7 PENTA: A phonetic model

Chapter 6 Intonation and meaning……..…….…..…
- 6.1 Introduction
- 6.2 Information structure

6.2.1 Information structure: Givenness, focus and topic
6.2.2 Prosodic and intonational reflexes of information structure
6.2.3 Focus and tonal scaling
6.3 Meaning of tones and tonal contours
6.3.1 Compositional analysis of tonal meaning
6.3.2 Partly holistic approaches
6.3.3 Holistic approaches
6.4 Paralinguistic aspects of intonation

Chapter 7 Tone and stress at the word level……...…….…..……
7.1 Introduction
7.2 Words without lexical stress and without lexical tone
7.3 Words with lexical stress and without lexical tone
7.4 Words with lexical stress and lexical tone: Pitch accent languages
7.4.1 Germanic pitch accent languages
7.4.2 Japanese, Basque and Turkish
7.5 Alignment
7.6 Words with lexical tone: Tone languages
7.6.1 Invariance of lexical tone in Cantonese
7.6.2 Tone invariance and tone sandhi in Beijing Mandarin
7.6.3 Variance of tones in Shanghai
7.6.4 Xiamen Min circle
7.6.5 Tonogenesis
7.7 Conclusion

Chapter 8 Sentence intonation in a typological comparison…….
8.1 Introduction
8.2 Intonation languages with lexical stress
8.3 Pitch accent languages
8.3.1 Stockholm Swedish
8.3.2 Tokyo Japanese
8.3.3 Turkish
8.4 Tone languages
8.4.1 Beijing Mandarin
8.4.2 Hong Kong Cantonese
8.4.3 Chichewa
8.5 Phrase languages
8.5.1 Seoul Korean
8.5.2 French

Chapter 9 The processing of intonation……...……
9.1 Introduction
9.2 Prosodic phrasing and speech comprehension
9.3 The role of pitch accents and rhythm
9.4 Implicit prosody

Chapter 10 Summary and conclusion……..…….……

References
Glossary
Index

Chapter 1
Introduction

1.1 Defining intonation and prosodic structure

In intonation patterns, not all parts of the melody are equally relevant: some parts of the melody are targets or turning points, and are highly significant for a phonological modelling, while some transitional parts, which are less crucial, 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 table 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 table 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, on *cars* or on *many*, the pitch accent. The syllable carrying the pitch accent is written with small caps. 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 a. and b. The graphs in table 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 table 1.2. The dots indicate the tonal targets, in the case of English, pitch accents and boundary tones. This terminology will be introduced in detail later.

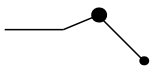
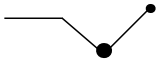
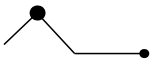
I saw many CARS  a.	Did you see many CARS?  b.	I didn't say a few cars, I said MAny cars  c.	I know that you saw cars, but did you see MAny cars?  d.

Table 1.1 Falling and rising tonal contours in English

A falling contour on *cars*, as in a., the leftmost panel, 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 b., and 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 a. 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 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 c. and d., where the falling or rising tone is now realized on *many*. In the melodic pattern illustrated in c., the

first syllable of *many* is accented and carries the falling contour that was illustrated in a. The word *cars* is now post-nuclear, deaccented, and it has a low and flat melody. In d., *many* has a rising intonation, comparable to b., and *cars* is again post-nuclear. 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 table 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.

MAny CARS	MAny CARS?	MAny CARS?	MAny CARS	JOnathan!
a.	b.	c.	d.	e.

Table 1.2 More tonal contours in English

The crucial observation is that *many cars* can have different contours without affecting the lexical meaning of the words: *cars* 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 invariant (with the exception of assimilation and sandhi phenomena that will be addressed in chapter 7). A monosyllabic word like *ma* can change its melodic pattern in the same way as *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 here in table 1.3. This syllable can be assigned all four full tones, and in each case, it is a different lexical item. The notations in table 1.3 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.

Basic Pitch Shape	Pitch Value	Examples
T1. High Level (H)	55	mā 'mother'
T2. Rising (LH)	35	má 'hemp'
T3. Low or Falling-Rising (L)	21(4)	mǎ 'horse'
T4. Falling (HL)	51	mà 'to scold, to blame'

Table 1.3 Beijing Mandarin lexical tones

The four tones are illustrated schematically in figure 1.1.

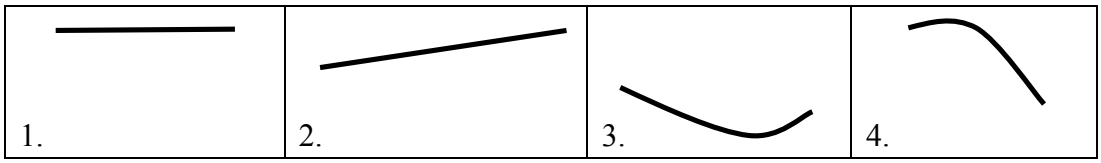


Fig.1.1 Beijing Mandarin lexical tones

In tables 1.1 and 1.2, a rudimentary and abstract annotation has been used to illustrate melodic patterns. The reader probably has 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 in the next chapter.

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 universal set of phonetic and phonological material, also holds for intonation. We all use the same variations of fundamental frequency (F0). 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 F0 throughout entire sentences, for instance. The falling and rising contours that we produce mostly occur at intervals of the size of one or two syllables. They are thus short in time. In most cases, F0 changes follow each other in rapid succession. Even a short sentence usually contains several melodic changes of different magnitudes. Moreover, the intervals in F0 are similar in all languages: we usually use only a portion of our voice range for speech, usually the one we are most comfortable with. The universal set of melodic material 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 above. 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. 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, like syntax and semantics. Typological comparison will play an important part. The typology used in this book considers not only 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.

The comparison between English and Mandarin illustrates that tones are assigned to the text in a systematic way in both languages, although the tonal system underlying tone assignment differs. The contour illustrated in table 1.1a, a falling pattern on *cars* in English, is similar to the one shown in figure 1.1.4, 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 often changes 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 tables 1.1 and 1.2), and the tones assigned to the words not only respect this pattern, but 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 see (1) for a definition, and chapter 2 for the phonetics of intonation.

(1) Intonation
Intonation is the tonal structure of speech making up the melody produced by our larynx. It has a phonetic part, the fundamental frequency (F0) and a grammatical (phonological) part.

As for the term 'prosodic structure', see (2) for a definition, it will chiefly refer to the prosodic domains arising from syntactic structure, information structure and pragmatic roles.

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

The correlates of these prosodic domains may be tonal, in which case intonation is used to delimit them, but they also can use prominence, duration or segmental factors as phonetic correlates.

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 has been used until the twentieth century to denote the musical accompaniment of the segmental parts of the words and sentences (or verses in poetry). It is only with Trubetskoy (1939) that the term entered linguistics and denoted other properties than word-level accentual features. According to Trubetskoy, prosodic features are referring to all rhythmic-melodic aspects of speech. Consequently, the term has been used as a synonym of suprasegmental phonology for several decades, until Bruce (1977), Liberman (1975) and Pierrehumbert (1980) separated intonation and tonal structure from the rest of prosody, especially the metrical structure or 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 & Vogel (1986), see chapters 3 and 4. In sum, prosody is characterized by intonational features, and intonation needs prosodic domains and its phonetic correlates, and for this reason, the overlapping of intonation and prosody is intrinsic to the organization of speech, and 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', 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).

In the past, intonation has been considered to be 'around the edge of language', to use Bolinger's (1964) well-known expression. It has been assumed that intonation differs from other parts of language in being divided into a grammatical and a non-grammatical component, the latter being 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 seventies, 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 as a separate module. Since 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 change under emotional stress and strain: the F0 range becomes smaller or larger, intensity decreases or increases, and the 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 under 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 seventies. Since then, the importance of the paralinguistic side of intonation has decreased over time due to 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 extra-linguistic component, but that this component is probably much smaller than often assumed. Depending on what pragmatics covers, almost all parts of intonation proper are language-dependent, and can thus be considered as part of the grammar.

1.2 Structure of the book

Chapter 2 (*Phonetic correlates of intonation*) introduces the phonetics of intonation. This short chapter provides basic knowledge about articulatory and acoustic aspects of intonation, and it also helps the novice reader to create and read pitch tracks. Section 2.1 explains how modal voicing and different types of phonation are produced, how the larynx functions, and how the melody of speech arises by the action of the vocal tract. Important terms and terminological distinctions used in the book are introduced here. Section 2.2 is dedicated to 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 these problems. 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, 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 they are assigned to. In other words, tonal cues need support to be realized, and this support is the text as organized in prosodic constituents. Chapter 3 (*Lower-level prosodic constituents*) 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 (SBU) and the tone bearing units (TBU) (section 3.2). The mora plays an important role in quantity-sensitive languages, since 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 which 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 (ω-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 & 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 morpho-syntactic constituents to prosodic ones.

Chapter 4 (*Intonation and syntax: The higher-level prosodic constituents*) is dedicated to the prosodic phrase (Φ-phrase) and the intonation phrase (ι-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, i.e. 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 are not recursive and that they obey the Strict Layer Hypothesis (Nespor & Vogel 1986, Selkirk 1984). 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 & 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 (*Models of intonation*) introduces several models of intonation. After a general introduction to the chapter in section 5.1, section 5.2 presents the Tone-Sequence (TS) model of Liberman (1975), Bruce (1977) and Pierrehumbert (1980), also called the Autosegmental-Metrical (AM) model by Ladd (1996). The term AM is not used in this book, because it assumes that all languages use prominence in their tonal system, which is not necessarily the case. 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) 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 (*Intonation and meaning*) is 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 above (tables 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 (3)a with the infelicitous one in (3)b, where small capitals stand for pitch accent and lower case for the lack thereof.

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

Sentence (3)b 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 (3)a is not felicitous. In the second part of the chapter, the task of assigning precise pragmatic meanings to specific tonal contours is investigated in section 6.3: Pierrehumbert & 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 Tables 1.1 and 1.2 above, 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 (*Tone and stress at the word level*) starts the typological overview by explaining the roles of tonal specifications at the lexical level. The view of intonation defended in this book is essentially compositional. Applied to intonation, this term denotes the resulting melodic contour that arises when tones originating at different levels of the prosodic hierarchy come together. We follow Laniran & Clements (2003), who view compositionality in tone production as a functionally motivated compromise between competing factors affecting individual tones in the melody of sentences. Besides the tones themselves, interactions between tones also play a role. In section 7.1, a primary distinction is made between stress, pitch accent and tone (see Hyman 2006 for this distinction). Section 7.2 reviews languages without lexical stress and without lexical tone. French and Korean are used as examples for this category. In section 7.3 languages with lexical stress but without tonal specification for stress are briefly discussed. English and German are prototypical instances of such languages. Section 7.4 takes a closer look at the so-called pitch accent languages, mainly the Scandinavian languages Swedish and Norwegian, as well as Japanese, although further languages, like Turkish, are also briefly addressed. Section 7.5 is dedicated to tonal alignment: given tones and texts, how exactly are the tones

associated with the syllables? Section 7.6 studies lexical tones in Asian languages. Cantonese has a limited number of tonal changes (sandhi effects) and Shanghai Chinese has a large number of simplifications in its tonal system. In this section tonogenesis is also examined: how did lexical tones arise in the history of tone languages? The chapter ends with a short conclusion.

Chapter 8 (*Sentence intonation in a typological comparison*) extends the typological review of intonational systems to the level of sentences. The chapter begins with an introduction in section 8.1. Section 8.2 discusses languages without tonal specification at the lexical level, but with lexical stress; German sentence intonation illustrates this type of language. In section 8.3, intonation of pitch accent languages is shown with Stockholm Swedish, Tokyo Japanese and Turkish. These kinds of languages best illustrate the compositional view of intonation, because they have tonal specification at three levels of the prosodic hierarchy: at the lexical level (ω-word), the phrasal level (Φ-phrase) and the sentence level (ι-phrase). Section 8.4 discusses tone languages with Beijing Mandarin, Hong Kong Cantonese and Chicheŵa as examples. These languages have lexical tonal specification and also at least some tonal specification at the sentence level (ι-phrase). Finally, in section 8.5, a new type of language is introduced, called 'phrase languages' to emphasize the role of phrasal tones in the tonal system of these languages. French and Korean, as well as Indian languages, are typical examples of this type of language. They have no lexical tonal specification or only scarce ones, and their sentence intonations are the products of higher-level specifications, at the phrasal level (Φ-phrase) and the sentence level (ι-phrase).

Chapter 9 (*The processing of intonation*) discusses the role of intonation from a psycholinguistic perspective. It is concerned with the role of prosody in the processing of spoken speech and silent reading, but is largely confined to English. The chapter begins with an introduction in section 9.1. Section 9.2 addresses prosodic phrasing and speech comprehension. Longer stretches of discourse are organized in smaller prosodic units. The syntactically motivated prosodic domains help speakers and hearers to structure the discourse in meaningful chunks. Consider the sentences in (4) from Kjelgaard & Speer (1999), which illustrate a local syntactic ambiguity. The noun phrase *the house* is preferably interpreted as the object of *leaves*, which is the intended reading in (4)a but not in (4)b. The question is whether a prosodic boundary after the verb forces the second reading.

(4) a. When Roger leaves the house, it's dark.
b. When Roger leaves, the house is dark.

In other cases, a structural ambiguity is not solved at the end of the sentence, as in (5) from Schafer, Carlson, Clifton, & Frazier (2000). The preferred reading of this sentence is the one in which *who is cold* is a relative clause. If *who* is a wh-word, it needs to be accented. Section 9.3 shows the role of pitch accents and rhythm in disambiguation.

(5) I asked the little girl who is cold.

Section 9.4 examines the role of implicit prosody, as formulated by Fodor (2002a,b), which predicts that in silent reading, a default prosodic contour is projected onto the stimulus that may influence syntactic ambiguity resolution.

Chapter 10 sums up the different aspects of intonation reviewed in the book and briefly considers other aspects that have been left out.

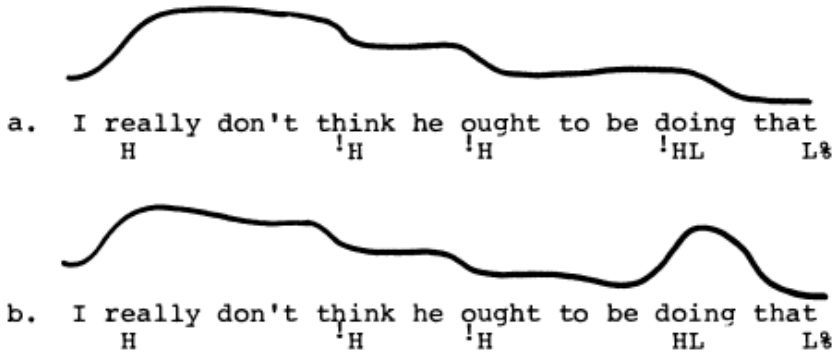
Annotated suggestions for further reading

Two books provide extensive introductions to intonation from a similar perspective to the one proposed in the present book, and also discuss subjects that are not covered in the present book. Gussenhoven's (2004) book sums up approaches like the Dutch IPO model. The second part develops an elaborate optimality-theoretic approach of several well-studied European languages and Japanese.
Ladd's book (1996/2008) presents a feature analysis of suprasegmentals to classify languages along the word level stress and accent (see also Beckman 1996). It also contains a critical discussion of some of the central issues of the Autosegmental-Metrical model, like alignment of tones, and how to represent downstep.

Discussion points

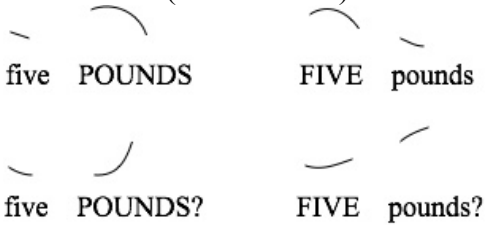
1. Find at least six different tonal contours for the expression *the nice month of September* in English. Where are the pitch accents in each case? And what does the end of the contour (the boundary tone) look like? If you speak another language, do the same with a similar expression in this language, and compare your results with English.

2. Compare the following tonal contours from Ladd (1983a:735). Reproduce them orally the best you can and describe the melodic differences between them.

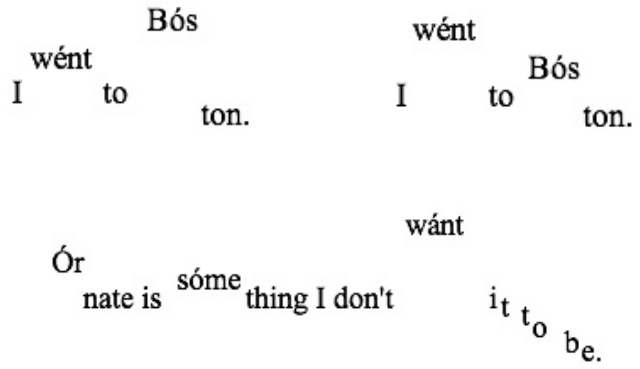


3. Different kinds of notation of tonal contours have been proposed in the literature, some of which are illustrated below. Describe each method and comment on the positive and negative aspects of each notation system. We will not use these notation systems in the remainder of the book.

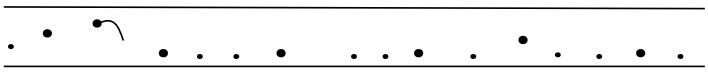
a. Line notation (Ladd 2008:7).



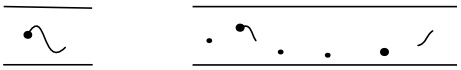
b. Bolinger's use of staggered lines of type, in which the ups and downs roughly indicate the melody of the voice. Here are three examples (1986:61).



c. Interlinear or *tadpole* notation used by Crystal (1969), O'Connor & Arnold (1973), and Cruttenden (1986); from Ladd (2008:48).



I hear Sue's taking a course to become a driving instructor.



Sue!? A driving instructor!?

d. Step notation in American structuralism: Trager & Smith (1951) and Pike (1945).

	no °3-1 ↓		no °3-1 ↑
Falling tone	↓	or	#
Rising tone	↑	or	‖
Level tone	˥	or	→

Chapter 2
Phonetic correlates of intonation

Intonation consists of changes in the acoustic signal, primarily in F0. To understand better how tones and tonal contours arise, it is necessary to understand the acoustic part of intonation, be it only to be able to refer to it in an informed way. This chapter introduces the basics of articulatory and acoustic phonetics, although it cannot replace a true introduction to phonetics. It briefly reviews the primary components of tones without entering into any detail of how they are produced, transmitted and perceived.

2.1 Articulatory correlates of intonation

Speech is transmitted by air in the shape of variations in air pressure that are registered by our eardrums in a continuous way. Language is primarily spoken, and the majority of speakers use language in its spoken channel either exclusively or most of the time. In fact, for a large part of the human population, spoken language is the only form of language known and used. With the exception of punctuation as a rough indicator of higher prosodic phrasing, intonation is not part of the written systems of language, and even when it is present in the writing system in the form of commas, semi-colons, full stops, or question marks, it does not necessarily impose a concrete intonation pattern; rather it gives an approximate idea of the syntactic structure. More specifically, a full stop does not tell the reader whether the sentence has to end with a falling or a rising melody. In the same way, a comma does not necessarily indicate a break in the flow of speech. The speaker may even choose to ignore punctuation at places. Thus, with *implicit prosody* the only exception, see section 9.4, intonation is exclusively concerned with spoken language. In this it differs from other components of grammar.

The phonetic components of intonation consist of pitch, intensity, and duration, as well as some segmental effects. Pitch or melody is transmitted by fundamental frequency (F0); see below for how the phonetic components are produced. Furthermore speech can use a so-called 'non-modal' kind of voice, e.g. creaky, breathy or murmured. Intensity or amplitude refers to the loudness of an utterance or of parts of it. Duration refers to the temporal organization of tones, segments, syllables, words, and sentences. It also includes small pauses between chunks of discourse. Spectral properties of vowels, the intrinsic pitch of vowels and some properties of consonants (aspiration for instance) may be used to enhance the prominence of a syllable for instance, reflecting its status within the prosodic structure. However, intonation is usually understood as the non-segmental part of speech.

The tonal part is the most important component of 'intonation', as the word itself suggests. F0 is the raw substance of the voice. It is one of the ways of rendering audible the air escaping or entering the vocal tract, and we begin by examining its physiological and anatomical properties.

In a typical egressive pulmonic speech sound, the air is exhaled, i.e. propelled out of the lungs (see figure 2.1). The air passes through the trachea and the larynx to exit through the mouth and/or the nose. An ingressive sound is produced by the inhaled air, i.e. air coming into the oral cavity instead of being exhaled.

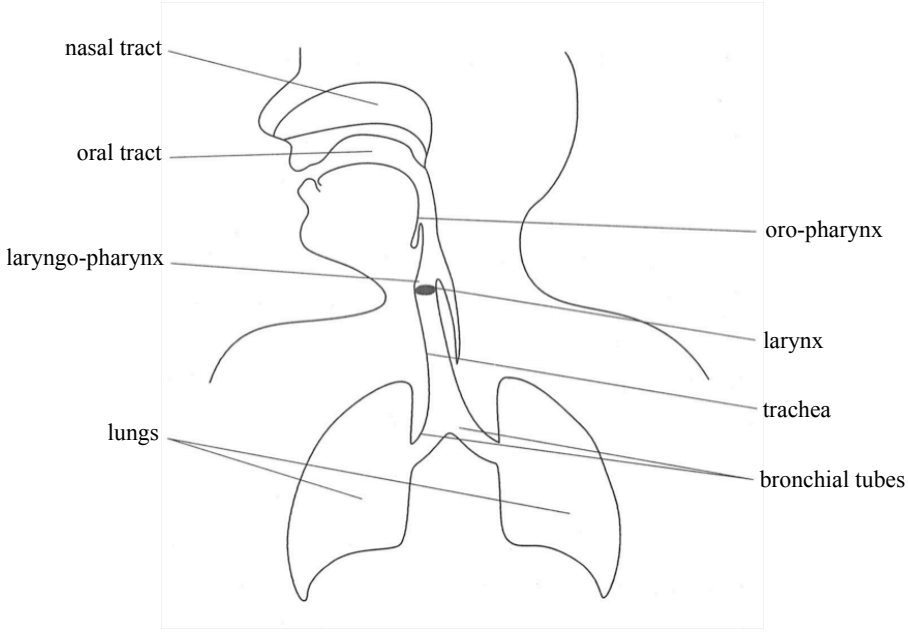


Fig.2.1 Vocal tract

Voice is produced by the closing and opening of the vocal cords (or vocal folds) in rapid succession. The airflow exiting the lungs is interrupted at the level of the larynx to produce air puffs. This happens when the vocal cords are brought together and stiffened. If the air puffs follow each other rapidly enough, more than 40 times per second, we hear them as continuous sound at a certain frequency or height, i.e. *pitch*, rather than as single puffs. In the source-and-filter model of speech production (Fant 1960), the action of the vocal cords is called the *source* of speech.

The vocal cords consist of two muscles located closely together at the front part of our neck by a so-called 'shield' cartilage, the thyroid cartilage, which is the cartilage protruding in many men and some women (see figure 2.2). At the back, they are attached to the arytenoid cartilages, and the cricoid cartilages below the arytenoids. The vocal cords are parallel to the ground when we stand. When we breathe and do not talk, the airflow circulates between the vocal cords, through the space called the glottis, and the vocal cords are not activated. When we talk, certain parts of the larynx are changing position constantly, mainly through the action of the arytenoid muscles, which open and close the vocal cords. The thyroid and cricoid muscles allow the voice to adjust by changing the rate of vibration, thus contributing to pitch, voice quality and timbre. During the voiced parts of speech, the glottis is narrowed through the action of the arytenoids, which can roll apart. When they are brought close together, the air pressure from the lungs pushes the vocal cords apart, leading to a drop in air pressure, which causes them to close again. The opening and closing of the vocal cords happen in rapid succession and periodically. The quicker the succession of closing and opening events, the higher the pitch of the voice. The second cause of a higher or lower pitch is the length and thickness of the vocal cords themselves, similarly to the strings of a musical instrument.

The airflow originating in the lungs and going through the larynx can be manipulated in different ways. In a *creaky* voice, it may happen that each air puff is heard individually. In this case, the vocal cords are fully relaxed; they hang against each other, and thus they cannot fully vibrate. Still the air passes through them in larger puffs, at a lower frequency (30-50 Hz), producing what is called glottalization. In many languages, there is a regular amount of creakiness at the end of sentences, as in Finnish, Hungarian and Georgian. Some languages use creakiness, or *vocal fry*, as a distinctive feature on

vowels, as for instance in Danish stød (see chapter 7). Reetz & Jongman (2009:96) cite Jalapa Mazatec, an Otomanguean language of Mexico, for creakiness as a vocalic feature.

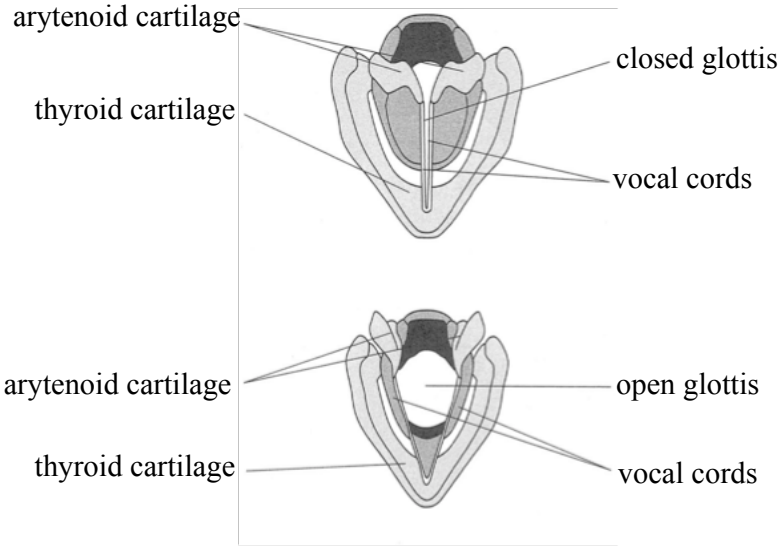


Fig.2.2 Larynx

In a different kind of phonation, the glottis can be slightly widened, allowing a non-turbulent flow of air through the glottis but no voice. The vocal folds are closely adducted towards each other and air is rapidly exhaled through the glottis. This happens in voiceless consonants or in breathy or murmured speech. Whispering more than one syllable consumes a large quantity of air, rendering this kind of phonation not very effective for the production of speech. As the readers can convince themselves, perceptively it is much less efficient than plain voicing.

Amplitude (loudness) is primarily the result of an increased subglottal air pressure. In order to produce louder speech, a speaker can increase the volume of air in the lungs and exhale more air at a certain time. Amplitude is usually measured in decibels (dB), a logarithmic and relative measure, used to express the perceived difference between sounds. An increased air pressure has a physical effect on the eardrums (tympanic membranes).

Duration is the time it takes to articulate speech sounds. A stressed vowel usually takes more time to be uttered than an unstressed one, and speech is typically slowed down at the end of sentences – this phenomenon is known as 'final lengthening', see Wightman, Shattuck-Hufnagel, Ostendorf & Price (1992). Further well-known factors affecting the duration of vowels include their intrinsic length: tense vowels are longer than lax vowels and voiced obstruents often lengthen the duration of the preceding vowel. Segments in longer syllables and words are usually shorter than segments in shorter syllables and words. In general, the duration of individual segments and syllables depends on the speaking rate.

The physical impression of stress in a word or in a sentence is caused by several articulatory gestures. A stressed syllable is typically produced with greater articulatory (muscular) effort, and thus with an enhanced activity of the laryngeal muscles and an increase in subglottal pressure, correlating with an increase in high-frequency energy in relation to low-frequency energy (spectral tilt), see Sluijter & van Heuven (1996). Thus the main acoustic correlates of stress are larger, i.e. higher and lower, F0 contours, increased intensity, longer duration and increased spectral values of vowels.

In the source-and-filter model of speech production, the voice (source) is further transformed (filtered) into speech segments through the action of the articulators. Articulation is called the *filtering* component of speech (see for instance Ladefoged & Johnson 2005-2011, Catford 1988, Reetz & Jongman 2009 for details). The articulatory gestures of consonants temporarily block or reduce the passage of air in the larynx, the pharynx or in the oral cavity. The air can escape through the mouth or through the nose, producing oral and nasal consonants and vowels. The articulators – the lips, the tongue and possibly the velum – can block the air entirely or produce constrictions, interrupting or changing the airflow in several ways to produce consonants. If they obstruct the airflow completely, stops are produced; if they only reduce the airflow, fricatives are produced. Vowel qualities arise through changing configurations of the articulators with hardly any constriction. As a result of the articulatory manipulation of the airflow, certain harmonics of the F0 produced by the larynx are blocked while others are enhanced. The enhanced harmonics are called *formants*, and it is the formants which provide the vowels with their typical characteristics, i.e. their quality.

2.2. Acoustic correlates of intonation

Returning to intonation proper, we have seen that the melodic component of speech is produced by the vocal folds, which vibrate at different speeds and in this way produce tonal contours. Pitch (F0) is measured in Hz, that is, the number of periods per second of an acoustic wave. A woman's voice is typically higher than a man's voice, because vocal folds are thinner and shorter in women and, as a result, they vibrate more quickly. On average, young women speak at 220 Hz and young men at 130 Hz. The vocal folds of children are even shorter, which means that children have even higher pitch. Very young children have an average pitch of 400 Hz (Reetz & Jongman 2009:218). Individual speakers have their own *register*: this is the F0 domain that their voice covers. The lowest value for a given speaker S is called the *bottom line*, sometimes called the low reference line, and the highest value is the *high reference line* or *top line*. All registers can be normalized. In a typical normalization, the highest F0 measure is calibrated at 1 and the lowest at 0. All values inbetween are transformed to values ranging between 0 and 1. In this way, different voices are made comparable: it does not matter if speaker S has a low or a high pitch. The normalization procedure allows researchers to perform statistical analyses with different speakers, regardless of the dissimilarities of their voice registers.

The term *pitch register* often means the same as *pitch range*, namely an interval in the F0 used by S in general, or in particular sentences. *Pitch span* (Ladd 1996:260) is used for concrete tonal events, and refers to the difference between two concrete values, a high one and a low one. It is also called *excursion size* ('t Hart, Collier, & Cohen 1990:75) or *key* (Cruttenden 1997:123); see Gussenhoven (2004:76-77) for an overview of this terminology. An illustration of 'pitch span' as intended in this book appears in figure 2.3 from Liberman & Pierrehumbert (1984:159). All seven realizations of the name *Anne* have a different span as a consequence of the fact that the highest tonal value differs in height. The seven contours were realized by the same speaker, and it is conspicuous that the lowest values are more or less at the same level. In other words, only the high values changed, while the low ones remained constant. Liberman & Pierrehumbert attribute the variation in height to different degrees of emphasis: more emphasis results in a higher peak.

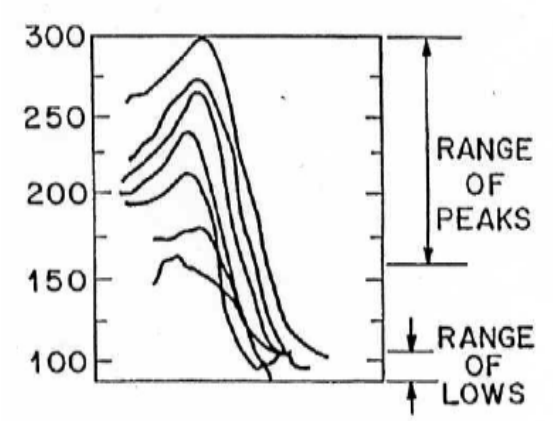


Fig.2.3 Realizations of the word *Anne* by the same speaker (from Liberman & Pierrehumbert 1984:159)

Compare figure 2.3 with figure 2.4 showing different German speakers with different voice registers pronouncing *Anna*.[1] In this case, the speakers not only have different high tones but also different low tones. The higher contour, produced by a child, is also shorter. No normalisation of pitch and duration was performed.

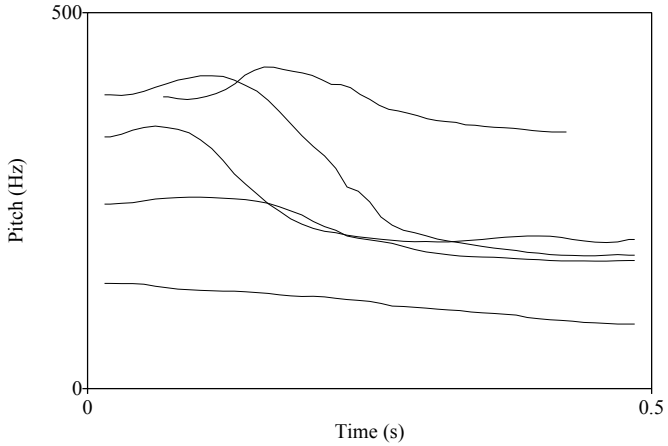


Fig.2.4 Realizations of the word *Anna* by different German speakers

The human auditory system perceives tones in a logarithmic way (instead of a linear way): 220 Hz and 440 Hz are two values of the musical tone A, the higher one being twice as high as the first. The next higher A is two times 440 Hz, thus 880 Hz, and the next one is 1760 Hz, two times 880. The logarithmic perception implies that a step of 50 Hz in the low domain of a male voice is perceived as a larger step than a 50 Hz step in a child's voice. One way of dealing with this perceptual difference is to use the more appropriate logarithmic *semitone* scale. A semitone is the difference between two adjacent tones on a piano, a perceptively stable interval. Twelve semitones always comprise an octave or a doubling of the F0. A semitone is always perceived in the same way, but a low semitone involves fewer Hz than a higher one. Certain other scales are used to interpret the perception of F0, like the psycho-acoustic Mel-scale, the Bark-scale, and the ERB-scale (see Reetz & Jongman 2009 for description and discussion of these scales).

It is also interesting to observe that F0 may be perceived even if it is not present in the sound wave (like in telephones where low frequencies are clipped). This is because of

[1] Here and throughout the entire book, the pitch tracks contributed by the author were created with Praat© (Boersma & Weeninck 1999-2015), a speech analysis software, freely downloadable at www.praat.org.

the presence of harmonics—or formants—of the vowels and other sonorant speech sounds. Since the distances between the formants correspond to the F0 (they are entire multiples of the F0), they deliver sufficient information about the fundamental frequency of the signal. Another property of our speech perception is that we do not need a continuous periodic signal to perceive the melodic contour of an utterance. It will be shown in the next section that the speech signal is typically interrupted by voiceless consonants or by stressless voiceless syllables. Even in breathy or murmured speech, information about melody is present although no voice and thus no F0 is present. This is due to the characteristic changes in the spectral properties of vowels. These changes imitate a melodic pattern.

That speech perception is a particularly robust cognitive property can also be concluded from the fact that we can effortlessly process speech in a noisy environment, as for instance when several persons are speaking simultaneously. We can follow one speaker even if the signal is strongly perturbed by interferences coming from the other speakers or from background noise.

As mentioned in the preceding section, the speech we produce and perceive is sound waves, i.e. very rapid changes in air pressure. Sound waves are characterized by frequency, measured in Hertz (Hz), amplitude, measured in decibels (dB), and duration, measured in seconds or milliseconds (a thousandth of a second).

Many of the illustrations in this book were obtained by means of pitch extraction algorithms or pitch trackers, using digitized forms of the acoustic signals produced by human voices. *Digitization* is the process of converting continuous information into a digital format, that is, into information provided by repeated measurements of the signal at regular intervals. The analog material is then transformed into discrete units of data (called bits) by means of an analog-to-digital conversion process. A handy illustration is provided by a comparison between an analog and a digital watch. While an analog watch provides information about the passing of time in a continuous fashion, the digital watch changes the information provided on the screen only once a minute, or once a second, thus at a regular interval. For measuring speech, it is crucial that the essential content of the analog signal is not altered in the digitization process. This can only be achieved by frequent measurements of the waveform because sound waves change very rapidly. The number of times per second that a waveform is measured is called the *sampling rate*. In the following pitch tracks, most measurements were taken 44,100 times per second, or at 44.1 kHz, and others at 48 kHz. A sampling rate of 44.1 kHz has become standard even for music.

Sound can be visualized. Figure 2.5 shows a periodic sound wave, namely the vowel [a] realized at a F0 of 220 Hz. Such a periodic sound wave illustrates how the same change in air pressure occurs in a similar way over and over again. The visible portion was pronounced in a tenth of a second, thus in 100 ms (milliseconds): 22 complete periodic peaks can be counted, that is, 22 negative peaks and 22 positive peaks. If this vowel were maintained for an entire second, there would be 22 x 10 = 220 such periods, and this is how Hz is calculated: the number of times the maximum (or minimum or any point) of the sound wave is reached in one second. This corresponds to the number of times the vocal folds are vibrating – or oscillating – in a second. Each oscillation is a *period*.

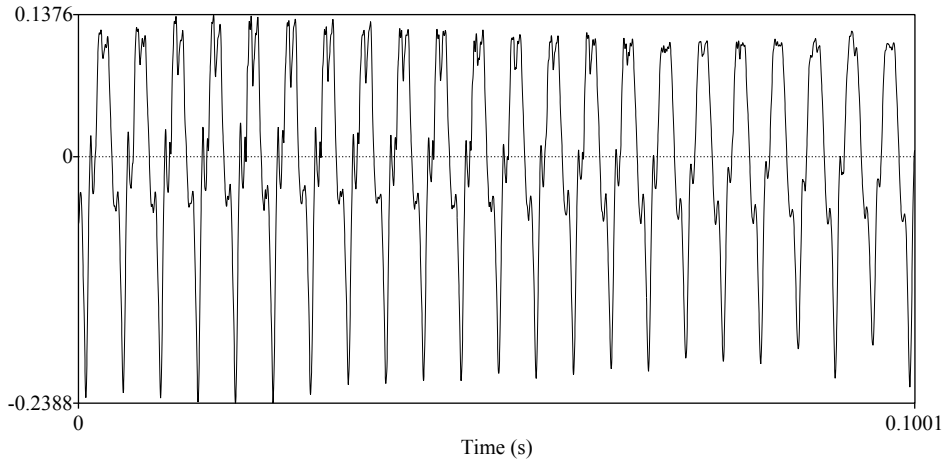


Fig.2.5 100 ms of [a] at about 220 Hz

The acoustic signal shown in figure 2.5 is a complex signal. It is quasi-periodic, which means that the periods repeat themselves (periodic), but they do not do so perfectly (quasi). Each period differs minimally from the neighbouring ones. This is because speech changes continuously. It can easily be seen in the figure that the amplitude of the periods is becoming smaller over time, and that they change their form minimally and continuously.

Compare figures 2.6 and 2.7, which illustrate the realization of the syllable *ba* at a frequency of about 133 Hz and at a frequency of about 220 Hz. In these figures, approximately 1.0 and 1.1 seconds of speech are visible, thus 10 times more than in figure 2.5.[2] This can be seen on the horizontal time axis (Time (s)), which goes from 0.0 to 0.1001 s in figure 2.5, from 0.0 to 1.117 s in figure 2.6 and from 0.0 to 1.026 s in figure 2.7. The result is that fewer details of the sound wave are visible in figures 2.6 and 2.7 than in figure 2.5, although it is still possible to count the oscillations corresponding to the individual pulses of air produced by the vocal cords. The sound wave appears at the top of the figure, and intensity can be calculated from it. The larger the individual peaks are, the louder they were produced. One can see that intensity is greater at the moment of release of the stop, thus at the beginning of the vowel, and that it decreases at the end of the vowel. That the voiced stop is truly voiced is visible from the sound wave and from the spectrogram below. The individual pulses produced by the vocal folds are present during the closure of the [b], although the formants are much weaker during consonants than during vowels.

[2] Figure 2.5 is extracted from the sound shown in figure 2.7.

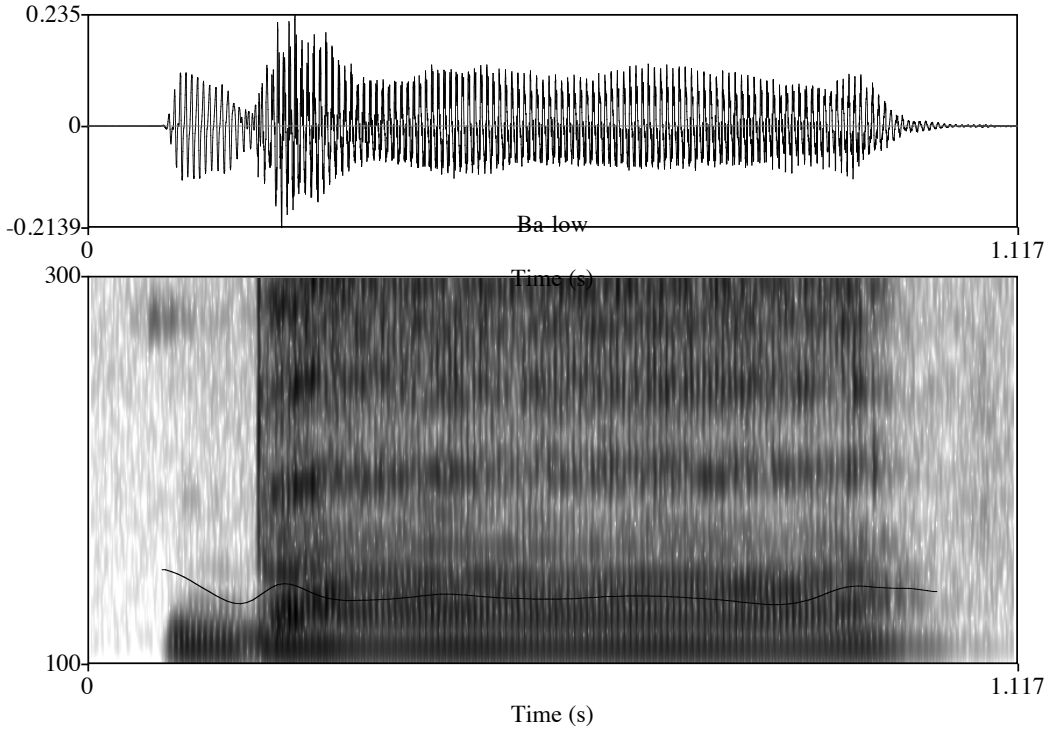


Fig.2.6 The syllable [ba] at about 133 Hz

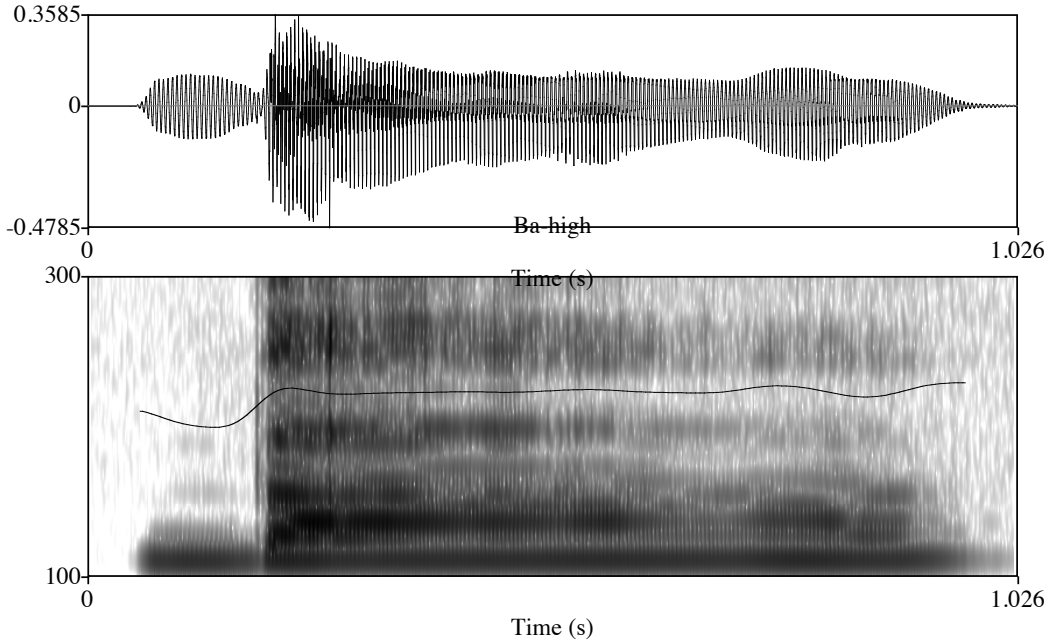


Fig.2.7 The syllable [ba] at 222 Hz

The F0 contour appears as a continuous line. In the spectrogram, each individual vibration of the vocal cords is visualized as a vertical dark line. On the vertical y axis of the lower part in the figures, the scale goes from 100 to 300 Hz in both cases. Additional information is contained in this spectrogram which is of little interest here but nevertheless crucial for speech, namely the horizontal dark regions: these are harmonics of the fundamental frequency, produced by the vowel [a]. As already mentioned above, the harmonics with high energy are called formants, and the formants determine the quality of vowels.

In figures 2.6 and 2.7, not only the vowel is visible but also the voiced onset [b]; the release of the consonant causes a small depression in the F0 contour. The perturbation, called the burst, caused by the opening of the lips, is visible in the pitch tracks, but also in the sound waves and in the spectrogram. The voice is nearly interrupted at this point. Before the release, during the realization of [b], there is voice, but nearly no formants, although these are weakly anticipated during the closing of the lips.

In naturally occurring speech, the F0 contour changes constantly. Figure 2.8 illustrates a semi-spontaneous expression of 2.7 seconds, taken from an unpublished experiment in which the speaker described a picture that she had seen a minute before, but that she no longer could see when uttering the sentence. The sentence *The man in front of the wall is pulling a woman* was realized as an answer to the question *In front of the blue wall is the man pulling a woman or a man?* The utterance is much longer than the pitch tracks shown so far, and due to the need to compress a longer utterance into the same space for visualization, a great deal of detail is lost. The advantage of pitch tracks of longer utterances is nonetheless evident: we can now visualize the intonation of longer stretches of spoken speech, and thus study their intonation. In this figure, the pitch track, the sound wave and the text are provided, allowing us to concentrate on the melodic content, the intensity and duration.

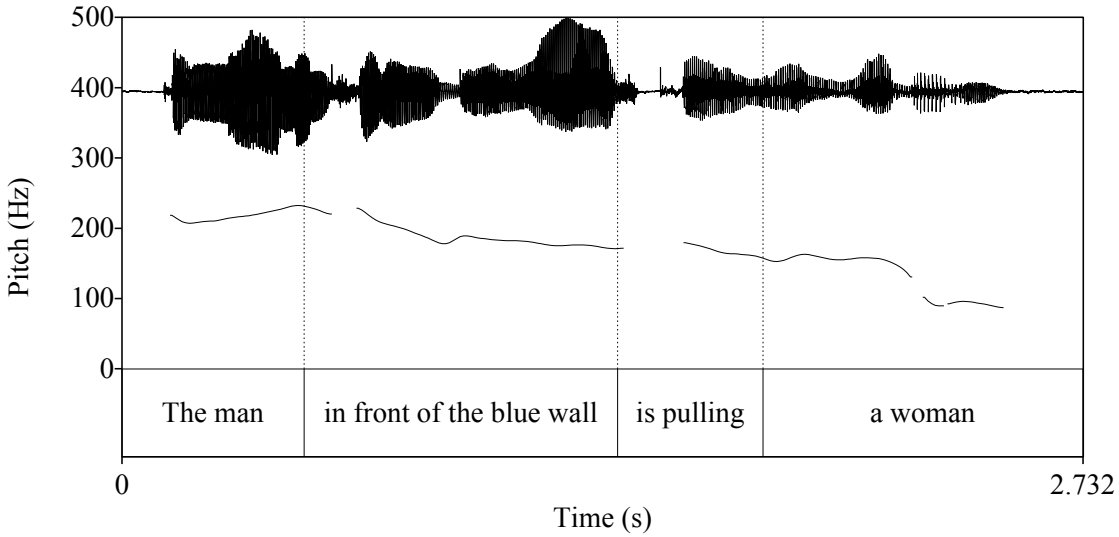


Fig.2.8 Pitch track of a semi-spontaneous experimental utterance (American English)

The sound wave is again shown at the top. It can be seen that besides the quasi-periodic parts made up by the vowels, the signal also consists of more complicated patterns, including noise, hisses and impulses with different amplitudes. We return to these complications in the description of the acoustic signal in the next section, and concentrate in the remainder of this section on the intonation pattern of entire sentences.

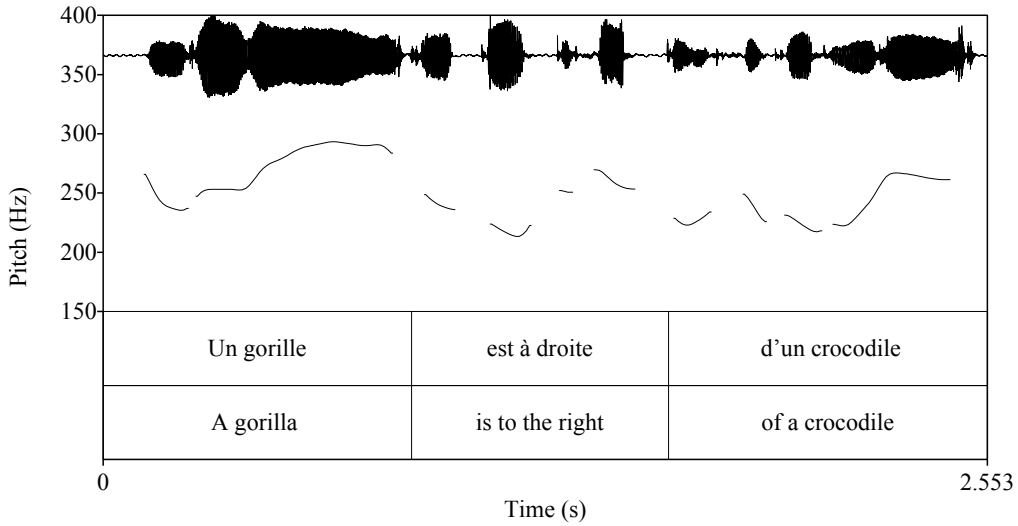
The sentence is uttered in three prosodic phrases (Φ-phrases; see chapter 4), corresponding to syntactic phrases: *the man*, a noun phrase, *in front of the blue wall*, a prepositional phrase, and *is pulling a woman*, a verb phrase, which itself consists of a verb and a noun phrase. The first Φ-phrase has a rising contour, a typical pattern for a topical subject. The other Φ-phrases are smoothly falling until the last word, at which point the pitch falls more steeply. The end of the sentence is realized with creaky voice, as is typical for a discourse-final declarative utterance of American English. The three parts of the sentence are *downstepped* relative to each other: the highest peak of the second part is lower than the highest peak of the first part (ignoring the beginning of the second part, which starts at the high level at which the first part ends), and the highest peak of the third

part is the lowest. Such a downstepped pattern is typical for a declarative sentence, not only in English, but in the majority of languages. We will return to the significance of downstep in later chapters, especially in chapter 5.

As before, the sound wave allows us to evaluate the intensity of the different parts of the utterance. Larger excursions correlate with louder parts. For instance the word *wall* is produced with a high intensity. As can be seen, the intensity diminishes throughout the course of the sentence, a familiar pattern. The pitch accent correlating with the focus on *woman* is not louder than the remainder of the sentence. Nevertheless it is perceived as accented.

The third correlate of intonation is duration. A syllable may take more or less time to be articulated, both in an absolute and in a relative sense. It may be noted that the duration of the word *woman* is quite long.

Finally, let us compare two occurrences of the bisyllabic word *gorille* 'gorilla' in French in the pitch tracks in figure 2.9.[3] The first instance of the word *gorille* (the long one) took 890 ms in an utterance of 2.5 s (a bit less than a third), and the second (short) one was 400 ms long in an utterance of 1.7 s (a bit less than a fourth). The final lengthening of the word in the first case occurs almost entirely in the last syllable. The syllable boundaries are easily recognizable in the sound waves. Perceptively, the difference is quite large, and correlates with a difference in the phrasing of the noun phrase of which *gorille* is the syntactic head. In the first case, it is the subject of the sentence and it forms its own Φ-phrase, while in the second case, it is in the middle of a VP, and thus in the middle of a Φ-phrase. Moreover *gorille* is the new and focused word in the first sentence, but in the second sentence, this role is taken by *vache* 'cow'. In this case *gorille* is discourse-given (see chapter 6 for information structure and intonation). Duration helps the participants to parse the constituents of the sentences in the right way.



[3] These sentences were half-spontaneous: the speaker had to describe the locations of toy animals relative to each other in any way she wanted. French is discussed in more detail in section 8.5.2.

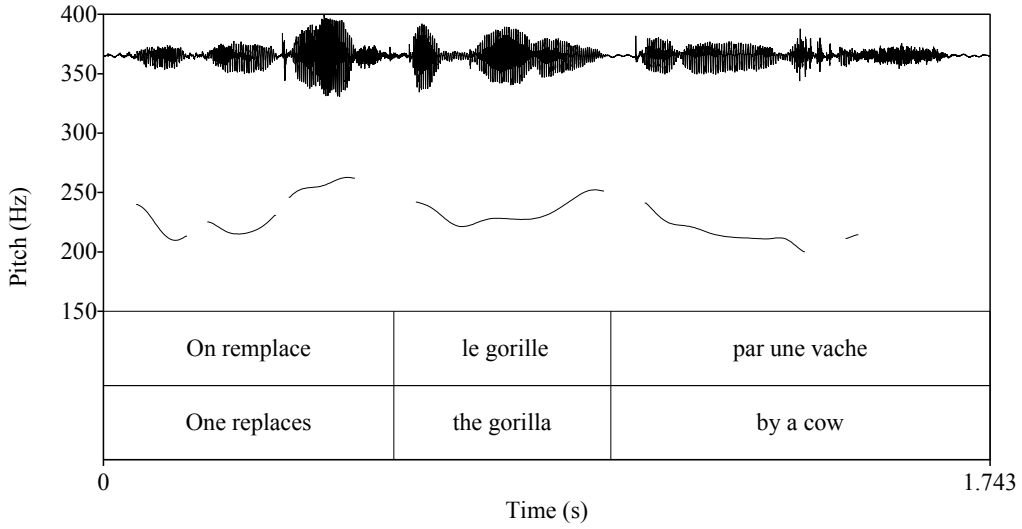


Fig.2.9 Duration of the French word *gorille* in two sentences

The three correlates of intonation discussed in relation to the figures of this section are part of the intonation patterns of all languages. However their roles vary in the grammars of the individual languages, pitch being the component that differs most. The role of intensity is not well studied, although it can be crucial for distinctions in some languages, especially those related to the expression of prominence. Longer duration is often associated with finality, but it can also express prominence.

2.3 Generating and reading pitch tracks

The most commonly used pitch tracker (algorithm generating pitch tracks) nowadays is Praat (Boersma & Weenink 1999-2015), which is freely downloadable from www.praat.org. There are also other pitch trackers implemented as computer algorithms, using different techniques for the extraction of pitch from sound waves.

Praat takes a digitized waveform as input that has been recorded or generated by a computer, or that comes from a different source, and it creates a Praat sound object or a sound file, which can be annotated with a TextGrid object. A TextGrid consists of as many tiers as needed, and allows the text to be aligned with the sound objects, in normal orthography, IPA or SAMPA, an ASCII code replacing the IPA symbols. ToBI tones can also be inserted (see chapter 5 for ToBI). Figures 2.8 and 2.9 contain TextGrids.

Praat allows the generation of all sorts of acoustic files, for instance pitch contour, intensity contour, spectral and formant analysis, and many others; see the illustrations above. With the help of these objects, it is possible to investigate different aspects of the acoustic signal. Here we concentrate on the effects of F0 and show some properties inherent to the speech signal itself or related to the pitch tracker algorithm.

2.3.1 Segmental effects (microprosody)

Some individual speech sounds have intrinsic properties that result in perturbations or distortions of the acoustic signal. These perturbations are called *microprosodic effects*. A pitch contour is interrupted by voiceless segments, since voice is needed for the production of F0, and none is present in voiceless segments. In figure 2.8 the [p] of *pulling* causes an interruption in the F0 contour. Voicing can also be a matter of degree, and it may be more or less regular in obstruents or in creaky voice. However, as already mentioned above, even when completely absent in the pitch track, the F0 contour can be

reconstructed on the voiceless portions, as a consequence of interpolation of the pitch contour. Praat uses an algorithm to generate pitch and, like all algorithms so far, this algorithm is not infallible. For instance, the algorithm may mistake friction (in fricatives or affricates) for voicing, and produce a pitch track for voiceless friction. Conversely, it may also interpret an irregular voice signal as voicelessness.

It can be seen in figures 2.6 to 2.8 that obstruents cause a small depression in the F0 contour just before the closure or the constriction. In the transition out of the obstruent and into the vowel, F0 tends to rise in a voiced obstruent (see [b] in figures 2.6 and 2.7) and to fall in a voiceless obstruent (see the effect of [p] of *pulling* in figure 2.8). The effect is stronger in stops than in fricatives, and also stronger in voiced obstruents than in voiceless ones.

The vowels also exhibit some microprosodic effects: a high vowel like [i] is 'intrinsically' higher than a low vowel like [a] by as much as 15 Hz. This intrinsic pitch effect is due to the position of the articulators in the production of different vowels, which may act on the rate of vibration of the vocal folds.

Segments are also subject to intrinsic duration, some segments being intrinsically longer than others. For instance [a] is longer than [i] and fricatives are longer than other consonants. As already mentioned, voiced consonants may lengthen preceding vowels. Furthermore, rapid speech can have an effect on segments, which may be reduced, or even disappear completely. In many cases, the native speakers of a language can reconstruct absent segments from the signal.

Additional problems may be caused by a poor recording quality. Background noise can be inappropriately calculated, with periodic background sounds causing the algorithm to report spurious F0 values.

2.3.2 Errors of the algorithm

One of the most common but also most obvious errors of the algorithm is due to the so-called octave jumps. Luckily, these errors are easy to detect since the pitch track shows an abrupt fall or rise in the F0 of half or double the value of the actual F0, without a corresponding auditory impression. This happens as a consequence of creakiness or more generally with irregular phonation because the airflow is weaker. When the energy of voicing is low, errors occur more frequently than when the energy is high. Many errors of the algorithm can be corrected manually, by inspecting the visual displays of F0. Typically, the pitch tracks must be corrected before they can profitably illustrate tonal patterns. Illustrations of tonal contours with octave jumps, faulty values and background noises should be avoided, since, in this case, it is impossible to know from the pitch tracks what is correctly rendered and what is the result of algorithm errors. And it is important that the range used in illustrations correctly render the range used by the speaker. If the range is too narrow, the pitch cannot be reproduced properly, and half values will be used in the pitch track. If the range is too wide, interesting tonal excursions may not be visible. It is thus important to optimize the display for the aim of the illustration.

2.3.3 What to measure?

Before the recent technological advances allowing everyone to perform prosodic experiments, phoneticians used so-called 'ear phonetics' to investigate intonation patterns, meaning that they were guided by their personal perception. Phoneticians mainly used their own intuitions to classify tones and tonal contours. This technique was often combined with functional classifications of the distinctive patterns. We will see some examples in chapter 6. Today ear phonetics is no longer a major component of research, even though good perception is certainly useful for getting first impressions about tonal patterns. Database and experimental studies are now the basis for intonational research.

A promising or successful experiment in prosody needs a representative sample of speech, best produced by a group of speakers who are comparable as to dialect, social background, and age for instance. It is necessary to have a set of predictions to be verified by the experiment. As an example, consider the effect of focus on duration at the end of the sentence in a specific language L. In order to find out whether focus increases final lengthening, it is necessary to compare sentences with and without focus in this position and to have a reasonable set of experimental items produced by at least 6 to 7 representative speakers. As the analysis of prosodic data is time consuming, the design of the experiments is extremely important. The effects of microprosody addressed above have to be taken into account. As for consonants, sonorants deliver a continuous F0 tonal pattern. Since vowels have a certain amount of intrinsic length, it may be important to use the same vowels in all examples, or to counterbalance different vowels. Measurements of duration concern segments, syllables or longer stretches of discourse. The proportions taken by a segment, a word or a syllable can be interesting, as illustrated above with the word *gorille* in French. And since speakers have different F0 registers, it is often necessary to normalize the data. A sound statistical analysis is necessary when the experiment reports results from several speakers, especially if the data come from both male and female speakers.

If the object of investigation is F0, as for instance when considering the effect of focus on pitch, measures are made of the F0 maximum, F0 minimum, pitch span, and F0 mean, as well as of the exact location in a vowel or a syllable where the maximum and minimum are located: the distance between the beginning of these entities and the point of measurement of the F0 can also be of significance (see chapter 7 for tonal alignment).

The objects of investigation in intonation research are varied, and it is necessary to consult the existing literature and to be inspired by previous research. The reader may profit by consulting the collection of articles in Sudhoff et al. (2006) addressing methodological issues in experimental prosody. Further recommendations appear under 'Annotated suggestions for further reading'.

Annotated suggestions for further reading

There are many books introducing articulatory phonetics and phonation. For instance Ladefoged & Johnson (2005-2011) and Reetz & Jongman (2009) are very useful for understanding phonation, the functioning of the larynx, and articulation in general.

Ladd (2008) comes with online material called 'Introduction to intonational phonetics' at http://www.cambridge.org/de/academic/subjects/languages-linguistics/phonetics-and-phonology/intonational-phonology-2nd-edition?format=PB which describes acoustic analysis of intonation for beginners.

Animated film of the vocal cords:
http://www.humnet.ucla.edu/humnet/linguistics/faciliti/demos/vocalfolds/vocalfolds.htm

Sudhoff et al. (2006) is a collection of papers on experimental methods and methods for analysing intonation and prosody.

The reader may want to consult Cohn, Fougeron & Huffman (2012), a collection of papers on laboratory phonology.

Discussion points

Learn Praat by do-it-yourself: Record your female and male friends and produce pitch tracks, spectrograms, intensity curves etc. Create TextGrids and segment the sentences. Compare the results. You may want to reproduce the illustrations in this book using Praat.

Chapter 3
Lower-level prosodic constituents

3.1 Indirect reference hypothesis, prosodic hierarchy and Strict Layer Hypothesis

This chapter and the next one introduce the prosodic constituents. Such an introduction is necessary because tones, both lexical tones and intonational tones, are understood as phonological objects assigned to different prosodic constituents, and thus play a role at different levels of the prosodic hierarchy. Moreover, they also associate in different ways with prosodic constituents, and are realized at specific places in the prosodified text.

The claim that tones are always associated with prosodic constituents is compatible with the *indirect reference hypothesis* (see Inkelas & Zec 1990 for this terminology as well as Liberman 1975, Liberman & Prince 1977, Hayes 1990, Nespor & Vogel 1986, Selkirk 1984, 1995, 2011, Gussenhoven 2004, Ladd 1996/2008, Féry 2011, 2013, Truckenbrodt 2015 and many others). The indirect approach conceives of the relationship between acoustics and meaning as being mediated by phonology, and in the case of intonation, by prosodic constituency and tonal grammar. From this perspective, intonational components are related to linguistic expressions organized in prosodic constituents. This approach is called indirect in comparison with the *direct reference hypothesis*, which assumes a direct relation between acoustic events and morpho-syntax and semantics, without going through phonology and prosodic domains. In the direct approach to intonation, pitch accents and boundary tones are directly mapped to syntactic or semantic constituents. In this conception of intonation, sets of acoustic features correlate with specific functional roles. This is the view adopted among others by Cooper, Eady, & Mueller (1985), Eady & Cooper (1986), Fry (1955), Lieberman (1967), Xu (1999) and Xu & Xu (2005). The indirect reference hypothesis posits that phonological rules do not access syntactic structure directly. Rather, syntactic structure plays a role in conditioning a prosodic constituent structure, and only this prosodic constituent structure may be referred to by the phonology. The phonetic implementation has no direct access to the syntactic structure either. Syntactic structure can only indirectly affect the phonetic correlates, to the extent that it leads to homomorphic prosodic structures between syntax and phonology, and that only the phonological structure bears on phonetics.[1]

As will be illustrated abundantly in this book, the main motivation for prosodic domains comes from the observation that phonological rules and processes need phonological domains in order to be interpretable. Intonation has its own categories, called autosegments, and its own grammar. On the one hand, the morpho-syntactic and semantic part of the discourse is mapped to prosodic constituents and a metrical structure is erected on the text. On the other hand, tones are assigned to this structure as a tonal interpretation of the prosodic and metrical structure, thus allowing us to think of and talk about intonation as a grammatical system separated from but related to syntax and semantics.

Let us now turn to the organizsation of constituents in a prosodic hierarchy. In the prosodic hierarchy, lower-level prosodic categories are dominated by higher-level ones. The prosodic hierarchy was formally introduced into phonology by Selkirk (1980, 1984) and Nespor & Vogel (1986). Morpho-syntactic units are mapped to prosodic units of different sizes, even if the mapping is not always strictly isomorphic. The version of the prosodic hierarchy used in this book is shown in (1). This hierarchy posits that moras are

[1] See footnote 9 in chapter 4 for homomorphy between syntax and prosody.

constituents of syllables, and syllables are grouped into feet, which are themselves grouped into words, etc.

(1) Prosodic hierarchy

υ	utterance	(corresponds roughly to a paragraph or more)
ι-phrase	intonation phrase	(corresponds roughly to a clause)
Φ-phrase	prosodic phrase	(corresponds roughly to a syntactic phrase)
ω-word	prosodic word	(corresponds roughly to a grammatical word)
F	foot	(metrical unit: trochee, iamb…)
σ	syllable	(strings of segments: CV, CVC, …)
μ	mora	(unit of syllable weight)

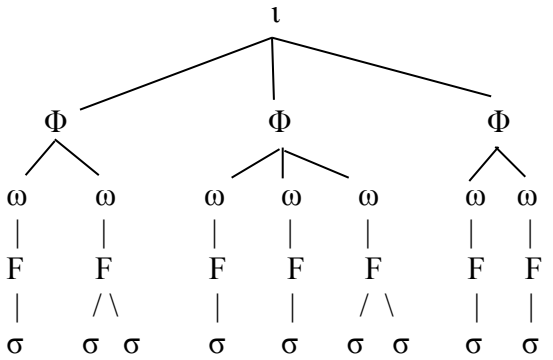
The highest level in the hierarchy, the utterance, typically contains more than one ι-phrase and organizes an intonationally structured paragraph or discourse. Utterances contain rhetorical relationships at the level of the discourse, rendered mainly by the duration and F0 register of entire ι-phrases. We will see some examples below, but this constituent will be largely ignored in the remainder of the book, which takes ω-words, Φ-phrases and ι-phrases as its object of study. One level down is the ι-phrase (see Liberman 1978, Pierrehumbert 1980 and Liberman & Pierrehumbert 1984), which usually encompasses entire clauses. The next level is the Φ-phrase (prosodic phrase), which is more or less isomorphic to syntactic phrasal constituents, like DPs, APs, PPs and VPs. Prosodic words or ω-words map grammatical words, but *function words* are often too light to form ω-words; see section 3.4.1 and are, as a result, included into adjacent ω-words. The higher prosodic levels ω-word, Φ-phrase and ι-phrase are 'interface levels'. These higher categories are not stipulated independently, but they derive from interface relations with morpho-syntax by mapping principles. How this mapping between morpho-syntactic categories and the corresponding prosodic categories takes place is the subject of section 3.4 of this chapter for ω-words and of the entire next chapter for Φ-phrases and ι-phrases.

The lower categories, mora, syllable and foot, are called 'rhythmic' categories because of their role in representing or assigning stress. The moras are constituents of syllables. Moras can also be direct constituents of feet. Syllables are strings of segments, one of which usually is a sonority peak. All other segments, if present at all, are of decreasing sonority. The most common type of syllable has a sonority peak or syllable peak, also called nucleus, an onset and more rarely a coda. Feet consist of syllables or moras. They are purely rhythmic units that are involved in the calculation of stress and they play a major role in morphological processes and rhythmic effects.

For several decades, linguists have adhered to the strict layer hypothesis (SLH) in (2). Principles 1 and 2 of the SLH contain several restrictions at once. They claim that constituents of level p exhaustively dominate constituents of the immediately lower level p–1 and that no level may be skipped. The constituents of level p–1 are parsed entirely in constituents of level p, which means that no part of the segmental string is left unparsed at any level of the hierarchy. Constituents of level p may not dominate constituents of the same level p (no recursion), or constituents of a higher level p+1.

(2) Strict Layer Hypothesis (SLH, Nespor & Vogel 1986:7)
Principle 1. A given nonterminal unit of the prosodic hierarchy, X_p, is composed of one or more units of the immediately lower category, X_{p-1}.
Principle 2. A unit of a given level of the hierarchy is exhaustively contained in the superordinate unit of which it is a part.

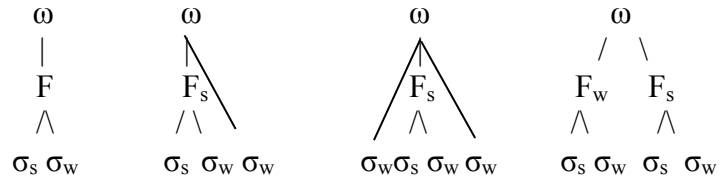
Ideally, a sentence is parsed exhaustively at all levels of the prosodic hierarchy. An example of exhaustive parsing is given in (3) for English.



(3) Miss Sa rah called Jim's stepson last week

It is furthermore assumed for English and numerous other languages that the prosodic and tonal components of grammar are associated with a metrical grid or tree, and mapped to the prosodic domains, following Liberman & Prince (1977), among others. The prosodic hierarchy is compatible with the principle of *culminativity*, which requires that each level have an obligatory and unique head.

In (4), a few words are illustrated with their metrical structure, first in the form of a metrical tree, and second in the form of a metrical grid. These words illustrate that the SLH cannot always be fulfilled, at least if some well-formedness constraints on the prosodic form of these words, like the trochaic form of feet, are to play a role. Trochees are feet of the form strong-weak. In the words *butter* and *camera*, the first syllable is stronger than the second one, and can be called *s* (for strong). The second one is then *w* (for weak). If feet are strictly trochaic, then in a word like *camera*, the last syllable is not footed, but rather it is directly adjoined to the higher level, the ω-word. The same happens to the first and the last syllables of *calamity*. These words violate the SLH. And in *introduction*, two feet are formed on the four syllables.



(4) a. butter b. ca me ra c. ca la mi ty d. in tro duction

The theoretical status of metrical strength has been instantiated in the form of a metrical structure which calculates the strength of the metrical positions from their level of embedding in a tree or from the number of beats in a grid (Liberman & Prince 1977, Selkirk 1984, Halle & Vergnaud 1987), as illustrated in (5). For the word *introduction*, it is shown that the same relation of prominence also exists at the level of the foot. The second foot is stronger than the first one, in line with primary and secondary stress. The metrical grid is also relevant for higher prosodic categories, as we will see in section 4.1.

×	×	×	×	ω-word
×	×	×	× ×	foot
× ×	× × ×	×× × ×	× × × ×	syllable
a. butter	b. camera	c. calamity	d. introduction	

(5)

In sections 3.2 and 3.3, we will review the main properties of the 'rhythmic' lower-level domains of the prosodic hierarchy: mora, syllable and foot. We will illustrate recursivity of ω-words in section 3.4, and we will return to the question of recursivity of the higher-level prosodic domains in the next chapter.

3.2 Syllables and moras

All languages have syllables, independently of the role of moras.[2] The syllable is a particularly important constituent for the investigation of intonation because stress and tones are usually expressed by changes in F0, intensity and duration at the level of the syllable. In this section, syllables are only considered in their functions as stress bearing units (SBU) and tone bearing units (TBU) and we largely ignore other properties of the syllable, like for instance segmental restrictions, which are beyond the concerns of this book (see Blevins 1995 for a review of syllables). In this section, syllables and moras are introduced in three languages: English, Swedish and Japanese. These languages play a special role in the remainder of the book, as they will be repeatedly used for illustrations of different aspects of prosodic structure and intonation. All three languages are quantity-sensitive, e.g. the internal structure of their rime (sometimes spelled rhyme) determines whether it may carry stress – and thus pitch accent. The rime includes the nucleus and the coda of the syllable, to the exclusion of the onset, see below for illustrations. However, English, Swedish and Japanese present different properties as to the role of these constituents for stress and tones. It is to be observed that a large number of languages are quantity-insensitive, that is the internal structure of syllables plays no role in stress and pitch accent assignment.

The mora (μ) is the smallest prosodic constituent: it is a constituent of the syllable used to calculate syllabic weight (quantity) or length (Hyman 1985, Hayes 1989), or both at once (Perlmutter 1995), and it plays a crucial role in the placement of lexical stress in quantity-sensitive languages. Hyman (1985), McCarthy & Prince (1986) and Hayes (1989) have demonstrated the advantages of using moras in the representation of syllable weight.[3] Heavy syllables are bimoraic or minimally bimoraic, and the bimoraicity comes either from a long vowel or from a short vowel plus a consonant. These syllables have a branching rime. By contrast, onsets generally play no role in syllable weight.[4] As proposed by Hayes (1989) and McCarthy & Prince (1986), an onset rule is formulated to the effect that a prevocalic consonant that can be syllabified as an onset is attached directly to the syllable node and not to the first mora; see the representations in (6).

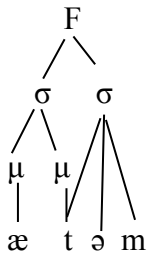
In English and German, syllables can be light, super-light, heavy, or super-heavy (see Kager 1989 for Dutch, Giegerich for English and Féry 1998 for German). The distinction between mono- and bimoraic vowels is related to a quality opposition in vowels: tense vowels are bimoraic and can thus be long (but need not be long), while lax vowels are monomoraic and are always short. Schwa syllables are non-moraic, that is, super-light. In English and German, only bimoraic vowels are allowed to form the rime of an open stressed syllable.

A non-moraic syllable appears in (6)a for English, the second syllable of *atom*: this syllable is non-moraic because its nucleus is a schwa, an unstressable non-moraic vowel. A bimoraic syllable is illustrated in the first syllable of (6)a and in (6)b. In (6)a, the bimoraicity is achieved by the fact that [t] in *atom* is ambisyllabic, i.e. it is at the same time coda of the first syllable and onset of the second one. As has been demonstrated by

[2] See Hyman (1985) who cites Gokana as a language ignoring the syllable in its phonology.
[3] Partly to replace the CV skeleton, which cannot distinguish light and heavy syllables without additional mechanisms, see Levin (1985) for the CV skeleton.
[4] But see Topintzi (2010) for rare exceptions showing that voiceless onsets may be moraic.

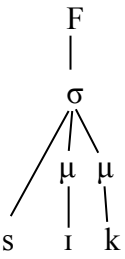
several authors in the past (Giegerich 1992, Perlmutter 1995, Hayes 1995), a stressed syllable is minimally bimoraic, and if the vowel is monomoraic as in *atom*, the second mora comes from the coda, see below. Syllables can also have more than two moras, although syllables are maximally trimoraic, and are then called super-heavy syllables. The third mora in *time* is sometimes analysed as a semi-syllable. The notion of the semi-syllable is a theoretically difficult one, related to latent syllabification and to weight, but also in some cases to extrametricality (see Hayes 1995 for extrametricality in a typological comparison). One way to account for the special status of the third mora is to syllabify it outside of the core syllable, and to add at the level of the foot, skipping the level of the syllable, see (6)c. In this way, all three moras count for weight and stress, and the core syllable is not heavier than the standard two moras.

(6) a. Bimoraic σ and non-moraic σ



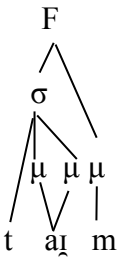
'atom'

b. Bimoraic σ



'sick'

c. Trimoraic σ

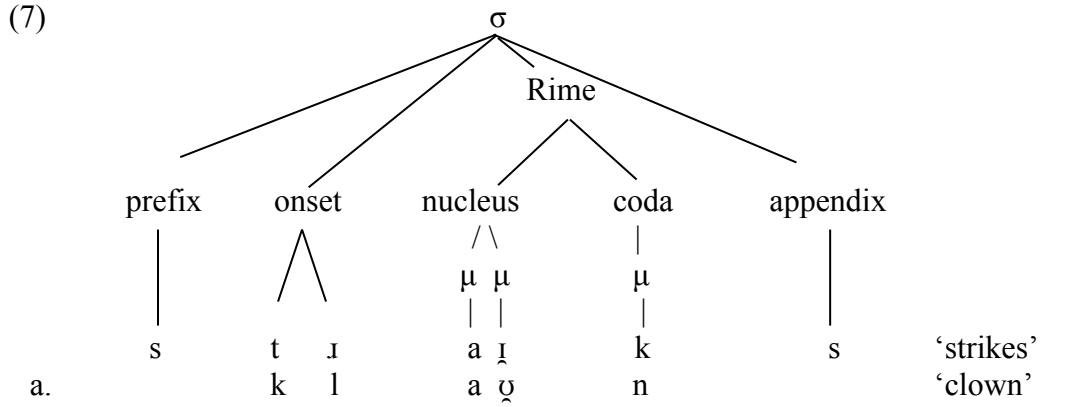


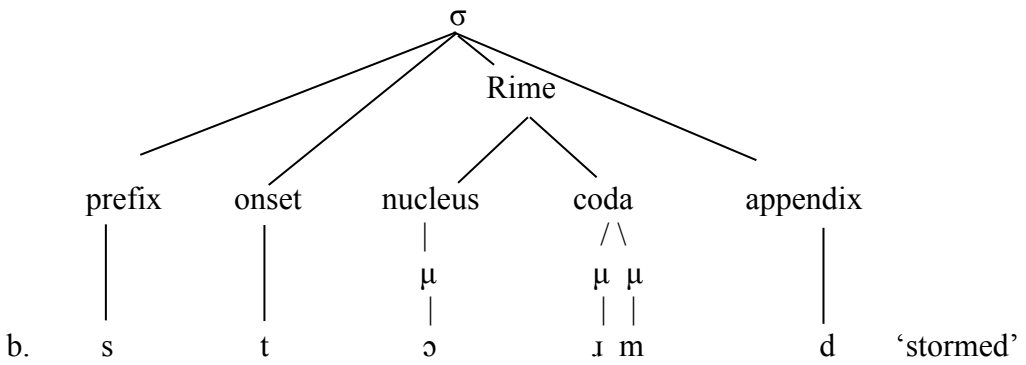
'time'

Perlmutter (1995) proposed that moras are in fact abstract weight units: a bimoraic syllable has the potential to be heavy but is not automatically heavy. This is exactly what we need for Germanic languages. It is their moraic structure which gives vowels the ability to lengthen. Tense vowels are bimoraic and can be long when stressed, but lax vowels are monomoraic and are never long. The rime of a syllable with a monomoraic lax vowel needs a closing consonant in order to acquire an additional mora, fulfilling in this way the bimoraic requirement for stressed syllables. Hayes formulated the Weight-by-Position (WBP) Rule, which posits that a consonant may acquire a mora by virtue of being in a coda.

Not all coda consonants are moraic, some final consonants are best analysed as appendical, weightless and moraless, especially when they are inflectional elements. The core syllable consists of maximally three moras, like in the syllable of *strikes, clown* and *stormed* in (7), and appendical consonants are part of the extended syllable. They are always anterior coronal obstruents in English (see Giegerich 1992). These consonants do not participate in syllabic weight and they also do not conform to the sonority sequencing principle, which requires segments not to increase in sonority from the nucleus to the margins (Sievers 1901). Additionally, there can also be a prefix to a syllable, nearly always [s].

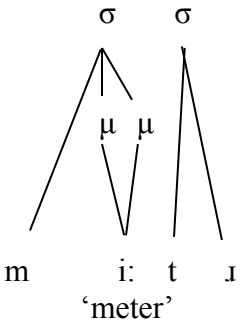
(7)



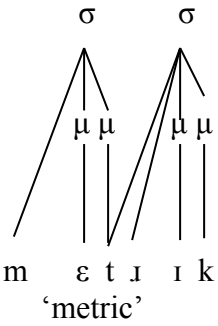


A different case involves consonants that are too sonorous to be included in the preceding syllable, like final [r], [m] and [l] in *meter*, *rhythm* or *double* in English, as illustrated in (8)a for *meter*. These final sonorants form super-light (semi-)syllables when they are nuclear and vocalic. In *metric, rhythmic* and *doubling*, the same segments are consonantal by virtue of being positioned in the onset of a syllable. This is illustrated in (8)b with *metric*. Ambisyllabicity of [t] guarantees bimoraicity of the stressed syllable.

(8) a. [ɹ] is syllabic



b. [ɹ] is non-syllabic



In sum, in English, moraicity comes primarily from the vowels, which are underlyingly mono- or bimoraic. Consonants are not moraic when in the onset, but coda consonants can become moraic in compliance with Weight-by-Position. Syllable weight correlates with lexical stress. Super-light syllables are never stressed. Bimoraic syllables are stressed when they are the strong member of a trochee, i.e. the first one, but not when they are the weak member, i.e. the second and last one; see next section. Super-heavy syllables are generally stressed; they form a foot by themselves, as shown in (6)c. Semi-syllables typically do not bear stress, pitch accent or tone, although they can be part of the foot of the preceding syllable. Stressed syllables play a crucial role in intonation. Because of their status as metrically prominent syllables, they serve as anchors for tonal pitch accents.

Swedish and Norwegian also have a ban on monomoraic stressed syllables and distinguish between open syllables with bimoraic vowels and syllables with closing moraic consonants. However, according to Kristoffersen (2000) for Norwegian and to Riad (2014) for Swedish, who also cites Eliasson (2010) and Eliasson & LaPelle (1973), moraicity in these languages is lexically marked on consonants, not on vowels.[5] A vowel in an open stressed syllable only becomes bimoraic when it is the only segment able to carry both moras. This implies that the moraic status of vowels is not part of the underlying representation of the vowels, but rather it is always derived. Two cases of closed syllables must be distinguished. Either the coda consonant is moraic and is pronounced as a long consonant, or it is not, and then it is again the vowel that lengthens. When moraic, the consonant is pronounced as a geminate coda, or, when there is a following syllable, the second part of the long consonant is the onset of the next syllable. The following contrasting pairs from Riad (2014:159-60) illustrate the effect of a moraic consonant in Swedish. In the left column of (9) and (10), the consonants are non-moraic, and the stressed vowels are necessarily long. In the right column, the long consonants in the words are underlyingly moraic, and the preceding vowels are short. *-a* is the infinitive suffix.

(9) Bisyllabic words

a. Non-moraic consonants				b. Moraic consonants			
läka	/lɛk-ɑ/	[ˈlɛːka]	‘heal’	läcka	/lɛkᵘ-ɑ/	[ˈlɛ́kːa]	‘leak’
såpa	/sop-ɑ/	[ˈsoːpa]	‘soft soap’	soppa	/sopᵘ-ɑ/	[ˈsɔ́pːa]	‘soup’
tiga	/tig-ɑ/	[ˈtiːga]	‘to keep silent’	tigga	/ˈtigᵘ-ɑ/	[ˈtigːa]	‘stagger’

(10) Monosyllabic words

	a. Non-moraic consonants			b. Moraic consonants			
ful	/fʉl/	[fʉ̟ːl]	‘ugly’	full	/fʉlᵘ/	[fʊlː]	‘full’
vin	/vin/	[viːn]	‘wine’	vinn	/vinᵘ/	[vɪnː]	‘win’

Besides being underlyingly moraic, long consonants may have two other sources. They can be the result of a simplification of a consonant cluster of coronal stops, the second one being an inflectional morpheme or part of it. This is illustrated in (11) for Swedish (Riad 2014:166). Or they may acquire their mora as a result of Weight-by-Position. The latter case applies to the first consonant of a postvocalic cluster, as shown in (12).[6]

(11) Assimilated moraic consonant

vitt	/vit-t/	[vɪtː]	‘white.NEUT’	(from vit [viːt] ‘white’)
bytte	/byt-de/	[ˈbʏtːɛ̠]	‘change.PAST’	(from byta [ˈbyːta] ‘to change’)

(12) Weight-by-Position

linda	/lind-ɑ/	[ˈlinːda]	‘to wind’
mjölk	/mjølk/	[mjølːk]	‘milk’

As in English and in German, long and short vowels have different qualities. However in Swedish, the vowel quality is a consequence of the length, rather than the other way round. What is relevant for intonation is that in both cases, moras add weight and heavy

[5] Eliasson assumes geminate consonants in place of moraic consonants. The interested reader is invited to consult the original texts.

[6] The facts have been greatly simplified here. In other cases of consonant clusters, the vowel is lengthened; see Riad (2014:165ff) for a careful motivation of the moraic status of consonants rather than vowels.

syllables can be accented more readily. In all cases, heavy syllables carry tones more often than light syllables, we return to Swedish stress and accent in section 7.4.

In Japanese, moras play a more important role than syllables as far as quantity, weight, length and distance between segments are concerned (see Trubetzkoy 1939). Kubozono (1999:31) cites the examples in (13), which are perceived by Japanese speakers as having the same length despite the different number of syllables. *Tokyo* has two syllables, *Amazon* three and *America* four. The crucial similarity lies in the number of moras: all three words consist of four moras made up by the nucleus and the coda of every syllable. In the following examples, mora boundaries are indicated by hyphens and syllable boundaries by dots.

(13)	a. to-o-kyo-o	too.kyoo	'Tokyo'
	b. a-ma-zo-n	a.ma.zon	'Amazon'
	c. a-me-ri-ka	a.me.ri.ka	'America'

Kubozono (1999) cites a number of phonological operations demonstrating the relevance of moras in Japanese: blends, compensatory lengthening, substitution errors and perceptual units are all mora-based. However, as Kubozono (1999, 2008) shows, the importance of moras does not cancel out the relevance of the syllable. That the syllable plays an important role is already visible from the fact that it is necessary to know which segments form the rime of the syllable in order to know which are moraic. Moreover, McCawley (1978) demonstrated that the syllable is the stress-bearing unit in Japanese, but the moraic structure of a word is nevertheless crucial in the placement of word stress. He formulated a rule in which the word stress is assigned to the syllable carrying the antepenultimate mora, and showed that this rule is responsible for stress assignment in loanwords. As a result, the stress can be located in the antepenultimate syllable, as in (14)a, or in the penultimate syllable, as in (14)b, depending on the mora count of the syllables involved. The data in (14) are from Kubozono (1989a). Kubozono (2015) proposed that McCawley's rule is responsible for a larger number of cases in the vocabulary of Japanese than just loanwords. For instance, regular stress in many compounds consisting of two parts can be accounted for by this rule. In fact, Kubozono assumes that McCawley's rule is the default rule for the entire Japanese vocabulary.

(14) Stress assignment in Japanese loanwords

a.	/dórama/	do.ra.ma	'drama'
	/rekóodo/	re.ko-o.do	'record'
	/sutoráiki/	su.to.ra-i.ki	'strike'
	/sutóorii/	su.to-o.ri-i	'story'
b.	/puráton/	pu.ra.to-n	'Plato'
	/pokétto/	po.ke-t.to	'pocket'

When the antepenultimate mora is a non-syllabic segment, like the underlined coda nasal in (15)a, the second part of a diphthong in (15)b, the second part of a long vowel in (15)c, or the first part of a geminate in (15)d, the syllable which this segment is part of is stressed, thus one mora leftward, showing the relevance of the syllable for this rule. All examples are from Kubozono (1999:43).

(15)	a. myá-n.ma-a 'Myanmar'	ró-n.do-n 'London'	wa.sí-n.to-n 'Washington'
	b. sá-i.pa-n 'Saipan'	su.pá-i.da-a 'spider'	de.zá-i.na-a 'designer'
	c. sá-a.ka.su 'circus'	pé-e.pa-a 'paper'	pí-i.na.tu 'peanut'
	d. só-k.ku.su 'socks'	de.rá-k.ku.su 'deluxe'	pí-t.tya-a 'pitcher'

Thus the role played by moras is crucial in the three languages examined, but in different ways. In Japanese, moras play a much more important role in stress assignment than in English and Swedish. In all three languages, elements of the rime are moraic, but in English, vowels are underlyingly moraic whereas in Swedish, consonants determine the moraic structure of the vowels. In some languages, codas are not moraic, as in Lardil, for instance. Only vowels are.

The role of syllables as Tone-Bearing-Units (TBUs) will be examined later in this book, especially in section 5.3. For now, only a few remarks are sufficient. The most striking fact is that tones combine with syllables at a late phonetic stage. The shape of syllables does not seem to play a crucial role in their association with tones. In other words, we do not find any syllable-tone association requirements like what can be observed for syllable-stress assignment, except of course when stress and tone coincide, as in Swedish; see also section 7.4. Tones often have larger scope than the syllables that carry them, and they sometimes undergo phonological rules and transformations: they double or spread, delete and float. They cause downstep and sometimes upstep, they are subject to sandhi effects, and they assimilate and dissimilate relative to adjacent tones, all concepts and processes that will be introduced in the following chapters; see in particular section 5.3 and chapters 7 and 8. In other words, tones obey their own rules and they associate to syllables in a much looser way than stress. By contrast, we saw that in quantity-sensitive languages, the position of stress is dependent on the weight of syllables, and stress is typically realized by a pitch accent on the syllable it is assigned to. There is no assimilation, dissimilation or floating behaviour. Stress does not spread or double, and if its bearer is deleted, stress is deleted with it, although there are some exceptions due to the effect of culminativity.

3.3 Feet

Moras and syllables are grouped into feet. The foot is the highest rhythmic prosodic domain. In many cases, it determines the position of stress. It provides a metrical template for accent placement and for morphological operations. Even in languages that have no lexical stress, effects of the foot are felt in the form of morphological operations using the foot – like hypocoristics or reduplication – and also in minimal word requirements (see below). Here we are only interested in the foot in its role as a metrical unit for stress and tone.

It is usually assumed that a specific language has only one kind of foot responsible for lexical stress, and that words that do not conform to the foot structure by themselves are either lexical exceptions or are made to comply with the obligatory foot structure by obeying general and specific principles of foot formation (see for instance Hayes 1995 for such a view). As an illustration, English is a trochaic language, and thus preferably forms feet consisting of a strong syllable followed by a weak syllable. As demonstrated in the last section, English is also a quantity-sensitive language. The consequence of this property is that a heavy (at least bimoraic) syllable attracts stress and conversely, stress feeds bimoraicity, by lengthening for example, or by rendering consonants ambisyllabic. Monosyllabic feet are also allowed on a regular basis. A word consisting of two bimoriac syllables is preferably realized as a trochee (*zéro*) and only rarely as an iamb (*Papá, bidét, Julý*), in which case it is often the result of lexicalization of a loanword. This does not hold for verbs and adjectives that have a regular final stress.[7] English also has words or

[7] Compare the following words, which vary in stress placement according to their status as nouns or as verbs: *permit, digest, convict, survey, torment, escort, abstract*.

parts of words with stress patterns in the form of a dactyl (strong-weak-weak), an anapest (weak-weak-strong) or an amphibrach (weak-strong-weak), as illustrated in (16).

(16) a. Trochee: récord$_{noun}$, zéro, úrgent, sólid
b. Iamb: raccóon, corrúpt, compléte, recórd$_{verb}$, appéal
c. Dactyl: wónderful, élephant, cápital, ánagram
d. Anapest: rendezvóus, colonnáde, commandánt
e. Amphibrach: illúsion, tomáto, aróma, horízon, amázing, triúmphant

The position of the stressed syllable is usually stated with respect to the end of the word. The words in (17)a have penultimate stress, those in (17)b have antepenultimate stress, those in (17)c have final stress and those in (17)d have preantepenultimate stress with a secondary stress on the penultimate syllable.

(17) a. párrot, átom
b. é.le.phant, a.spá.ra.gus
c. ra.ccóon, ba.bóon
d. á.lli.gà.tor, dán.de.lì.on

Regardless of their superficial structure, it may be assumed that the words in (16) and (17) all have a trochaic structure. In the default case, trochees consist of two relatively light syllables; see (17)a-b and (18)a. In such cases, the stressed syllable acquires its second mora by ambisyllabicity of the following consonant, recall *atom* in (6)a. Trochees may also consist of a heavy syllable followed by any syllable, parsed in a foot, as in (18)b, or even of a heavy syllable alone, as in (17)c and (18)c. Syllables preceding main stress, which are too weak to form a foot, like the first syllable in *illusion* in (18)b, are not parsed in a foot. If they were, they would form a degenerate foot, which is allowed in some languages, but not in English: final syllables which are not part of the primary foot are not parsed (as in *camera* above, or *elephant* and *America* in (18)a), and they remain reduced and unstressed.

(18) a. (récord), (párrot), (éle)phant, A(méri)ca
b. a(génda), il(lúsion), (August)
c. (hóle), bal(lóon), in(ténse), cor(rúpt), rac(cóon)

Other facts also speak for a basic trochaic metrical structure of English. First, in long words, not only the last two syllables are organized in a trochee, but the preceding syllables also often are, thus avoiding a sequence of more than one unstressed syllable. This is illustrated in (19)a. The second piece of evidence comes from complex words. The examples in (19)b show that complex words are also preferably organized in trochaic feet.

(19) a. (Mìssi) (síppi), (pára)(mòunt), (níghtin)(gàle)
b. (músic)/mu(sícian), (átom)/a(tòmi)(zátion), (mìli)ta(rístic)[8]

But of course not all languages have a metrical system comparable to English. Although many have a trochaic foot system, some use iambs. A few languages use ternary feet: Cayuvava for instance uses dactylic feet; see Hayes (1995) and Kager (1989). Some forms of feet (anapest, amphibrach) are rarely used as a default foot, if at all. These feet

[8] An alternative for such pentasyllabic words is to assume recursive foot structures ('layered feet'), as proposed by Davis (2005) for (Wìnne)$_F$(pe(sáukee)$_F$)$_F$ or (mìli)$_F$(ta(rístic)$_F$)$_F$. The third syllable of these words is not completely reduced.

only exist in lexicalized patterns or in complex words, as illustrated above for English. Hayes (1980) introduced parameters involved in a foot typology, as shown in (20). If conceived as violable constraints, as in Optimality Theory (OT) (see chapter 5), these parameters allow for a variable and flexible explanation of metrical facts.

(20) Parameters involved in a foot typology

1) Dominance: Feet are first or end headed and the head determines the strong side.[9] In binary feet, this parameter distinguishes first-headed trochees and end-headed iambs. The dominant side is labelled s and the recessive side w.
2) Directionality: Feet are formed starting from the beginning or from the end of the word. In an OT terminology, feet are aligned to one edge of the prosodic word. If all words of a trochaic language start with a stressed syllable followed by an unstressed one, this indicates that feet are assigned from beginning to end. Assignment of feet may be iterative or not. In an optimality-theoretic (OT) framework, this is regulated by the ranking between a constraint requiring parsing of all syllables in feet (iterativity) and one demanding perfect alignment (non-iterativity).
3) Quantity-sensitivity: As already discussed above, some languages count moras and some do not. In quantity-sensitive languages, moras determine whether a syllable is heavy or not. In quantity-insensitive languages, each syllable has the same status, regardless of the number of segments contained in the syllable.
4) Extrametricality: Some constituents, generally word-final syllables or consonants, can be extrametrical. In such a case, feet are assigned as if the extrametrical element were not there. In OT, extrametricality is expressed by a constraint forbidding a foot or a syllable to be final in a certain domain: It is always a language-dependent property.[10]
5) A universal foot inventory has been proposed by Hayes (1995) and McCarthy & Prince (1986) among others, which singles out syllabic trochees, moraic trochees and iambs as the only possible feet for lexical stress; see (21).

(21) Universal foot inventory (Hayes 1995, McCarthy & Prince 1986)

a. Syllabic trochee: (canonical form: σ́ σ)
b. Moraic trochee: (canonical form: μ́ μ)
c. Iamb: (canonical form: σ σ́)

In many languages, the foot is the minimal part of a prosodic word. If a stem is too short, it needs to be augmented, as in Lardil, which has a disyllabic foot requirement (Hale 1973, Wilkinson 1988, Prince & Smolensky 1993a). In Arabic and in all Germanic languages, English being no exception, there is a minimality condition on ω-word formation anchored in the prosodic hierarchy. In these languages, ω-words share the property of culminativity with all other prosodic constituents from the syllable on: they must have a head. Culminativity at the level of a ω-word implies that they bear a primary stress, and this implies that every ω-word minimally contains a foot, the bearer of stress. Since a foot is at least two moras heavy (a stressed syllable is bimoraic), every ω-word has to be minimally bimoraic. As a result, all ω-words are minimally bimoraic. The minimal foot requirement is absent in some languages, as for instance in French, which

[9] It has become the standard in linguistics to portray *spoken language* in written form from left to right in terminological conventions like *left-headed*, which assume that writing systems are naturally oriented from left to right, which is clearly not the case.

[10] Extrametricality is not the same as unparsed syllables, as was illustrated with the first syllable of *calamity* or *illusion*. In the latter case, the initial syllable is not extrametrical, but just unparsed.

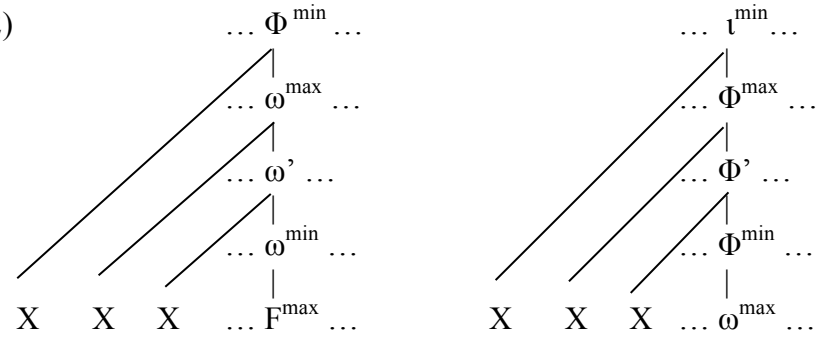
can have monosyllabic words with a short vowel in open syllables (*là* [la] 'there', *trou* [tχu] 'hole'). The absence of a minimality requirement for French correlates with an absence of culminativity in the prosodic hierarchy.[11]

3.4 Prosodic words

3.4.1 Min-max model and compounds in Japanese

From the ω-word level on, prosodic domains are interface levels: the higher prosodic domains interact with morphology, syntax and semantics. As a point of departure, let us take Ito & Mester's (2007, 2013, 2015) model, called *min-max* model of prosodic domains in the following, shown in (22) with ω-words and Φ-phrases. The min-max model assumes recursivity of the higher prosodic domains (ω-word, Φ-phrase and ι-phrase). Recursivity characterizes the fact that a category of level p can dominate a constituent of the same category p: here a ω-word dominates a ω-word, and a Φ-phrase dominates a Φ-phrase. As a result, the model does away with the strong principle of non-recursivity implemented in the Strict Layer Hypothesis in (2). See next chapter for more on recursivity of prosodic domains. At each level of the tree in (22), adjunction of further material is possible, and it is the adjunction operation that motivates recursivity. The optional adjoined material is noted as X in the trees.

(22)



An important property of the model is that it distinguishes between different levels of the same category: a ω^{min} can differ from a ω^{max}. The domination relationship between these constituents, illustrated in (22) is rendered explicit in (23).

(23) Maximal ω (ω^{max}): ω not dominated by ω.
Minimal ω (ω^{min}): ω not dominating ω.
ω': ω dominated by ω and dominating ω.

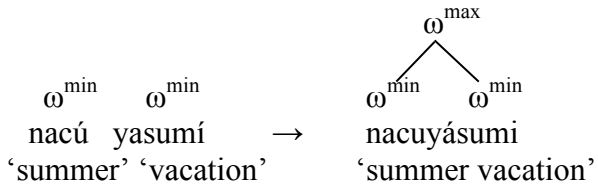
Ito & Mester (2007, 2013, 2015) assume that not only the ω-word and the Φ-phrase but also the ι-phrase are recursive and organized in minimal and maximal constituents. A maximal ω-word is dominated by a minimal Φ-phrase. According to this representation, only one ω-word can be minimal and only one can be maximal. The third variant, ω', can appear an indefinite number of times.

The first motivation for the min-max model was a formal account of the prosodic structure of Japanese compounds. Ito & Mester (2007, 2015) used accent placement, deaccenting and *rendaku* 'sequential voicing', all rules sensitive to compounding, as diagnostics for recursive structure in prosody. Rendaku is a dissimilation rule, which voices the first voiceless obstruent of a non-initial word except if this word already contains a voiced obstruent. As for accent, there is at most one culminative accent per ω-

[11] This does not mean that feet are completely absent in French. In fact, iambic feet regulate reduplication, see section 7.2 for examples.

word in Japanese, both in simplex and in complex words. Compounds comprising two words always form a ω^{max} together, each of the elements of the compound having a ω^{min} status, see (24). There is only one accent for the entire compound, often located on the first syllable of the second word– at least when this element consists of three or four moras–the so-called junctural accent; the other accent is deleted (see Kubozono 2015 for a detailed analysis of accent position in Japanese words). This happens regardless of the original accent structure of the word. For instance the compound *nacuyásumi* 'summer vacation' in (24) consists of the words *nacú* and *yasumí*, the example comes from Vance (2006:187ff). The first word is deaccented, and the second word is accented on its initial syllable.

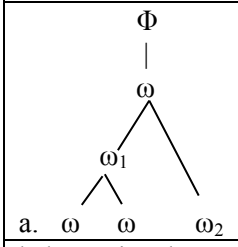
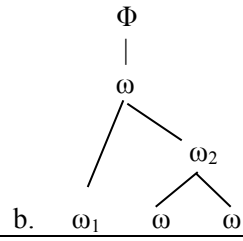
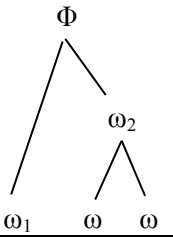
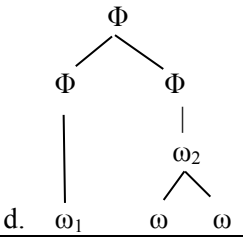
(24)



For compounds consisting of three words, Ito & Mester (2007) distinguish four different categories in Japanese: two are ω-word-based and two are Φ-phrase-based, as illustrated in (25), with the following characteristics:

(25)a is a left-branching ω-word, with the first ω-word (ω_1) itself comprising two ω-words: such compounds have junctural accent on the third word (ω_2), they deaccent the first (complex) member of the compound, and they are subject to rendaku in both ω_1 and ω_2 (*kaisya* → *gaysia* and *hanare* → *banare*).
(25)b is a right-branching ω-word with the second ω-word (ω_2) comprising two ω-words: junctural accent is on the first word of ω_2, and the first word is deaccented. Rendaku does not apply in ω_2.
(25)c is a right-branching Φ-phrase formed by two ω-words, with the second ω-word (ω_2) comprising two ω-words: there is a junctural accent, but only in the embedded compound. As a result the third word is accented, and the other words are deaccented. Rendaku does not apply in ω_2.
(25)d is a right-branching Φ-phrase comprising two Φ-phrases, the second of which is a compound. In this latter case, there is again a junctural accent only in the embedded compound, and no deaccenting, which is due to the fact that the structure itself consists of two Φ-phrases, each with its own accent. Rendaku does not apply in ω_2.

(25) Compounds in Japanese[12]

Word compounds		Phrasal compounds	
		Monophrasal	Biphrasal
 a. ω ω ω_2	 b. ω_1 ω ω	 c. ω_1 ω ω	 d. ω_1 ω ω
hoken-gaisya bánare insurance company departure 'movement away from insurance company'	genkin fúri-komi cash swing push.away 'cash transfer'	hatsu kao-áwase first.time face align 'first face-to-face meeting'	zénkoku kaisya-ánnai nation-wide company guide 'nation-wide corporate guide'

According to Ito & Mester (2007), accents in Japanese are head features associated with minimal Φ-phrases. In (25)a, b and c there is only one Φ-phrase and thus one accent. In (25)d, there are two Φ-phrases and thus two accents.

As exemplified with Japanese, it is possible to go a long way with a restricted number of prosodic categories, a very welcome result. F However, the Japanese model of compounding cannot always be straightforwardly adapted in other languages. Let us examine the prosodic structure of complex words in English that show a different pattern from Japanese.

3.4.2 Prosodic word in Germanic languages

In Germanic languages, prosodic words are roughly equivalent to lexical grammatical words: they can be simplex (monomorphemic) or complex (polymorphemic). Riad (2014) modifies Ito & Mester's adjunction model for Swedish ω-words and claims that the two classes of ω-words traditionally derived by cyclic derivations of different morphological processes can be translated into ω^{min} and ω^{max}, a suggestion that is taken up in this section, albeit in a modified version and for English.[13] A simplex ω-word can be called ω^{min}. It is at the same time a ω^{max} if it is not inflected or derived or part of a compound, see (26). A logical consequence of this representation is that a ω^{max} obligatorily contains at least a ω^{min} but a ω^{min} is not necessarily isomorphic to a ω^{max}. The ω^{max} of *atom* is at the same time the ω^{min} of *atom*, but *atom* can be extended to *atomic*, *atomicity*, *atomless* etc., and in such cases, the ω^{max} does not correspond to the ω^{min}, see also Ito & Mester (2013).

(26) Simplex $\omega^{min}/\omega^{max}$: *atom, music, green, alpaca, rhinoceros, hippopotamus, but*

Extensions of ω-words come from the concatenation of morphemes, typically a stem plus an affix, or more than one stem. The two traditional cyclic classes of English morphological derivation go together with a classification of the operations responsible for them. The classification of the affixes in two classes as well as the order in which the affixes are attached to the stems lie at the heart of Lexical Phonology, a derivational approach intermingling morphology and phonology (Kiparsky 1982, Mohanan 1986). Class 1 affixes are sometimes called 'stress-shifting' and class 2 affixes 'stress-neutral'. Stress-shifting affixes (class 1) are adjoined to the stem and they build a tight

[12] I am grateful to Shin Ishihara for supplying additional information on the lexical meaning of the words in the compounds and on the application of rendaku.

[13] Riad (2014) only considers prosodic structure and does not allow morphology to interfere with prosody. Thus morphologically complex words, for example words containing a stem ω^{min} stem can also be ω^{min}. I am grateful to Sara Myrberg (p.c.) for clarifying this point for me.

phonological entity together, as if they were not complex; see (27)a. The same holds for inflected words, as in (27)b. They are domains for syllabification, foot structure and stress placement.

(27) Complex ω-words consisting of one ω^{min}
  a. Derivation with stress-shifting (or class 1) affixes: *atom-ic, music-al-ity, music-ian, picture-esque, Canton-ese, lemon-ade*
  b. Inflection: *nut-s, boy-s, bush-es, laugh-ed, intend-ed, laugh-ing*

The stress-neutral suffixes (class 2) are ω^{min} and are adjoined to the stem in a looser way. They share a number of properties with compounding, e.g. syllabification and stress see (28)a-b. Each part of a phonologically and morphologically complex word has separate syllabification domains for each morpheme. They do not carry primary stress in the complex word even though they sometimes form super-heavy syllables. Furthermore they can often be deleted in coordination and, in this way, they behave identically to members of compounds (*home- and pennyless*). Because of culminativity, which is a principle active in both ω^{min} and ω^{max}, there is only one primary stress in every ω^{max}. In compounds, this means that some stresses present in ω^{min} are deleted; see (28)b and (32).

(28) Complex ω-words consisting of more than one ω^{min}
  a. Derivation with stress-neutral (or class 2) affixes: *penni-less, happi-ness, neighbour-hood, human-like*
  b. Compounding: *bláckbird, pétticoat*

In (29), it is illustrated for English how complex ω-words consisting of one ω^{min} are organized in prosodic constituents. A monomorphemic content word comprises at least one foot, and there is a primarily stressed/strong syllable; the other ones are unstressed and weak or secondarily stressed. The monomorphemic word *nation* projects a trochaic foot on its two syllables. The complex word *nationality* contains the stem and two suffixes of class 1 (*-al* and *-ity*). These suffixes are bound morphemes, i.e. they need a host to adjoined to. They are feet or syllables, and only form a ω^{min} together with the stem. The words in (29)d-e show that an inflectional suffix is just added to the stem. Inflectional suffixes are syllables or just appendical segments. These suffixes do not affect the stress pattern, although the syllable structure may change, as in *bushes*. The words in (29)a-c illustrate that class 1 affixes may be part of the main foot of the word.

(29) Foot structure of ω^{min}
  a. $((\text{nátion})_F)_\omega$ → $((\text{nàtio})_F\ (\text{náli})_F\ \text{ty})_\omega$
  b. $((\text{átom})_F)_\omega$ → $(\text{a}(\text{tómic})_F)_\omega$ → $((\text{àto})_F\ (\text{míci})_F\ \text{ty})_\omega$
  c. $((\text{músic})_F)_\omega$ → $(\text{mu}(\text{sícian})_F)_\omega$ → $((\text{mùsi})_F(\text{cáli})_F\ \text{ty})_\omega$
  d. $((\text{bóy})_F)_\omega$ → $((\text{bóys})_F)_\omega$
  e. $((\text{búsh})_F)_\omega$ → $((\text{búshes})_F)_\omega$

In (30)a, the pre-stressing suffix *-'ian* is added to the stem *music* causing stress to move to the second syllable of *musician*. In (30)b, the inflected suffix has no effect on the stress pattern of the word it attaches to. In both cases, there is only one primary stress after the word has become complex.

(30) a. × mu.sic + -'ian → × mus.si.cian
  b. × boy + s → × boys

Let us now take a look at the prosodic structure of complex ω-words consisting of more than one ω^{min} in (31) in more detail. In (31)a, the simplex word *happy* forms a trochaic foot, and this trochaic foot remains intact when class 2 suffixes, themselves ω^{min}, are attached. Even though these affixes are bound morphemes, they often resemble free morphemes, to which they are homophonic, i.e. *less, hood, like*. In (31)b, a class 2 derivation is illustrated with an English example. It is shown that the word *meter* appears in its underived form, with a syllabic [r] (compare with *metric* in (8)). Both parts of a compound are at least ω^{min}.

(31) Prosodic structure of complex ω-words consisting of more than one ω^{min}
a. $((\text{háppy})_F)_\omega + ((\text{ness})_F)_\omega \rightarrow ((\text{háppy})_F)_\omega\,((\text{ness})_F)_\omega + ((\text{less})_F)_\omega$
b. $((\text{meter})_F)_\omega + ((\text{like})_F)_\omega \rightarrow\ (((\text{meter})_F)_\omega\,((\text{like})_F)_\omega)_\omega$

In ω^{max}, morphemes are just concatenated and the original syllabification of the parts of the word is kept intact. Stress structure is only changed in the interest of culminativity (superfluous stresses are deleted), but there is no shift of accent. This is illustrated in (32)a for stress-neutral affixes and in (32)b for compounds.[14] The stem carries the primary stress, and in compounds, the first word is primarily stressed.

(32) a. × × × / na.tion + less → na.tion.less
b. × × × / black + bird → black.bird

The min-max model as introduced for Japanese compounds is not adequate for accounting for Germanic ω-words. This is because both kinds of Germainc ω-words are recursively organized: There can be more than one class 1 and more than one class 2 suffixes. Examples (33) show complex words consisting of different kinds of suffixes.

(33) Prosodic structure of complex ω-words (ω^{min} and ω^{max})
a. $(\text{nàtion})_\omega + (\text{al})_\sigma + (\text{ity})_F + (\text{less})_\omega \rightarrow (\text{nationálityless})_\omega$
b. $(\text{síster})_\omega + (\text{hood})_\omega + (\text{like})_\omega \rightarrow (\text{sísterhoodlike})_\omega$

As demonstrated above, class 1 words can be called ω^{min} and this level is recursively organized (which is contrary to Ito & Mester's definition). In the same way, class 2 words are ω^{max} and again this level is recursive. If both class 1 and class 2 words can be intermediate ω' (the only truly recursive level in Ito & Mester's model), we do not expect any difference between the two. An approach locating the distinction between the two complex ω-word categories in the different kinds of adjunction – one delivering a constituent ω^{min} and another one delivering a ω^{max} – does not solve the problem. We still need different intermediate levels of ω-words, one for ω^{min} and one for ω^{max}. Instead, both ω^{min} and ω^{max} are recursive and there is no need for an intermediate ω'.

However, it remains that ω-words are recursively organized. Another insight of the min-max model that is retained here is that all ω-words are part of the same category: both ω^{min} and one for ω^{max} are ω-words, a culminative prosodic category located between the foot and the Φ-phrase.

[14] Prefixes resemble class 2 suffixes, although they behave differently. On the one hand, prefixes can be stressed (*ínconie*) or not (*unháppy*). In this respect, they are comparable to suffixes.

Complex ω-words may also arise in syntax, by concatenation of function words or clitics to ω-words or in combination with each other. Function words, like auxiliaries, pronouns or prepositions typically do not form ω-words by themselves. They may be too weak to form a foot, and are thus not candidates for being ω-words. Alternatively they may be heavy enough to form a ω-word by themselves, but still they are often unaccented, and prefer to be cliticized to an adjacent ω-word. Examples of syntactic ω-words appear in (34). Observe that the voicing of the clitic *s* and *d* in the first two examples is in line with the general voicing assimilation of suffixes in English, showing that the two elements truly form one syllable together.

(34) $(\text{Pat's})_\omega$ $(\text{they'd})_\omega$ $(\text{wanna})_\omega$ $(\text{doesn't})_\omega$ $(\text{she'll})_\omega$

In the next chapter, we turn to the higher interface categories, the Φ-phrase and the ι-phrase.

Annotated suggestions for further reading

Nespor & Vogel (1986) and Selkirk (1984) are the classical books on the prosodic hierarchy. Ghini (1993) extends the approach with rhythmic facts, mostly from Italian.
Hayes (1989) explains why the mora is inevitable in all aspects of suprasegmental phonology. See also McCarthy & Prince (1986, 1993a) for prosodic morphology.
Blevins (1995) remains one of the best overviews on syllable structure in a typological comparison.
Hayes (1995) is an excellent book to learn about metrical theory and foot structure. With 200 languages, it covers an amazing number of facts, and is theoretically well-thought out.
Prosodic word: McCarthy & Prince (1986) and Booij (1995) have revealed the importance of the prosodic word for phonological theory.
A useful collection of papers is the *The Blackwell Companion to Phonology*. Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice. Oxford: Blackwell Publishing. There are some chapters on prosodic constituents.

Discussion points

1. Moras have been given different roles. What are they? Show how all words mentioned in this chapter, especially those of section 3.3, are organized in feet and moras.

2. Discuss quantity-sensitivity and the role it might play in intonation: quantity-sensitive languages have stress and pitch accents. What do you expect from quantity-insensitive languages? Do they have pitch accents?

3. Several authors argue against recursivity in prosodic words, as for instance Downing & Kadenge (2014) and Vigario (2014). How are the complex words of this chapter analysed prosodically without recursivity? Discuss arguments pro- and against recursivity at this level.

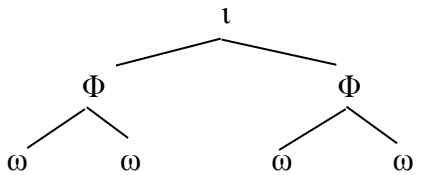
Chapter 4
Intonation and syntax: The higher-level prosodic constituents

4.1 Introduction

The higher-level prosodic categories Φ-phrase and ι-phrase are located at the interface between syntax and phonology: They are phonological in nature, i.e. part of the sound system of languages, but they also reflect the syntactic structure. The Φ-phrase roughly corresponds to a syntactic maximal projection (a VP, a DP, an AP) and the ι-phrase roughly corresponds to a clause (root clause or embedded clause). Two somewhat controversial assumptions are made here: first, a relatively simple syntactic constituency in the form of bare phrase structure (Chomsky 1995) is taken for granted and second, prosodic structures map pre-existing syntactic structures. These assumptions are in line with the mainstream view on the syntax-prosody interface in a generative grammar framework, although they may turn out to be in need of refinement. In fact, some evidence for a prosodic influence on syntax will be provided in chapter 9. In the meantime, the question pursued in this chapter can be summed up as follows: if we consider syntactic structure to be pre-existing, how is prosodic structure mapped to it? It must be observed at the outset of the chapter that most researchers define two levels of prosodic phrasing between ω-word and ι-phrase, called by a variety of names, as for instance 'Accentual Phrase' and 'Intermediate Phrase' (Beckman & Pierrehumbert 1986, Gussenhoven 2004, Jun 2005, 2014), 'Minor Phrase' and 'Major Phrase' (Poser 1984, Selkirk 1986), or 'Clitic Group' and 'Phonological Phrase' (Nespor & Vogel 1986). In the same way as was demonstrated for the ω-word in the preceding chapter, a unique category corresponding to syntactic phrases is assumed here, the Φ-phrase. In line with Féry (2010, 2011), Ito & Mester (2007, 2013), Ladd (1992, 2008), Selkirk (2011) and Zerbian (2007), as well as with the evidence presented in section 3.4, the interface categories are recursive. An additional motivation for adopting a recursive model of prosodic constituency is that in most cases, two levels of constituency between ω-word and ι-phrase are simply not sufficient.

Let us start the chapter with an illustration of the Φ-phrase and the ι-phrase. Consider an ι-phrase comprising a whole sentence and consisting of two Φ-phrases, which themselves correspond to a DP and a VP, as in (1). *Good novels* is a DP containing two words, an adjective and a noun. *Read easier* is a VP also consisting of two ωwords, the verb and the adverb. Together the subject and the VP form a sentence.

(1)



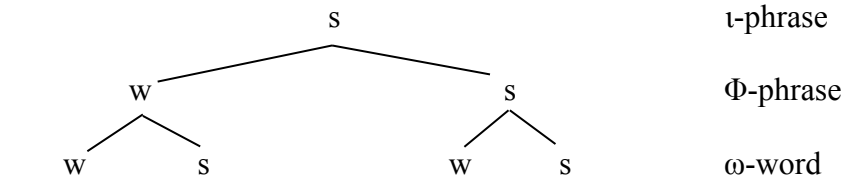
[[Good novels]$_{DP}$ [read easier]$_{VP}$]$_{CP}$

Consider now how this syntactic structure is mapped to a prosodic structure. The ι-phrase is the domain of intonation proper, in the sense that the entire tone sequence of this domain must be well formed in a language-dependent way. In Germanic languages, all higher prosodic domains, from the foot on, are prominence-based (see below for illustration). In a language like English, tones are associated with prosodic structures supplemented by a metrical structure that determines how a prominence relationship is established between syllables. As a result, every domain from the foot on has a prominent constituent, and every ι-phrase has at least one pitch accent and one boundary tone, taken

from an inventory of pitch accents and an inventory of boundary tones (Pierrehumbert 1980, see chapter 5).

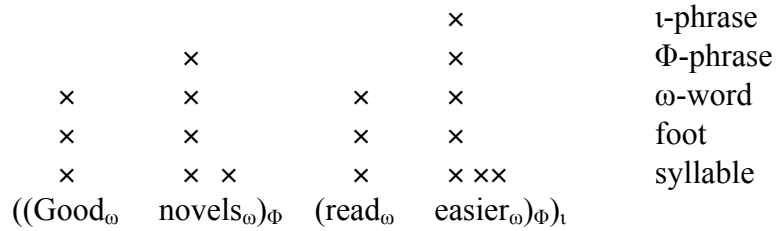
In (2), it is shown with the labels s for *strong* and w for *weak* that not only the words have a metrical structure, as illustrated in chapter 3, but also higher prosodic constituents. The metrical structure was introduced in sections 3.1 and 3.3 in the form of a metrical tree and a metrical grid, and it is extended in (3). The first syntactic phrase, the DP in (1), is mapped to a Φ-phrase in which the second ω-word, corresponding to the noun, is strong and the ω-word corresponding to the adjective is weak. Thus the ω-word of the noun is the head of the Φ-phrase formed on the DP. The VP is mapped to a second Φ-phrase. There the adverb is strong and the verb is weak. The ω-word of the adverb is the metrical head of the second Φ-phrase. At the ι-phrase level, a prominence relationship between the constituents, this time the Φ-phrases, is also established. The second Φ-phrase is the strongest, and it is the metrical head of the ι-phrase; see below for extensive motivation for this conclusion. In this simple example, the constituency and strength relations agree with the requirements of the Strict Layer Hypothesis (SLH) introduced in chapter 3.

(2)



The same relationship is expressed in the form of a metrical grid, as in (3), where levels of metrical structure strictly correspond to prosodic domains and a one-to-one correspondence between metrical beats and prosodic domains can be established. At each level, there is a prosodic head for each constituent, indicated as a node labelled s in the tree and as a higher grid column in the grid. The head of an ι-phrase has a stronger metrical position than the head of a Φ-phrase, which is itself stronger than the head of a ω-word. In the example, the nuclear stress or nuclear accent, i.e. the strongest position in the sentence, is the last word. It corresponds to the highest level in the grid. Such a metrical grid is easier to interpret than a metrical tree, since it shows all strong-weak relationships at once, and, for this reason, it is often used in the literature. Ideally, the tree and the grid correspond to each other one-to-one (see Liberman & Prince 1977 and Halle & Vergnaud 1987 for mapping principles).

(3)



In the same way as was illustrated in chapter 3 for lower-level constituents, a rigid application of the Strict Layer Hypothesis has proven to be impracticable for Φ-phrases and ι-phrases as soon as sentences become more complex: parsing is not always exhaustive, prosodic levels may be skipped, and the same prosodic constituent may appear more than once: a constituent of level n may dominate constituents of the same level n. The example in (4) illustrates the need for recursivity of Φ-phrases. The PP *with a basket full of eggs* is complex: it consists of two Φ-phrases recursively embedded into one larger Φ-phrase. And the VP is a Φ-phrase as well, thus three levels of Φ-phrases are

needed.[1] It goes without saying that the subject might be complex in the same way, and that ι-phrases can also be recursive, as in the case of embedded or coordinated clauses; see below for some examples. If the min-max model, introduced in section 3.4, is right – and we will assume that it is right for the higher prosodic levels, there is a Φ^{min} and a Φ^{max} level, as well as an indeterminate number of intermediate levels, each of which has a culminative head. In regard to the next level, the ι-phrase, Φ^{max} is relevant for prominence.[2]

(×)	ι-phrase
(×) (×)	Φ^{max}
(×) (×) (×)	Φ'
(×) (×) (×)(×)	Φ^{min}

(4) Miss Martin went to the market with a basket full of eggs

The basic idea of all models accounting for syntax-prosody mapping is that the syntactic component is submitted to an algorithm – a set of rules or constraints – the aim of which is to map a prosodic structure to it. Theoretical issues relate to the way this correspondence is formulated as well as to the resulting prosodic constituency. The aim of this chapter is to review some of the known algorithms.

The theoretical frameworks used for the mapping between syntax and prosody can be classified in two main strategies, themselves divided into sub-strategies. One main class of strategies encompasses the *relation-based, edge-based* and *alignment approaches*, discussed in sections 4.2 to 4.4. In this group of approaches, syntactic constituents are mapped to prosodic constituents at one edge only, allowing prosodic constituents to be smaller or larger than syntactic ones, and predicting the emergence of mismatches between syntax and prosody on a regular basis. Although the edge-based approach is older than Optimality Theory (see especially M. Chen 1987 and Selkirk 1986), the emergence of OT has given it a new impulse. OT has a property that many other grammars do not have, namely it allows symmetrical influences of syntax and prosody. In this theory, not only prosody can be influenced by syntax, but the reverse is also possible: syntax can be influenced by prosody. An important property of this kind of approach is the following: edge-based models favour non-recursive prosodic phrases because they partition the text into sequential domains of equal importance. Avoidance of recursivity is intrinsic to this kind of approach, even though recursivity can be implemented in these models, and is thus not completely forbidden.

---

[1] A reviewer observes that it is the ι-phrase which could be recursive, as shown in (i). This remark points to the difficulty of unambiguously naming higher prosodic constituents. In many cases, ι-phrase and Φ-phrase are difficult to distinguish from each other, especially if one uses intonational features. In this chapter, we rely on the syntax to decide. Only clauses are mapped to ι-phrases.

(i) ((Miss Martin went to the market)$_\iota$ (with a basket full of eggs)$_\iota$)$_\iota$

[2] Beside culminative prominence, the phonological cues that languages have at their disposal to signal Φ-phrases vary a great deal. They can be segmental like *raddoppiamento sintattico* 'syntactic doubling' in Italian, see below. They can be signalled by sandhi rules or absence of them, like *liaison* in French. Languages may also use other correlates for phrasing, like vowel deletion or vowel coalescence (see Frota 2000 for Portuguese and Frascarelli 2000 for Italian), lenition, palatalization etc., but also phrasal boundary tones, as in Hindi, West Greenlandic, English, French and German. Φ-phrases may also be marked by an increased duration on the final or penultimate syllable—final lengthening, as we saw in chapter 2. In Chichewa and other Bantu languages, the penultimate syllable is systematically lengthened to indicate the end of a Φ-phrase, see chapter 8 for references. Since we are interested in this chapter in the way prosodic Φ-phrases emerge from syntactic structure, we will only address the phonological cues and phonetic realizations of prosodic boundaries other than tonal ones when needed for the argumentation. Some of the correlates for phrasing will be described in detail in chapter 8.

The second strategy is called the *containment strategy* (Ladd 1992) or *Match* (Selkirk 2011). In this kind of approach, syntactic constituents are mapped to prosodic constituents in their entirety: small morpho-syntactic constituents are mapped to small prosodic constituents and larger syntactic constituents are mapped to larger prosodic constituents. Containment strategies are intrinsically recursive. They require smaller constituents of one category to be entirely contained in larger constituents of the same category. In section 4.5, evidence for recursivity in prosodic constituency is provided, and section 4.6 concludes the chapter by comparing edge-based accounts with containment accounts of the syntax-prosody interface. Evidence and counter-examples for these theoretical approaches will be discussed, with data from several languages.[3]

4.2 Relation-based model of the syntax-prosody interface

The first formal approach to the syntax-prosody interface in a generative framework came from Chomsky & Halle (1968, henceforth SPE), who used boundaries of different strengths to explain how different phonological processes apply in different domains. Chomsky & Halle showed that some processes apply across morpheme boundaries but not across word boundaries, like assimilation, vowel laxing, syllabification, stress assignment and the like. Morpho-syntactic boundaries were first translated into phonological boundaries, and ordered rules were cyclically applied in phonological domains or at their boundaries, from smaller to larger units of speech. Brackets between domains were erased when leaving one level and reaching the next level.[4]

Selkirk (1980, 1984), Nespor & Vogel (1986), Hayes (1990) and others extended and sharpened Chomsky & Halle's use of phonological domains to formulate principles of Φ-phrase formation and assignment of accents.

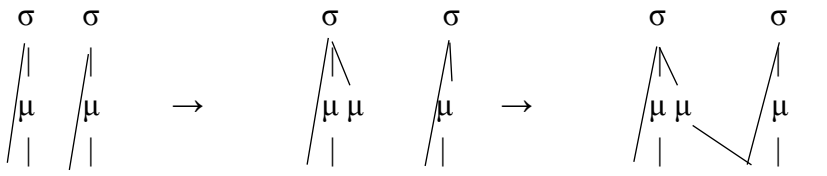
Let us motivate their approach with *raddoppiamento sintattico* (RS) in Italian, a right-branching language, as presented by Nespor & Vogel (1986). RS is a gemination process applying in Tuscan Italian, among other dialects: a consonant C_1 becomes long in a context where a short vowel in a word-final open syllable carries the main stress of a Φ-phrase, and it is followed by a word starting with C_1. The motivation for gemination is that lexically stressed syllables must be bimoraic (or otherwise bipositional) when they are accented. This implies that the monomoraic vowel of a word-final stressed syllable must become long or, if this solution is unavailable, it must be followed by a consonant. Consider the following example. A postnominal adjective is part of the same Φ-phrase as the head noun, since it is on the non-recursive side of the lexical head. In ω_1, the head noun *caffè* 'coffee' in (5) ends with a short monomoraic vowel and the next word ω_2, the adjective, begins with a consonant (C_1 = [k] in *caldo* 'hot'). This consonant geminates so that the first part of the consonant becomes the coda of the stressed syllable of ω_1. In this way, the second mora of the final stressed syllable of *caffè* acquires a consonantal realization. This is illustrated in (6).

[3] Without going into the details of all the theories, it is conspicuous that both syntactic and phonological theories have shaped the interface theories, often in a parallel way. Compare for instance cyclic derivational and relation-based theories (Selkirk 1984, Nespor & Vogel 1986), principles and parameters and edge-based theories (M. Chen 1987, Selkirk 1986), Optimality Theory and alignment (McCarthy & Prince 1993a,b, Truckenbrodt 1999, Selkirk 2000, Samek-Lodovici 2005), minimalism, phase and spell-out (Ishihara 2007, Kahnemuyipour 2009, Kratzer & Selkirk 2007), and recursive accounts of syntax and Match approaches (Féry 2011, Ito & Mester 2012, Selkirk 2011).

[4] An example will be discussed in section 4.4 of this chapter.

(5) Raddoppiamento sintattico in Italian
(ω_1 ω_2)$_\Phi$ [caffè$_N$ [caldo]$_{AP}$]$_{NP}$ → [caffè$_N$ [kk]aldo]$_{NP}$ 'hot coffee'
(ω_1 ω_2)$_\Phi$ [tè$_N$ [fréddo]$_{AP}$]$_{NP}$ → [tè$_N$ [ff]redo]$_{NP}$ 'cold tea'

(6)



From the data in (7) and below, it appears that RS is sensitive to syntactic constituency. It is blocked in (7), where the head of the adjective phrase is on the right: the adverb *molto* 'very' on the left side, the recursive side, cannot be included in the same Φ-phrase as its head.

(7) No raddoppiamento sintattico in (ω_1)$_\Phi$ (ω_2)$_\Phi$
[molto]$_{Adv}$ [cáldo]$_{Adj}$ → (molto)$_\Phi$ (cáldo)$_\Phi$ → molto [k]aldo (*molto [kk]aldo)
'very warm'

The Φ-phrase formation algorithm that accounts for RS appears in (8). It is adapted from Nespor & Vogel (1986:168).[5]

(8) Default Φ-phrase Formation
The domain of Φ-phrase formation consists of a lexical head X and all the material on its non-recursive side up to the next head outside of the maximal projection of X.

When ω_1 belongs to a syntactic constituent that excludes ω_2, RS is also blocked. In such a case, they are in different Φ-phrases. The generalization that surfaces is that RS applies in a Φ-phrase and does not apply across Φ-phrase boundaries. This is illustrated with further examples from Nespor & Vogel (1986) and Ghini (1993) in (9), where ω_1 is the subject of a sentence and ω_2 is the verb. The subject is on the recursive side of the verb. In this case, the gemination process is blocked and the final stressed syllable of the subject remains open.

(9) No raddoppiamento sintattico in (ω_1)$_\Phi$ (ω_2)$_\Phi$
a. [Papà]$_{NP}$ [mangia]$_{VP}$ → (Papà)$_\Phi$ ([m]angia)$_\Phi$ (Papà *[mm]angia)
'Daddy eats.'
b. (La veritá)$_\Phi$ ([v]ínce)$_\Phi$ (La veritá *[vv]ínce) 'The truth wins.'
c. (La solidarietà)$_\Phi$ ([k]resce)$_\Phi$ (La solidarietà *[kk]resce) 'The solidarity increases.'

Compare the RS environment and the blocking environment in (10). The difference between the two conditions in the prosodic structure lies in the presence of a Φ-phrase boundary in (10)a and its absence in (10)b.

[5] Nespor & Vogel use the term Phonological Phrase, translated here as Φ-phrase, and they posit an intermediate level between the Prosodic Word and the Phonological Phrase called the Clitic Group. Assuming recursive domains, the Clitic Group may be considered equivalent to ω^{max}.

(10) Comparison between recursive and non-recursive side

a. [caffè$_N$ [caldo]$_{AP}$]$_{NP}$ [kk] b. [Papà]$_{NP}$ [mangia]$_{VP}$ *[mm]

[ω_1 [ω_2]$_{AP}$]$_{NP}$ [ω_1]$_{NP}$ [ω_2]$_{VP}$

( )$_\Phi$ ( )$_\Phi$ ( )$_\Phi$

The generalization that emerges from the blocking environments is that RS is blocked when ω_1 is bounded to its right by a Φ-phrase boundary. In (11), a further minimal pair, from Nespor & Vogel (1986), illustrates that it is the syntactic structure that is responsible for the formation of a Φ-phrase and the resulting application of RS rather than just sequences of words. In (11)a, *perché* 'why' and *Carlo* (name) are in the same Φ-phrase, and RS applies, but in (11)b the same words are separated by a Φ-phrase boundary and RS is blocked.

(11) Minimal pairs with raddoppiamento sinttatico

a. (Perché Carlo)$_\Phi$ (non é venuto?)$_\Phi$ → Perché [kk]arlo…
why Carlo not is come
'Why didn't Carlo come?'

b. (Che c'é un perché)$_\Phi$ (Carlo lo sa)$_\Phi$ → perché [k]arlo…
that it.is a why Carlo it knows
'Carlo knows there is a reason.'

Nespor & Vogel assumed that right-branching languages are phonologically right-headed, and that the recursive side of such languages is to the left (and consequently that the non-recursive side is to the right). Conversely, left-branching languages are left-headed and have their recursive side to the right. They formulated the principle of Relative Prominence as in (12) to express this generalization.

(12) Relative Prominence (Nespor & Vogel 1986:168)

In languages whose syntactic trees are right-branching, the rightmost node of a Φ-phrase is labelled s (for strong); in languages whose syntactic trees are left-branching, the leftmost node of a Φ-phrase is labelled s. All other nodes are labelled w (for weak).

The principle of Relative Prominence relates the direction of branching in syntax to the position of heads in prosody. In a right-branching language like German, words can be added to the right of the head (the non-recursive side in (8)) and still be part of the same prosodic constituent as the head. Everything being equal, the main accent is final.

Under certain circumstances, expressed in (13) by Nespor & Vogel, and illustrated schematically, two Φ-phrases can be grouped into one. Nespor & Vogel called this process *restructuring*.

(13) Φ-phrase restructuring (optional)

A non-branching Φ-phrase which is the first complement of X on its recursive side is joined into the Φ-phrase that contains X.

..... X .... X .....

( )$_\Phi$ ( )$_\Phi$ → ( )$_\Phi$

Examples of restructuring appear in (14). In a first step, every lexical head projects a Φ-phrase, and in a second step, Φ-phrases can be restructured, as shown in (14)a-b. Example (14)c illustrates the limitations of restructuring. In this case, the subject of the sentence is

not on the recursive side of its verbal head. Rather it is on its non-recursive side. The indicated accents are part of the orthography of Italian, and do indicate sentence accents.

(14) Phrasing according to (13)

a. (Il caribú)$_\Phi$ (nani)$_\Phi$ (sono estinti)$_\Phi$ → (Il caribú nani)$_\Phi$ (sono estinti)$_\Phi$
the caribous dwarf are extinct
'Dwarf caribous are extinct.'

b. (Daró)$_\Phi$ (un libro)$_\Phi$ (a Gianni)$_\Phi$ → (Daró un libro)$_\Phi$ (a Gianni)$_\Phi$
I.will.give a book to Gianni
'I'll give a book to Gianni.'

c. (Papà)$_\Phi$ (mangia)$_\Phi$ no restructuring *(Papà mangia)$_\Phi$
Daddy eats
'Daddy is eating.'

Ghini (1993) showed that Φ-phrase restructuring may also have its origin in rhythm. He proposed that in a sequence of Φ-phrases in which the second one is a complement of the preceding one, Φ-phrases are parsed into units of the same length. The Φ-phrases are preferably symmetrically weighted, or, if this is not possible, they are organized in increasing units in right-branching languages (and in decreasing units in left-branching languages); see one of his examples in (15)b in which rhythm overwrites syntactic structure. Example (15)d shows the only restructuring of (15)c predicted by Nespor & Vogel. Because *di cioccolata* is branching according to them, it cannot be restructured into the preceding Φ-phrase. However, Ghini's rhythmic principles allow restructuring as in (15)e. He recognizes that the function word *di* is too light to block restructuring.

(15) a. (Mangia)$_\Phi$ (molti frutti)$_\Phi$ (Nespor & Vogel)
eats many fruits
'S/he eats lots of fruits.'

b. (Mangia molti)$_\Phi$ (frutti di bosco)$_\Phi$ (Ghini)
eats many fruits of.the woods
'S/he eats lots of forest fruits.'

c. Ho mangiato [dei pasticcini [ripieni di cioccolata]$_{AP}$]$_{NP}$
I.have eaten some pastries filled of chocolate
'I ate pastries filled with chocolate.'

d. (Ho mangiato dei pasticcini)$_\Phi$ (ripieni)$_\Phi$ (di cioccolata)$_\Phi$ (Nespor & Vogel)

e. (Ho mangiato dei pasticcini)$_\Phi$ (ripieni di cioccolata)$_\Phi$ (Ghini)

In sum, Nespor & Vogel (1986) predicted syntax-based prosodic constituents, as well as optional restructuring operations severely restricted by syntax. Ghini (1993) augmented Nespor & Vogel's mapping principles with rhythm-based restructuring principles, which improve the syntax-based prosodic structure in regard to balanced prosodic structure. He showed that not all restructurings are possible, but rather only those which do not contradict strong syntactic boundaries.

Raddoppiamento sintattico (RS) is a good example for illustrating the superiority of the Indirect Reference Hypothesis in prosody (Inkelas & Zec 1990); see section 3.1 for discussion. If RS is to be directly applied on the syntactic structure, as would be the case in the Direct Reference Hypothesis, a number of different syntactic configurations triggering RS would have to be listed, some of which were illustrated above: noun followed by an adjective, numeral followed by a noun, adverb followed by an adjective, complementizer followed by a noun etc. Additionally, these sequences are located in the same syntactic phrase, as made clear in (8) and (13). There is no simple definition of the

syntactic structures relevant for the application of RS. By contrast, the Indirect Reference Hypothesis only needs to let RS apply in a Φ-phrase, and it is this prosodic constituent that is mapped to the syntactic structure in different ways.

A further important aspect of the relation-based approach is the distinction between mapping rules between syntax and prosody, like the one in (8), which takes a syntactic constituent and translates it into a prosodic constituent, on the one hand, and rules like (13), which allows rhythmic restructuring of prosodic constituents on the other hand. It will be shown in section 4.5.2 that the division of labour between the two kinds of rules goes much further than simple rhythmic readjustment. In fact, in a syntax-prosody mapping algorithm, it is crucial to distinguish between the syntax-prosody mapping proper and the prosodic principles acting only on prosody. The division is also needed to understand what is sometimes described as erratic variation in prosodic realization. A sentence can often be prosodically realized in more than one way, according to tempo and style of speech, or according to the individual preferences of speakers. This variation should not lead us to believe that prosody is not ruled by grammar. Rather, the variation is the consequence of different well-formedness conditions or rhythmic effects, beside performance factors, which can complicate the pattern.

A well-known phonological rule applying inside of the Φ-phrase is the so-called *Rhythm Rule*. In (16), the metrical grid resulting from morpho-syntax is shown to the left of the arrow. *Fourteen* has its lexical stress on the final syllable, and the following noun *women* has its lexical stress on the first syllable. As a result, two lexical stresses are adjacent in the Φ-phrase. The second accent is the strongest, in agreement with the Nuclear Stress Rule (Chomsky & Halle 1968, see section 4.4). On the right side of the arrow, a metrical change, the Rhythm Rule, has taken place, the effect of which is that the main stress of *fourteen* is now on the first syllable of this word. The stress shift can only land on a secondarily stressed syllable in the same domain, i.e. the Φ-phrase. The main pitch accent of the Φ-phrase is never shifted. Moreover, the application of the Rhythm Rule is bounded by edges of Φ-phrases. See Liberman & Prince (1977), Nespor & Vogel (1986), Hayes (1995), Bolinger, (1965, 1986), Shattuck-Hufnagel, (1989), Ladd & Monaghan (1987), Horne (1990), Gussenhoven (1991) and many others for the Rhythm Rule.

(16)

			×					×	
		×	×			×		×	
	×	×	×	×		×	×	×	×
	fourteen		women		→	fourteen		women	

The Rhythm Rule has been shown to apply in a variety of languages besides English, like German, Dutch, Swedish, and Italian to name a few. In a language like French, no Rhythm Rule can ever apply, since this language has no lexical stress, although there may be a similar restriction against two adjacent high tones. In French as well high tones are preferably separated by a low tone.

4.3 Edge-based approach

The edge-based approach to the syntax-prosody interface has similarities to the relation-based theory. In both approaches syntactic constituents and prosodic constituents correspond to each other at one edge only. The first edge-based analysis of the syntax-prosody interface was proposed by M. Chen (1987) for Xiamen, a South Chinese Min language.

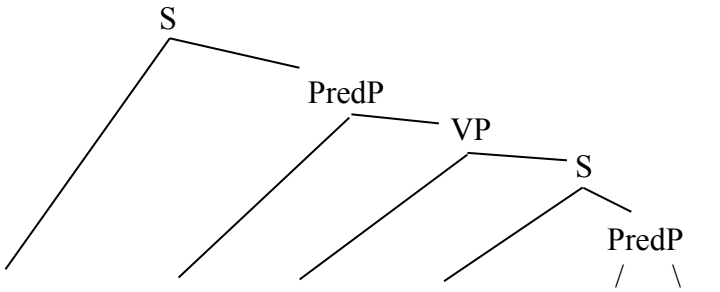
In this language, the prosodic phrasing at the level of the Φ-phrase often corresponds to syntactic structure, as shown in (17) and (18). The former example

displays a short VP consisting of the verb and a following complement, and the latter one shows a longer VP that consists of two Φ-phrases: the verb plus a complement in one Φ-phrase and an additional complement in another Φ-phrase. Chen showed that Φ-phrases (also called TG for *Tone Groups*) are identifiable on the basis of tone sandhi, a phonological process limited to Φ-phrases. Every lexical tone in a Φ-phrase participates in a process of tone sandhi except for the last tone of the Φ-phrase, which retains its original lexical tone (see section 7.6.4 for more on tone sandhi in Xiamen). The tones are not shown here

(17) [V [NP]$_{NP}$]$_{VP}$
(pang hong-ts'e)$_{Φ}$
fly kite
'fly a kite'

(18) [V [NP]$_{NP}$ [NP]$_{NP}$]$_{VP}$
(hoo yin sio-ti)$_{Φ}$ (tsit pun ts'eq)$_{Φ}$
give his brother one CL book
'give his brother a book'

The edge-based approach allows the formation of Φ-phrases that do not correspond to syntactic phrases. In (19), *ying-ko* 'parrot' is included in the preceding Φ-phrase, although from a syntactic point of view, it is the subject of the following verb. Nespor & Vogel's relation-based model cannot account for this mismatch, because there, each lexical head defines a new Φ-phrase, which implies that the preceding verb is in a separate Φ-phrase.



[NP]$_{NP}$ [[Neg V NP]$_{NP}$ V VP]$_{VP}$
(19) (lao tsim-a-po)$_{Φ}$ (m siong-sin ying-ko)$_{Φ}$ (e kong-we)$_{Φ}$
old lady NEG believe parrot can talk
'The old lady does not believe that the parrot can talk.'

M. Chen (1987, 2000:459) formulated an edge-based principle called TG formation, positing that the right edge of every X^{max} (XP in M. Chen's 1987 formulation) falls together with a Φ-phrase boundary, except where X^{max} is lexically governed.[6]

(20) Xiamen Φ-phrase formation (or TG formation)
{Right, X^{max}}, X^{max} not lexically governed.

Selkirk (1986), Hale & Selkirk (1987), Selkirk & Shen (1990) and Selkirk & Tateishi (1991) further developed the edge-based model and applied it to other languages (English, Chimwiini, Tohono O'odham – or Papago, Shanghai Chinese, Bengali and Japanese). Selkirk (1986:389) proposed a universal edge-based model of the syntax-prosody

[6] In the original formulation, adjuncts were exempted from the Φ-phrase formation. This principle has been re-formulated as in (20) by Lin (1994) and as been adopted by M. Chen (2000).

interface. The end parameter for Φ-phrase setting (called Phonological Phrase in the original version), can be formulated as in (21).[7]

(21) End parameter setting
The left/right edge of a Φ-phrase is located at the left/right edge of a Maximal Projection, thus $_{Xmax}[$ or $]_{Xmax}$.

In the next sub-section, the optimality-theoretic version of the edge-based model, the alignment approach, is examined in more detail.

4.4 Alignment approach

Before we turn to the technicalities of alignment, let us first take a look at an important motivation for the formalization of the syntax-prosody interface. Chomsky & Halle (1968) considered the distribution of nuclear accents as crucial for the relationship between syntax and prosody in English. They claimed that the relative prominence of words in a sentence is determined by a recursive algorithm that translates syntactic phrase structures into phonological categories, and that this phonological transcription includes relative prominence. They made a distinction between two stress rules. The first one, the Nuclear Stress Rule (NSR), formulated in (22)a, assigns stress to the last full vowel of a phrase. The second one, the Compound Stress Rule (CSR) in (22)b, assigns stress to the first vowel of a compound. In sequences of words like *black* and *bird* or *world* and *wide*, the location of the stress disambiguates between a phrase and a compound reading. In the following, we concentrate on the NSR, since it is the active rule of a Φ-phrase.

(22) Chomsky & Halle's stress rules

a. Nuclear Stress Rule: $V \rightarrow$ [1 stress] / $VX \_\_ Y]_P$ P = Phrasal Category
(*black boárd*)

b. Compound Stress Rule: $V \rightarrow$ [1 stress] / $\_\_ XVY]_L$ L = Lexical Category
(*wórldwide*)

The NSR is a cyclic rule, applying first to a small phrase and then to successively larger ones. The rule is paired with a Stress Subordination Convention (SSC), which has the effect of reducing by one step the stresses already obtained by previous applications of the NSR. The SSC assigns a word a weaker stress when another stressed word follows: [n stress] is downstepped to [n+1stress] (a larger number is equivalent to a lesser stress).

An example from Gussenhoven (1992) illustrates the cyclic nature of the NSR and SSC. In a first step (line 1 in (23)), each word is assigned a main stress, indicated with 1. In a second step (line 2), the NSR assigns a 1 stress to the final word of the object *telegraphic communication*, and the SSC downgrades the first word of this nominal phrase by one point. In a third step (line 3), the NSR assigns [1 stress] to the final word of the entire sentence. This has two effects: the word *telegraphic* is downgraded by another point and the verb is downgraded by one point relative to the highest number in the object. As a result, main stress is on the last word, the verb has the second stress and the adjective modifying the noun has the weakest stress of the sentence.

[7] She also proposed a similar setting for prosodic words that is not addressed here.

(23)		[They	established	[telegraphic	communication]$_{NP}$]$_{S}$
Line 1:	Word stress	1		1	1
Line 2:	NSR+SSC	1		2	1
Line 3:	NSR+SSC	2		3	1

Even if the NSR goes a long way in accounting for sentence accent in English, several aspects of the syntax-prosody interface led linguists to conclude that the NSR needs further amendments. The first important observation is illustrated by sentences such as (24), an example from Newman (1946) also discussed by Bresnan (1971, 1972), Berman & Szamosi (1972), Bolinger (1972), Schmerling (1976), Halliday (1970) and Gussenhoven (1983b, 1992), among others. In the reading suggested in (24)a, in which George leaves plans behind, *plans* carries the main stress of the sentence without being the last word. It is the pre-verbal argument of *leave*. In the reading of (24)b, in which George wants to leave, the last word carries the main accent: *plans* is the argument of *has*, and the infinitive *to leave* modifies *plans*.

(24) a. George has PLANS to leave.
b. George has plans to LEAVE.

To understand the relevance of this observation, consider sentence accent placement in German. In this language, the verb is often sentence-final, but still the main accent of the sentence is on the pre-verbal argument of the verb. In English, by contrast, an argument is generally post-verbal and if the nuclear stress is assigned to it, or to a following adjunct, this is indistinguishable from a rule assigning nuclear stress to the final word.

(25) Barbara hat gestern ihre BLUMEN gegossen.
Barbara has yesterday her flowers watered
'Yesterday Barbara watered her flowers.'

Because of the regularity of sentence accent placement on the last argument of the verb, the NSR loses its generality, and it must be replaced by a rule placing the accent on the last argument of the verb—or the most embedded argument in Cinque's 1993 framework.

Cinque (1993) proposed that the position of the main accent relates to the depth of syntactic embedding. German NPs are head-initial and right-branching in syntax: complements are to the right, and main accents are final.

(26) a. die [$_{N'}$ Entdeckung [$_{NP}$ des [$_{N'}$ IMPFstoffs]]]
the discovery of.the vaccine
b. die [$_{N'}$ Landung [$_{PP}$ auf [$_{NP}$ dem [$_{N'}$MOND]]]]
the landing on the moon

In Cinque's direct approach to sentence accent, the main accent is thus on the most embedded constituent. In line with the model developed in SPE, Cinque assumed a cyclic accent assignment: every syntactic node is a cyclic node with a metrical constituent: the strongest constituent becomes even stronger because it is assigned a new metrical head on each level; see (27) for an abstract illustration of cyclic accent assignment in a right-branching syntactic structure. The metrical grid formation applies from the bottom up. In (27), the bottom level has three instances of ω-words. Due to binary syntactic bracketing, constituents are organized in a binary way, and become less numerous until the clause-level is reached, at which level only one head has survived, the most embedded one. In

Cinque's model, syntactic constituents carry stress directly (without the intermediary of prosodic domains), and the syntactic levels are directly translated into metrical structure.[8]

(27)
```
          ×
( .  .    ×  )
( . ( .   ×  ))
(ω (ω ( ω )))
```

The second important observation minimizing the impact of Chomsky & Halle's Nuclear Stress Rule relates to intransitive sentences. There is a type of intransitive verb which, in certain circumstances, has no final sentence accent. Schmerling (1976:41-42) compared (28)a with (28)b. These are sentences that she had heard within a few weeks of each other.[9] The difference between the two sentences was elicited by the predictability of Truman's death as opposed to the unexpectedness of Johnson's death, which conferred upon it a whole-event nature. She explained the accent pattern in (28)a in the following way: '[…] Johnson's health was not on people's mind as Truman's had been, and when his death came it was a surprise.'

(28) a. JOHNSON died.
b. Truman DIED.

Consider also Halliday's (1970:38) well-known example in (29), a warning posted in the London Underground. With the accent pattern in (29)a 'the man in the London Underground was worried because he had no dog.' But of course, the reading (29)b is intended, which can be paraphrased as 'if you have a dog with you, you have to carry it.'

(29) a. DOGS must be carried.
b. DOGS must be CARRIED.

Sentences with an accent pattern as in (28)a and (29)a have a global interpretation. They express whole events, and as such are eventive (or thetic) sentences. Sentences like (28)b or (29)b, on the other hand, are so-called 'categorical sentences', which consist of a topic and a comment (see chapter 6 for discourse-related issues). In short, a strict application of the NSR is not sufficient to account for sentence accent in English, and syntactic and semantic considerations play a role as well.

Gussenhoven (1983b, 1992), Truckenbrodt (2006), Kratzer & Selkirk (2007), Féry (2011) and many others formulated sentence accent principles for English, German and Dutch taking into account the predicate-argument relationship of the constituents and basing assignment of stress on phrase formation. Kratzer & Selkirk (2007) assumed a minimalist approach such that syntactic constituents that are defined in terms of phase-based spell-out are "cyclic domains" for phonological realization. According to them, the default phrase stress is a consequence of prosodic phrasing, which is itself a consequence of prosodic spell-out on phase-dependent spell-out domains. Féry (2011) proposed that in German sentences equivalent to (28)a and (29)a, the accent pattern is regularly assigned to the unique argument of the verb, whereas in (28)b and (29)b, the topic requires a separate Φ-phrase (see chapter 6 for a definition of topic). In such short sentences, the verb is the only constituent in the VP that may carry the nuclear accent. Since in

[8] Because of Cinque's insight, there is some confusion in the literature as to the meaning of the NSR. In this book, it is used in the original meaning, as a rule assigning main accent to the last (major) *word* of a domain (not the last *argument*).
[9] Harry S. Truman died on December 26, 1972, and Lyndon B. Johnson on January 22, 1973.

Germanic languages, every Φ-phrase needs a head, and a Φ-phrase's head is realized with a pitch accent, there are two more or less equally strong pitch accents in these cases.

Some examples illustrating the sentence accent assignment rule appear in (30)a-c adapted from Gussenhoven (1992, 2004). According to him and to Truckenbrodt (1995, 2006), metrical structure and prosodic structure are parts of the same representation. Every Φ-phrase is mapped to a syntactic maximal projection and has a head represented as a grid-mark, and, conversely, every grid-mark is the head of a Φ-phrase.

(30) a. [The CAT]$_{DP}$ [GRACEFULLY]$_{AP}$ [KILLED it]$_{VP}$
b. Sie hat [GEOGRAPHIE studiert]$_{VP}$
she has geography studied
'She studied geography.'
c. Sie hat [in ALGERIEN]$_{PP}$ [STUDIERT]$_{VP}$
'She studied in Algeria.'
d. [She loaned her ROLLERBLADES]$_{VP}$ [to ROBIN]$_{PP}$

In the framework of Optimality Theory, Truckenbrodt (1995) introduced a constraint called STRESS-XP, formulated in (31), which can be held responsible for the distribution of phrasal stresses in (30).

(31) STRESS-XP
Each XP must contain a phrasal stress (where 'phrasal stress' is the head of a Φ-phrase).

Truckenbrodt (1995, 2006) assumed Nespor & Vogel's Relative Prominence Principle in (12), but used alignment constraints to define the position of accents resulting from the directionality of branching, in the spirit of Selkirk's (1986) edge-based account of the syntax-prosody interface described in section 4.3. Right-branching languages assign prominence rightmost within a Φ-phrase (as in English, Italian, Chimwiini, Chichewa, Kimatuumbi etc.) and left-branching languages assign prominence leftmost within a Φ-phrase (as in Japanese and Shanghai Chinese; see Selkirk and Shen 1990). Truckenbrodt's (1995) constraint ALIGNR requires that the head of a Φ-phrase be aligned with the right edge of the Φ-phrase, and ALIGNL requires left alignment; see (32). Alignment constraints require two categories, that can be syntactic, morphological or prosodic, to fall together at one edge, left or right. In (32), the constraints address a prosodic category (Φ-phrase) and a metrical category (a culminative head), and requires that the metrical head is aligned with the edge of a Φ-phrase (of which it is the head).

(32) a. ALIGNR(Φ, head of Φ): the head of a Φ-phrase is aligned with the right edge of the Φ-phrase (right-branching syntax)
b. ALIGNL(Φ, head of Φ): the head of a Φ-phrase is aligned with the left edge of the Φ-phrase (left-branching syntax)

In Truckenbrodt's account, STRESS-XP forces a maximal projection in syntax to correspond to an accent at the level of the phrase and ALIGNR forces the accent to be aligned with the right edge of the Φ-phrase, as illustrated in (33). As a result, a necessary feature of the prosodic structure is the presence of a pitch accent on a syntactic constituent.[10] A third constraint WRAP-XP requires that the prosodic domain be at least as large as the relevant syntactic phrase. Example (33), for instance, adapted from

[10] The prosodic structure ignores traces and empty projections (Nespor & Vogel 1986), as well as functional projections (but see Elfner 2013 for data from Connemara Irish showing otherwise).

Truckenbrodt (1995), illustrates a syntactic structure consisting of three ω-words, with a syntactic maximal projection ending after the second ω-word [*The former president*]$_{XP}$. The third ω-word (*slept*) does not belong to the first XP. In the optimal candidate shown here, which fulfils all three constraints, the second word ω2 (*president*) carries a phrasal accent fulfilling STRESS-XP: ω2 is not only final in the Φ-phrase, fulfilling ALIGN-R, but it is also final in the corresponding syntactic domain XP (here a DP). WRAP-XP is also fulfilled because a Φ-phrase wraps the entire XP. NON-RECURSIVITY, not shown in (33), is a further active constraint prohibiting recursion. NON-RECURSIVITY forces the largest syntactic phrase to be mapped to a Φ-phrase.

(33)

	WRAP-XP	STRESS-XP	ALIGNR
(x)$_{\Phi}$ ω1 ω2]$_{XP}$ (ω3) *the former president* (*slept*)	√	√	√

Notice that the ALIGN-R constraint (32) is responsible for the assignment of main stress at one edge of a Φ-phrase. The left edge is the result of WRAP-XP, which requires that the entire syntactic phrase be contained in a prosodic phrase. The role of WRAP-XP is only visible when it is higher-ranked than the ALIGN constraints. It renders the Φ-phrase at least as large as the largest syntactic XP. In other words, WRAP-XP does not require that the Φ-phrase have exactly the same size as the syntactic phrase. A candidate with a larger prosodic domain than the syntactic phrase, for instance one including ω3 in (33), fulfils WRAP-XP as well, since in this case, the XP is contained entirely in a Φ-phrase. However, if the accent is on ω2, ALIGNR is violated in this case.

To sum up so far, Nespor & Vogel's relation-based model, Chen's and Selkirk's edge-based model and Truckenbrodt's alignment model differ from each other in various respects, but they also have important similarities. They all map prosodic constituents to syntactic constituents at one edge only and let the other edge be assigned by other constraints or principles. Recursion in prosodic structure is forbidden in Nespor & Vogel (1986) and Selkirk (1986) and avoided and penalized in later works, like Selkirk (1996) and Truckenbrodt (1995, 1999), who use violable constraints. In the next section, it is shown that the prohibition on recursive prosodic constituency could only survive so long because the focus of analysis was put on very simple sentences, like a main clause containing one or two arguments and/or an adjunct, or even smaller constituents. As soon as longer and embedded sentences and other sentences with complex syntactic structures are the object of study, non-recursivity is no longer tenable.

4.5 Recursivity in prosodic structure

4.5.1 Evidence for recursive prosodic structure

Hauser, Chomsky & Fitch (2002) defined recursion as the basic operation that allows the generation of a potentially infinite array of discrete expressions out of a finite set of elements. There are at least two different ways in which a spoken linguistic string can be recursive, sometimes called direct and indirect recursion: the first (direct) one is iterativity, and the second (indirect) one is embedding. In the following, a principled distinction is made between iteration and embedding of prosodic domains *n*. There is no doubt that morphological and syntactic structures may be recursive in both senses of this term. As far as morphology is concerned, we saw examples in chapter 3. Cyclic derivation and multiple compounds were cases in point. We concentrate here on larger prosodic domains to show how prosody can be recursive in the same way.

The Strict Layer Hypothesis claims that prosody is intrinsically different from syntax in having a flat, essentially non-recursive structure. A rigid application of the SLH and of the edge-based models implies that a sentence is parsed into a number of consecutive and non-overlapping prosodic domains, for instance one containing the subject, and another one containing the VP, so that iteration of prosodic domains is the rule, but embedding is prohibited. Such a pattern was illustrated in (2). This section explores the alternative view that prosodic structure is essentially recursive. The proposal amounts to a conception of prosody as homomorphic to syntax, but more restricted than syntax, due to the smaller number of prosodic constituents. Homomorphism is not to be confused with isomorphism. Isomorphism predicts a one-to-one bijective correspondence between syntax and prosody, which is rarely fulfilled and homomorphism is a structure-preserving map between two categories. It does not assume that the correspondence is one-to-one. In the case at hand, the syntactic structure is preserved in the prosodic one, but without all the details of the syntax being recognizable in the prosody: prosody is simpler than syntax. In this case, even the sentence (2) has a recursive prosodic structure, as shown in (34).

		×		ι
	×	×		Φ’
	×	×		Φ
×	×	×	×	ω
×	×	×	×	foot
×	× ×	×	× ××	syllable

(34) ((Good$_{\omega}$ (novels$_{\omega}$)$_{\Phi}$)$_{\Phi}$ (read$_{\omega}$ (easier$_{\omega}$)$_{\Phi}$)$_{\Phi}$)$_{\iota}$

Starting with iteration in syntax, it is easy to construct sentences containing several domains of the same category, like lists; see (35)a. A further way of generating iterative constructions is by stacking non-restrictive relative clauses for example, as illustrated in (35)b with a literary example. Nothing intrinsic in the grammar blocks infinite generation of stacked constituents and thus infinite iteration. Iteration is universally admitted in the literature on prosodic structure, as well. In such structures, the prosodic domains do not overlap; see among others Nespor & Vogel (1986) and Liberman & Pierrehumbert (1984) for the prosodic realization of lists, and especially how downstep is realized in lists (see chapter 5 for more on this issue).

(35) a. Anna did some errands and bought many things: (a bottle of orange juice)$_{\Phi}$, (an apple)$_{\Phi}$, (sugar)$_{\Phi}$, (butter)$_{\Phi}$, (a pair of socks)$_{\Phi}$

b. ‘Hanna, (who puts on her stockings in the kitchen)$_{\iota}$, (who holds the terry towel in front of the bath tub)$_{\iota}$, (who rides the bike with her skirt wafting)$_{\iota}$, (who stands in my father’s study)$_{\iota}$, (who looks at me in the swimming pool)$_{\iota}$, Hanna, (who listens to me)$_{\iota}$, (who talks to me)$_{\iota}$, (who smiles at me)$_{\iota}$, (who loves me)$_{\iota}$.’[11]

The second type of recursion, on which we concentrate in the remainder of this chapter, involves embedding of prosodic constituents within each other, like in embedded subordinate clauses. This is illustrated in (36) with relative clauses. Example (36)a is a case of sequential relative clauses. Example (36)b is the test case because the main clause is interrupted by center-embedded relative clauses.

(36) a. [[Jonathan]$_{DP}$ [met [a girl [who wore a shirt [that she just bought]$_{CP}$]$_{CP}$]$_{DP}$]$_{VP}$]$_{CP}$.

[11] From ‘The Reader: A novel’ by Bernhard Schlink (1997). Vintage International. New York, translated into English by Carol Brown Janeway.

b. [[The girl [who wears [a shirt [that she just bought]$_{CP}$]$_{DP}$]$_{CP}$]$_{DP}$ [wants to become a lawyer]$_{VP}$]$_{CP}$.

If prosody is iterative, there should not be any difference between the prosodic structures of these sentences. According to the SLH, both (36)a and (36)b are expected to form iterative Φ-phrases. They should both be realized as a sequence of higher-level prosodic domains, downstepped relative to each other. However, if prosody is recursive, the prosodic structure of (36)b should differ from the one of (36)a. In a structure like (36)b, two parts of a single syntactic constituent are separated by another constituent of the same kind: two center-embedded clauses occur in the middle of a main clause, which, as a result, is divided into two parts: the subject *the girl* is one part and the VP *wants to become a lawyer* is the second part. If syntax and prosody are homomorphic, a prosodic domain of level n may be contained in another larger domain of the same level n, ι-phrases in this case, as illustrated in (37)a. As a consequence, the entire clause *the girl wants to become a lawyer* is a single ι-phrase, containing smaller ι-phrases recursively embedded into each other. The alternative, illustrated in (37)b, avoids recursion. In this case, the sentence is divided in a sequence of non-overlapping prosodic domains.[12]

(37) a. ((The girl (who wears (a shirt (that she just bought)$_{\iota}$)$_{\Phi}$)$_{\iota}$)$_{\Phi}$ (wants to become (a lawyer)$_{\Phi}$)$_{\Phi}$)$_{\iota}$.
b. (The girl who wears a shirt that she just bought)$_{\iota}$ (wants to become a lawyer)$_{\iota}$.

See Féry & Schubö (2010) for recursive structure of embedded relative clauses in German showing exactly a prosodic structure corresponding to (37)a. In German sentences like the one illustrated in (38), the second part of the first embedded clause is reset relative to the most embedded relative clause *der naturbelassen ist* 'which has been left in a natural state'. The notion of reset is introduced in detail in the next chapter. At this point it is sufficient to understand that reset is a phonetic clue in this case. It shows that an ι-phrase can be interrupted by another ι-phrase. After reaching the end of an embedded ι-phrase, the F0 returns to a higher level indicating a return to a previous register level, as illustrated in figure 4.1 on the word *leben* 'live' (see also Ishihara 2004, 2007 and section 8.3.2 for results showing embedding of prosodic domains in Japanese, and resulting reset).

(38) (Die Bären, (die im Wald, (der naturbelassen ist)$_{\iota}$, leben)$_{\iota}$ sind friedlich)$_{\iota}$
the bears that in.the forest which nature-left is live are peaceful
'The bears that live in the forest which has been left in a natural state are peaceful.

[12] The phrasing shown in (37)b reflects a proposal by Truckenbrodt (2005) according to which a relative clause and its antecedent are phrased in a single ι-phrase.

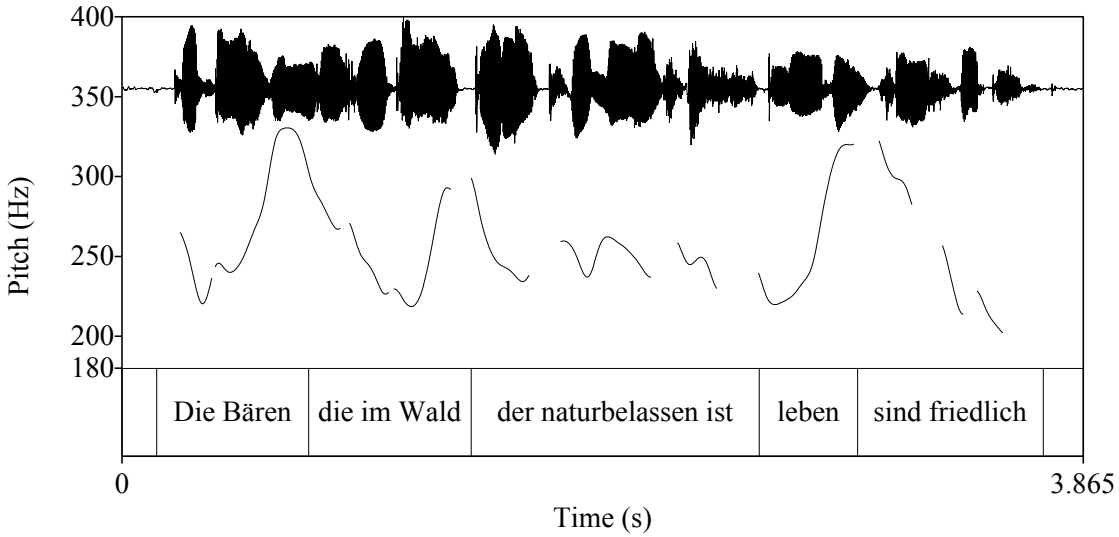
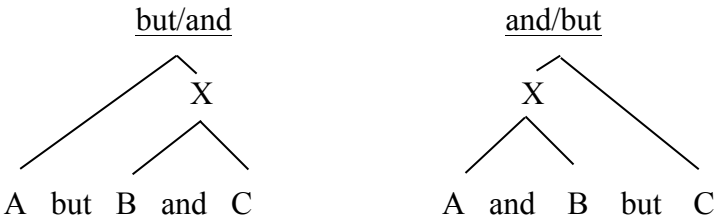


Fig. 4.1 An experimental sentence from Féry & Schubö (2010) with center-embedded relative clauses and upstep of the second part of the first relative clause.

Ladd (1990, 1996/2008) showed that recursion in the intonation of English is not exceptional and that it should be allowed on a regular basis. He called recursive prosodic phrases 'compound domains' to emphasize the observation that they behave like recursive compound nouns.

He compared complex sentences with two different structures as in (39), where A, B, and C are clauses of similar rhythmic and syntactic design, each with clause-internal downstep. An example of each case is given in (40). The clauses contained one of three subjects (*Governor Allen [GA]*, *Senator Warren [SW]* and *Congressman Ryan [CR]*) and one of three predicates (*has a lot more money*, *has more popular policies* and *is a stronger campaigner*). Ladd hypothesized that the different hierarchical structures induced by the conjunctions *but* and *and* would be reflected in the tonal scaling, i.e. they would translate into downstep, where downstep could affect larger domains, even though each of these larger constituents contains internal sequences of downstepped pitch accents. Due to the internal complexity of the clauses A, B, and C, such effects would be non-local on the surface.

(39)



(40) a. <u>but/and</u> condition [A *but* [B *and* C]$_X$]: GA has a lot more money [*but* SW has more popular policies *and* CR is a stonger campaigner]$_X$

b. <u>and/but</u> condition [[A *and* B]$_X$ *but* C]: [GA is a stronger campaigner *and* CR has more popular policies]$_X$ *but* SW has a lot more money

In both conditions, *but/and* and *and/but*, there is one large ι-phrase, the entire sentence, consisting of a simple ι-phrase and a complex one, called X. Ladd's prediction was that the scaling of the initial high tones of the non-initial ι-phrases would differ in both conditions, see figure 4.2. The black dots show the height of the first pitch accent in each

clause, and the grey dots show the non-initial pitch accents in each clause, see section 5.2.4 for more on complex downstep patterns. The black and grey lines show the highest F0 values in each clause. Considering the black dots, downstep applies in each clause, but differs across clauses. In the *but/and* condition, clauses B and C are both downstepped relatively to the preceding one. In the *and/but* condition, clause B is downstepped relatively to clause A, but clause C is only downstepped relatively to the entire preceding ι-phrase, A and B, and is thus approximately at the same level as B. This prediction was confirmed experimentally. In other words, not only the smaller constituents, but also entire clauses contained internal downstep. The phonetic height of the clauses relative to each other proved to be sensitive to the hierarchical differences between the two experimental conditions, showing non-local sensitivity to higher-level prosodic structure, as predicted.

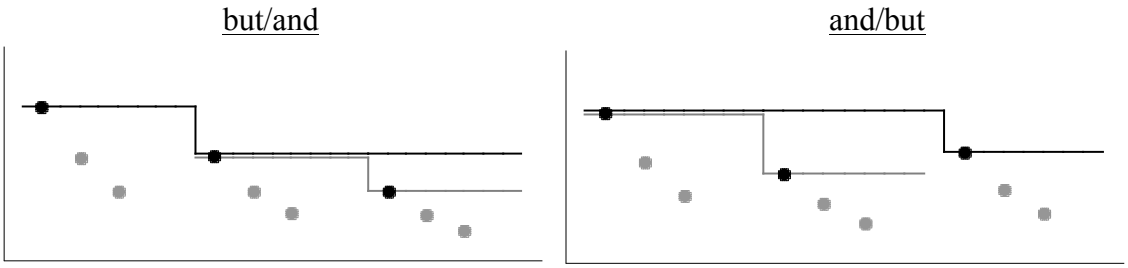


Fig. 4.2 Illustration of the results of Ladd (1990), from Féry & Truckenbrodt (2005)

Féry & Truckenbrodt (2005) and Truckenbrodt & Féry (2015) reproduced Ladd's results for German. Even though there was a great deal of individual variation, the average results for the five speakers recorded in a production experiment showed a difference in the relation between the first high tones of each sentence. In the AX (*A but (B and C)*) condition, all sentences were in a downstep relation to each other. In the XC ((*A and B) but C*) relation, by contrast, sentence B and sentence C started at approximately the same height.

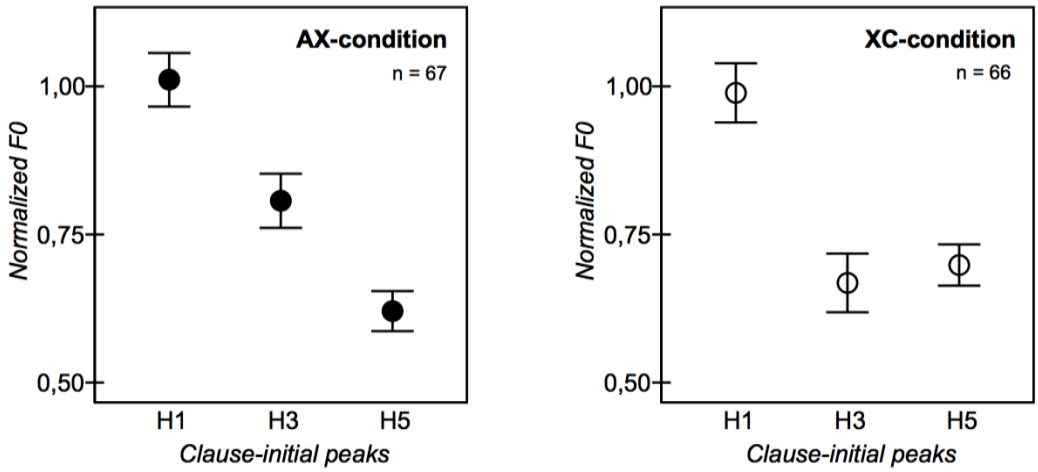


Fig. 4.3 Results of Féry & Truckenbrodt (2005): *but-and* (AX) condition in the left panel; *and-but* (XC) condition in the right panel

The solution proposed in Féry (2010) and Truckenbrodt & Féry (2015) for German was that each sentence, A, B and C, is an ι-phrase, each X (A and B or B and C) is an ι-phrase

and the entire coordinated sentence is also an ι-phrase.[13] The results in figure 4.3 are compatible with an embedded model of prosodic domains such as in figure 4.2. At each level of embedding there is one step of downstep. There are three such steps in AX and only two in XC. The reason for the absence of downstep on C in the XC condition is that both B and C are downstepped from A in a similar way. To express it in a different way, C is scaled to the entire constituent X and thus is blind to the level of B. Again, if prosody were non-recursive, we would expect the two conditions tested by Ladd (1990) and by Féry & Truckenbrodt (2005) to deliver identical results.

Other researchers assuming recursive prosodic structures in experimental data are Elfner (2013), Ishihara (2004), Myrberg (2010, 2013), Wagner (2005), Féry & Ishihara (2009, 2010) and Kentner & Féry (2013).

Kentner & Féry (2013) examined the phonetic correlates of embedded Φ-phrases for pitch and duration. They used German data like those illustrated in (41) and (42), thus a set of three names grouped in three ways and a set of four names grouped in six ways. They proposed that each grouping is an instance of a recursive Φ-phrase. Moreover, each name is included in its own Φ-phrase and each grouping of names, simplex or complex, is also a Φ-phrase. They obtained different F0 and duration results for all groupings, reflecting the recursive structure of each configuration of names, confirming that prosody is closely mapped to syntactic structure and that syntactic recursion is mirrored by prosodic recursion. A very interesting result of this study is that right-branching structures have a default prosody, while left-branching structures mark the end of the grouping in a different way. Thus the right-branching structures in (41)c and (42)b,d present downstep of high tones, similar to the baseline conditions in (41)a and (42)a. But the left-branching structures in (41)b and (42)c,e, have a marked prosody, with high boundary tones and long duration at the right edge of the groupings.

(41)	a. N1 or N2 or N3	(Nino or Willi or Mila)$_{\Phi}$
	b. (N1 and N2) or N3	((Nino and Willi)$_{\Phi}$ or Mila)$_{\Phi}$
	c. N1 or (N2 and N3)	(Nino or (Willi and Mila)$_{\Phi}$)$_{\Phi}$
(42)	a. N1 or N2 or N3 or N4	(Nino or Willi or Mila or Susi)$_{\Phi}$
	b. N1 or N2 or (N3 and N4)	(Nino or Willi or (Mila and Susi)$_{\Phi}$)$_{\Phi}$
	c. (N1 and N2) or N3 or N4	((Nino and Willi)$_{\Phi}$ or Mila or Susi)$_{\Phi}$
	d. N1 or (N2 or (N3 and N4))	(Nino or (Willi or (Mila and Susi)$_{\Phi}$)$_{\Phi}$)$_{\Phi}$
	e. ((N1 and N2) or N3) or N4	(((Nino and Willi)$_{\Phi}$ or Mila)$_{\Phi}$ or Susi)$_{\Phi}$
	f. (N1 and N2) or (N3 and N4)	((Nino and Willi)$_{\Phi}$ or (Mila and Susi)$_{\Phi}$)$_{\Phi}$

As for the analysis, Kentner & Féry (2013) used two principles regulating the different realizations, called 'Proximity' and 'Similarity'. These principles account for two effects of syntax on the prosodic realization of groupings, one having to do with the proximity of constituents inside of phrases (constituents in one grouping are prosodically proximate and those separated by a syntactic boundary are prosodically distant, see Watson & Gibson 2004 for a different implementation of this idea), and the other having to do with the levelling of constituents at the same syntactic level, see Wagner (2005) for levelling as a result of syntax. The effects of Proximity and Similarity are mirrored by phonetic correlates of different strengths.

[13] In Féry & Truckenbrodt (2005), a non-recursive solution was offered that is ignored here.

4.5.2 Match constraints and well-formedness constraints

The Match constraints in (43) are adapted from Selkirk (2011), see the collection of papers in the special edition of Phonology 32.1 edited by Selkirk & Lee (2015) for illustrations. These constraints assume a perfect isomorphism between morpho-syntax and prosody. A syntactic clause is matched by an ι-phrase, a syntactic phrase by a Φ-phrase and a simplex (monomorphemic) or complex lexical word by a ω-word.[14]

(43) Match Constraints

a. MATCH CLAUSE
   A clause in syntactic constituent structure must be matched by a corresponding ι-phrase in prosodic representation.

b. MATCH PHRASE
   A phrase in syntactic constituent structure must be matched by a corresponding Φ-phrase in prosodic representation.

c. MATCH WORD
   A word in syntactic constituent structure must be matched by a corresponding ω-word in prosodic representation.

The Match constraints provide the syntactic structure with an isomorphic prosodic structure, as sometimes envisaged in syntactic accounts of prosody (Chomsky & Halle 1968, Cinque 1993, Zubizarreta 1998, Kratzer & Selkirk 2007). According to these constraints, since syntactic constituents are embedded into each other, prosodic constituents must also be embedded into each other. At least this is the result of a straightforward application of the Match constraints. Notice though that since they are violable OT constraints, they can be dominated and violated. All syntax-prosody interface constraints potentially conflict with the so-called 'well-formedness constraints'. These latter constraints influence the form of the prosodic structure, and if ranked higher, can cancel the effects of the Match constraints. As an example, a pronominal object, even if it syntactically is a DP, and thus can potentially be a Φ-phrase, will generally not form one because it is prosodically too light and does not fulfil the minimality requirement for a Φ-phrase. Some well-formedness constraints can be understood as elements of the Strict Layer Hypothesis (SLH), but now expressed as a set of essentially violable universal tendencies that prosodies of all languages tend to fulfil (see also Feldhausen 2010, Ito & Mester 2013 and Myrberg 2013 for this view).

The first constraints of this kind in (44) reproduce the components of the SLH in the form of constraints. They were formulated as violable constraints by Selkirk (1996:443). NON-RECURSIVITY militates against recursion: a constituent of level n cannot dominate a constituent of the same level n (e.g. no foot dominates a foot). LAYEREDNESS posits that a prosodic constituent may not dominate another one belonging to a higher level (e.g. no syllable dominates a foot). HEADEDNESS requires that constituents have heads. EXHAUSTIVITY posits that no level is skipped (e.g. no ω-word immediately dominates a syllable).

[14] According to Selkirk (2011:452) Match(Clause,ι) is a family of two constraints. The first one is responsible for the "standard clause" (CP), which may be syntactically embedded, whether as a complement to a verbal or nominal head, or as a restrictive relative clause within determiner phrase, or in other positions. The second one, the "illocutionarv clause", is the highest syntactic projection of the sentence and it carries its illocutionary force, which determines its appropriateness in a discourse context: [Force^0 [illocutionary clause]]$_{\text{ForceP}}$. Such a clause may be embedded within the larger sentence. Its meaning does not contribute to the "at issue" meaning of the surrounding sentence (Potts 2005). Typical cases are non-restrictive relative clauses, parentheticals and appositions.

(44) SLH with constraints (where C_n = some prosodic category)

a. NON-RECURSIVITY: A prosodic constituent C_n does not dominate another constituent of the same level C_n.

b. LAYEREDNESS: A prosodic constituent C_m does not dominate a constituent of a higher level C_n, $n > m$.

c. HEADEDNESS: A constituent C_n dominates a constituent of the immediately lower level C_{n-1}. (A prosodic constituent has a head on the immediately higher level.)

d. EXHAUSTIVITY: No C_n immediately dominates C_{n-2}. (No prosodic constituent is skipped.)

Further constraints regulate the form and rhythm of prosodic constituents; see (45). CULMINATIVITY requires singularity of heads, see Hyman (2006). MINIMALBINARITY (MINBIN) requires that a prosodic category dominate at least two constituents.[15] Moreover, if a category dominates two categories, then preferably categories of the same kind, as posited by EQUALSISTERS (from Myrberg 2013:75). Notice that violation of EQUALSISTERS normally implies violation of EXHAUSTIVITY (but the reverse is not true).

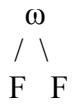
(45) a. CULMINATIVITY: Every prosodic constituent has a single head.

b. MINIMALBINARITY: A prosodic constituent C dominates at least two Cs. (A prosodically binary constituent is better balanced than a simple one.)

c. EQUALSISTERS: Sister nodes in prosodic structure are instantiations of the same prosodic category.

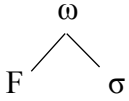
It is the nature of OT constraints that none of the well-formedness constraints are truly inviolable, even though some authors, like Selkirk (1996) assume that LAYEREDNESS and HEADEDNESS cannot be violated; see the next section for some examples of violations of LAYEREDNESS. In (46)b-f, configurations are shown which violate the constraints in (44) and (45).

(46) a. Immediate dominance, no recursivity: all constraints are fulfilled



b. Violation of MINBIN

c. Violation of EQUALSISTERS



d. Violation of LAYEREDNESS

[15] MAXBIN has also been proposed, limiting the size of a constituent to at most two constituents.

e. Violation of EXHAUSTIVITY

f. Violation of NON-RECURSIVITY

To sum up this section, Match constraints assume a perfect isomorphism between morpho-syntax and prosody. Match constraints are much too rigid to be usable all by themselves, and need well-formedness constraints to allow corrections in the mapping and to regulate the form of the prosodic constituents. However, note that this is the case of all models of the syntax-prosody interface.

4.5.3 Comparing ALIGN and MATCH

Having reviewed the predictions made by an essentially non-recursive account like ALIGNMENT and an essentially recursive account like MATCH, we can now compare them directly. Recall that ALIGN constraints favour iteration of prosodic domains by aligning Φ-phrases with the edges of syntactic phrases at their beginning or end, creating a sequence of non-overlapping prosodic domains, and in this way avoiding recursivity of prosodic domains. Even though ALIGNMENT does not impose non-recursivity, the models using it have assumed a high-ranking constraint against recursivity, an exception being a category of data in Kimatuumbi, as analysed by Truckenbrodt (1999). In other words, alignment partitions sentences into sequence of Φ-phrases or ι-phrases corresponding to the edges of syntactic phrases. As was shown in section 4.4, Truckenbrodt proposed a constraint WRAP-XP, whose function is to wrap entire constituents in single prosodic phrases, and to eliminate all possible Φ-phrases inside of the larger one. If WRAP is higher ranked than ALIGN, the effect of ALIGN is cancelled, and only one big phrase encompassing the highest syntactic phrase is the result (see Truckenbrodt 1999).

The main differences between MATCH and ALIGN constraints are that the former ones match prosodic constituents with syntactic constituents at both edges, and do not avoid embedded syntactic constituents. Since syntactic structures are regularly embedded into each other, prosodic structure is assumed to allow embedding as well. There is a simple way to reproduce the effect of the MATCH constraints with ALIGN, and this is to let the ALIGNMENT constraints be effective at both edges of the syntactic constituents. If ALIGN-R and ALIGN-L are ranked equally, then the effect of both ALIGN constraints is indistinguishable from MATCH. Moreover NON-RECURSIVITY should then be ranked so low that it has no effect. But obviously, in this case, MATCH constraints are more economical (or for that matter WRAP constraints without the additional effect of NON-RECURSIVITY). The question that arises is whether it is necessary to address the two edges by means of different constraints or not, and what is the role of NON-RECURSIVITY.

To compare the predictions of ALIGN with those of MATCH, consider extraposition of complement clauses in German (see Féry 2015) in (47). The version with extraposition in (47)a is much better than the version with an in-situ complement clause (47)b, which is barely grammatical, even though a verbal complement is pre-verbal in German, at least when the verb is final in the sentence, as is the case in (47).

(47) a. [(Sie hat niemandem t_i erzählt)$_\Phi$]$_\iota$ [dass sie an dem Tag spät nach Hause kam]$_\iota$
she has nobody told that she on that day late to home came
'She didn't tell anybody that she came home late on that day.'
b. *[?][Sie hat (niemandem$_\omega$, [dass sie spät nach Hause kam]$_\iota$ erzählt$_\omega$)$_\Phi$]$_\iota$

Figure 4.4 illustrates that due to MATCH CLAUSE, the complement clause is an ι-phrase, and, following MATCH PHRASE, the verb plus its argument *niemandem* 'nobody' form a Φ-phrase by virtue of being a VP. In (47)a, the main clause forms an ι-phrase as does the extraposed complement clause. In (47)b the Φ-phrase matched by the VP is interrupted by the ι-phrase of the complement clause. The result is an ill-formed structure violating the well-formedness constraint LAYEREDNESS, called a 'prosodic monster' in Féry (2015). This result is based on the assumption that a prosodic phrase of a certain type is always created on the argument and the verb, and that everything intervening between the argument and the verb also intervenes in prosody. In other words, argument and verb are prosodically related to each other across intervening material. Avoidance of prosodic monster can be considered the cause of extraposition, at least in this case, and this apparent from an account in terms of MATCH constraints.[16]

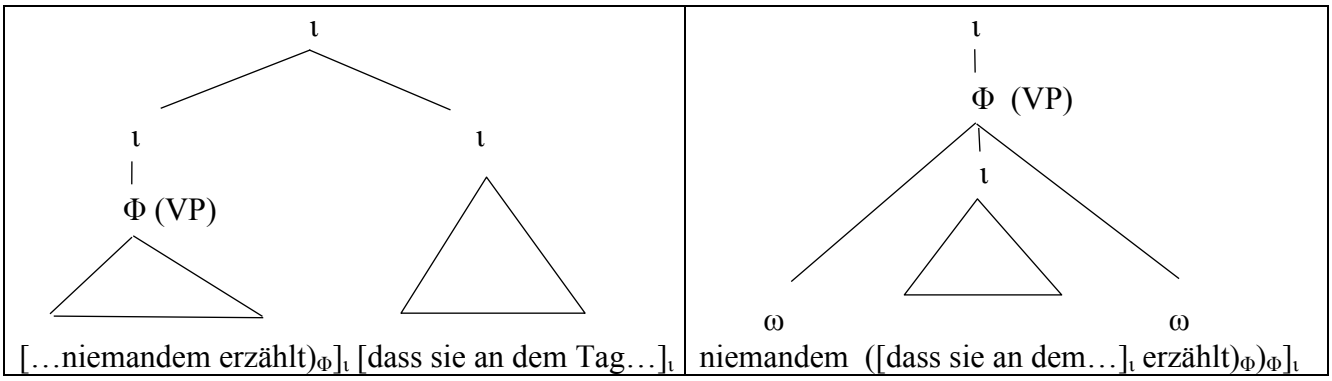


Fig. 4.4 Extraposition as avoidance of a prosodic monster (*dass*-complement)

An analysis using the ALIGN model, with both ALIGN and WRAP constraints as formulated in (48) cannot explain the preference for extraposition in this sentence. ALIGN takes any syntactic sequence and delivers a well-formed sequence of prosodic phrases. The result of extraposition is the same as before, a sequence of ι-phrases. In the case of in-situ position of the complement clause, an ALIGN model can only decide on where to place the prosodic boundaries.

(48) a. ALIGN-CP-R: The right edge of a CP must coincide with the right edge of an ι-phrase.
b. WRAP-CP: Each CP is contained in a single ι-phrase.

In the option shown in (49)a, ALIGN ranks higher than WRAP. This is the option offered by Truckenbrodt (2005:273), who suggested that there is a prosodic ι-phrase boundary at the right edge of an embedded clause, and no boundary at the left edge. In his own terms: "The right edge of embedded clauses, though not the left edge, leads to an intonation phrase boundary." The result is a first large ι-phrase consisting of the first part of the main clause plus the complement clause, and a second ι-phrase consisting only of the second part of the main clause, in this case just the final verb. The second option is the result of

[16] In Féry (2015), it is shown that relative clauses are also preferably extraposed, but that they are better tolerated in situ than complement clauses. The reason for this difference in acceptability is that a relative clause prefers to be adjacent to its antecedent.

the opposite ranking, when WRAP is higher-ranked than ALIGN. In this case, the whole sentence is wrapped in one big ι-phrase, as shown in (49)b. Both phrasing options are too restrictive and not preferred in German.[17]

(49) a. [Sie hat niemandem dass sie an dem Tag spät nach Hause kam]$_{\iota}$ [erzählt]$_{\iota}$.
b. [Sie hat niemandem, dass sie an dem Tag spät nach Hause kam, erzählt]$_{\iota}$.

It could be considered an advantage for ALIGNMENT that it allows that only one edge of a syntactic constituent corresponds to a prosodic edge. MATCH always assumes that both edges are addressed at the same time. However, as we saw all along this chapter, edges are asymmetric due to the branching direction of syntactic structure. A right-branching language does not mark its right edges as much as its left-edges, and the reverse is true for left-branching languages. The question is whether this is to be accounted for in the phonetic component or in the syntax-prosody interface. If it is just a predictable consequence of directionality of syntax, MATCH wins again, but if the asymmetry is pervasive and leads to mismatches between syntax and prosody, ALIGN may be better after all.

To sum up, MATCH and ALIGNMENT differ in their predictions, and in the role played by NON-RECURSIVITY. Recursivity is more or less obligatory in MATCH, but has been systematically avoided in models using ALIGNMENT. The same can be said for WRAP: WRAP has been introduced by Truckenbrodt (1995) as a constraint that eliminates embedded phrasing, but this requirement is not contained in WRAP itself. ALIGNMENT has the potential to imitate all phrasings accounted for by MATCH, if it addresses both edges of syntactic phrases at once. And it has the advantage that it can also address only one edge at a time. Mismatches like the ones illustrated for Xiamen in 4.3 above can be explained by ALIGNMENT but not by MATCH. What is needed, then, is an ALIGNMENT model, augmented by WRAP or MATCH, which is more flexible than has been intended in the past, allowing recursivity, and mitigated by well-formedness constraints like those introduced in 4.5.2.

4.6 Conclusion

The relationship between syntax and prosodic structure is essential to understand intonation. An optimality-theoretic account is probably the best we can do at the moment, because it mitigates the effects of a strict syntax-prosody mapping with well-formedness constraints. And it also allows an influence of prosody on syntax, be it only in choosing between alternative syntactic configurations, as has been shown with extraposition of complement clauses in German.

We have seen in this chapter that syntax-prosody interface can be approached with ALIGNMENT, i.e. constituents are aligned to each other at one edge only, or with matching of entire constituents at both edges at once. It has also been shown that ALIGNMENT favours non-recursivity of prosodic constituents in partitioning utterances into constituents of the same level. A constraint like WRAP, if it is accompanied by NON-RECURSIVITY, by contrast, only lets the larger prosodic constituents survive and

[17] Hartmann (2013) claims that the reason for the ungrammaticality of (49)b is that the verb *erzählt*, being only a ω-word, is too light after an ι-phrase. However, her proposal makes wrong predictions in other cases: It also excludes in-situ relative clauses, as well as other similar prosodic structures, like in (i).

(i) Anna erreichte, (nachdem sie drei Tage lang gefahren war)$_{\iota}$, (das Ziel)$_{\omega}$.
Anna reached after she three days long driven was, the destination.
'After having driven three days long, Anna reached the destination.'

eliminates all smaller ones. As a result, it is too coarse to account for the subtleties of prosodic constituency.

A promising avenue seems to be a departure from the prohibition on recursivity of prosodic constituents which has dominated the field for so long, and admitting recursivity as a regular feature of the syntax-prosody interface. Prosody is simpler than syntax because it uses fewer constituents, which are intrinsically layered and which obey well-formedness constraints. Embedding is more constrained in prosody as compared to syntax. The right relation seems to be one of homomorphism, one that maps the complex syntactic structure to a reduced set of constituents, preserving structure, or part of it, in doing so. If several syntactic structures are competing for the same content, prosodic well-formedness may bear on the decision as to which candidate is chosen, as we have seen with extraposition in German.

However, we have to bear in mind that languages may differ in the way they map syntax and prosody. Intonation languages like English and German may be recursive both in syntax and in prosody, while other languages may have recursion only in syntax, or may express it differently. Chapters 7 and 8 will discuss some aspects of this issue.

Annotated suggestions for further reading

Nespor & Vogel (1986) is the seminal book in which prosodic structure was given the first formal account.
Gussenhoven (1992) lays down the relation between syntax and prosody in English. In this paper, he examines a number of syntactic structures in relation to their accent pattern. He proposes that the prosodic phrasing goes hand in hand with metrical structure.
Cinque (1993) establishes a relation between embedding in syntax and prominence in prosody.
Selkirk (1986, 1995, 2011) presents different approaches to the syntax-prosody interface, illuminating the theoretical changes.
Truckenbrodt (1999, 2006) presents a simple and elegant model of the syntax-prosody interface in an OT framework.
The special issue of Phonology 32.1 (2015), edited by Selkirk & Lee and called *Constituent structure in sentence phonology* contains papers addressing syntax-prosody mapping. Most of them are coached in the Match model of syntax-prosody interface.

Discussion points
1. Take any sentence (an example: *White pelicans are flying through Bulgaria*) and erect a prosodic tree at all levels on it.

2. Discuss the necessity of adding prosodic domains.

(×)	(ι-phrase)
(×) (×)(×) (×)	(Φ-phrase)
(×) (×) (×) (×)(×) (×)(×)	(ω-word)
(×) (× .) (×) (×)(× .)(× .)(×) (×)(×)	(foot)

(1) Miss Martin called her uncle on the phone last week

3. In most work on the syntax-prosody interface, the syntax is given, and the prosody just maps to it. There is, however, an alternative conception in which syntax and prosody

work hand in hand and alternative syntactic structures are possible options. What advantages does each of the methods have?

4. Discuss the merits and drawbacks of the different theories of syntax-prosody interface introduced in the chapter.

Chapter 5
Models of intonation

5.1. Introduction

It was shown in chapter 1 that beside the syntax-prosody interface detailed in the preceding chapter, the tonal variation in English is shaped by the information structure and the pragmatics, and that in Mandarin, tonal contours are mostly the result of lexical tones. Because of this difference, a formal approach is needed which is capable to account for tone assignment in each language, and for the variation across languages. In English, the choice of tones in a specific case is a consequence of the kind of sentence, e.g. a declarative sentence, a wh-question, a polar question, an imperative, but the location of tones, e.g. to which specific syllables tones are associated, is dependent on the syntax and the metrical structure. It is thus important that the formal model of intonation take into account the influence of other parts of grammar. Just assigning tones or contours at different places of a sentence is not enough. The relation of tones to their grammatical origin is needed as well.

Consider the examples in (1). The sentence *Baxter will come back* can be interpreted as simple declarative sentence (1)a, an all-new sentence. In the right context, it can also be a question, as in (1)b, an order (1)c or a menace (1)c. Moreover the information structure can vary, as in (1)e. In all cases, the intonation may differ.

(1) a. Did you hear the news: Baxter will come back.
b. Is it true? Baxter will come back?
c. The boss gives us no choice: Baxter will come back.
d. Be prepared: Baxter will come back!
e. It is not Smith I am talking about: BAXTER will come back.

In this chapter, the *tone-sequence* model, i.e. the model of intonation that will be used in the remainder of the book, as well as a few other models are presented. The tone-sequence model describes how tones are assigned to prosodic domains and prosodic constituents, sometimes just called the *text*. This model has also been called 'autosegmental-metrical model' (AM) by Ladd (1996), a name that is widely used in the literature. The term autosegmental-metrical refers to the separation of tones and stress on different planes of the phonological representation. The metrical grid (or tree), which represents accents and prominence, and tones are combined with each other in the phonetic component. Because the metrical part of intonation is not obligatory in all languages, we will use this term only for the languages in which both tones and metrical structure play a role. Tones are present in all languages, and a prosodic structure is obligatory as well, but metrical structure may be absent in some languages. Tones do not always need prominence to be assigned.

The tone-sequence model as we know it today was first proposed by Pierrehumbert (1980) for English, who was herself inspired by the work of Gösta Bruce (1977) for Swedish, and Liberman's (1975) model for English. In her model, tones have their own grammar and tonal sequences have specific properties allowing them to be well formed in their prosodic domains. However, the tonal part of intonation is more than just a collection of well-formed tone sequences. Other components of the grammar of intonation are the tone inventory, tone scaling, tone alignment, tone transformations and variations, tone spreading, and tone simplifications, all aspects that will be addressed below. Moreover the tonal component of intonation can interact with other aspects of intonation such as metrical structure, duration and intensity.

Phonologically oriented researchers largely agree on an analysis of tones as abstract phonological objects, in the same way as features, segments and syllables are abstract. One of the main roles of abstract phonological objects is to account for distinctions in the linguistic system. The segmental feature [voice], for instance, does not include in its definition how exactly voicing is realized, i.e. by which actions of the vocal folds or of the laryngeal muscles controlling them. Rather this feature means that the segments carrying it could be [–voice] or that they could be unspecified for [voice]. If segments are distinguished by it, it is part of the feature inventory. Furthermore features may be supressed or added. [+voice] may arise on a segment through assimilation from one segment to an adjacent one. [+voice] can be deleted as the result of final devoicing or laryngeal neutralization. In other words, feature specifications form a system of oppositions in a structuralist spirit.

In the same way, if we choose to use tones as abstract phonological objects, we do not specify what it means in the phonetics for a tone to be high (H) or low (L): the opposition just accounts for the distinction between H and L. At the same time, phonological distinctions need to be grounded in the phonetics. We expect a high tone to be phonetically realized in the higher part of speakers' registers and low tones in the lower part, but how exactly these tones differ from each other and how they are associated with the text is not specified by abstract H and L. According to Pierrehumbert, the phonological representation of the tonal structure first has to be translated into phonetic values, and this translation takes place in a different part of the model.

The tone-sequence model introduced in section 5.2 for English largely uses the concepts introduced by Pierrehumbert (1980) for this language, the best-studied language so far. At the same time, some basic phenomena of the tonal structure are introduced as well. Section 5.3 introduces some of the main properties of the autosegmental representation for African tone languages. Section 5.4 introduces ToBI (Tones and Break Indices), an annotation method derived from the tone-sequence model. Section 5.5 describes how the tone-sequence model can be accounted for in an optimality-theoretic (OT) framework. In this section, Gussenhoven's (2004) OT approach is summed up. Besides being a derivational approach to tones, in the same ways as the autosegmental model of intonation is derivational, the OT model of intonation is one of the few approaches so far allowing a representation of tones in which the underlying tones differ from the surface tones, and for this reason, it is an important approach.

The later parts of this chapter briefly address alternative models. Section 5.6 is dedicated to the British school and section 5.7 to the parallel encoding and target approximation model (PENTA) of Xu (2005) and Xu & Xu (2005).

5.2 Tone-sequence model of intonation

As explained in chapter 2, melodies consist of changes in F0. But the melodies by themselves do not tell us much about the message conveyed. It is the tonal structure associated with the text, as well as the temporal alignment between the two – the t*ext-to-tune* component of the model – that allows the hearers to reconstruct the prosodic structure underlying the phonetic signal produced by the speaker, and ultimately enables them to reconstruct the syntactic and pragmatico-semantic component of the tonal structure. Nevertheless, in the tone-sequence model of intonation, tones and tonal contours have their own grammar. Tones are grammatical atoms, associated with prosodic constituents. In English, they need a metrified and prosodified text, and the association between prosodic constituents and text takes place in a specialized component of grammar, as was shown in chapters 3 and 4. According to Bruce (1977), Gårding (1983), Pierrehumbert (1980), and many researchers after them, each significant turning point in

the F0 plot is evidence for the presence of a tone that defines this turning point. In Pierrehumbert's original model, tones in English come in three variants: pitch accents, phrase accents and boundary tones. This inventory was modified in Beckman & Pierrehumbert (1986) and Pierrehumbert & Beckman (1988): phrase accents became boundary tones of *intermediate phrases* – roughly equivalent to the Φ-phrases used in this book.

The inventory of Pierrehumbert for English is shown in (2)a. There are two monotonal and five bitonal pitch accents. Bitonal tones always consist of a starred tone plus another non-starred tone, which can precede the starred tone (a leading tone) or follow it (a trailing tone). We will see below that the starred tone is associated with a metrically strong syllable.[1]

(2) a. Pitch accents: H*, L*, H*+L$^-$, H$^-$+L*, L*+H$^-$, L$^-$+H* and H*+H$^-$
b. Phrase accents: H$^-$ and L$^-$
c. Boundary tones of ι-phrases: H% and L%

Not all authors use + between the two tones and some dispense with the additional diacritic $^-$, thus H*+L$^-$ can be written H*L, a convention adopted in the following, except in citations. Moreover, boundary tones can be identified by a subscript, Φ or ι, to indicate whether they are assigned to Φ-phrases or to ι-phrases, thus H_Φ and L_Φ and H_ι and L_ι, a proposal adapted from Hayes & Lahiri (1991). Pierrehumbert proposed that the possible tone sequences appearing in an English ι-phrase are best accounted for by a finite state grammar (1980:29), as shown in figure 5.1, reproduced in her original notation.[2] An intonation phrase (ι-phrase) starts with an optional boundary tone. It is followed by one or more pitch accents (see the arrow leading back) and it ends with an obligatory phrasal tone (or phrase accent) and an obligatory boundary tone.

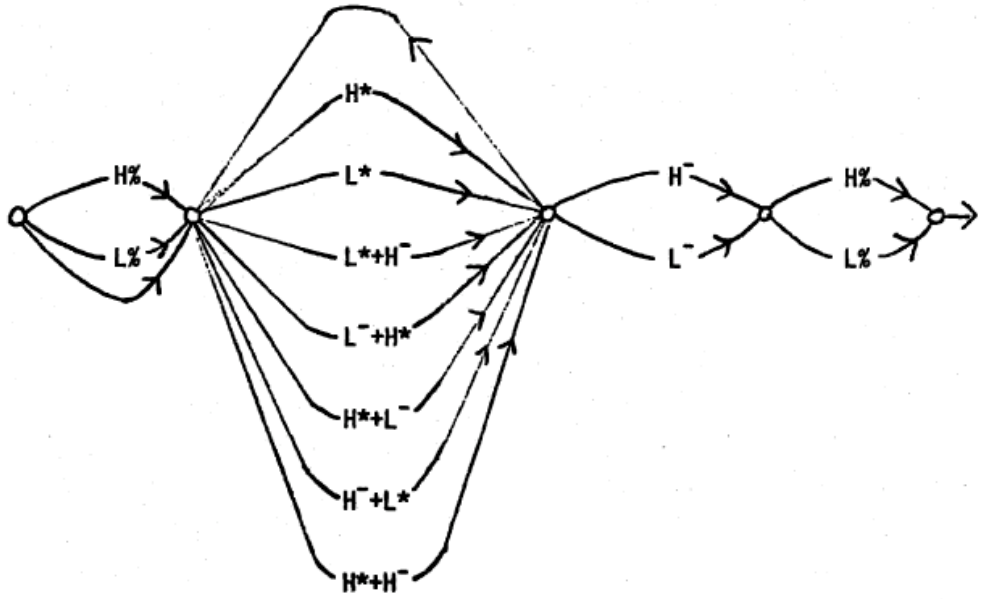


Fig. 5.1 Finite state grammar of tone sequences (from Pierrehumbert 1980:29)

Pierrehumbert's finite state grammar can be written more compactly if it is independently posited which tones are part of the inventory. In (3), T stands for H or L. It is proposed that leading tones and trailing tones are independent of each other and that they can both

[1] H*+ H$^-$, which generates a F0 plateau, was eliminated from the inventory of Beckman & Pierrehumbert (1986).
[2] In this section, Pierrehumbert's original examples are used whenever possible.

be optionally present. There is an optional initial boundary tone and the ι-phrase may consist of more than one Φ-phrase, each with its own boundary tone. There may be between 1 and n Φ-phrases in an ι-phrase, as indicated by the sub- and superscripts 1 and n after the curly brackets. Moreover, there may be more than one pitch accent in each Φ-phrase. As a result, only one pitch accent and the final boundary tones are obligatory in an ι-phrase, everything else is optional.

(3) (T_ι) $\{\{(T)T^*(T)\}_1^n$ T_Φ $\}_1^n$ T_ι
Boundary tone Pitch accent Phrasal tone Boundary tone

In the following sub-sections, the basics of the tone-sequence model are introduced by elucidating the following aspects of the model: discrete tones (5.2.1), text-tone association (5.2.2), continuous phonetic melody (5.2.3) and downtrends (5.2.4).

5.2.1 Discrete tones

Tones are targets with high and low values in the pitch range of speakers. In general, the F0 of high tones is higher than the F0 of low tones although there are cases in which a high tone at the end of an ι-phrase is lower than a low tone at the beginning of the same ι-phrase (see figures 5.4 and 5.5a-b for examples). Most of the tones can be interpreted straightforwardly, as for instance in bitonal tone sequences: a low tone followed by a high tone iconically stands for a rising contour. But some of the tone sequences have different functions, some of which will be examined below.

Let us start by looking at the tone sequences in figures 5.2 and 5.3 from Pierrehumbert (1980). The starred tones L* are associated with the stressed syllables of *Mánitowoc* and *bówling* respectively.[3] Boundary tones are high, and phrase accents between L* and H% define a high (H^-) or a low (L^-) plateau depending on their value. In figure 5.3, the utterance begins high, and the initial high target is analysed as a high boundary tone.

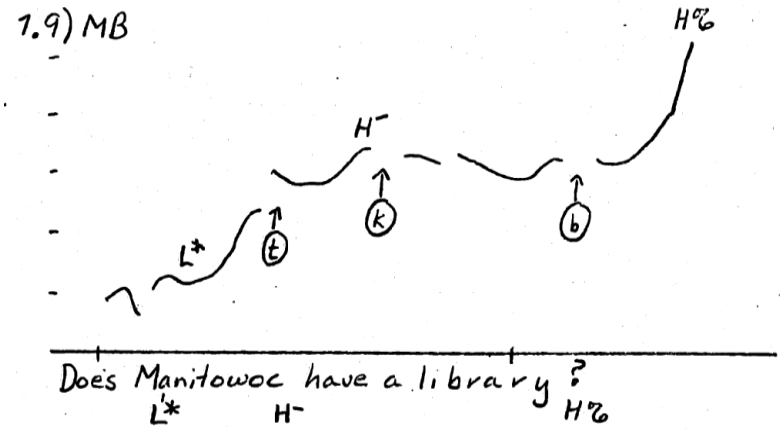


Fig. 5.2 *Does MANITOWOC have a library?*: L* H^- H% (from Pierrehumbert 1980:265)

[3] In the examples in this chapter, the accent structure is given in the legend. It is important to note that they are not always the default accent structure associated with an all-new utterance (see chapter 6).

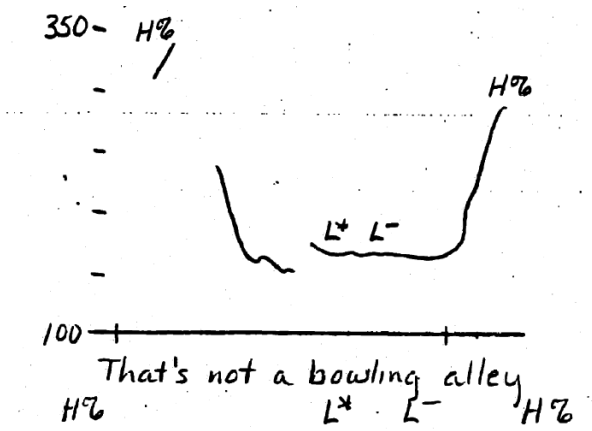


Fig. 5.3 *That's not a BOWLING alley*: H% L* L⁻ H% (from Pierrehumbert 1980:385)

Compare next the two pitch tracks of figure 5.4. The first one is again straightforward. There is a high pitch accent H* on the stressed syllable of the word *Mánny*, and this high tone is followed by two low tones, a phrase accent and a boundary tone, defining a steep fall on the second syllable of *Mánny*.

The second pitch track shows a vocative contour. It also has a peak on the stressed syllable, but this time the falling contour is accounted for by a bitonal tone H*+L⁻. H*+L⁻ L⁻ is not intrinsically different from H* L⁻. Both consist of a high peak followed by a falling contour. The reason for the different analysis has to do with the end of the phrase. In the first case, the fall reaches the bottom of the speaker's register, and is typical for a declarative intonation. In the case of the vocative, the lowest value only reaches mid-register. In Pierrehumbert's model, the latter pattern is accounted for by a final sequence H⁻ L%, a high phrase accent causing upstep on the following tone, plus a low but upstepped boundary tone. H⁻ is responsible for the fact that L% remains at mid-register instead of going all the way down in the same way as in the first pitch track.

Moreover, bitonal tones like H*+L⁻, containing a low tone, account for downstep. The trailing tone L⁻ has an effect on the value of the following H: H is lower than a similar H preceding L⁻. Downstep characterizes the fact that in a sequence of two H tones, the second one is lower than the first one. In Pierrehumbert's (1980) account, H*+L⁻ can account for downstep, but H* cannot. This is because in her analysis, downstep of a high tone relatively to a preceding high tone is triggered by a bitonal tone containing a low tone.

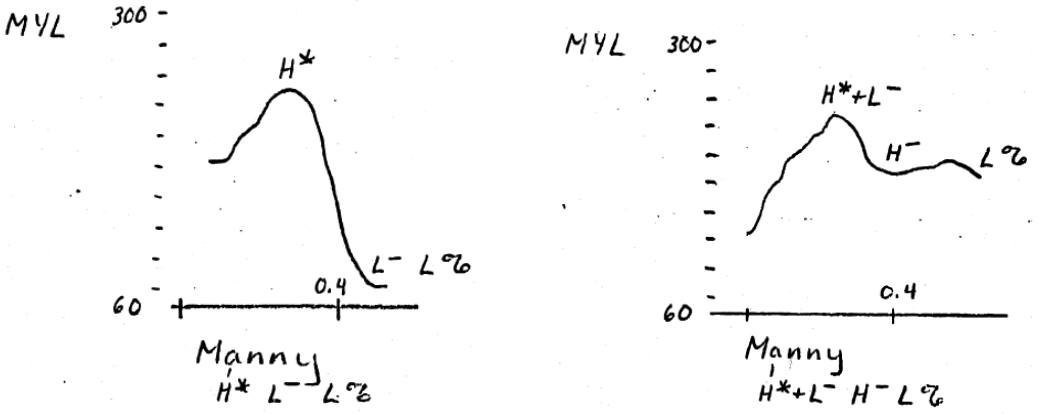


Fig. 5.4 *MANNY* in two tone sequences, H* L⁻ L% and H*+L⁻ H⁻ L% (from Pierrehumbert 1980:273)

In other words, the L between the two H tones is necessary to trigger downstep. But not all L tones can trigger downstep. It must be part of a bitonal pitch accent, like H*+L⁻, H⁻+L*, L⁻+H* or L*+H⁻, i.e. in any combination of a bitonal tone containing an H and an L. Figure 5.5 illustrates different patterns of downstep all triggered by an L tone between two high tones, the second one being downstepped relatively to the first one.

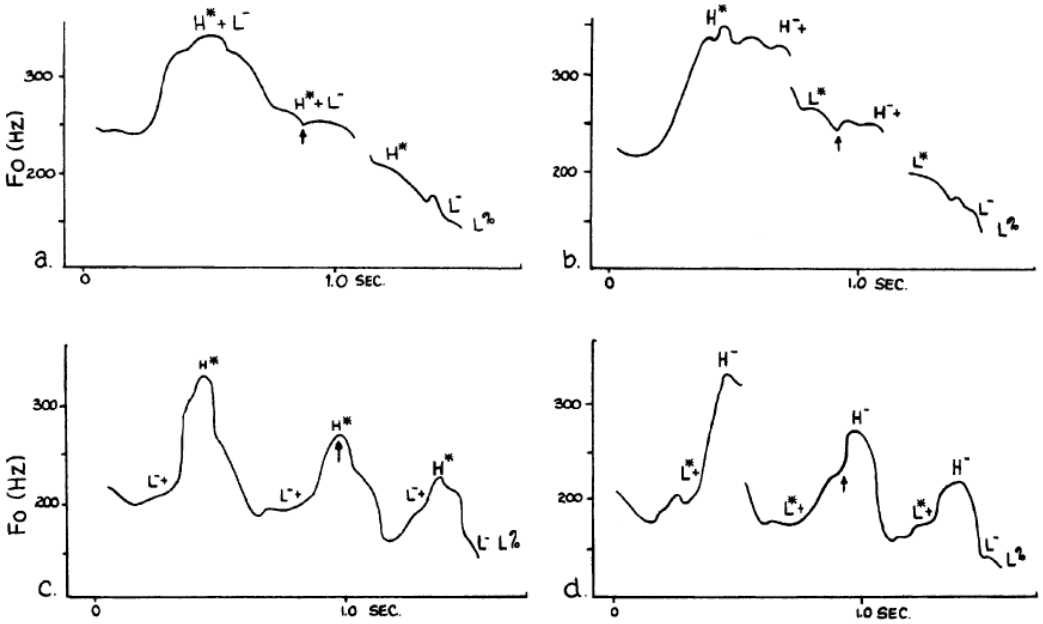


Fig. 5.5 From Ladd (1983a) reproducing four pitch tracks from Pierrehumbert (1980): *There are* MANY INTERMEDIATE LEVELS in four variants with downstep. The arrows show the location of [d] of *intermediate*.

In Beckman & Pierrehumbert (1986), downstep was generalized to additional contexts. It applied to all high tones after any bitonal pitch accent, also those in which the L part is not between the two H tones. The illustration in figure 5.6 shows that the leading tone H of H+L* on *Marianna* can be downstepped relatively to an immediately preceding H, itself part of a bitonal L*+H on *really*. Here, the two H tones are not separated by an L tone.[4]

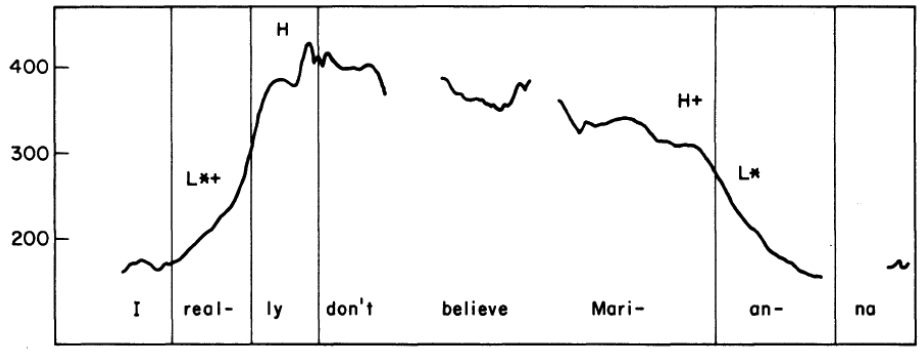


Fig. 5.6 *I* REALLY *don't believe* MARIANNA: downstep of H after L*+H (Beckman & Pierrehumbert 1986:276)

To sum up, not only tones have a role to play in melodic contours, but also sequences of tones. This remark also holds in Pierrehumbert & Hirschberg (1990), where it is shown

[4] Notice also that in Pierrehumbert & Hirschberg (1990), H*+L⁻ was simplified to H*+L.

that meanings of intonation arise most clearly in sequences of tones, in a compositional way (see chapter 6 for more on the meaning of tones). We will return to downstep and other downtrend tendencies in section 5.2.4.

5.2.2 Text-tone association

Not every syllable is associated with a tone, and the question arises as to how tones end up being associated with specific syllables in the text, and which syllables end up being associated with a tone (and sometimes more than only one tone). As an answer to these questions, Pierrehumbert formulated rules for text-to-tune association, which are paraphrased in (4).

(4) Phonological rules for text-to-tune association
a. Pitch accents associate with (the strongest) stressed syllables within their prosodic domain.
b. Boundary tones associate with the boundary of the prosodic domain for which they are diacritically marked.

As an application of (4)a, a pitch accent such as H* or L* seeks the stressed syllable of a ω-word, and the starred tone associates with the stressed syllable, as in (5); see also the examples in chapter 3. The primarily stressed syllable has the highest column of grid positions in the word (it is the head of the ω-word). If there is a secondary accent (the head of a foot), it is also indicated in (5). This is the metrical part of the autosegmental-metrical (AM) model of intonation.

		×				×				×				×	
		×		×		×				×				×	
(5)	a.	(Mary)$_{\omega}$		b. ((Isa)$_{F}$(belle)$_{F}$)$_{\omega}$?				c. ((Ale)$_{F}$ (xandra)$_{F}$)$_{\omega}$				d. ((Aber)$_{F}$(nathy)$_{F}$)$_{\omega}$?			
		| |		| \			|	|			| |			
		H*L		L*H				H*	L			L*H			

The association of starred tones with metrically prominent positions may take place at different levels of the prosodic hierarchy, this will be the subject of chapter 7. In English, the association between prominent metrical positions and tones does not take place at the level of the ω-word, but only later, at the Φ-phrase or the ι-phrase, when syntax and pragmatics decide on the tone sequence.[5] If present, leading tones and trailing tones are associated with adjacent syllables or if there is no adjacent syllable, to the same syllable as the one carrying the starred tone, like the trailing tone H in *Isabelle* in (5)b. Pierrehumbert (1980:40) suggested that the trailing tone in an L*H pitch accent follows the starred tone at 'a given time interval.' Subsequent work (see Arvaniti, Ladd, & Mennen 2000 and Dilley, Ladd, & Schepman 2005) has suggested that this claim may be oversimplified. Moreover, there are differences between languages in the way tones are aligned with the text. We return to alignment of tones in chapter 7.

As for (4)b, the boundary tones are primarily associated with boundaries, but phonetically they are also realized on syllables. Since (6) is not an example from Pierrehumbert, the notation used in the remainder of this book is used here, which also serves to illustrate how the tones introduced by Pierrehumbert can be translated into the prosodic approach used in this book. The starred tones are associated with metrically prominent syllables; this remains unchanged. The phrase accents of Pierrehumbert are regular phrasal tones H_{Φ} and L_{Φ}, as proposed in Pierrehumbert & Hirschberg (1990), and

[5] Other languages, like Swedish, Japanese or Mandarin do associate tones at the level of the ω-word.

are associated with boundaries of Φ-phrases. Finally, H_ι and L_ι are associated with the final boundary of the ι-phrase. In (6), as well as in many other examples, L_Φ and L_ι are both associated with the last syllable of the sentence. Bitonal tones are replaced by a sequence of a pitch accent plus a boundary tone in most cases, although true bitonal tones also exist, thus bitonal tones in which there is a starred tone and another tone that is not a boundary tone. Part of the scaling difference between the tones is explained by the reference lines of prosodic domains; see section 5.2.4 for more on this.

								×			ι-phrase
			×					×			Φ-phrase
	×		×		×			×			foot/ω-word
	×	×	×	×	×	×	×	×			syllable
(6)	((Hum	pty$_\omega$	Dum	pty$_\omega$)$_\Phi$	(sat$_\omega$	on	a	wall$_\omega$	)$_\Phi$	)$_\iota$	
	|		|	|	|			|	|	|	
	H*		H*	L_Φ	H*			H*	L_Φ	L_ι	

5.2.3 Continuous phonetic melody

How does the melody of the sentence arise given that not all syllables are assigned tones by the rules in (4), and given that all syllables participate in the melody of a sentence, also those which do not carry a phonological tone? The problem can be paraphrased as follows: once the tones are associated with syllables, they still have to be interpreted by the phonetics to deliver a continuous melody.

According to Pierrehumbert, there are two kinds of rules responsible for translating the phonological representation into a phonetic representation. The first ones assign concrete values to tones, and the second ones determine the contour between one target and the following one.

As for the first kind of rules, tones are best understood as targets to which the voice is directed as best it can be: speakers target specific values, calculating from preceding specific values (see also Pike 1945, as well as Trager & Smith 1951 for the notion of tonal targets). High and low tones need a context to be perceived as such. A tone is only high relative to a reference line defined by the speaker's register, or relative to another preceding tone (or to a preceding sequence of tones). In the same way, intensity or duration can only be interpreted comparatively. Context-sensitive rules assign a tone its F0 value according to the following three parameters (see also Liberman & Pierrehumbert 1984):

a) Its relation to a bottom line (or baseline), i.e. the deepest value in the speaker's register, which is fairly invariant for each speaker.
b) The degree of prominence that the speaker assigns to the utterance or to a local accent: the F0 range increases with emphasis, as shown in figure 2.3 in chapter 2.
c) Its relation to the preceding tones. According to Pierrehumbert, every tone is calculated on the basis of the preceding tone, with the help of the so-called *tone mapping rules*, which calculate the phonetic value of each tone relative to the preceding one.

The second kind of rules responsible for the text-to-tune association implement the F0 contour between two tones. The melody between tones has two possible sources: interpolation rules or tone spreading. In interpolation, the melodic contour between two syllables specified for tones is just smoothly going from the first tonal value to the second one, across all syllables in-between that are unspecified for tone, i.e. not associated with any tone. In figure 5.7 (the first contour of figure 5.5, but this time in the original version

of Pierrehumbert) the syllables *inter* of *many intermediate* are unspecified for tone and interpolation takes place. The contour interpolates between syllables associated with H*L accents and smoothly falling contours are the result. Interpolation can also result in rising contours, or flat ones. There may be additional assumptions. For instance, Pierrehumbert predicted that there is a small dip between two H tones, but not between two L tones. In this latter case, the voice remains flat.

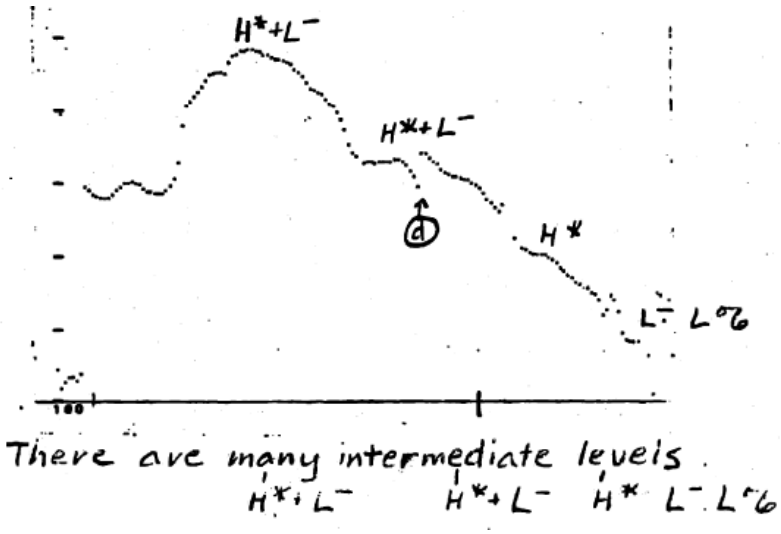


Fig. 5.7 *There are* MANY INTERMEDIATE LEVELS: interpolation between L⁻ and H* and between H* and L⁻ (from Pierrehumbert 1980:329)

Tone spreading is the second method for assigning a melody to syllables unspecified for tones. But differently from interpolation, spreading produces tonally specified syllables. Only H⁻ or L⁻ (leading or trailing tones and phrase accents) can spread in Pierrehumbert's (1980) model and this is the reason why these tones share a diacritic. Starred and boundary tones do not spread. Spreading is illustrated in figures 5.2 and 5.8. A specific tone is associated not only with the syllable from which it originates, but to all following syllables up to the next specified tone. All syllables sharing the same tone are realized on the same level. In figure 5.2, H⁻ spreads until the final H%, forming a plateau before the F0 rises again for the high boundary tone. In figure 5.8, all syllables of *-mon bread was palata(ble)* are associated with the phrase accent L⁻ (until the last syllable of *palatable*, which is associated with the high boundary tone). The result is a low and flat contour on the syllables sharing a low tone.

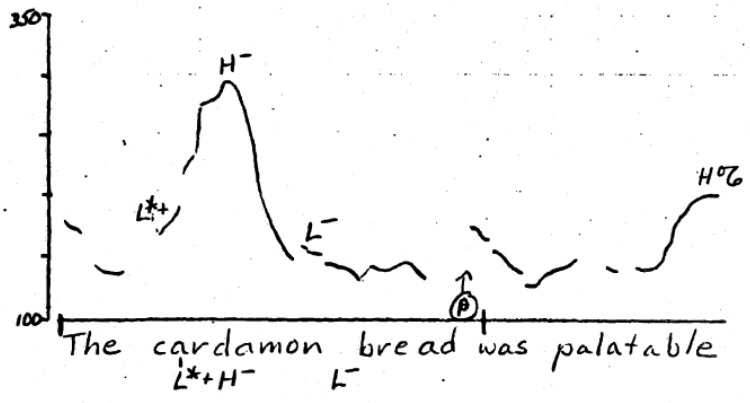


Fig. 5.8 *The* CARDAMON *bread was palatable*: spreading of L⁻ (from Pierrehumbert 1980:371)

To sum up so far, the melody of an utterance is determined by the concatenation of the individual tones, as well as the way they interpolate and spread to define the melody between the specified tones. In short, the global contour (tune) is a product of the local specifications and the implementation rules. For instance, if in a global tune the whole contour is falling, this may be due to the local specifications of the single tones that are downstepped relative to each other, or because there is interpolation between a high and a low tone. The assumption that all melodies are the results of tonal decomposition into single tones is central in Pierrehumbert's phonological model: the tones are morphemic, but not the melody. As a result, differences in the configurations of the intonation pattern are primarily the consequence of the single tones and the way they are combined.

Melodies are constructed through the concatenation of these tones, and the meaning of intonation is also compositional. The choice of the tones as high or low is dependent on the prosodic structure mapped to the syntax as demonstrated in the preceding chapter, but also on the semantics and pragmatics (see Pierrehumbert & Hirschberg 1990, Hobbs 1990, Bartels 1997, 1999, Gunlogson 2003 and chapter 6).

5.2.4 Downtrends, reset and upstep

Downtrends is a cover term for the F0 tendencies to decrease over the course of an utterance. We distinguish between declination, downstep (or catathesis) and final lowering.

Declination is a continuous decline of F0. It has the utterance as its domain and can be understood as a physiological process. It is usually thought to be the result of falling subglottal pressure, causing a concomitant gradual fall in F0. It is a natural process associated with the relaxation of the muscles during expiration. As the air flows out during speech, there remains less and less air in the lungs and trachea and the subglottal pressure is bound to decrease. The result is a slower vocal fold vibration rate, and a lower F0. Pierrehumbert (1980) visualizes declination as a downward tilt and narrowing on a graph on which the values predicted by the phonology-phonetics mapping are plotted, thus something like the illustration in figure 5.9. Thus, an H value realized at 1.5 units above the speaker's baseline will have a higher absolute value earlier in the sentence.

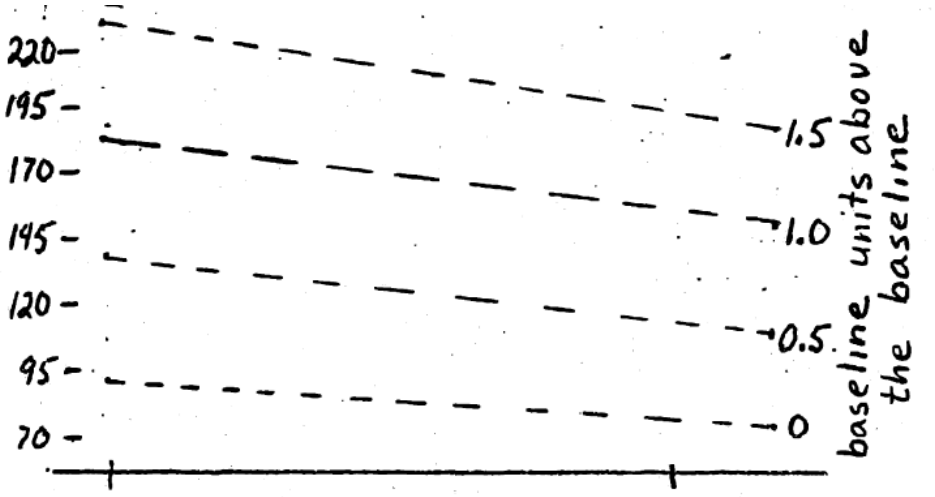


Fig. 5.9 Declination as downward tilt and narrowing on a graph (Pierrehumbert 1980:270)

The need to account for more downtrend tendencies than just declination in a language like English is illustrated by the following observation. In a declarative proposition containing two pitch accents, the second one is lower than the first one if both accents are intended to be roughly equivalent in strength; see the first example in figure 5.10.

Crucially in the second example in figure 5.10, the second accent is felt to be more prominent than the first one, as an instance of narrow focus (see chapter 6 for focus). In this case, upstep has taken place. This shows that the choice of downstepping a pitch accent, or any high tone, is within the speaker's control.

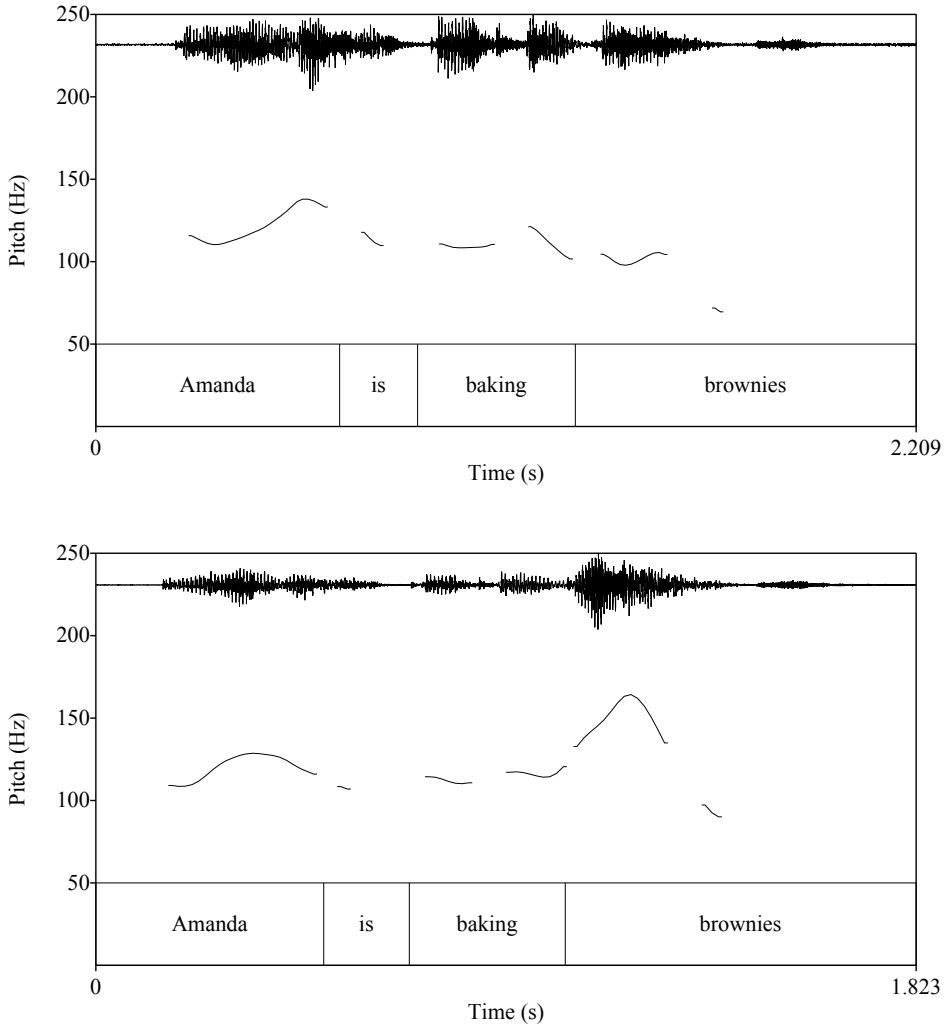


Fig. 5.10 Downstep and upstep between *AMANda* and *BROWnies*

Pierrehumbert's treatment of downstep has a phonological and a phonetic side. Let us first turn to the phonetic side, since this is the easier and uncontroversial part. Liberman & Pierrumbert (1984:186) proposed a model of downstep with a decaying exponential based on the mean F0 of the five peaks of twenty renderings of a list of two to five berry names by four speakers; an example for five names appears in figure 5.11.

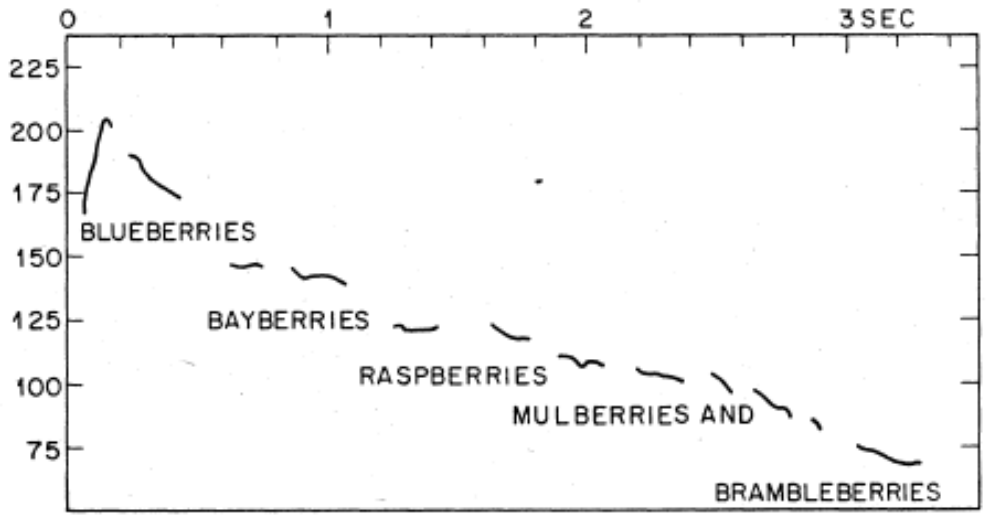


Fig. 5.11 List intonation (from Liberman & Pierrumbert 1984:171)

The formula they use is in (7). P_i is a value in Hz at a certain point in time, while d is a downstep factor between zero and 1. Every peak is the result of a multiplication of the F0 of the previous peak by d, which causes each downstep to be smaller than the one before. To express the fact that the steps do not go down beyond a minimum value, a constant r is included to represent the reference line. For instance, in a succession of peaks (or pitch values) P1 to P5, if r = 75, P1 = 200 Hz and d = 0.5, then P2 = ((200 − 75) x 0.5) + 75 = 137.5 Hz, P3 = 106.25 Hz, P4 = 90.6 Hz and P5 = 82.8 Hz. A much used value for d is 0.7.

(7) $P_i+1 = (d \times (P_i - r)) + r$

In the same article, Liberman & Pierrehumbert (1984:186) showed that the last high tone in a series of downstepped high tones is lower than predicted by the algorithm calculating the non-final downstepped values in (7). This is visible in figure 5.11 and is further illustrated in figure 5.12. They called this effect 'final lowering'. If P5 is the last peak, its value is not 82.8 Hz, but it is lower, closer to 75 Hz, the reference line.[6] There are several explanations for final lowering. The most straightforward one is that the combined effect of different low tones at the end of a falling ι-phrase, e.g. of the two adjacent L tones in H*L and $L_ι$, lets the low value of the final L go lower than expected from the formula in (7). Truckenbrodt (2004) proposed an alternative explanation for German: high tones followed by downstep are realized higher than high tones not followed by downstep: a rule of raising before downstep does not apply in the final position, since no downstepped tone is following. The ! diacritics in figure 5.12 indicate that the high tones are subject to downstep.

[6] Final lowering has been documented by Laniran (1992) for Yoruba, by Prieto, Shih, & Nibert (1996) for Mexican Spanish and by Féry & Kügler (2008) for German, among others.

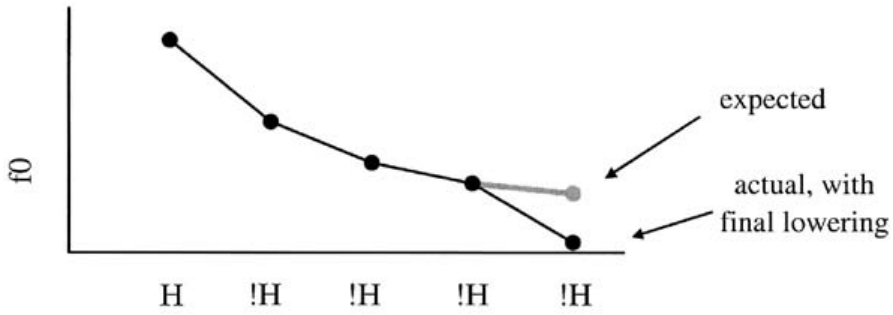


Fig. 5.12 Final lowering (from Truckenbrodt 2004:314, adapted from Liberman & Pierrehumbert 1984:187)

As for the phonological side, it was shown around figure 5.5 above that Pierrehumbert (1980) used a particular configuration of tones for predicting downstep in phonology: an H tone is downstepped relatively to a preceding H if they are separated by an L tone. Her proposal was very much influenced by how Africanists described downstep in African languages; see section 5.3 for more on this issue.

Ladd (1983a,b, 1996) proposed that downstep is morphemic in English, with a meaning like finality or completeness, and that this meaning is orthogonal to the meaning of other intonational choices. His view is that Pierrehumbert's suggestion to account for downstep with a specific sequence of tones "obscures the fact that the sequence of tones has another function, independent of downstep" (Ladd 1996:90-91). In other words, H*+L is just a falling tone, which can be downstepped or not. He proposes that "downstep must be treated as an independently selected 'feature' of each accent that can be present or absent independently of the choice of the accent itself. For example we might have two H* accents in sequence, either with downstep or without it." He proposes the writing !H* for a downstepped high tone, following the Africanist tradition; see figures 5.12 and 5.13 for this notation. Ladd's proposal that downstep is a feature affecting individual accents has been adopted in ToBI (see section 5.4).

A third way to account for downstep, which is adopted here, originally comes from Clements (1979), who analysed downstep in terms of a phonetic register; see also Ladd (1986, 1988, 1990, 1992), Odden (1995), van den Berg, Gussenhoven, & Rietveld (1992), Truckenbrodt (2002), Féry & Truckenbrodt (2005) and Féry & Ishihara (2009, 2010) for similar views. Register is the interval on the vertical F0 scale used in a stretch of discourse, with some H tones fixed to the top of the interval, and some L tones fixed to the bottom of the interval. In this model, downstep is analysed as a narrowing of the register interval in successive domains, entailing lowering of the H tones fixed to the top of the register. Notice that this model of downstep is not incompatible with Pierrehumbert's conception of declination illustrated in figure 5.9. It may well be that downstep implies not only gradual narrowing of the register (declination), but an additional categorical narrowing of the register used in a specific stretch of discourse (downstep). In van den Berg et al.'s (1992) version, downstep is a property of entire prosodic phrases, although it is primarily implemented on the H tones. The model allows us to elegantly account for embedding of downstep in smaller domains within downstep in higher domains ('wheels within wheels', as van den Berg et al. 1992 call this effect), recall also the discussion around figure 4.2 in chapter 4 form Féry & Truckenbrodt (2005). In Figure 5.13 from van den Berg et al. (1992), there are two large prosodic domains. The second one is downstepped relative to the first one: the first high tone of the second domain is lower than the first high tone of the first domain. This is indicated with "!(". Moreover the high tones in each domain are in a downstep relation to each other. The first high tone of the second domain is upstepped relative to the last high tone of the

first domain. In sum, different downtrend tendencies may interact to deliver a more complex pattern than simple declination or simple downstep would predict.[7]

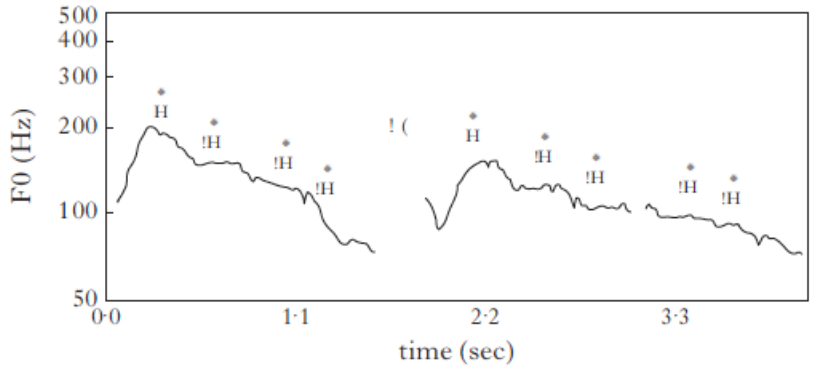


Fig. 5.13 Partial reset: van den Berg et al (1992)

I propose that all sequences of high tones are in a default downstep relation with each other when they are in a sequence of domains of the same kind. The black dots which are at the beginning of Φ-phrases are in a downstep relation with each other. Inside of each Φ-phrase, the high tones are also in a downstep relation with each other. The result is downstep within downstep. The red dot shows the effect of final lowering.

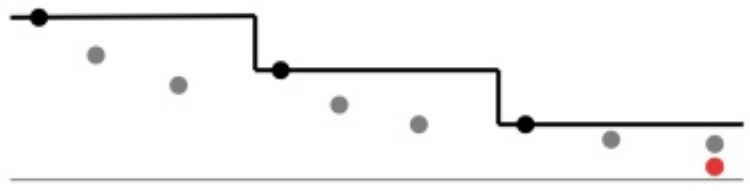


Fig. 5.14 Downstep within downstep and final lowering

Reset refers to the return of the F0 to a high level after a prosodic domain of the size of an ι-phrase ends and a new one begins, as illustrated in figure 5.15.

Reset can also be partial (Ladd 1988, van den Berg et al. 1992, Laniran & Clements 2003) as illustrated in figures 5.13, 5.14, and 5.15. If each domain is a prosodic domain with default downstep, the passage from one domain to the next one implies a partial reset of the pitch range, relative to the last grey dot of the preceding domain. Consider the black dots in figures 5.14 and 5.15, placed at the beginning of Φ-phrases. Reset between the first and the second black dot is partial because the second dot is higher than the preceding high tone (in grey) but lower than the preceding black dot.

A complete reset represents a complete return to the utterance-initial F0 height, as illustrated in the third black dot in figure 5.15.

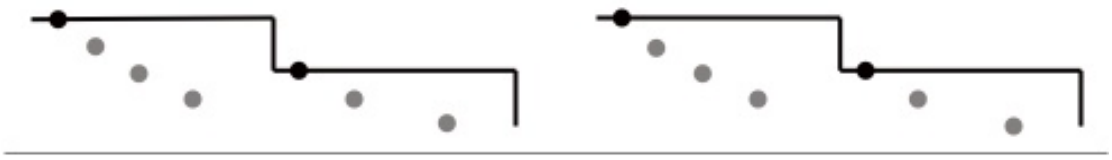


Fig. 5.15 Reset

[7] Beckman & Pierrehumbert (1986:299) claimed that the domain of downstep is the intermediate phrase.

The notion of upstep according to Pierrehumbert (1980) was illustrated in figure 5.4 with the vocative contour on *Manny*: L% was raised due to the preceding H¯. The effect of H¯ on H% is similar: H% is raised (compare figure 5.2). But notice that the term *upstep* also has another meaning. It refers to the raising of a pitch accent, as illustrated in figure 5.10b.

Still a different use of this term has been proposed by Truckenbrodt (2002) for German. In this language, the last H of an ι-phrase is sometimes much higher than an unmarked downstep pattern would lead us to expect. Truckenbrodt (2002) found this effect at the final position of a medial ι-phrase in a sequence of ι-phrases and he proposed analysing this phenomenon as the return of a high value to the high reference line of the domain. In Féry & Kügler (2008), it was shown that the phenomenon is also found in final pitch accents, in a declarative falling contour.

Pierrehumbert's tonal model was influenced by two precursors, the analysis of lexical tones in the Africanist tradition and Bruce's analysis of Swedish. We will discuss the intonation of Swedish in chapters 7 and 8. Some elements of the Africanist tradition are presented in the next subsection.

5.3 Tones in African tone languages

The autosegmental part of the tone-sequence model described in the preceding section has been crucially influenced by insights coming from the Africanist tradition. The autosegmental framework was first proposed by Leben (1973, 1978), Goldsmith (1976) and Williams (1976) for African languages like Mende, Igbo and Ewe. Tones are typically represented by H, L and combinations of H and L, although occasionally M (for mid) has been proposed as well.[8] Leben (1973) showed that words in Mende, a Mande language from Sierra Leone, can have high (H), low (L), falling (HL), rising (LH) and rising-falling (LHL) melodies, and that the distribution of these melodies is quite restricted, as illustrated in (8).[9] He proposed that the contours in fact originate from sequences of individual tones, and that, for this reason, not all tone sequences are allowed. The melodies in (8) are properties of words, regardless of the number of syllables appearing in the words. A rising sequence, for instance, is rendered by a sequence LH in all words of (8)d: *mbă*, with the entire melody on one syllable, *fàndé*, where the low tone is on the first syllable and the high tone on the second syllable, and *ndàvúlá*, where the rising melody is also on the first two syllables, and the last one has a high tone, have the same word melody. Notice the notation conventions for tones: ˊ stands for a high tone, ˋ for a low tone and ¯ for a mid tone; ˆ indicates a falling tone,ˇ a rising tone,᷄ a rising-falling tone,˝ a super-high tone and ̏ a super-low tone.

(8)	a. H	kɔ́	'war'	pɛ́lɛ́	'house'	háwámá	'waistline'
	b. L	kpà	'debt'	bɛ̀lɛ̀	'trousers'	kpàkàlì	'tripod chair'
	c. HL	mbû	'owl'	ngílà	'dog'	félàmà	'junction'
	d. LH	mbă	'rice'	fàndé	'cotton'	ndàvúlá	'sling'
	e. LHL	mba᷄	'companion'	nʲàhâ	'woman'	nìkílì	'groundnut'

To account for the distribution of tones in words, Leben formulated tone mapping

[8] There is an asymmetry in the specification for tones, L tones being unspecified, or less specified than H tones in many languages. A privative system (H, Ø) seems warranted for some languages (see Philippson 1998:321 for Giryama and also Nash 1992-94). See Hyman (2011) for an overview of tone underspecification. Some African languages, in contrast, require not only specification for H and L, but also for M.

[9] In his 1978 paper, Leben added a downstepped H (!H) to this inventory.

conventions to the effect that the first tone is associated with the first syllable, the second tone with the second syllable, and the third tone with the third syllable etc.; see examples of associations in (9). Any remaining syllable is associated with the last tone, as in (9)b and any remaining tone is associated with the last syllable, as in (9)d-e. For different formulations of the mapping conventions, the reader may want to consult Williams (1976) and Clements (1984). As a consequence of the mapping conventions, an accumulation of tones at the end of a word is not an infrequent pattern. This has been called tonal crowding in the literature, see (9)d-e.

		L		H		LHL		L HL		LHL
		|		/ \		| |		| \/		|/
(9)	a. kpà		b. pɛ́lɛ́		c. nìkílì		d. nʲàhâ		e. mba᷄

The Obligatory Contour Principle (OCP), as proposed in Leben (1973), disallows successive identical tones and forces dissimilation of adjacent tones.[10] As a result, the tone pattern (8)d for a word like *ndàvúlá* 'sling' is as shown in (10)a. It cannot have the tone pattern in (10)b. We will see below that a complete prohibition of adjacent tonal elements is too strong and that the tone association conventions have to be relaxed.

		L H			L H H
		| / \			| | |
(10)	a. ndàvúlá	'sling'		b. *ndàvúlá	

As an illustration of the effect of the OCP, consider Meeussen's (1970) rule in (11), a dissimilatory rule, which lowers a high tone after a preceding high tone. Compare the data from Shona in (12) in which a clitic *né* 'with' induces lowering of a following high tone across a morpheme boundary. The absence of tonal marking on a syllable means that the syllable is realized with a (default) low tone.

(11) Meeussen's rule

$$H \rightarrow L \ / \ H\ (\#)\ \_\_$$

(12)	a. né # bwá	→	né bwa	'dog, with a dog'
	b. né # hóvé	→	né hove	'fish, with a fish'
	c. né # hákáta	→	né hakata	'with the bones'
	d. né # bénzíbvunzá	→	né benzibvunzá	'(with a) curious fool'
	e. né # badzá	→	né badzá	'axe, with an axe'

In (12)b-d, two syllables from the stem change from high to low as a consequence of Meeussen's rule, showing that only one H tone is involved. Exceptions and special cases are possible. For instance Meeussen's rule sometimes only applies to one syllable in a word, the following ones remaining high. And in (12)d-e, the last syllable is H, but the first ones in (12)d have been changed to L. The fact that the tones' behaviour can be described as largely independent of the syllables carrying them has been called 'tone autonomy'.

Tone autonomy finds further instantiations, some of which are illustrated here. First, autosegmental association of tones can account for the phenomenon called 'stability

[10] In Leben's (1973:66) original formulation: "because contours such as HHL and HLL have the same 'melody', which is represented suprasegmentally as HL, it follows that a language in which tone is suprasegmental in underlying forms will not be able to have distinct contours HHL and HLL in underlying representations."

of the tones', illustrated with the following data from Etsako (or Yekhee), a Kwa language spoken in the Bendel state of Nigeria, in which vowels can be merged (or coalesced) between words. The data in (13) from Odden (2005:303) show that, when two words form a single ω-word, the last vowel of the first word is deleted, but the corresponding tone is not. The resulting *floating* tone docks to the first vowel of the following word, which may become complex as a result, see (13)c-f.[11] This is illustrated in (14) with an autosegmental representation.[12] For instance, the floating tone H resulting from the deletion of the final vowel of *òké* in (12)d docks to the following vowel with a low tone, which becomes falling.

(13)

a. ídzé élà	→	ídzélà	'three axes'	
b. èké élà	→	èkélà	'three rams'	
c. údzé òkpá	→	údzôkpá	'one axe'	
d. òké òkpá	→	òkôkpá	'one ram'	
e. áwà áwà	→	áwŏwà	'every house'	(textually: 'house house')
f. ídù ídù	→	ídwĭdù	'every lion'	(textually: 'lion lion')

(14)

L H L H		L H L H		HHL H		
| | | |		| | |		| | |		
òké òkpá	→	òk òkpá	→	òkôkpá		

Second, tones can be floating from the start and have effects on the surrounding vowels even if they are not directly associated with a tone bearing unit (TBU). Some particular morphemes are only tonal. Odden (2005:315) illustrates morphemic floating tones with case-marking and modified nouns in Angas, a Chadic language from Nigeria; see (15). Case-marking on a noun is indicated by a floating H which links to the final vowel of the noun, as illustrated in the second column of (15). When the noun ends with M or L underlyingly (first column), this tone becomes a rising tone under case-marking, i.e. when it is at the end of the subject or object NP, for example, see (15)c-e. When a noun is followed by an adjective, they form a Φ-phrase together. In this case, a floating L tone is suffixed to the last syllable of the noun, resulting in a falling contour tone when the underlying last tone of the noun is M or H (third column of (15)a-c).

(15)

Stem	Case-marked +H	Modified +L	
a. téŋ (H)	téŋ	têŋ	'rope'
b. nʲí (H)	nʲí	nʲî	'elephant'
c. ʔās (M)	ʔăs	ʔâs	'dog'
d. màs (L)	măs	màs	'locust bean'
e. ʔàs (L)	ʔăs	ʔàs	'tooth'

Some languages also have morphemes without tones, which acquire their tone by assimilation or spreading from the preceding word (see Leben 1973, 1978 for examples from Mende).

Floating tones also exist in Kikuyu, a Bantu language spoken in Kenya. Clements (1984) showed for this language that there is a systematic tone shift of the underlying lexical tones so that they appear one syllable later than the one they originate from. Additionally, morphemic tonology also creates floating tones at the end of words in this language. Low tones in this position are not necessarily realized as such, but they trigger

[11] Floating tones are unassociated in the underlying or intermediate representation.
[12] Similar examples come from Tangale, a Chadic language (Kenstowicz & Kidda 1987:230).

downstep of a following high tone. This phenomenon is not restricted to Kikuyu but takes place in a number of African languages.

Welmers (1959) called languages with downstep triggered by a floating L 'terraced-level'. In a typical terraced-level language, all syllables bear a tone (H or L), and a syllable bearing H is realized at a lower pitch than a preceding such syllable whenever L intervenes, as was illustrated for English in the preceding section. Stewart (1966) called this kind of downstep *automatic downstep* and distinguished it from *non-automatic downstep*, which appears in a sequence H !H, where the second H is lower than the preceding one even though no L intervenes. Clements & Ford (1980) showed that the distinction between automatic and non-automatic downstep is not always warranted. A non-automatic downstep may be induced by a floating L, thus an L unassociated with a TBU. Moreover, it is very common that both associated and floating L tones trigger downstep, as is the case in Baule, a Kwa Tano language of Côte d'Ivoire (Ahoua 1996) and in Chumburung (Snider 1999). Still, even if automatic and non-automatic downstep may be indistinguishable in some languages, it is obvious that some other languages do have true non-automatic downstep: a high tone is lower than a preceding one without an intervening L tone, see for instance Akan, as analysed by Dolphyne (1988) and Genzel (2013). As a result, downstep need not be triggered by an L tone, floating or associated.

It was shown in (8) with Mende that all syllables of a word can share a single tone. However, in other languages, two consecutive syllables of a word may have two separate H tones. In such a case, languages can downstep H after H in the same word. As an example, consider the following data from KiShambaa (Shambala) (Odden 1982, also mentioned in Gussenhoven 2004). This language has a three-way surface contrast after H tones, as illustrated by the words in (16). In *n̩óká* 'snake', both syllables share a single H realized at the same level; in *ngó!tó* 'sheep' each syllable has its own H (in violation of the OCP), triggering a downstep on the second H; finally *kúì* 'dog' has a sequence HL: The L tone in *kúì* is lower than the downstepped H in *ngó!tó*. Downstep *within* words is triggered in the same way as downstep *across* words, i.e. by a preceding H. An H-toned word like *dú* 'only' is always downstepped after another H-toned word; see (16)b and d. Notice that an !H is still an H and not a different kind of tone. Downstep is just an instruction on how to realize an H tone.

(16)

H !H | /\ a. ní !n̩óká 'it is a snake'	H !H /\ | b. n̩óká !dú 'only a snake'
H !H!H | | | c. ní !ngó!tó 'it is a sheep'	H !H !H | | | d. ngó!tó !dú 'only sheep'
HL || e. *kúì* 'dog'	

This concludes the illustration of the autosegmental representation of tones and their phonological autonomy. Tones in languages in which lexical accents play no role for tone assignment are associated with TBUs, but are at the same time independent of them in a much more obvious way than in English. In other words, they show an evident *tone autonomy* with respect to the text they are associated with. There are many more phenomena that could be illustrated here showing that tones have an independent representation from the text, such as tone doubling. We will return to tone languages in chapters 7 and 8. For the time being, it suffices to know that tones can in certain languages be highly autonomous, and that African languages have more tone autonomy than other languages, tonal or not.

5.4 ToBI

Besides the profound influence that Pierrehumbert's work has had on all subsequent work on intonation, the main visible offspring of Pierrehumbert's model nowadays is ToBI, (Tones and Break Indices), an annotation convention for intonation. ToBI was developed and publicized from the nineties on (Silverman et al. 1992). Researchers working with ToBI have applied it to a large number of languages (see for instance Jun 2005, 2014 for a review of some of them). ToBI was developed by a group of phonologists who were interested in a simple and convenient notation for English intonation using the AM approach. ToBI integrates some of the components of Pierrehumbert's model and proposes a surface oriented representation of intonation.

The notation consists of basic tones: pitch accents, phrase accents and boundary tones, as shown in (17), see section 5.2 for the equivalent notation in Pierrehumbert's (1980) model; other labels are shown in (18). These include notations for downstep, which is taken to be a property of individual tones following Ladd (1983a), for uncertainty – when the annotator does not know for certain which label she should choose for a certain tonal event for instance, and for delayed or early peaks. Finally, note the break-strength symbols from 0 to 4 in (19). The values are meant to correspond to phonological breaks, but in real speech, they may also stand for hesitations. According to Calhoun (2010:5), they are "useful annotation conventions, not indicators of qualitatively different phrase types." She goes on with the following comment: "This position neatly explains long-noted discrepancies between phrasing structure and syntax structure— that is, syntactic boundaries are marked by prosodic boundaries at SOME level of prosodic structure, rather than each syntactic phrase type being matched by a particular prosodic phrase type." We return to this or a similar perspective on prosodic breaks in section 9.2. The list is not exhaustive as there are more diacritics, which can differ from language to language. Moreover, languages differ in their inventory of tones.

(17) pitch accents: L*, H* (!H*), L+ H* (L+!H*), L*+H (L*+!H), H+!H*
phrase accents: H⁻ (!H⁻), L⁻ (obligatorily placed at every Break Index = 3 and higher)
boundary tones: H%, L% (obligatory at every Break Index = 4)
%H (marginal, at the beginnings of some intonational phrases after a pause)

(18) a. downstep: e.g. !H*, L+!H*, !H⁻
b. uncertainty: *?, ⁻?, %? (uncertainty about occurrence)
X*?, X⁻?, X%? (uncertainty about tone type)
c. phonetic events transcribed in careful labelling:
delayed tone <; early tone >

(19) Basic break index values:
0 (very close inter-word juncture)
1 (ordinary phrase-internal word end)
2 (a stronger or weaker disjuncture as expected, depending on the level)
3 (intermediate phrase end, with phrase accent)
4 (intonational phrase end, with boundary tone)

ToBI's labelling tools are extremely useful for the development of synthetic speech or for automatic tonal annotation of large spoken corpora. They are be considered as tools for non-specialists. For specialists of intonation, they represent a surface oriented annotation,

the result of a phonological analysis. The annotation surface cannot distinguish between different kinds of underlying tones. Lexical tones, default nuclear accents, pitch accents arising from a narrow focus etc. are all represented as starred tones. A development of ToBI could be to integrate metrical structure, syntactic structure and information structure, since up to now there is no practical way to address these issues directly in the annotation.

5.5 Optimality-theoretic model of intonation

Optimality Theory (OT) was elaborated in the nineties by Prince & Smolensky (1993) and McCarthy & Prince (1993a,b). It is a theory assuming that speech (output) is based on underlying representations (input) and that the relationship between the two is determined by a language-specific ranking of universal constraints. For this section, some knowledge of Optimality Theory is required, as the basic concepts of this model are assumed. Readers with no knowledge of OT are invited to skip this section.

In his 2004 book, Gussenhoven proposes an OT model of intonation for different languages: Basque, Tokyo Japanese, English, Scandinavian languages, French, and Franconian dialects of German and Dutch, all based on a tone-sequence analysis of intonation. Hyman (2011) also proposed an OT model of intonation although his aim was more limited. He wanted to show the effect of the OCP in tone languages.

In this section, only the main points of an OT approach to intonation are reviewed. Both Gussenhoven and Hyman make the usual OT distinction between markedness and faithfulness constraints. The markedness constraints, like those listed in (20), support unmarked patterns, while the faithfulness constraints, see (21), require identity between input and output. Two of the markedness constraints, NoCrossing and OCP, are general constraints, and are not specific for tonal structure: NoCrossing forbids crossing association lines in the autosegmental model, and OCP, which was introduced in section 5.3, prohibits adjacent identical elements. The other constraints in (20), NoContour and NoRise/NoFall, militate against complex tonal structure.

(20) Markedness constraints:
- a. NoCrossing: Association lines do not cross.
- b. OCP (Obligatory Contour Principle): Adjacent identical tones are disallowed.
- c. NoContour: No complexity in tonal representations: tones are monotonal.
- d. NoRise: *$(LH)_{\sigma}$
- e. NoFall: *$(HL)_{\sigma}$

The faithfulness constraints, some of which are listed in (21), are the usual ones: no deletion, no insertion, identity between input and output and no metathesis.

(21) Faithfulness constraints:
- a. Max(Tone): Deletion of tones is prohibited.[13]
- b. Dep(Tone): Insertion of tones is prohibited.
- c. Ident(Tone): A tone in the output is identical to its correspondent in the input.
- d. Linearity: The sequencing of the tones in the input and in the output is identical.

Consider an abstract input consisting of two moras and two tones, L and H in this order. The two moras should be taken as elements of a single syllable. The markedness constraint considered here is NoRise, which is in conflict with the faithfulness constraints

[13] 'Tone' stands for H or L.

MAX(Tone), DEP(Tone) and IDENT(Tone). NORISE forbids a rising contour within a single syllable. If NORISE is undominated, the output cannot be faithful to the input. Such a situation is illustrated in the two optimality-theoretic tableaux T1 and T2, adapted from Gussenhoven (2004: 147-148). The candidates b to f fulfil NORISE, but they do so at the price of violating one or more faithfulness constraints. In T1, candidates b and c each delete a tone, and violate MAX(Tone), the constraint against the deletion of tones. Candidates d-f replace one or two tones, and violate the IDENT constraints. In this tableau, IDENT(H) and IDENT(L) are not ranked relatively to each other. The optimal candidates are d or e, which violate the IDENT constraints only once, but in different ways. Candidate f cannot win the competition because it violates IDENT twice. Alternatively, it violates LINEARITY, not shown in the tableaux. The ultimate winner depends on the ranking of IDENT(H) and IDENT(L), as well as other constraints not shown here.

T1: NORISE >> MAX(Tone) >> IDENT(Tone)

(L H), μ μ	NORISE	MAX (Tone)	IDENT(H)	IDENT(L)
a. $(\mu\ \mu)_\sigma$ | | L H	*!			
b. $(\mu\ \mu)_\sigma$ | L		*!		
c. $(\mu\ \mu)_\sigma$ | H		*!		
d. $(\mu\ \mu)_\sigma$ | | ☞ H H				*
e. $(\mu\ \mu)_\sigma$ | | ☞ L L			*	
f. $(\mu\ \mu)_\sigma$ | | H L			*	*!

In Tableau 2, the input is the same as in T1, and NORISE is still undominated. However, the ranking of the faithfulness constraints is changed: IDENT(H) and IDENT(L) are now ranked above MAX(Tone). In this case, the candidates that have deleted one of the input tones are the winners, thus b and c.

T2: NoRise >> Ident(Tone)>> Max(Tone)

(L H), $(\mu\ \mu)_\sigma$	NoRise	Ident(H)	Ident(L)	Max(Tone)
a. $(\mu\ \mu)_\sigma$ | | L H	*!			
b. $(\mu\ \mu)_\sigma$ | ☞ L				*
c. $(\mu\ \mu)_\sigma$ | ☞ H				*
d. $(\mu\ \mu)_\sigma$ | | H H			*!	

Consider next the following tone association constraints (from Anttila & Bodomo 2000).

(22) TBU → T: TBUs are associated with tones.

(23) T → TBU: Tones are associated with TBUs.

(24) NoCrowding: A TBU is associated with at most one tone.

The constraints in (22) and (23) together have the effect that a tone is associated with a TBU and a TBU is associated with a tone. T → TBU should be understood as a family of constraints. For instance, consider the natural hierarchy: H → TBU >> L → TBU. This amounts to associating H first and L second. Example (25) illustrates three options concerning the ranking of (22), (23) and (24) in a situation in which two tones are present, but only one syllable.

(25)

a. σ	b. σ	c. σ
|	/ \	|
H L	H L	H L
NoCrowd, H → TBU >> L → TBU	H → TBU, L → TBU >> NoCrowd	NoCrowd, L → TBU >> H → TBU

Specific tones associate with specific syllables (TBUs) by virtue of being of a certain kind: for instance a starred tone associates with a stressed syllable. If a syllable is metrically strong, it is associated first, as predicted by a natural hierarchy like H* → TBU >> H → TBU.

Turning now to OT alignment,[14] boundary tones are aligned with edges of prosodic boundaries: for instance, a Φ-phrase boundary tone is aligned with the right edge of a Φ-phrase. In such a case, alignment is fulfilled by a tone when it appears on the last element of the prosodic domain able to carry a tone, usually a syllable. Alignment of a prosodic domain with a boundary tone is accounted for with alignment constraints like those in (26). Alignment of a boundary tone with the right edge of a Φ-phrase is fulfilled in (27)a-b, but not in (27)c-d. The preference for high or low boundary tones is regulated by the pragmatic component of intonation; see chapter 6.

[14] See section 7.5 for a different meaning of tone alignment.

(26) a. Align-R(Φ,L_Φ): Align the right edge of a Φ-phrase with an L_Φ.
b. Align-R(Φ,H_Φ): Align the right edge of a Φ-phrase with an H_Φ.
c. Align-R(ι,L_ι): Align the right edge of an ι-phrase with an L_ι.
d. Align-R(ι,H_ι): Align the right edge of an ι-phrase with an H_ι.

(27)

a.	b.	c.	d.
σ σ σ]$_\Phi$	σ σ σ]$_\Phi$	σ σ σ]$_\Phi$	σ σ σ]$_\Phi$
|	/\	|	
H_Φ	H_Φ L_Φ	H_Φ	

Association and alignment are sometimes in a complementary distribution relationship. In H*L, the H* is associated with a stressed syllable, but L is not. It is just aligned to H* and as a result realized next to H*. L's left edge is aligned with H*'s right edge. This is accounted for by CONCATENATE, formulated in (28)a (adapted from Gussenhoven 2004). Finally, the constraint NOSPREAD in (28)b is responsible for the fact that tones are preferably associated with one syllable only. Again this is a family of constraints rather than a single constraint. A distinction must be made between T* and T_Φ for instance. A starred tone does not spread, and a T_Φ does in some circumstances, as was shown in section 5.2.

(28) a. CONCATENATE: The tonal components of bitonal tones are aligned with each other.
b. NOSPREAD: A tone is associated with at most one TBU.

This section has only touched upon the basic constraints, notions and application of a tonal OT. The reader should consult Gussenhoven (2004) for more constraints, and concrete applications of OT to a number of languages and facts. To close this section, it should be noted that OT needs an additional phonological representation of intonation, like the tone-sequence one, in order to allow for the formulation of adequate constraints. As always, OT is to be seen as a meta-theory, as a tool for implementing a phonological system in a constraint-based framework.

Besides its ability to account for typological variation with violable constraints, one of the main strengths of OT resides in its generative component. Both input tones and output surface representations coexist and are related to each other by means of faithfulness constraints. Between input and output, markedness constraints are operational, forcing the input to be transformed in specific ways. This aspect already existed in Pierrehumbert's model, but there, it was not accounted for by a specialized grammar.

5.6 Nuclear tone tradition (British school)

The nuclear tone tradition has its origins in Britain and is still widely used nowadays. The intention of its creators was in their own words to develop a tool for the learning of English intonation by foreign students. For this reason and because of the simplicity of the notation and theoretical concepts, it does not require any theoretical background. Moreover it proposes a straightforward relation between form and meaning. Its main proponents are Palmer (1922), O'Connor & Arnold (1961, 1973), Crystal (1969) and Cruttenden (1997). Utterances consist of units called *word groups* (O'Connor & Arnold 1973), *sense groups* (O'Connor & Arnold 1961), *intonation-groups* (Cruttenden 1997), or *tone-units* (Crystal 1969), which are roughly equivalent to ι-phrases, though these domains are not defined in grammatical terms. In this section, we introduce the model by

following O'Connor & Arnold's (1973) proposal. In each sentence or sense group, there is a nuclear tune, taken from the set in (29).

(29) Nuclear tunes of O'Connor & Arnold (1973:8-9)
- a. Low Fall: the voice falls during the word from a medium to a very low pitch.
- b. High Fall: the voice falls during the word from a high to a very low pitch.
- c. Rise-Fall: the voice first rises from a fairly low to a high pitch, and then quickly falls to a very low pitch.
- d. Low Rise: the voice rises during the word from a low to a medium pitch or a little above.
- e. High Rise: the voice rises during the word from a medium to a high pitch.
- f. Fall-Rise: the voice first falls from a fairly high to a rather low pitch, and then, still within the word, rises to a medium pitch.
- g. Mid-Level: the voice maintains a level pitch between high and low, neither rising nor falling.

The tune (or nuclear tone) starts on the nucleus of a sentence, defined as the stressed syllable of the last accented word of the sentence. The nucleus is the most prominent part of the sense group. It is followed by the tail (or nuclear tail for Crystal 1969). Together, nucleus and tail carry the nuclear tune. Compare figure 5.16, which gives an overview of the nuclear tunes on a single syllable, *two* for example. The examples are illustrated with the original notation style of the British school, called the *tadpole* notation for obvious reasons. In this notation, each syllable is represented by a dot; the stressed syllables are provided with larger dots than the unstressed ones. Two horizontal lines show the place of the dots in the register used in each sentence: they correspond to 'the normal high and low limits of the voice.' (p.7). In figure 5.16, the bottom and top limits of the cells stand for the limits of the register. Alternatively, tonetic stress-marks are used, which are symbols placed before the words and which specify the contour of the word. For instance ´`*two* indicates that the word is realized with a rise-fall, and `*two* that it has a high-fall. Since these diacritics have a different function in the present book, this notation is not used in this section.

Low Fall	Rise-Fall	High Rise	Mid-level
			
High Fall	**Low Rise**	**Fall-Rise**	
			

Fig. 5.16 Nuclear contour on nuclear monosyllables

The nuclear tune is often extended over several syllables comprising the nucleus and all following syllables up to the end of the word group, the *tail*, as illustrated in (30)a. Before the nucleus, the *head* begins with the stressed syllable of the first accented word (providing it is not also the nucleus) and ends with the syllable immediately preceding the nucleus, as in (30)b. The pre-head contains the material preceding the head. These additional divisions are also illustrated in table 5.1. The 'nucleus' is the only obligatory part of the sentence; pre-head, head and tail are facultative.

(30) a. [I could have] [KICKED] [myself] (O'Connor & Arnold 1973:18)
Pre-head Nucleus Tail

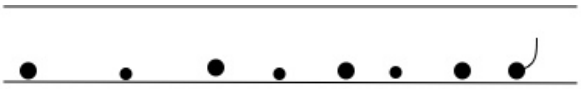
b. [It was an UN][USUALLY DARK][NIGHT] (O'Connor & Arnold 1973:18)
Pre-head Head Nucleus

The tail has no contour of its own since it is always part of the nuclear tune. Table 5.3 illustrates pre-heads, heads and nuclear tunes on several syllables.

Prehead	Head	Nucleus	Nuclear tune Tail
I	WANT to be ABSOLUTELY	SURE	about it
low	high	Low Fall	
What on	EARTH did you do	THAT	for, Peter
low	high	High Fall	
I	QUITE A	GREE	with you, my dear chap
low	high	Rise-Fall	
	WOULD you like to COME to the	THEA	TER with us?
	high	Low Rise	
It' s	NOT	FAIR	did you say
low	high	High Rise	
I	DOUBT whether	THAT	would be any good
low	falling	Fall-Rise	
IM	MEDIATELY I TOLD him the	TRUTH	about it (he relented)
low	high	Mid-Level	

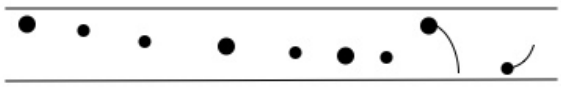
Table 5.3 (from O'Connor & Arnold 1973: 13-14)

In table 5.1 the pre-head is always low and the head is always high, except in one case. However, pre-heads can be high or low. Moreover, there are four types of head: low, high, falling and rising. In a low head, all syllables of the head are on the same low pitch, as illustrated in (31).



(31) SOMEONE'S BOUND to COME ALONG SOON.

The first syllable of a falling head is rather high in pitch and the following syllables carry a gradually lower pitch, as shown in (32).



(32) EVERYONE'S BOUND to SEE it SOMETIME.

The rising head is the opposite of the falling head: its first syllable is low in pitch and any following syllables gradually carry the pitch higher. This is shown in (33).

(33) HOW did you MANAGE to do THAT?

Pre-heads can be low, as in table 5.1, or high, as in (34).

(34) It was AMAZING.

In a second component, the British school intonation system proposes meanings for individual tunes. For instance, all statements associated with a falling nuclear tone sound definite and complete. We will return to the meaning of intonation in chapter 6.

5.7 PENTA: A phonetic model

Since this book defends a phonological view of intonation, it may be interesting to take a closer look at how it diverges from a non-phonological model, like the Parallel ENcoding and Target Approximation model (PENTA) of Xu (2005) and Xu & Xu (2005).

Xu's view of Beijing Mandarin Chinese intonation is fairly traditional: tones are associated with syllables at the lexical level, so that each syllable has its own tone. Recall from chapter 1 that Mandarin Chinese specifies nearly all syllables with one of the tones listed in (35). The only unspecified syllables are those carrying the neutral tone, which changes its tonal specification according to the environment. However, the four tones of Mandarin Chinese influence each other because of tonal coarticulation.

(35) Beijing Mandarin lexical tones

Tone 1	55	(H, **H**igh)	˥	mā	'mother'
Tone 2	35	(LH, **R**ising)	˧˥	má	'hemp'
Tone 3	21(4)	(L, **L**ow, Falling-rising)	˨˩˦	mǎ	'horse'
Tone 4	51	(HL, **F**alling)	˥˩	mà	'to blame'

The phonetic motivation for Xu's PENTA model of Mandarin tones is illustrated in figure 5.17 from Xu (2005:223, adapted from Xu 1999). In all panels of the figure, the first syllable is High (H), the second syllable has one of the four lexical tones, and the following tone in the third syllable is High in (a), Rising in (b), Falling in (c) and Rising again in (d). And in all slots, the last syllable is High. Panel (d) differs from (a-c) in having an L tone in the fourth syllable. The figure illustrates the third syllable's idealized trajectory (shown by the dashed line). In each case, its trajectory is heavily influenced by the preceding tone. For instance, it is visible from (a) that when a low tone appears in the second syllable, the low target is reached at the end of this syllable, and the high target of the third syllable appears only at the end of the third syllable. Because of the slowness of our articulatory organs, it takes some time to go from one target to another.

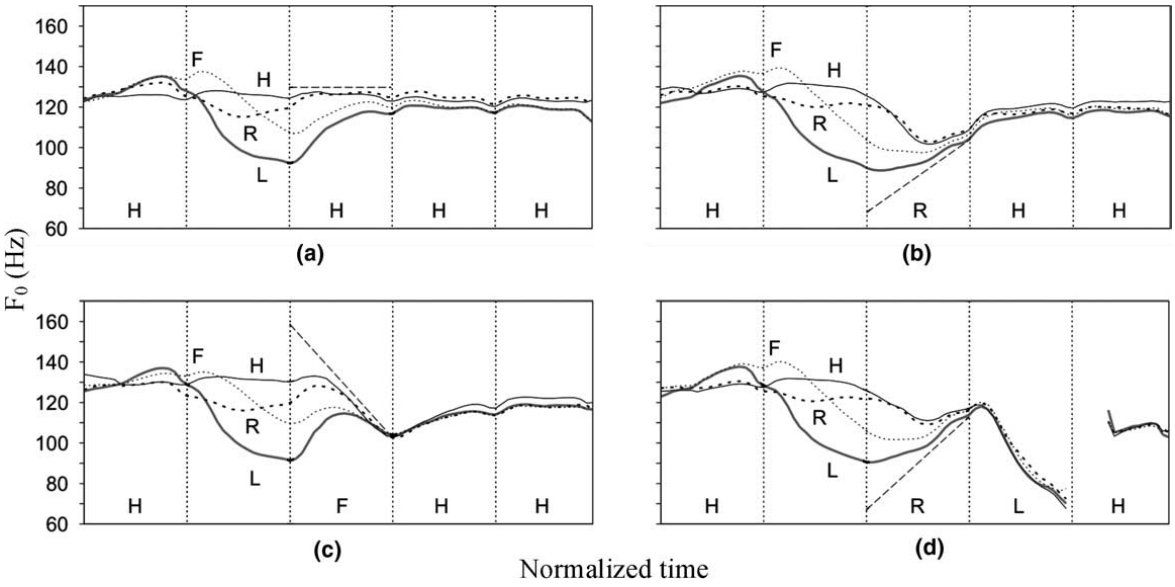


Fig. 5.17 Syllable 3 has different tones and is preceded by different tones. Vertical lines indicate syllable boundaries. The short dashed lines depict pitch targets. (from Xu 2005:223)

On the basis of the differences in the contours of the third syllable, Xu concluded that there is no invariant surface acoustic correlate of tones. Rather what remains constant is "an underlying trajectory that is continually approximated throughout the syllable to which the tone is associated" (Xu 2005:223). And he further proposed that tone and syllable are implemented simultaneously and at the expense of full implementation, since tones are associated one-to-one with syllables (see Gårding 1987, Shih 1988, Bao 2011, Yip 2002 and others for the same view).

More questionable is whether this one-to-one association between syllables and tones can be transposed to a language without lexical specification of tones, like English. In other words, does English also need a full specification of the trajectory of each syllable? Xu & Xu (2005) suggested that it does. According to them, an invariant pitch target is assigned to each syllable, exactly as in Mandarin Chinese, but in English, this does not happen at the lexical tone level. The PENTA model applied to English consists of two basic components of intonation, the (melodic) communicative functions and the articulatory system used to transmit the information, understood as tools for the transmission, as shown in figure 5.18 from Xu & Xu (2005:190). Multiple communicative functions (lexical, sentential, focus, grouping, and others) are conveyed in parallel by varying the melodic primitives.

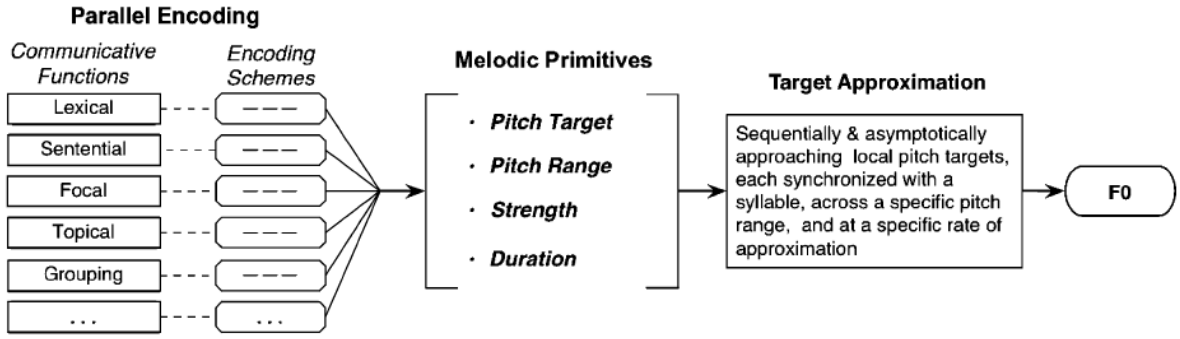


Fig. 5.18 (from Xu & Xu 2005:190)

The communicative functions listed on the left of figure 5.18 influence F0 and other intonational features by means of the encoding schemes, the empty boxes to the right of the communicative functions. The encoding schemes specify the values of the melodic primitives of the melody, which consist of local pitch targets (high, low, rising, falling, mid), pitch range (wide, narrow, normal), articulatory strength (strong, weak, normal) and duration (short, long, normal). The communicative components of speech melody are thus defined based on their function rather than on their form. An important part of the model is the strictly syllable-based generation of F0. Each syllable is assumed to have its own target, as described in the box called *target approximation*. Because it is syllable-based, this model has some similarities with the tone-sequence model described in section 5.2.

The major strength of the model lies in the detailed phonetic links between the components of speech melody and surface F0 contours regulated by an articulatory process of syllable-synchronized sequential target approximation. It is evidently an issue for every model to explain how the intonational events are realized and how they interact. As already mentioned above, pitch movements are constrained by articulators and our articulatory organs are rather slow. This leads to phenomena such as undershoot, truncation and compression of the local pitch targets. In the PENTA model, all melodies are treated as mechanical instructions to the articulators by communicative functional roles. The communicative functions of figure 5.18 are taken as primitives and are assumed to have different directly measurable effects on the phonetics of intonation (see also Cooper & Sorensen 1981, Cooper, Eady & Mueller 1985, and Eady & Cooper 1986 for a similar view).

However, it is not quite clear how to define communicative function in such a purely phonetic model, in which there is no phonological entity like prosodic domain or prosodic boundary. One communicative function is called 'focus', which plays an important role in Xu & Xu's analysis of the tonal structure of English. Focus is understood as an instruction to the articulators to expand the pitch range. In Xu & Xu (2005:160f), the functional role of focus is defined as an answer to a wh-question (a single-word informational narrow focus), but in the melody, it is understood as the last accent of a sentence. Notions like focus projection, givenness and focus domain are absent (see chapter 6 for information structure). In fact, all phonological notions are absent from this model. Phonetic realization of focus in English is described in the same terms as in Mandarin, as a local expansion of the pitch range, followed by compression. There is also no possible variation as to the effect that 'focus' can have in English and Mandarin. In English it is described as follows: "An F0 peak often occurs after the offset of a stressed syllable if the syllable duration is much shorter than 200 ms" (Xu & Xu 2005:187).

In the next chapter, it will be shown that focus cannot be reduced to phonetic

instructions to the articulators. We will propose instead that focus has to be defined in semantic and pragmatic terms and that it can be realized in a multitude of ways in different languages, also in English and in Mandarin. In a model of intonation taking the syntactic and semantic influence on intonation seriously, focus has a phonological effect: in English, it changes the position and/or the height of the nuclear pitch accent. It can raise or lower it depending on different factors, as for instance the illocutionary act associated with the sentence under consideration. Further factors influencing its realization are the position of the focus in the sentence and the kind of sentence it is in (declarative or question for instance). In all cases, focus is mediated by the phonological structure of the sentence. Mandarin Chinese is of course different from English since it is a language without pitch accent. Treating the two languages in the same way hides important and interesting typological differences.

A further point of concern is related to the fact that English like Mandarin needs a tonal target on each syllable. In such a framework, there cannot be any strategy of interpolation, which is specifically needed for syllables without tonal targets. Pierrehumbert (1980) claimed that when a tonal target is reached, the next one is aimed at, and this next tone can be located several syllables ahead. The idea of interpolation, which is not rendered faithfully by Xu & Xu's (2005) criticism, is that these syllables do not interfere in the trajectory between one target and the next other.

To conclude this chapter, several very different models of intonation have been proposed in the past, and the tone sequence model, which allows phonology and phonetics to be treated in different but interacting modules, has been the most successful one. Importantly for the remaining of this book, syntax and semantics must be taken as exerting an influence on the tonal structure, but only through the prosodic structure and the tone sequence, not as direct instruction to the articulators to modify the acoustics.

Annotated suggestions for further reading

Pierrehumbert (1980) is an important work as it lays down the principles of the tone-sequence model.
The annotation conventions for American English ToBI can be found at the url: http://www.ling.ohio-state.edu/~tobi/ There are also links to ToBI annotation systems for other languages to be found on this site.
For ToBI models of a large number of languages, the reader can profitably consult the papers gathered in Jun (2005, 2014).
Gussenhoven (2004) presents an OT analysis of the intonation of several languages.
Xu (2005) and subsequent works by the same author introduces the PENTA model of intonation for Mandarin Chinese.

Discussion points

1. Discuss the merits and drawbacks of the tone-sequence model of intonation for English and for Mandarin Chinese.

2. Cruttenden (1997) 2nd edition is an overview of the British school model of intonation. It concentrates on English intonation. Try to translate the terminology used in table 5.1 into a ToBI notation to get a feeling of the systems. How are boundaries express in both systems? And pre-nuclear pitch accents?

3. And how would you write in the tone-sequence notation the following nuclear contours of the British School:
fall, fall-rise, high rise, low fall, rise-fall, low-rise with high head, low-rise with low head?
Consult Ladd (2008:91) for a systematic comparison.

Chapter 6
Intonation and meaning

6.1. Introduction

The aim of this chapter is to examine the relationship between intonation on the one hand and pragmatics, semantics and information structure on the other. The relation between meaning and intonation is at least twofold. First, the tonal contour of a sentence is shaped by information structure. The informational roles investigated in this chapter are focus, givenness and topic, as proposed by Krifka (2008) for example, see also the collection of chapters in Féry & Ishihara (2016). These roles influence the prosodic phrasing and the tonal pattern of sentences by assigning special tones or tonal sequences to bearers of the information structural roles. Lexical stresses and boundaries can be more or less prominent depending on their status: information structural roles can be conveyed by scaling of tones and scaling of reference lines. Moreover, in many languages, prosodic phrasing is changed as a consequence of special information structure.

Second, the choice of specific tones and the way they combine is essential in communication. In intonation languages, meanings and illocutionary roles influence the tonal patterns of sentences, and conversely, the choice of tones has an influence on the pragmatic interpretation of sentences. Intonation has a different role from the other grammatical modules, without being completely separated from other parts of grammar; see below for references. Rather the part played by an intonation contour interacts with other grammatical components to produce specific nuances.

The question of how intonation relates to meaning, which has been central in the discussion of intonation, especially in the models preceding the tone-sequence model, is given a simple answer in this chapter. As was demonstrated in chapters 4 and 5, a prosodic domain is realized intonationally by tones with specific roles, like pitch accent, lexical tones and boundary tones. From this perspective, the melody emerging from the concatenation of tones (in the prosodic domain) is associated with specific pragmatic meanings. The semantic effect of tones is thus mediated through the prosodic categories they appear in. In other words, it is not a 'rising tune' that is meaningful, but rather the pragmatic interpretation of a sentence is determined by specific tones in a prosodic domain, for instance a low pitch accent followed by a high Φ-phrase boundary tone. In sum, meanings arise as a consequence of the association of individual tones to specific prosodic domains and of the way these tones are concatenated in a particular text, see the examples in section 6.3.1 for concrete examples. This view differs from what can be called the standard view of the meaning of tones represented by Gussenhoven (1984, 2004), Pierrehumbert & Hirschberg (1990), Büring (1997), Bartels (1999) as well as later work by Gunlogson (2003) and Truckenbrodt (2012), which only looks at tones, and not at the prosodic domains to which they are associated.

In section 6.2, the role of prosody as a reflex of information structural roles is examined. In section 6.2.1 focus, givenness and topic are defined, section 6.2.2 reviews some of the prosodic and intonational reflexes of information structure and section 6.2.3 shows the effect of focus and givenness on tonal scaling. Section 6.3 addresses the meaning of tones and tonal sequences in English. Section 6.3.1 examines in some detail the compositional approach to intonational meanings. Section 6.3.2 introduces different non-compositional approaches to intonational meanings that have at least a compositional component. Section 6.3.3 briefly shows some examples of holistic approaches. Finally, section 6.4 addresses the claim that intonation also has a paralinguistic role, independent of its role in grammar.

6.2 Information structure

6.2.1. Information structure: Givenness, focus and topic

Consider the following dialogue, in which speaker A wants to know whether speaker B will come to her birthday's party. She starts with a question of B's disponibility on that day.

(1) A: {Do you have to work next Friday?}
B: No, I always stay home on Fridays.
B': No. Next Friday, I stay home.

In her answer B, the speaker repeats *Friday*, which has been uttered by Speaker A, this is now 'given' information. B does not answer the polar question with yes or no, but rather chooses to answer the question by adding the information that she always stays home on this day. This is the focus of the sentence, the part answering A's question. An alternative answer is B'. In this case, *next Friday* is the topic of the sentence. It is repeated from A's question, but assigned a special status. Speaker B has a reason to place *Friday* at the beginning of the sentence, for instance, she wants to highlight the fact that this Friday is special, and that on all other Fridays, she would be at work. Or she wants to contrast Friday with the other days of the week. In both cases, the information is going further than what speaker A is asking, although in both cases, the question under discussion (QUD), namely whether B is free on Friday, is answered.[1]

In accordance with a major part of the research, information structure is introduced in this section as coming in these main categories: givenness, focus, and topic.[2] Parts of sentences typically carry different informational roles and this partition may be transmitted by prosody and intonation. The way grammar encodes information structure is sometimes referred to as 'information packaging' following Chafe (1976). A sentence is 'packaged' in different parts, according to their roles as focus and topic. The given parts are the remnant parts of the sentence, and can be scattered at different places: they are not 'packaged'. Chafe (1976) understood information structure as the receiver's assumptions about the status of the referent of each linguistic expression, as it is represented in the mind of the receiver at the moment of utterance. Information structure is indifferent to the content, i.e. the lexical or propositional content of a sentence, around which grammar usually centres. Krifka (2008:343) claimed that information structure notions should be grounded in theories of how communication functions: 'The basic notions of Information Structure (IS), such as Focus, Topic and Givenness, are not simple observational terms. As scientific notions, they are rooted in theory, in this case, in theories of how communication works.' Variation in information structure is often reflected in syntax, morphology or prosody, while the information itself, i.e. propositional content and truth conditions, remains unaffected.

In defining the notions of information structure, it is important to keep in mind that for many authors, they have ambivalent meanings (see Kuno 1972, Chafe 1976, Prince 1981, Lambrecht 1994 and many others). On the one hand, they may denote the extra-linguistic, cognitive or mental states of referents, actions, locations, and temporality. On

[1] I consider the notion of Question Under Discussion (QUD), that investigates the contexts in which a particular occurrence can be uttered, as part of information structure, and do not go into the detail of QUD here. Anyway, prosody has not been included in the phenomena looked at under the label QUD.

[2] 'New' is the complement of given, 'background' the complement of focus and 'comment' is the complement of topic, see below.

the other hand, the notions also refer to the formal and communicative aspects of language for the expression of information structural roles.

The information-structural notions used below can all be reconstructed with the help of the concept of Common Ground (CG, Stalnaker 1974). The CG is the background against which the knowledge shared by speaker and addressee is updated during a conversation. It allows us to make a distinction between 'presuppositions, as requirements for the input CG', and 'assertions or the proffered content, as the proposed change in the output CG' (see Krifka 2008:245). Let us examine givennes, focus and topic in more detail.

Givenness: The status of referents can be new (inactive at the point of their introduction into the discourse) or given (active in the consciousness of the conversation partners). In the example (1)B, *Friday* has been activated by A. The referents can also be accessible from the context, the text or simply by inference. According to Clark & Haviland (1977:3), given information is 'information [the speaker] believes the listener already knows and accepts as true,' and new is 'information [the speaker] believes the listener does not yet know.' Givenness is divided into text-givenness (previously mentioned in the discourse) and context-givenness (contextually salient). In the definition proposed here, givenness can be discourse or textual givenness.

(2) A given constituent is entailed by the preceding discourse, or it is salient in the context.

Schwarzschild (1999) claimed that a given constituent is entailed by the preceding discourse. For him and for Halliday (1967/68), givenness is not a gradient property of referents. By contrast, Prince (1981), Gundel, Hedberg & Zacharski (1993), Chafe (1994), Lambrecht (1994) and Baumann & Riester (2012) have proposed hierarchies of givenness. We will see in section 6.3.1 that Pierrehumbert & Hirschberg (1990) also proposed some gradience in the concept of givenness. The hierarchy in figure 6.1 comes from Lambrecht (1994): going from left to right, givenness increases. Baumann & Riester identified a number of factors acting on the kind of givenness. Additionally, they made a distinction between referential givenness and lexical givenness, similar to the contextual and textual givenness of Prince (1981) and others.

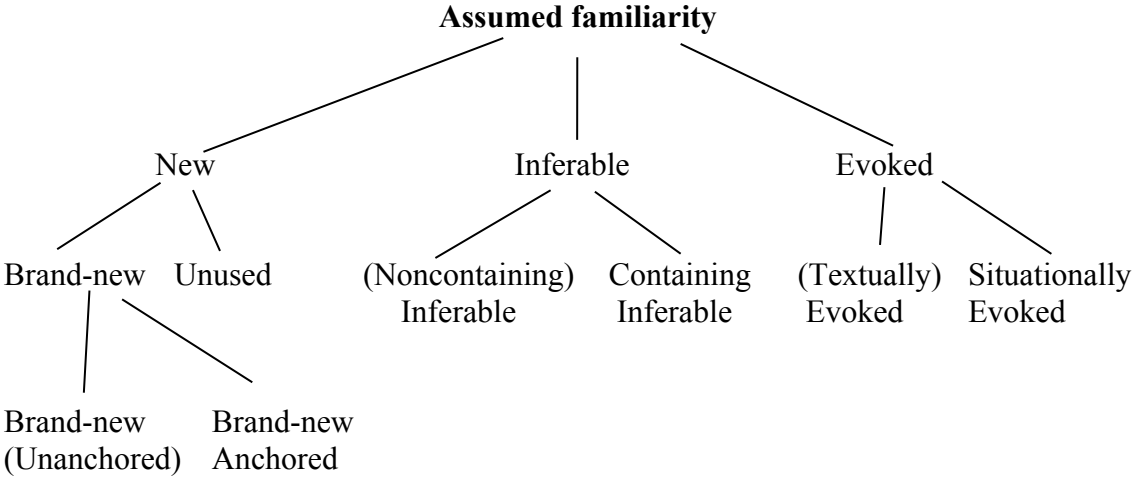


Fig. 6.1 A hierarchy of givenness (Prince 1981:237)

A much cited example of accessibility (or inferability in Prince's terminology, or

bridging) from Prince (1981:233) appears in (3)a. It goes without saying that a bus has a driver, and by mentioning *bus*, *driver* becomes cognitively available (inferable or accessible). The subscripted G indicates givenness.

(3) a. I got on a bus yesterday and [the driver]$_G$ was drunk.
b. I got on a bus yesterday and I immediately noticed [the driver]$_G$.

Focus: A first definition of focus appears in (4) in Krifka's (2008:247) wording, following Rooth's (1985, 1992, 2016) insight that a focus elicits a set of alternatives that are relevant for the interpretation of the focused expression, as well as for the discourse it is inserted in.

(4) Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions.

The classic use of focus is to highlight the part of an answer that corresponds to the wh-part of a constituent question, as illustrated in (5).[3] Focus is generally sub-sentential, that is it is realized on one part of the sentence, other parts being given, or otherwise backgrounded. All parts of a sentence can be narrowly focused, from a very small part to a larger part of a sentence. In other words, the size of focus is variable. In the following examples, the subscripted F indicates focus and capital letters indicate pitch accents.

(5) {Who ate a lemon pie?}
a. [MARY]$_F$ ate a lemon pie.
{What did Mary eat?}
b. She ate [a LEMON PIE]$_F$
{What happened?}
c. [Mary ate a LEMON PIE]$_F$

Focus is used in the sense of Rooth (1985, 1992) as the part of the sentence that introduces alternatives. Rooth's Alternative Focus Semantics assumes that there are (at least) two levels of interpretation, the ordinary level and the level of alternatives. Besides the normal semantic value present in each expression, a *focus semantic value* (fsv) is a set of alternative denotations that potentially contrast with the ordinary semantic value; by definition the ordinary semantic value is always contained in this set. A pitch accent in English is often the reflex of the second semantic value (the fsv), and is thus understood as indicating focus intonationally. The fsv interacts with other semantic and pragmatic processes. The 'normal' semantic component of a grammar associates denotations (semantic values) with the corresponding syntactic phrases. These values can be

[3] Krifka (2008:248) distinguished between *denotation focus* and *expression focus*. The definition in (4) is refined, as in (i):
(i) A property F of an expression α is a Focus property iff F signals
a. that alternatives of (parts of) the expression α (expression focus) or
b. alternatives of the denotation of (parts of) α (denotation focus) are relevant for the interpretation of α.
Expression focus is illustrated in (ii). Here the second speaker (B) corrects the accent location on *Berlin*, wrongly put on the first syllable by speaker A.
(ii) A: {They live in [BERlin]$_F$.}
B: They live in [BerLIN]$_F$!
Artstein (2005) proposed that *mites* and *tites* in (iii) are denotation foci, because they can be construed as meaningful elements.
(iii) We only saw stalag[MITES]$_F$ in the cave, no stalag[TITES]$_F$.
The alternatives discussed in this chapter are always alternatives of denotation.

propositions, properties, individuals etc. In (6), *like*(*m,s*) is a proposition, and *m* and *s* are individuals.

(6) Mary likes Sue: *like*(*m,s*)

The alternatives to the focus play a central role when it comes to interpreting focus semantically. In (7)a, $[\text{Mary}]_F$ is the focus. In the case of an explicit contrast, as in (7)b, it is the contrasting element *Anna* which fills the variable x in 'x likes Sue'. The alternatives are contextually restricted, as for instance to the individuals in (7)c. The set consists of the individuals who may like Sue in the particular context of this utterance. The question or the context presupposes that somebody likes Sue. *Likes Sue* is the non-contrastive part of the sentence, the *background*, which is complementary to the focused parts. Rooth's intuition is that the right focus contrasts with all possible right and wrong answers to a contextually given question – its 'Hamblin denotation'. Curly brackets stand for a context for the following sentence.[4]

(7) a. {$[\text{MARY}]_F$ likes Sue.}
b. No, $[\text{ANNA}]_F$ likes Sue.
c. Alternative set: {BILL likes Sue, MARY likes Sue, ANNA likes Sue}

If the accent is not on the item asked for by a wh-question, the result is infelicitous, as illustrated in the dialogue (8), where *Clara* is accented and *got a dog* is deaccented.[5]

(8) A: {What did Clara get for birthday?}
B: #CLARA got a dog.

In a spoken sentence, the hearer knows which element is focused because the context usually specifies what element is new and highlighted in the sentence. Because of this, a pitch accent on the focused constituent may seem to be superfluous, and indeed by far not all languages systematically mark the focus in grammar; see Zerbian (2006), Fiedler et al. (2010) and Downing & Pompino-Marschall (2013) for relevant examples. The reason for adding a pitch accent on the focus, or for using any grammatical reflex for focus, like changing word order or using a morphological marker, may be related to attentional factors. It is helpful to highlight important words and constituents, as this may help a listener to process what is being said and to extract the information in an efficient way.

Focus comes in different guises. We saw an informational focus in (5) and a corrective focus in (7)b. In some cases, the focus is already suggested, and the speaker can select her answer from a list of items specified in the question, as in (9). Or the focus is a correction as in (10)a, or a confirmation as in (10)b.

(9)Selection focus
{Was it Mary or Peter who ate the lemon pie?}
$[\text{MARY}]_F$ ate it.

[4] In this particular example, curly brackets are also used for the alternative set.

[5] In many cases, the question is covert rather than overt. In the case of a covert question, the coherent discourse is structured by implicit questions, as has been proposed by Klein & von Stutterheim (1987), Roberts (1996) and Büring (2003), among others. In both overt and covert questions, the answer relates an expression to the immediately preceding context. As pointed out in Rooth (1992), Rooth's alternatives form the Hamblin denotation (Hamblin 1973) of the background question.

(10) Correction and confirmation focus
{Mary ate a cookie.}
a. No, [PETER]$_F$ ate it!
b. Yes, [MARY]$_F$ ate it.

Another use of focus is to highlight parallels in interpretations. In all cases, a focus indicates alternatives, as in (11). (11)b is an example discussed by Rooth (1992).

(11) Parallel focus
a. [MARY]$_F$ ate a [LEMON PIE]$_F$ and [PETER]$_F$ ate a [PIZZA]$_F$
b. An [AMERICAN]$_F$ farmer talked to a [CANADIAN]$_F$ farmer,. . .

In this catalogue of different focus types, the broad informational focus (or all-new sentence), as in (5)c, is the weakest focus, followed by an informational narrow focus, for instance elicited by a specific wh-question, as in (5)a. If such a narrow focus has an exhaustive interpretation, as in (9), the elicited focus is slightly stronger. Next, we find associated focus (Jackendoff 1972, Krifka 1998), which is a focus in the scope of a focus operator like *only*, *also* or *even*, as in (12) below. Stronger kinds of focus are parallel foci, as in (11). And the strongest focus of this hierarchy is a corrective focus, illustrated in (10)a; see Zimmermann (2008). The probability that a focus is realized with a marked syntactic or prosodic structure increases with the strength of the focus, see Calhoun (2010:19) and Féry (2013) for this claim. However, even though the different kinds of foci can differ in strength according to the way they are elicited, it is assumed that they can all be analysed semantically as triggering alternative sets.

Focus can also be associated with focus-sensitive particles like exclusive *only*, additive *also* and scalar *even*, among other focus-sensitive elements. In such a case and following Jackendoff (1972), we talk of 'association with focus'. Particles like these are semantic operators whose interpretational effects depend on focus.

(12) a. In October, Eduardo *also* saw [Véronique]$_F$
b. In December, Eduardo *even* travelled to [Florida]$_F$

(13) {What happened?}
a. [Mary ate a lemon PIE]$_F$, and [Sue did TOO]$_F$ (ate a lemon pie).
b. [Mary]$_{TOP}$ [ate a lemon PIE]$_F$, and [Sue]$_{TOP}$ [did TOO]$_F$

In (12) and (13), the association with focus does not contribute to the (communicated) truth conditions of the sentence but adds a 'side condition' (conventional implicature). With or without *too*, the second part of the sentence means that Sue ate a cookie. In addition to its pragmatic meaning, the particle *too* may fulfil the role of a carrier of pitch accent (Féry 2012) if everything else in the sentence is given. Observe that in the second part of (13), *ate a lemon pie*, the entire VP is thus given since it just repeats what has already been expressed in the first part of the sentence. In fact, it is readily deleted in such a context. By virtue of being given, it has to be deaccented (Féry & Samek-Lodovici 2006), and, as a result, it cannot carry the pitch accent. The same is true of the placeholder *did*. In both cases, *too* provides a locus for the focus on the affirmative part of the sentence. In (13)b, where the subjects *Mary* and *Sue* are (contrastive) topics, *too* is obligatorily present.

In some cases, a focus is truth-conditional, in the sense that the position of the pitch accent changes the meaning of the sentence. This is illustrated in (14), and in (15), an example from Dretske (1972). These sentences are clear cases of alternatives triggered by

the place of the focus. In one case Clyde has to marry a specific person to be eligible for the inheritance, and in the other case it is left open whether he could marry somebody other than Bertha to achieve the same aim.

(14) This summer, Eduardo *only* met [Monique]$_F$

(15) a. Clyde had to marry [BERTHA]$_F$ (and nobody else) in order to be eligible for the inheritance.
b. Clyde had to [MARRY]$_F$ Bertha/someone in order to be eligible for the inheritance.

To sum up, in the Roothian version of the theory, focus in English has a uniform semantic import related to 'contrast' within a set of alternatives. The key is an interpretation principle introducing a variable (a contrasting element). This variable can be anaphoric to a variety of pragmatic and semantic objects, resulting in a variety of focus-sensitive effects.

Topic: A 'topic' is analysed here as a referent which the remainder of the sentence is about, possibly contrasting with other referents under discussion, and crucially followed by a comment, which itself contains a focus constituent.[6] The topic has often, but not necessarily, been previously introduced into the discourse. A topic constituent is characterized as in (16); see also Reinhart (1981), Jacobs (2001) and Krifka (2008:265) for alternative definitions.

(16) A topic is a denotation of a referential expression about which the remainder of the sentence expresses a proposition.

Compare the following examples with *aboutness* topics. In (17)a, the topic is *Angelina Jolie* and in (17)b, it is *Brad Pitt*. *Top* stands for topic, and *Com* for comment. In a sentence with a topic, the remainder of the sentence is the comment, the counterpart of the topic, which also contains the focus of the sentence. In (17), the focus can be the entire VP, the verb, or the object.

(17) a. [Angelina Jolie]$_{TOP}$ [married Brad Pitt]$_{COM}$
b. [Brad Pitt]$_{TOP}$ [married Angelina Jolie]$_{COM}$

Besides aboutness topics, a distinction usually made in the literature between different kinds of topics is that of frame-setting and contrastive topics.[7] A large part of the literature on topics addresses the contrastive topic. Contrastive topics resemble focus; consider the example in (18).

(18) A: What do your children do?
B: [My [DAUGHTER]$_F$]$_{TOP}$ [studies law]$_{COM}$, and [my [SON]$_F$]$_{TOP}$ [wants to travel to Brazil]$_{COM}$.

According to Büring (2003, 2015), Tomioka (2010) and Constant (2014) among others, a contrastive topic elicits a set of alternative propositions, which are explicitly not used for an exhaustive answer. Büring (2015) wrote: 'From the mere existence of such non-excluded alternatives, a hearer can deduce that the speaker must find these alternatives

[6] Steedman (2000) speaks of theme and rheme which are largely equivalent to topic and comment.
[7] Familiarity topics are also sometimes mentioned. However, these are equivalent to given aboutness topics, and do not form a category by themselves in the present classification.

potentially relevant, and at least possible (otherwise she would have explicitly excluded them, i.e. have used a focus instead).' And '[w]hereas [focus] relates a declarative sentence to alternative propositions, [contrastive topic + focus] relates it to alternative questions.' In other words, the typical reason why the presence of alternatives is highlighted is to indicate that the actual utterance does not deliver all the information that is expected. This is why we often find contrastive topics indicating a strategy of incremental answering in the Common Ground management, as exemplified in (19), in which an issue is split into sub-issues. This has been assumed to be the function of contrastive topics in Roberts (1996) and Büring (1997, 2003). In the following case, the contrastive topic accommodates a more general question. 'CTop' stands for 'contrastive topic'.

(19) {Who was where at the time of the murder?}
[I]$_{\text{CTOP}}$ [was [at HOME]$_{\text{F}}$]$_{\text{COM}}$

Such a partial answer to a question may be the best the speaker can come up with in a particular context. An answer like (20)B with *my son* as a topic and the affirmative part of the answer as the focus clearly implies that Mary does not speak French, at least assuming that the speaker knows whether Mary speaks French and has no reason to conceal this fact to the hearer. Another option is to realize a focus on *son*, and deaccent *does*.

(20) A: {Does Mary speak French?}
B: [My SON]$_{\text{TOP}}$ [[DOES]$_{\text{F}}$]$_{\text{COM}}$
C: [My SON]$_{\text{F}}$ [does]$_{\text{G}}$

A contrastive topic resembles a focus in indicating the presence of alternatives, in this case, alternative topics. Both focus and contrastive topic are marked by an accent, but in the case of topic, the accent is not the main one of the sentence. Rather the main accent is on a constituent of the comment; in (20)B, it is the affirmative part of the answer.

As mentioned above, another kind of topic is the frame-setting topic (Chafe 1976, Jacobs 2001, Krifka 2008). Following Chafe (1976), frame setting is used 'to limit the applicability of the main predication to a certain restricted domain'; see the examples in (21) and (22) ('FrTop' stands for 'frame-setting topic'). They relate to CG management, as they imply that there are other aspects for which other predications might hold. In this they are similar to contrastive topics, as they too split up a complex issue into sub-issues.

(21) A: How is Pamela?
B: [Healthwise/As for her health]$_{\text{FrTop}}$, she is [FINE]$_{\text{F}}$

(22) A: How is business going for the car industry?
B: [In GERMANY]$_{\text{FRTOP}}$ the prospects are [GOOD]$_{\text{F}}$
but [in FRANCE]$_{\text{FRTOP}}$ they are [losing MONEY]$_{\text{F}}$

Topics are usually sentence-initial, and this in most languages. But in some languages, and in some cases, they can be non-initial, as well. Consider (23), an example from Japanese, where the topic constituent is located after the quantified phrase (from Tomioka 2010).[8]

[8] Topics in Japanese are usually initial and followed by the particle *wa*.

(23) [(Daremo-ga)$_{\Phi}$ (dezaato-wa)$_{\Phi}$ (aisu-o tabeta)$_{\Phi}$]$_{\iota}$
everyone-NOM dessert-TOP ice-cream-ACC ate
'As for dessert, everyone ate ice cream.'

Other languages can have 'anti-topics' as in Cantonese (24)a and French (24)b, that are located at the end of the sentence, but have a similar information structural role as sentence-initial topics.[9]

(24) a. [(Go loupo)$_{\Phi}$ (nei gin-gwo gaa)$_{\Phi}$, (ni go namjange)$_{\Phi}$]$_{\iota}$
CLF wife 2.sg see-exp PTC this CLF man.MDF
'The wife you have seen, of this man.'
b. [(Pierre)$_{\Phi}$ (l' a mangée)$_{\Phi}$, (la pomme)$_{\Phi}$]$_{\iota}$
Peter it-acc has eaten, the apple
'Peter has eaten the apple.'

Finally, sentences may have no topic constituent at all, in which case they are called *thetic*, following Marty (1918); see the examples in (25).

(25) a. [The HOUSE is on fire]$_{\text{COM}}$
b. [My WALLET is gone]$_{\text{COM}}$
c. [The CAR broke down]$_{\text{COM}}$

Thetic sentences have the property to have only one sentence accent, and are often illustrated in the literature with unaccusative intransitive constructions, such as those in (25). But in fact, they need not be intransitive. Fanselow & Lenertová (2010) show that in many languages, the highest argument of the verb can be fronted and accented in a wide focus reading, like in their example (26). Notice that *ein Buch* 'a book' is not a topic in this sentence, as it does not need a referent. The focus is the entire VP *read a book*.

(26) {What do you want to do in your holiday?}
[Ein BUCH]$_{i}$ würde ich gerne t_{i} lesen.
a book would I with.pleasure read
'I would like to read a book.'

Some authors (see for instance Erteshik-Shir 1997) have called the unexpressed spatio-temporal contextual component of such sentences topical, but I prefer to call such sentences topicless, in order to set them apart from their counterparts with a realized topic.

After this rather cursory introduction to the information-structural notions, the prosodic and intonational means of expressing them are elaborated in the next subsection.

6.2.2 Prosodic and intonational reflexes of information structure

Information-structural roles have a variety of grammatical reflexes in different languages, only some of which have to do with intonation.[10] In Féry (2013), it was proposed that focus seeks to be aligned with the edge of a Φ-phrase or an ι-phrase. In other words, a focus seeks to be phrase-initial or -final. Another way to understand the tendency for a focus to be aligned with a prosodic domain is to interpret it as the need for a focus to be phrased as independently from the remainder of the sentence as possible. This tendency is

[9] Anti-topics resemble constructions called right-dislocation.
[10] Changes in word order, pronominalization, changes in morphology, and discourse particles are further common ways to express information structure.

even stronger for a topic, which tends to be phrased separately. The given or backgrounded part does not show the same tendency, and the result is that given parts of a sentence (in the information-structural sense) may both intervene between topic and focus and be located at the beginning or end of the sentence.

In this section, it will be shown how information structure affects prosody and intonation in West Germanic languages. The grammatical reflexes of information structure differ from language to language within this family, but there are evident similarities among them, especially in comparison to other families.[11]

In a nutshell, focus increases the prominence of a pitch accent and givenness decreases it, more radically so in the post-focal position of the sentence than in the pre-focal position. An aboutness or contrastive topic is generally the first element of the sentence, and it is prosodically separated from the following part of the sentence. In German, a topic is generally located in the sentence-initial position. It is bounded by a boundary tone, generally with a rising contour.

Let us begin with focus, and come back to the examples in (5) for illustrations of the pitch accent strategy. These examples are reproduced and expanded in (27).

(27) {Who ate a lemon pie?}
a. [MARY]$_F$ ate a lemon pie.
{WHAT did Mary eat?}
b. Mary ate [a LEMON PIE]$_F$
{What did Mary do?}
c. Mary [ate a LEMON PIE]$_F$
{What happened?}
d. [Mary ate a LEMON PIE]$_{AN}$

In (27)b-d, the object carries the pitch accent of the sentence. This location of the sentence nuclear accent is ambiguous: the focus may be on the object as in (27)b, on the VP as in (27)c, or even on the whole sentence as in (27)d. The latter version is called 'all-new' or 'broad' or 'wide' focus. The subscripted letters AN stand for 'all-new'. The different focused parts of the sentence illustrate the variable size of a focus, which is not always inferable from the position of the pitch accent. The predictable location of the pitch accent in larger constituents, also observed by Chomsky (1971), is called *focus projection*, and is part of the theoretical syntax-prosody mapping in different frameworks (see Selkirk 1995 for focus projection rules; and also Arregui 2015 for an overview). In (27)a, when the nuclear pitch accent is on the subject, the focus can only be located on the subject. In this case, it cannot project onto larger constituents, but it is limited to the constituent carrying the sentence main accent.

A similar case of accent placement in two different focus structures and two different presupposition structures is illustrated in (28): the first reading can be paraphrased in the following way: when she eats, Zoe eats candies; and the second reading means: if somebody eats candies, then it is Zoe.

(28) a. Zoe always eats CANDIES.
b. ZOE always eats candies.

[11] German, Dutch, Greek, Slavic, Baltic and many other European languages present more or less the same properties, so that English is used as an example for all these languages. Many European and non-European languages use other strategies, one of which is alignment of focus to the right or to the left of a prosodic phrase, or morphological markers. Most languages use different strategies for diferent cases. We will address the typological diversity of prosodic means in chapter 8.

Different scopal relationships between quantifier and negation in German may be expressed by means of pitch accent assignment, as illustrated in (29): in the first case, when one accent is realized on *beide* 'both' and another one on the negation *nicht*, the sentence is interpreted with the negation taking scope over the quantifier *beide*, and thus expresses that at most one of the two came. In (29)b, there is only one accent, viz. on the quantifier *beide*, and the remainder of the sentence is deaccented. In this case, the quantifier has scope over the negation and the sentence means that neither one came (see Krifka 1998 for a semantic analysis). In such a case, pitch accent assignment not only relates to the focus-givenness relationship, but expresses scopal dominance as well.

(29) a. BEIDE sind NICHT gekommen.
both are not come
'Both didn't come.'
b. BEIDE sind nicht gekommen.

A special case of focus has been called 'verum focus' by Höhle (1992); see also Lohnstein (2015). Verum focus, which must be distinguished from *predicate focus*, denotes a constellation in which only the affirmation in a sentence is focused. In the other case, the predicate of a sentence is focused. In English and German, verum focus is realized with a pitch accent on the finite part of a sentence, as in (30).

(30) A: {Bill told me that Mary ate 56 lemon pies. This can't be true!}
B: Yes, she DID (eat 56 lemon pies).

A peculiarity of verum focus is that, in order to be perceived at all, it usually excludes pitch accents on other constituents. If in (30) a second pitch accent is realized on *lemon pies*, this accent may obscure the accent on the auxiliary.

Turning to the intonational expression of givenness in English and German, languages have at their disposal a large number of specific anaphoric expressions and other connectors for the expression of givenness: personal pronouns, clitics, demonstratives and definite articles; see (31) with *she* and *it*. The use of such devices also helps to establish coherence in discourse. Intonationally they are often deaccented.

(31) Mary_i saw a lemon PIE_j and she_i decided to EAT it_j.

An accent on a pronoun implies an unusual anaphoric relation between pronoun and antecedent, as in Halliday's (1967) example in (32), also cited in Lakoff (1971). The parallel principle, which is not fulfilled in this example, posits that '[a] speaker assumes that the hearer will predict, unless there is evidence to the contrary, that (a proper part of) a new (conjoined) construction will be parallel/equivalent in some semantic/pragmatic way(s) to the one just processed' (from Prince 1981:228). If the pronouns *he* and *him* are accented in (32) and the verb *insulted* is deaccented, a hearer assumes that the anaphoric relationship between the pronouns and antecedents is reversed.

(32) John called Bill a Republican and then HE insulted HIM.

The most well-known device to mark constituents as given is deaccentuation. Some examples appeared above, as in (27), (28)b and (29)b. In (33), adapted from Ladd (1980:52), the default nuclear pitch accent placement is on *towns*. By mentioning *Brussels*, *towns* is rendered accessible or inferable by speaker A (see (3) above). As a result, it is not the best carrier of pitch accent for speaker B. The other accentable element

in the Φ-phrase formed on the VP is the verb, which, as a result of the unaccentedness of *towns*, gets the nuclear accent. Notice that, for the sake of nuclear accent assignment, it does not matter whether *towns* is discourse-given, context-given, or accessible. As far as accent assignment is concerned, it is unstressable (see Wagner 2005 and Constant 2014 for interesting restrictions on what kind of constituents can be deaccented in such a context).

(33) A: {Does John know *Brussels*?}
B: John [doesn't VISIT [towns]$_G$]$_\Phi$

Other grammatical devices, such as word order changes and deletion, are also available to mark constituents as given. The answer can be just a sentence fragment (see Merchant 2004, Vicente 2006, Winkler 2014 among others).

From these remarks, it can be gleaned that, if an element is deaccented or deleted that would be accented if the sentence were all-new, it is interpreted as given. The reverse does not hold. Thus, it is not the case that, if an element is given or accessible, its prosody and accentuation can be predicted: discourse givenness has no invariant prosody. A given referent can still be a topic or a focus, as in example (34), adapted from Schwarzschild. Even though *Mary* is already mentioned in the preceding question in (34), it is still the focus in the answer, because it answers the wh-question. Here the focus wins and the constituent is accented.

(34) A: {Who did Mary's grandmother greet first?}
B: She greeted [MARY]$_{G/F}$ first.

In other situations, as for instance in Second Occurrence Focus (SOF), where an element is also both focused and given, givenness may win in certain circumstances. The example in (35) is adapted from Partee (1999: 215-216).

(35) A: {Jane knew that Mary only eats [vegetables]$_F$}
B: Even [Paul]$_F$ knew that Mary only eats [vegetables]$_{SOF}$.

When the SOF is in the post-nuclear position, as in (35), it is deaccented (see Beaver et al. 2007, who found no pitch accent in this position in English).[12] The difference between deletion of pitch accent in (35) and retention in (34) has to do with the fact that in (34), the given element is the sole focus of the sentence, whereas in the case of SOF, there is another (new) focus that carries primary accent. In Féry & Ishihara (2009), it is shown that in a language with free word order, like German, the SOF can also be in the pre-nuclear position, in which case, it is pitch accented. This illustrates the asymmetry in the deaccenting of given elements in the post-nuclear position vs. retention of pitch accent in the pre-nuclear position.

In a German all-new sentence such as (36), a sentence from the experiment reported in Féry & Kügler (2008), every adjunct and every argument is mapped to a Φ-phrase, and the last one includes the predicate (see chapter 4). This holds for English as well, but German, being verb-final, illustrates the integration of the predicate in the adjacent Φ-phrase better. Because the verb is post-nuclear, it is unaccented, and as a result, its F0 is low and flat. In the pitchtrack of this sentence in figure 6.2, every high tone is downstepped relatively to the preceding one. The first (non-final) two arguments have a rising contour analysed here as a pitch accent L* associated with the lexically

[12] They found that the SOF is 6 ms longer than the same word when it is just given.

stressed syllable of each ω-word, the first syllable in the animal names, and a high boundary tone H_Φ. The last pitch accent is the nuclear one, and it has a falling pattern in a declarative sentence. In figure 6.2, the nuclear tone is an instance of an early peak: the highest tone is realized on the syllable preceding the nuclear accent, here on the determiner *den* 'the.ACC'. The pitch accent is analysed as HH* and it is followed by a low boundary tone L_Φ. The last tone of the sentence is a low L_ι. Between these two low tones, the F0 on the verb remains flat and low. The lower-ranked boundary tones L_Φ and H_Φ are aligned with the syllables immediately following the pitch accents, see section 8.2 for metrical structure and assignment of tones in German intonation.

```
                          L*  HΦ            L*  HΦ    H      H* LΦ                 Lι
                          |   |             |   |     |      |  |                  |
(36) [(weil     der       Löwe)Φ (dem      Hammel)Φ ( den    REIHER  vorgestellt  hat)Φ]ι
     because  the.NOM   lion     the.DAT  sheep     the.ACC heron   introduced   has
     'because the lion introduced the heron to the sheep'
```

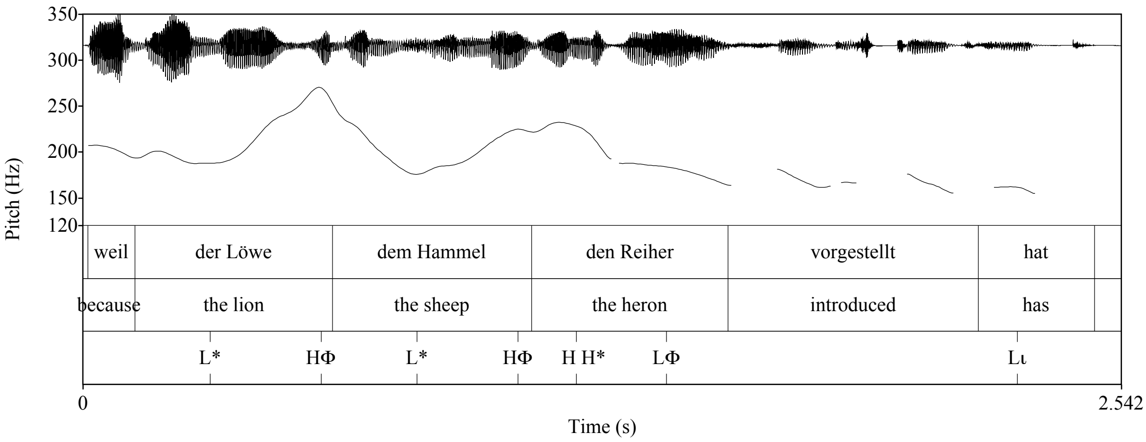


Fig. 6.2 Pitch track of an all-new German sentence

6.2.3 Focus and tonal scaling

Turning now to the effect of information structure on tonal scaling, focus has a raising effect on the F0 of the focused word; see the Focus Rule of Jackendoff (1972:247) in (37), which expresses F0 raising in terms of 'highest stress', and see also Truckenbrodt (1995) for the notion of focus prominence.

(37) If a phrase P is chosen as the focus of a sentence S, the highest stress in S will be on the syllable of P that is assigned highest stress by the regular stress rules.

However, the effect of focus is not limited to assigning highest stress on the focus. It also changes the relationship between the different parts of a sentence.

Let us compare the different effects of focus and givenness on tonal scaling. We saw in chapter 4 that the syntactic structure of a sentence has an influence on the formation of prosodic phrases. What was missing there is the tonal scaling of the prosodic domains relative to each other, as a consequence of information structure. First, the neutral pattern of an English and a German sentence looks like in figure 6.3a, illustrating a sequence of two Φ-phrases. Each curved line stands for the main bitonal pitch accent of its Φ-phrase, and the thick 'staircase' line shows downstep between the Φ-phrases in an ι-phrase. This is represented by downstepping the high reference line of each Φ-phrase, as discussed in chapter 5. The first pitch accent and/or the main pitch accent of each Φ-phrase reaches the high reference line. Figure 6.3b illustrates F0 raising as a consequence

of narrow focus in the first Φ-phrase and F0 lowering because of givenness in the second Φ-phrase; figure 6.3c shows raising because of narrow focus in the second Φ-phrase and F0 lowering because of givenness in the first Φ-phrase. It must be noted however that post-focal compression in b. is more radical and complete than pre-focal reduction of register.

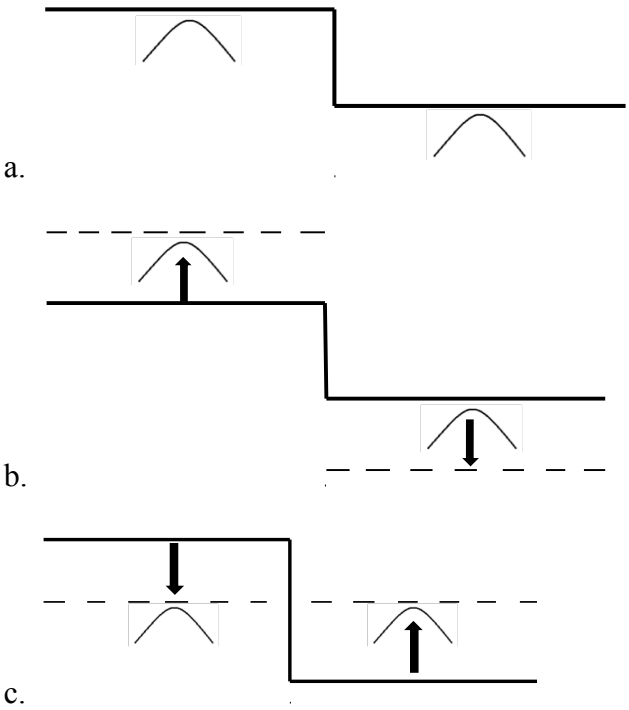


Fig. 6.3 Tonal scaling in focused and given Φ-phrases

The next sentence in (38) illustrates the effect of an early focus, here on the subject, on the post-nuclear material. The reference line of a post-nuclear given Φ-phrase is lowered to a minimum, as illustrated in figure 6.3b As a result, the register is reduced and the lexical stresses are reduced and deacented (see Féry & Samek-Lodovici 2006 and Kügler & Féry 2016). This is visible in figure 6.4.

(38)

		H* $L_Φ$						$L_ι$
		| |						|
[(weil	der	Löwe)$_Φ$	(dem	Reiher)$_Φ$	(den	Hammel	vorgestellt	hat)$_Φ$]$_ι$
because	the.NOM	lion	the.DAT	heron	the.ACC	sheep	introduced	has

'because the lion introduced the sheep to the heron'

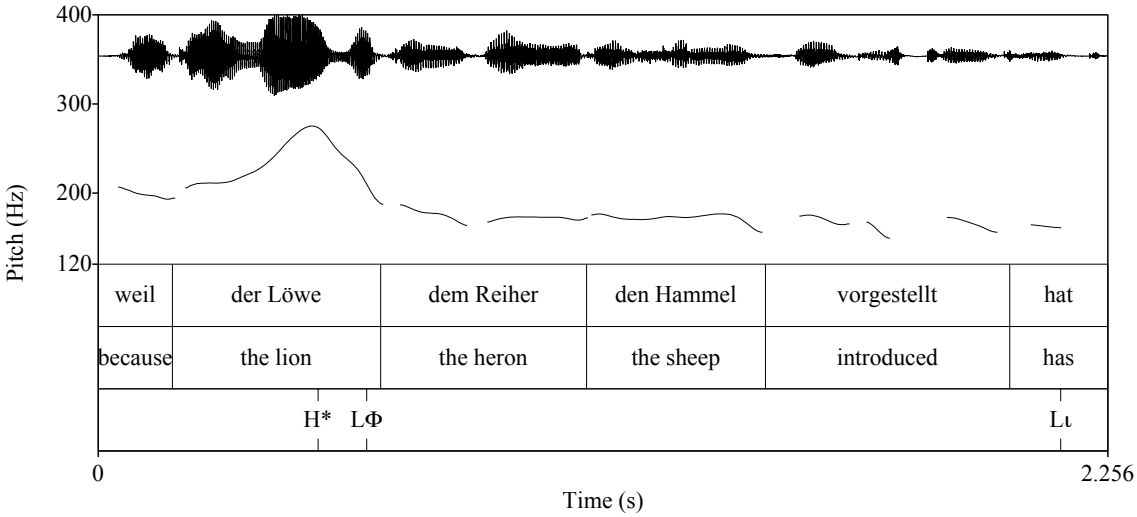


Fig. 6.4 Pitch track of a German sentence with narrow focus on the subject and deaccented post-nuclear material

By contrast, in the pre-nuclear part of a sentence, no deaccenting applies. The prenuclear accent may be less prominent than in an all-new realization, but not necessarily so (see chapter 4). In example (39), the pre-nuclear accent on given material is not deleted, as illustrated in figure 6.5, see Féry & Kügler (2008).

(39)

		L* H_Φ		H* L_Φ			L_ι
[(weil	der	Hammel)$_\Phi$	(dem	Rammler	nachgelaufen	ist)$_\Phi$]$_\iota$	
because	the.NOM	sheep	the.DAT	buck	followed	is	

'because the sheep followed the buck'

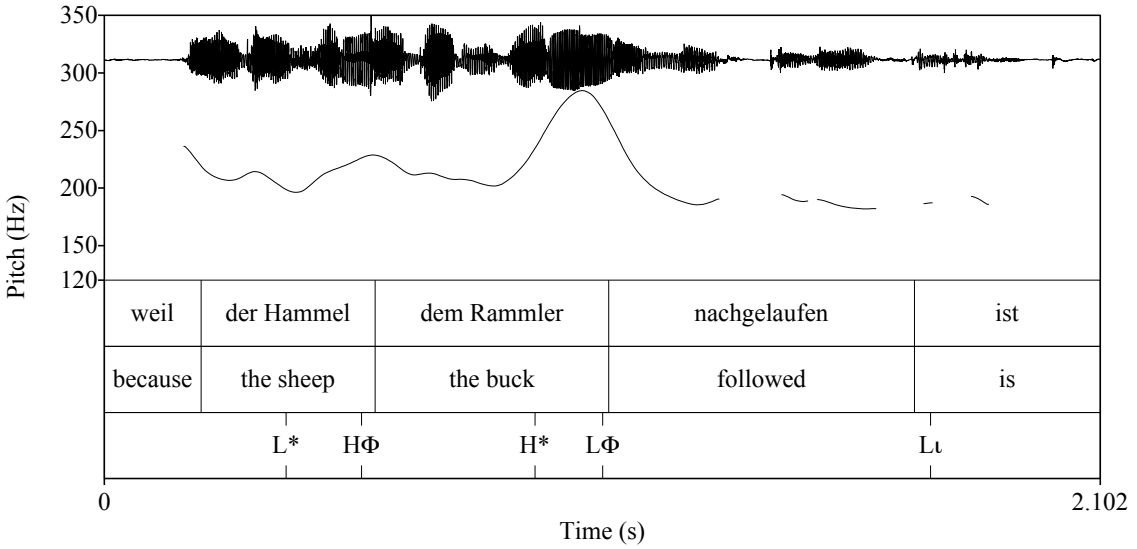


Fig. 6.5 Pitch track of a German sentence with pre-nuclear accented given material

The asymmetry in the deaccenting strategy of languages such as German is accounted for by the alignment effect of focus. Metrical heads are preferably aligned with the right edge of an ι-phrase (see Selkirk 2000, Truckenbrodt 2007, Féry 2011, 2013). If they are not, for example if there is additional given material after a narrow focus, deaccenting of the post-

nuclear material occurs, as illustrated in figure 6.4. The pre-focal material is left untouched or only slightly compressed because additional accents in this position do not affect right alignment, as illustrated in figure 6.5.

Information structure has a packaging effect that can be felt in syntax and prosody. It has been shown that focus and givenness affect the prosodic domains mapped to the syntactic constituents bearing the information structural roles, and pitch accents only adjust to the changed register of these Φ-phrases, as illustrated in figure 6.3. This view is best compatible with an indirect reference approach to the syntax-prosody interface, since the information structure changes the tonal scaling of independently created prosodic domains.

6.3 Meaning of tones and tonal contours

How exactly a speaker chooses to intonate a sentence is important for the communication of what is being said, in addition to the content of the sentence. For instance, the melodic pattern conveys additional information as to the modality of the sentence, its declarative or interrogative status, and as to the speaker's attitude. And conversely, the type of illocution and speech act conveyed by the speaker affects the choice of contours and individual tones, the melody thus being partly governed by interclausal dependencies (coherence of what is being said) and by interactive attitudinal aspects. To cite just one example that will be elaborated below, it has been repeatedly proposed in the literature – though in different guises – that rising vs. falling intonation on declaratives is responsible for attribution of the commitment of what is being said to the addressee vs. the speaker, respectively (see Gussenhoven 1984, Bartels 1999, Pierrehumbert & Hirschberg 1990, Gunlogson 2003 and Truckenbrodt 2012 among others). Moreover, intonation also provides additional information on the mood and emotional state of the speaker, like cheerfulness, anger, joy, fear or sadness (see section 6.4). In sum, the same sentence can come in a variety of nuances according to its tonal pattern. The way tones are shaped by pragmatic features is rather complex, and this chapter can only give some elements of the influence of pragmatics on intonation, without going into the details of a semantic and pragmatic analysis. It is sometimes claimed that tones are morphemic and are the smallest bearers of contrast in the tonal system.[13] From this perspective, a single tone can change the meaning of the entire contour. Tones combine with each other to form bi- or tritonal contours. In English, tones are not lexical in the sense that they do not primarily reflect the speaker's attitudes all by themselves (see also Pike 1945, O'Connor & Arnold 1961, Sag & Liberman 1975, Liberman & Sag 1974, and Ward & Hirschberg 1985 for this insight). Nor do they carry emotive meanings all by themselves (Bolinger 1982). They need a text, a context and a compositional analysis in order to be meaningful.

In the discussion about the meaning of tones and tonal contours in this section, a distinction will be made between 'compositional' approaches and other approaches. According to compositional approaches, tones have morphemic meanings and the meanings of tone sequences arise from the way they are combined. Among the other approaches, some are 'holistic' in the sense that entire contours are taken to be meaningful, and some can be considered intermediate between compositional and holistic.

[13] This is for instance the view of Liberman (1978:94). Liberman found that intonation consists of ideophonic 'words' or morphemes: 'The meanings of these [ideophonic] words are extremely abstract properties, which pick out classes of situations related in some intuitively reasonable, but highly metaphorical way: the general meaning seems hopelessly vague and difficult to pin down, yet the application to a particular usage is vivid, effective, and often very exact.'

6.3.1 Compositional analysis of tonal meaning

Pierrehumbert & Hirschberg (1990) proposed the first compositional analysis of melodic contour meanings within the tone-sequence framework. They claimed that the meaning of sentence intonation is the result of the concatenation of individual tones, each of which contributes its own semantic content. However, it is fair to say that even the strictest tone-sequence approach, like the one of Pierrehumbert & Hirschberg, which assigns meanings to every individual tone, needs sequences of tones (and appropriate texts) to illustrate these meanings, as will become clear in this section. The basis of Pierrehumbert & Hirschberg's compositional analysis is that tunes of the same shape convey the same relationship between the proposition expressed by the textual sentence and what they call *mutual beliefs*, following Clark & Marshall (1981) and Joshi (1982), that is, the knowledge assumed by the speaker to be shared (or 'believed') by speaker and addressee. In order to be consistent with the remainder of this chapter, the *Common Ground* (CG, Stalnaker 1974, 2002) will replace mutual beliefs (even though there are differences in general). The CG forms the background – the information that is mutually known (or believed) to be shared by speaker and addressee – and to which new information is added. As a result, the CG is continuously being modified in communication (see section 6.2 on information structure, where the CG has already been mentioned).

A crucial aspect of Pierrehumbert & Hirschberg's analysis is the distinction between pitch accents, phrase accents, and boundary tones. Pitch accents are decomposed into their atomic components: H* and L* mark the accented element; H* stands for *addition* to the Common Ground,[14] while L* *excludes* the accented element *from the predication*. More precisely, H* indicates information that should be treated as 'new' in the discourse in that 'intonational phrases whose accents are all H* appear to signal to [the] H[earer] that the open expression is to be instantiated by the accented items and the instantiated proposition realized by the phrase is to be added to [the CG]' (p.289f). L* 'marks items that [the] S[peaker] intends to be salient but not to form part of what [the] S[peaker] is predicating in the utterance' (p.291).

Let us examine some of their examples. The notation is adapted to the one used in this book. In the original, phrase accents are not indicated as such, and it is not always immediately clear whether a tone without a diacritic is a trailing tone of a bitonal tone or a boundary tone. The examples below will be disambiguated in this respect. In (40), *x likes y* is an open proposition, and the variables *x* and *y* are replaced by appropriate nouns. There is an H* accent on *Georges* and an H* on *pie*, conveying that they are new information (additions to the CG). Adding the low boundary tones L_Φ and L_ι renders the whole sentence a neutral declarative sentence; see below for the meaning of boundary tones.

(40) a. [(GEORGES likes PIE.)$_\Phi$]$_\iota$
H* H* L_Φ L_ι

b. *x likes y*
x (H*) *x = Georges*
y (H*) *y = pie*

c. H*: new information

d. H* L_Φ L_ι: neutral declarative sentence

An H* accent is also possible in vocatives and chanted contours, as for instance on the preposed vocative in (41)a. This contour is chosen if the speaker does not already have

[14] For Gussenhoven, it is the complete falling nuclear tone that expresses addition (see below). For Truckenbrodt (2012:2039) H* 'requires a salient proposition that the speaker is adding to the common ground.'

the hearer's attention. In postposed vocatives, as in (41)b, the contour is flat, and is used if the speaker does have the hearer's attention.

(41) a. [(ANNA)$_{\Phi}$ (your LUNCH is READY.)$_{\Phi}$]$_{\iota}$
H* L$_{\Phi}$ H* H* L$_{\Phi}$ H$_{\iota}$
b. Your LUNCH is READY ANNA.
H* H* L$_{\Phi}$ L* L$_{\Phi}$ H$_{\iota}$

Since the pitch accent L* indicates that the element with which it associates is prominent but not part of the predication, it is commonly used in yes-no questions, as in (42), my example.

(42) [(Is it RAINING?)$_{\Phi}$]$_{\iota}$
L* H$_{\Phi}$ H$_{\iota}$

The next example with L* followed by H$_{\Phi}$ in (43) is to be imagined in the context of a joke: an old Bolshevik is forced to confess publicly and formulates his confession in the form of rising declaratives. Every Φ-phrase has a rising pattern; see the similar example in (56).

(43) [(I)$_{\Phi}$ (was WRONG)$_{\Phi}$ (and STALIN)$_{\Phi}$ (was RIGHT)$_{\Phi}$]$_{\iota}$ [(I)$_{\Phi}$ (should APOLOGIZE)$_{\Phi}$]$_{\iota}$
L*H$_{\Phi}$ L* H$_{\Phi}$H$_{\iota}$ L* H$_{\Phi}$ L* H$_{\Phi}$H$_{\iota}$ L*H$_{\Phi}$ L* H$_{\Phi}$H$_{\iota}$

The L* can also be 'extrapropositional' when used in ritualized greetings and vocatives, such as illustrated in (44). The speaker is probably supposed to already have the attention of the hearer. There is no new content that the hearer could add to the CG.

(44) [(GOOD MORNING.)$_{\Phi}$]$_{\iota}$
L* L*L$_{\Phi}$ H$_{\iota}$

As illustrated in chapter 5, bitonal pitch accents consist of the starred pitch accent plus a leading tone or a trailing tone. LH can be L*H or LH*. In both instances, LH 'conveys the salience of some scale linking the accented item to other items salient in the H[earer]'s mutual beliefs' (p.294), as illustrated with L*H in (45). A word associated with this pitch accent is an element of a scale, but it is not part of a predication: it expresses uncertainty. In (45), an example originally from Bing (1979), 'B expresses uncertainty about whether being a good badminton player provides relevant information about degree of clumsiness.'

(45) A: Alan's such a klutz.
B: [(He's a good BADMINTON player.)$_{\Phi}$]$_{\iota}$
L*H L$_{\Phi}$ H$_{\iota}$

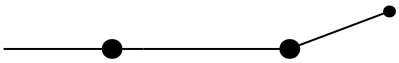
An example for LH* appears in (46). Pierrehumbert & Hirschberg (1990:296-7) commented upon it as follows: 'The meaning assigned to this exchange is also "contrastive" – something like "As for the beans, Fred ate them. As for the other food, other people may have eaten it". Here, B's answer is felicitously produced in two phrases, *Fred* and *ate the beans*; the second, representing the topic, has a fall-rise pattern on *beans*.' They went on to comment on 'S's commitment to a particular instantiation of an open expression with an item chosen from a salient scale – here, a set of salient foods.' We saw in the preceding section that such a non-exhaustive answer is typical for a

contrastive topic (see also Steedman 2014 and the discussion of Bolinger's accents A and B in the next section).

(46) A: What about the beans? Who ate them?
B: [(FRED)$_{\Phi}$ (ate the BEANS.)$_{\Phi}$]$_{\iota}$
H* L$_{\Phi}$ LH* L$_{\Phi}$ H$_{\iota}$

H*L and HL* are not distinguished from H*, the predication is the same in both cases. The only difference is that the element carrying HL should be directly or indirectly inferable from the CG, and the inference is pragmatic in character. The addressee is supposed to know the inference path. The original examples are difficult to interpret, and I do not reproduce them here.

Turning next to phrase accents and boundary tones, these relate to boundaries of prosodic phrases and intonation phrases and are thus not associated with accented syllables. They can be high or low. In (47), what Pierrehumbert & Hirschberg called 'a high rise question' in the British school style is terminated by two H tones. It is equivalent to the rising declarative discussed by Gunlogson (2003, see section 6.3.2), with a declarative syntax and a question intonation.



(47) [(My NAME is Henry WATSON)$_{\Phi}$]$_{\iota}$
H* H* H$_{\Phi}$ H$_{\iota}$

A distinction between an alternative and a disjunctive reading of questions containing two elements can be expressed by a phrase accent located between two phrases. A high phrase accent (H$_{\Phi}$) indicates that the current phrase is part of a larger interpretive entity with the following phrase. By using an H$_{\Phi}$ on *apple juice* in (48)a, 'the S[peaker] emphasizes that *apple juice* and *orange juice* form an entity, namely, the set of available juices' (p. 302). This is the alternative reading. The low phrase accent (L$_{\Phi}$) emphasizes the separation of the current phrase from the following phrase. By using L$_{\Phi}$ in (48)b, '[the] S[peaker] emphasizes the separate status of each type of juice and thus does not evoke a larger interpretive entity' (p.302). This is the disjunctive reading.

(48) a. [(Do you want APPLE juice)$_{\Phi}$ (or ORANGE juice?)$_{\Phi}$]$_{\iota}$
H* H$_{\Phi}$ H* L$_{\Phi}$ L$_{\iota}$

b. [(Do you want APPLE juice)$_{\Phi}$ (or ORANGE juice?)$_{\Phi}$]$_{\iota}$
H* L$_{\Phi}$ H* L$_{\Phi}$ L$_{\iota}$

Due to the role of H$_{\Phi}$, in (49) a causal link is readily established between the opening of the door and the pouring of the rain. If a low phrase accent is used instead, this causal link is not available.

(49) [(I OPENED the DOOR)$_{\Phi}$ (and the RAIN POURED down.)$_{\Phi}$]$_{\iota}$
H* H* H$_{\Phi}$ H* H* L$_{\Phi}$ L$_{\iota}$

In the same way as phrase accents express relationships between prosodic phrases, boundary tones express relationships between intonation phrases, thus whether the present intonation phrase has to be interpreted with respect to the preceding or the following

intonation phrase. In (50), *annoying* refers to the unreadable character of my car manual or to the long time spent in figuring out how to use the jack. The direction of the tune on the middle sentence according to a high or a low boundary tone helps the addressee to make the right inference.

(50) My new car manual is almost unreadable.
It's quite annoying ($H_ι$ or $L_ι$).
I spent two hours figuring out how to use the jack.

In other words, the typical L* $H_Φ$ $H_ι$ yes-no question contour should be interpreted with respect to what follows, e.g. the answer. The typical H* $L_Φ$ $L_ι$ statement contour, on the other hand, predicates new information.

In general, it can be observed that the meaning of tones, especially of boundary tones, is usually assigned in relationship to the other tones of the sentence.[15]

A second compositional analysis comes from Bartels (1999), who, like Pierrehumbert & Hirschberg (1990), was also concerned with individual tones and their pragmatic roles in a sentence. However she restricted her in-depth analysis to the phrasal tones $L_Φ$ and $H_Φ$.[16] $L_Φ$ expresses the interactive cognitive feature *assertiveness*, thus a completely different meaning from the one assigned to this tone by Pierrehumbert & Hirschbert (1990). When a speaker realizes assertiveness through $L_Φ$, she expresses an instruction to the addressee to commit herself publicly. When $L_Φ$ is absent, the speaker reveals her uncertainty. Boundary tone $H_ι$ conveys the meaning of *concessive continuation dependence*, which may lead to pragmatic inference of various kinds, for instance a question, a following rectification in the case of a contradiction, or the reinforcement of a scalar implicature in the case of certain fall-rise declarative statements. Some of Bartels' examples in which a final rise may change the speech act type are reproduced in (51).

(51) a. Paul arrived all by himself. ($L_ι$ or $H_ι$)
b. Do you like the cake? ($H_ι$ or $L_ι$)
c. Yes-no questions: Did Mary finish her homework? ($L_ι$ or $H_ι$)
d. Wh-questions: Who did you talk to last night? ($H_ι$ or $L_ι$)

An alternative question like the one illustrated in (52), can only end in a low tone, in the same way as most declaratives.

(52) Does this elevator go up or down? ($L_ι$)

Declarative questions with a final rise were described by Bartels as properly assertive with respect to the proposition. Technically, however, they behave like questions in conveying uncertainty regarding the addressee's ability to take up the assertion.

6.3.2 Partly holistic approaches

Bolinger's (1958, 1982) approach is anchored in a tradition of associating tunes with meanings (see also Pike 1945), which is perpetuated today by some authors. But Bolinger also sought to assign meanings to parts of tunes. His partly holistic approach is best

[15] For Truckenbrodt (2012:2039), 'the H- tone (often found in a rising contour) requires a salient proposition that the speaker is putting up for question'.

[16] As for the pitch accents, the L* pitch accent marks an entity as being (believed by the speaker to be) 'already evoked and salient for the listener' whereas an H* 'has the purpose of making the entity salient' (p.57).

illustrated with his discussion of the so-called 'Contradiction Contour' (CC), in figure 6.6, from Liberman & Sag (1974) and Sag & Liberman (1975), and discussed at length by Ladd (1980).

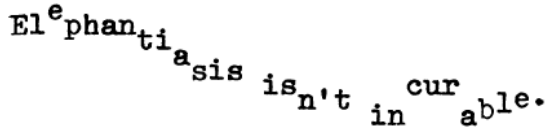


Fig. 6.6 *Elephantiasis isn't incurable* (in Bolinger's 1982:506 notation)

Several authors have suggested that this contour is not typical for contradiction and that it can be used in completely different contexts. Pierrehumbert & Hirschberg (1990:292) mentioned the sentence *Well, I'd like a Pavoni* uttered in a situation in which the speaker is asked to supply a list of things he would like for his birthday. If this sentence is said with the same contour as in figure 6.6, the speaker assumes that that the hearer already knows that he'd like a Pavoni. No contradiction is intended.

Bolinger made a division between several constitutive parts of the CC, in a compositional manner. He proposes that the CC consists of the following parts:
(a) An initial rise, cueing the hearer to the concern, interest etc. of the speaker.
(b) A high pitch on an unaccented syllable, marking the concern etc. as 'broad focus', i.e. as applying to the utterance as a whole.
(c) An immediately succeeding stepped or gradual fall, showing the tension to be under control, and therefore intended.
(d) An accented syllable at low pitch, de-emphasizing the referent of the word and contributing to the restraint of the down-motion.
(e) A terminal rise, leaving the utterance 'open' to further comment or to continuation within a larger utterance. The gradient extent of the terminal rise augments the effect of the initial rise.

According to him, intonation – and the CC is an example of this – participates on an emotive scale. Bolinger (p. 524-5) went so far as to state that the CC is proof 'that whatever else intonation may be, it cannot be divorced from the tension and relaxation of our bodies.'

Bolinger has played an important role in the discussion about the meaning of intonation, not only because of his physical and emotional interpretation of intonation, but also for his distinction between different kinds of pitch accents. According to him (see for instance 1958:147), pitch accents (as tunes) are morphemic and have a meaning. He distinguished between Accents A, B and C. Accent A is a plain fall, and is called 'assertive'. Accent B is a rise, called 'connectedness' or 'incompleteness'. Accent C is described as an 'anti-Accent A' and is qualified as 'anti-assertive'.[17]

The distinction between accent A and B has been continued by Jackendoff (1972), and by Liberman & Pierrehumbert (1984), as well as by many authors after them. In Liberman & Pierrehumbert's approach, Accent A is still a plain fall, but B is now a fall-rise. Accents A and B are illustrated in figure 6.7 with the sentence 'Anna came with Manny' as a response to 'What about Manny? Who came with him?' (left panel) and as a response to 'What about Anna? Who did she come with?' (right panel). In each case, Accent A provides the new information (a focus) and Accent B realizes the background information (aboutness topic or contrastive topic).

[17] Besides the different kinds of accents, an important property of intonation for Bolinger is the fact that it is gradient, as manifested by the post-accentual contour. For instance after an A accent, there is a gradient difference between a level and a fall. See Ladd (1980) for an in-depth discussion of Bolinger's three accents and Ladd (2014) for a lengthy discussion of Bolinger's notion of gradience in intonation.

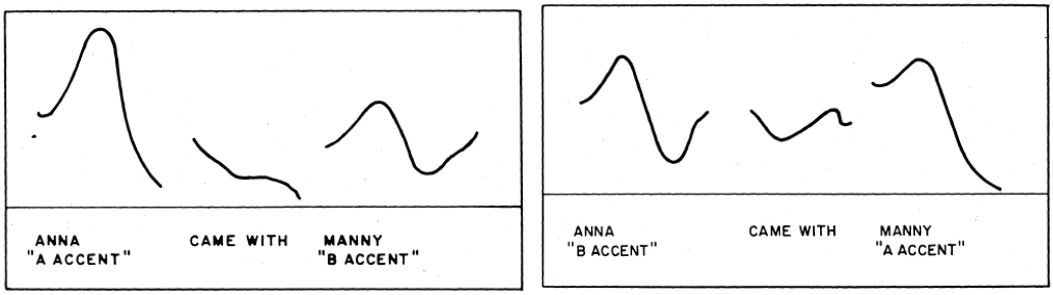
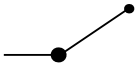
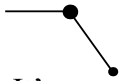
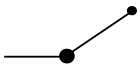


Fig. 6.7 *Anna came with Manny.* As responses to 'What about Manny? Who came with him?' (left panel) and as a response to 'What about Anna? Who did she come with?' (right panel) (from Liberman & Pierrehumbert 1984:168)

Except for the sentences in figure 6.7, the examples we have discussed so far have all been different. Each contour was illustrated with a different text. However, to investigate the import of intonation for meaning, it is also important to understand how different melodies on a single text can change its meaning. The crucial question is then: what kind of nuances are added by varying the contour on a given text? This is by far the most common strategy to investigate meanings in intonation. In (53), the same utterance varies in two ways: it can end in a rising contour, as in (53)a and (53)c, or in a falling contour, as in (53)b and (53)d. The other variation involved in (53) is the order of the subject and the auxiliary. In (53)a-b, the subject comes first (declarative syntax), and in versions (53)c-d, the verb is first (interrogative syntax). This latter order unambiguously signals that the sentence is a question. A question can have a rising or a falling intonation. (53)b is also straightforward, since it has both the syntax and the contour of a declarative sentence. The interesting case is (53)a with the declarative word order, but an interrogative contour.

(53) a.  It's over? b.  It's over. c.  Is it over? d.  Is it over?

Gunlogson (2003) calls versions (53)a-b 'rising and falling declaratives' respectively. A typical case of a rising declarative appeared in (47), which illustrated the sentence *My name is Henry Watson?* Henry Watson can utter this sentence when he presents himself at the counter of a medical consultation office. In such a context this contour may inquire about the following: 'Am I right here, do you find me on your list, do I have to do something in order to register?' These questions do not relate to the literal content of the sentence, itself a straightforward declarative. In other words, Henry Watson does not question his name or his identity, which are the part of the message stated in the syntactic part. Rather the interrogative function of declaratives arises through the interaction of sentence type, intonation, and context. Gunlogson proposes that a declarative form (subject first) expresses commitment to the propositional content of the sentence, though the kind of commitment elicited in each case is subject to variation: rising intonation in declarative sentences, as in (53)a, (53)c and (47), expresses the speaker's commitment to a proposition, and at the same time marks the proposition or the question as contingent on ratification by the addressee.[18] Falling contours on the other hand, as in (53)b and (53)d, assign the commitment to the speaker. Thus in uttering a falling declarative, the speaker commits herself to the content of the proposition, while in uttering a rise, she attributes

[18] According to Truckenbrodt (2012), the high boundary tone H_Φ puts up a contextually chosen salient proposition for question.

the commitment to the addressee. Because of these properties, declarative questions are restricted to contexts in which the requirements are met for their interpretation as questions. In English, declarative questions are only one instance of questions that introduce or assume a bias.[19] Since (53)a and (53)b constitute a minimal pair differing only in intonation, systematic differences in character and function between them – in particular, the relative 'naturalness' of (53)a as a question as compared to (53)b – must be located in the contrast between the fall and the rise.

Gunlogson (2003) further observes that rising declarative questions cannot be used 'out of the blue', without any relevant preceding context, as interrogatives can be. The straightforward interrogative in (54)a is felicitous as an initial remark, without any preceding discussion of persimmons, while the rising declarative in (54)b is odd in the same context. Example (54)c is also odd in such a context, since it presupposes that the eater does not know the name of the fruit she is eating.

(54) [to coworker eating a piece of fruit]
   a. Is that a persimmon?
   b. #That's a persimmon?
   c. #That's a persimmon.

In the same way, (55) can only be uttered in a context in which the speaker already has evidence that it is raining, for instance upon seeing somebody entering the room with a wet umbrella.

(55) It's raining?

The following example in (56) illustrates that declarative questions are not appropriate in situations where the questioner is supposed to be impartial or uninformed, as in a courtroom or committee hearing:

(56) [at a committee hearing]
   a. Are you a member of the Communist party?
   b. #You're a member of the Communist party?
   c. #You're a member of the Communist party.

Gussenhoven (1983b, 2004:297-8) also attributed meaning to contours and proposes the following translation of the three nuclear contours in the terminology of the British school. His approach is adapted from Brazil's (1975) and Bolinger's (1958) similar proposals. Gussenhoven's model incorporates elements of Pierrehumbert's tone-sequence model. He distinguished the three nuclear contours illustrated in (57) and in figure 6.8. A fall (H*L L_t) is called 'Addition'. 'The speaker refers to knowledge, or wishes to be understood as referring to knowledge, already shared by him and his listener.' When producing a fall, the speaker 'adds' an element to the background of the discourse. Addition refers to the commitment of the message to the discourse model. The second contour associated with a meaning, 'Selection', is the fall-rise (H*L H_t). The speaker selects and activates an element already in the discourse. Finally, a rise is called 'Testing'. It can be a high rise (H*H_t), thus starting high in the register and going even higher, or a low rise (L*H H_t), starting from low in the register. 'Testing' leaves it up to the listener to decide whether the message is to be understood as belonging to the background. The meaning explains the contour's ready interpretation as an interrogative: the speaker

[19] Negative polar interrogatives and tag questions are further instances of questions assuming a bias, while polar (yes-no) questions are neutral.

invites the listener to resolve the issue.

(57) Nuclear tones (Gussenhoven 2004)
a. H*L L_t 'Addition' (British school: fall)
b. H*L H_t 'Selection' (British school: fall-rise)
c. L*H H_t /H* H_t 'Testing' (British school: low rise or high rise)

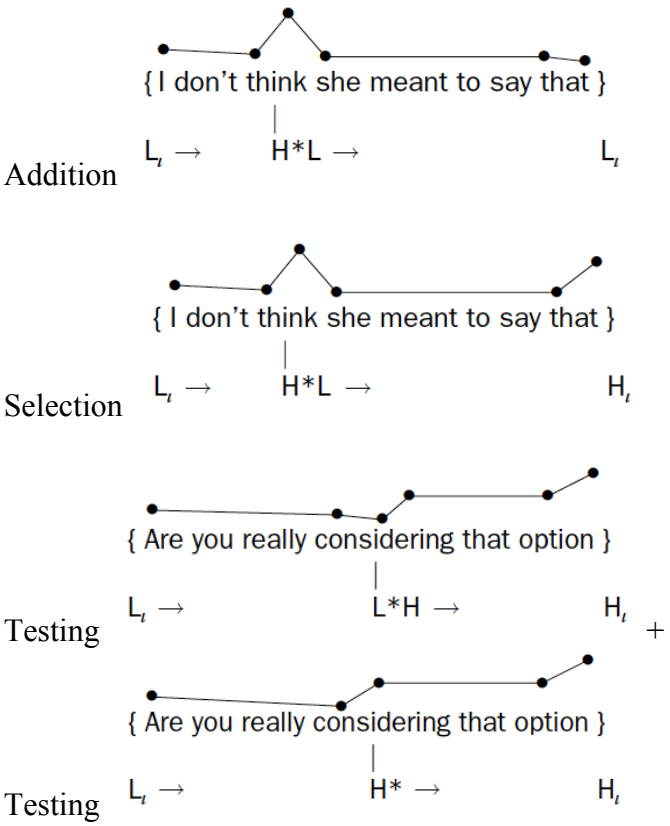


Fig. 6.8 Addition, selection, and testing in Gussenhoven's (2004) model. The arrow stands for tone spreading.

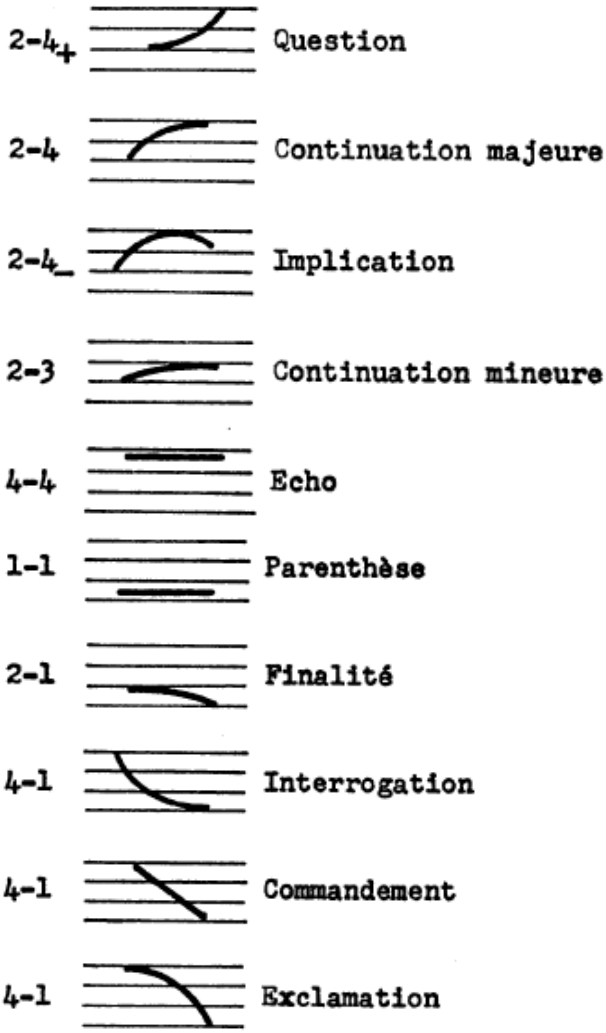
Gussenhoven's description of his three nuclear tunes has evolved from purely holistic (1983a), in the style of the British school, to compositional (2004). The tunes in figure 6.8 arise as consequences of individual tones, but the description he provided for the contours in his 2004 book are couched in the British school tradition of providing meanings to tunes.

Summing up this section, the proposals that have been discussed show that sequences of tones and/or their relationship with the text are essential in the interpretation of what is being said. The pragmatic nuances transmitted through intonation only come to their full expression when tones arise in sequence. Not only the individual tones are assigned meanings, but also sequences of tones, in a compositional way.

6.3.3 Holistic approaches

Some proposals of how to assign meanings to tonal contours were already widespread in the work of phoneticians and phonologists at the end of the 19th and during the 20th century before the development of easy-to-use speech analysis computer algorithms; see for instance the work of Delattre (1966) for French, and of O'Connor & Armstrong (1973) for English. Figure 6.9 illustrates the 10 basic intonation contours of French according to Delattre. They are conceived as forming a system of significant oppositions,

in the structuralist spirit. These contours should be understood as associating with the nuclear part of a sentence (see section 5.6).



Si ces oeufs 'If these eggs'	Continuation mineure 'Minor continuation'
Étaient frais, 'were fresh'	Continuation majeure 'Major continuation'
J'en prendrais. 'I'd take some.'	Finalité 'Finality'
– Qui les vend? 'Who sells them?'	Interrogation 'Interrogation'
C'est bien toi, 'Is it you'	Question (oui? non?) 'Yes-no question'
Ma jolie? 'my pretty?'	Echo 'Echo'
- Evidemment, 'Sure'	Implication 'Implication'
- Monsieur. 'Sir.'	Parenthèse 'Parenthesis
- Allons donc! 'Well then'	Exclamation 'Exclamation'
Prouve-le moi. 'Prove it to me.'	Commandement 'Request'

Fig. 6.9 From Delattre (1966:4)

It is typical for the holistic approach that the examples are chosen in such a way that the contours confirm the meaning provided by the specific text, as shown in figure 6.9. O'Connor & Arnold (1973) also gave an inventory of 'tone groups' for English, composed of pre-head, head, nucleus and tail (see section 5.6). They distinguished between different kinds of contours, applied to statements, commands, wh-questions and yes-no questions, among others.

When confronted with such associations between melodies and meanings, native speakers usually agree that indeed intonation contributes to communicative purposes. However, one has to be careful to keep apart the role of the contour in a certain context, in its association with a piece of discourse, and the contour itself. It is in general a simple exercise to find other options of how different meanings are assigned to the same tonal contour, depending on the context. In other words, counterexamples to 'necessary' associations between contours and meanings are plentiful, and contexts can be easily found where a tonal pattern assumed to be associated with a specific meaning has a completely different function. For instance, in the French examples in figure 6.9, all final contours (*finalité, commandement, exclamation*) are interchangeable. We have also seen that a question can have a rising or a falling contour. To this date, it has not been shown that a tone or a sequence of tones is obligatorily associated with a specific meaning all by itself. Rather, it must be associated with the text or be licenced in the context in a certain way to be meaningful.

To sum up this section, different authors have proposed different meanings for tones and tunes in English. Whether these meanings are consistently used in everyday speech is questionable, and it is up to future research to confirm or falsify the well-foundedness of such one-to-one correspondences between tones and meanings. We have reason to be optimistic that at least in some parts of the discourse, tunes do have specific pragmatic meanings. In a corpus survey of contours in American English questions, Hedberg, Sosa & Gorgulu (to appear) did find some consistency in the choice between rises and falls, partly confirming Gussenhoven's, Bartels' and Pierrehumbert & Hirschberg's results. For example, they find that what they call the 'low-rise nuclear contour', in the terminology of the British school is the 'unmarked' question pattern in American English and it is also the most frequent pattern. Both falling and high rise patterns are used for 'non-genuine' questions (the speaker already knows the answer to the question or does not desire an answer to the question).

Furthermore there is the question of the universality of such meaning assignments, to which we return in section 6.4.

6.4 Paralinguistic aspects of intonation

Besides the grammatical part of the semantics-intonation interface reviewed in sections 6.2 and 6.3, intonation is usually assumed to have a paralinguistic – or metalinguistic – component. This section is an attempt to define this role and to eliminate some misconceptions about what exactly is paralinguistic about intonation.

When going into the details of what may be considered paralinguistic in intonation, it turns out that some of them are physiological, others are part of the communication of emotion, and still others are best classified as language-dependent pragmatic components of intonation. This latter class contains the expressions of attitudes and modalities. In other words, some aspects are straightforwardly non-linguistic, some may be considered to be at the boundary between linguistic and non-linguistic, and some are definitely linguistic, at least if pragmatics is part of grammar. Let us address these three aspects in turn.

Intonation has an undeniably physiological component, related to phonation, rendering intonation more subject to individual variations than other aspects of phonetics and phonology, as for instance the production of vowels, due to the large individual variation in formant values as shaped by our articulatory organs. The evident individual physiological characteristics of intonation need to be clearly separated from communicative and emotion-related aspects on the one hand, and pragmatic uses of intonation on the other. We saw in chapter 2 that speakers' F0 register relates to the length and thickness of their vocal folds, and that in general children have higher pitch than

women, and women have higher pitch than men. Because the size of the vocal folds is also related to the physical strength of speakers, and their gender and age, some researchers have established a direct relationship between higher pitch and intrinsic power; see for instance Gussenhoven (2004:82). However, this relationship is not support by any anthropological or psychological studies, and even though physical strength implies power in many societies, the direct connection between low pitch and authority is not clear. Persons with a higher register than average (like Bismarck for example, or older men in many societies) may be more powerful than those with a lower pitch.[20] Since this debate is not related to grammar, this is not the place to discuss it. However, physiology has a role to play in the production and perception of language in general, and a better understanding of these aspects of language will certainly be helpful in the long run.

Turning to the expression of universal basic emotions, such as happiness, sadness, surprise, fear, disgust and anger, we are prone to speak louder when excited and softer when sad, and altogether more effort may be involved in the act of speaking when we are emotionally under stress (Crystal & Quirk 1964, Scherer 2000, Scherer et al. 2003). When happy, speakers may extend their range, and in sadness, register and loudness are expected to be somewhat reduced. This kind of effect of emotion on pitch and voice quality is truly paralinguistic, and is probably universal since it is related to physiology. Importantly, the tonal grammar is not affected by emotion: in other words, intonation does not change in substance under the influence of emotion. From a sociological or psychological perspective, we may change our intonation patterns when we speak to different persons, and these changes are part of socio-pragmatics. On an official mission, as a professional or in public, we do not use tonal contrasts in exactly the same way as when we speak to close friends or to small children. In his discussion of paralanguage, Ladd (2014:91ff) discussed a number of communicative modalities that he considers paralinguistic: physical distance between the speakers, gaze direction, facial expression (e.g. smiling), pitch range and voice quality.

However, in some cases, the line between paralinguistics and grammar is more difficult to draw. Chen, Gussenhoven & Rietveld (2004) speculated that paralinguistic correlates of intonation are universal. They wanted to confirm the existence of 'biological' aspects of intonation, expressed in terms of 'codes'; see Ohala (1983) and Gussenhoven (2004). If paralinguistic aspects of intonation are truly biological, speakers of different languages should use the same intonational cues for the expression of emotion. In order to test this hypothesis, they conducted an experiment with British English and Dutch speakers and investigated emotions and attitudes with attributes such as 'confident', 'friendly', 'emphatic', and 'surprised' by manipulating pitch height and pitch register. However, it turned out that the differences perceived by speakers of one language were not transferable to speakers of the other language. Evidently more research is needed in this domain. For the time being, and before results of valid experiments are available, caution is required in the vocabulary used for describing allegedly paralinguistic roles of intonation, especially when applied to different groups of people.

Finally, the last aspect of paralinguistic intonation mentioned above is its use to express attitudes and modalities, both as universal and as language-dependent features. An example often cited is the necessary relation between a question and a rising contour, as postulated by Lieberman (1967), Bolinger (1978) and the defenders of the biological codes. Ladd (1980) called this position the 'universalist view of intonation'. Recall that for Bolinger, intonation cannot be dissociated from the tensions and relaxations of the body, and thus intonation has an intrinsic non-linguistic component, reflected in the

[20] In operas, the most powerful figures are often tenors, not baritons. This is due to the fact that in the 18th and 19th centuries, powerful men were supposed to have high voices (How to Listen to and Understand Great Music, The Teaching Company, Robert Green).

grammatical use of intonation. Ladd showed that the parallelism between rising intonation and question is not necessary. Many languages do not use rises for questions, even though they are in direct contact with languages that do. For instance, Greek, Hungarian and Romanian questions are expressed by a low tone on the accented part of the question, followed by a rising-falling contour, rather than by a final rise (see Ladd 2008:145f). More recently, Rialland (2007, 2009) and Clements & Rialland (2007) also showed that many African languages use what they call a lax intonation for questions, consisting of a final falling contour among other properties. In the same way, a number of Japanese dialects use falling intonation for question intonation (Kubozono 2016). Furthermore, many tone languages end a question with whatever lexical tone is last in the sentence.

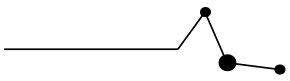
Turning now to the language-dependent aspect, Pike's often quoted description sums up what is usually considered to be the role of paralinguistic intonation (Pike 1945:22, partly cited by Ladd 2008:35 and 2014:95).

> [T]he hearer is frequently more interested in the speaker's attitude than in his words – that is, whether a sentence is 'spoken with a smile' or with a sneer ... If one says something insulting, but smiles in face and voice, the utterance may be a great compliment: but if one says something very complimentary, but with an intonation of contempt, the result is an insult. A highly forceful or exciting statement in a very matter-of-fact intonation may, by its lack of balance, produce one type of irony. Lack of balance between intonation and word content may be deliberate for special speech effects.

Notice that Pike addresses an asymmetry between lexical content and intonation. In his description, both are misaligned. It is not the privilege of intonation to be 'paralinguistic' in this way. This remark holds for all cases of 'the way it is said' advanced by Ladd and others to illustrate paralinguistics. An example would be to say something like 'It would be my greatest pleasure to fulfil your wish, my dear sir' instead of simply 'Yes'.

It still remains to pin down what exactly could fall under the third class of paralinguistic uses of intonation. Not only do languages grammaticalize different pragmatic roles, but they do so in different ways, only some of which involve intonation. For instance, the 'declarative question' that was addressed in some detail in section 6.3.2 is firmly anchored in grammar and cannot be considered as paralinguistic.

What about Kohler's (1987) analysis of 'early peak' in German.[21] Is this an example of paralinguistic use of intonation? In this language, a high pitch accent is usually realized on the stressed syllable of a focused word. In some cases, however, the high tone falls on the syllable preceding the stressed one and then the stressed syllable itself is much lower in pitch, resembling an unstressed syllable in other contexts. Kohler used the sentence *Sie hat ja gelogen* 'Indeed she lied' to demonstrate three patterns of nuclear accenting (early peak, neutral peak and delayed peak). In the early peak variant, illustrated in (58), the second syllable of the participle *gelogen* 'lied' is the stressed one but the high tone is on the first syllable of this word, an unstressed (and unstressable) schwa syllable.

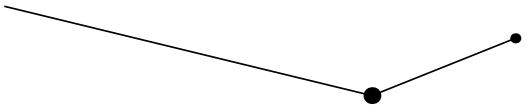


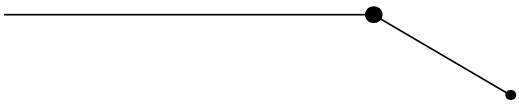
Sie hat ja geLOgen.
she has PART lied
(58) 'Indeed she lied.'

[21] An additional example, illustrated with a pitch track, appeared in figure 6.2.

Kohler analysed the early peak as an intonational variant of the neutral H*L and proposed that it adds an attitude component as compared to the neutral declarative role of H*L. He proposed that the special meaning of an early peak can be summed up as 'predictability'. It is also often used when the speakers do not involve themselves, as in reported speech, journalistic style, and in announcements at airports and railway stations. In fact the early peak is best analysed as a different category altogether, separated from the neutral H*L. It has a fall starting on the preceding syllable and can be called H!*H, the notation used in German ToBI. In the same way, the delayed peak is better characterized as a rising tone L*H, followed by a low boundary tone. As for the declarative question, the use and alignment of the participating tones in this example is an inherent part of the grammatical system of German, and is part of the grammar of this language.

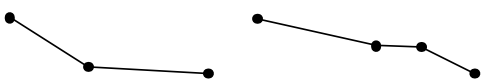
Ladd (1980:123ff) also cited some examples which show that different languages use different contours for the same aim. He compared the contours of a polite and an abrupt question by an American speaker, illustrated in (59)a and (59)b respectively. Note that politeness and rudeness are typically listed as paralinguistic use.

(59) a. 
How do I get to the Empire State Building?

b. 
How do I get to the Empire State Building?

Ladd goes on to show that the same difference between the two American English contours can also be made by words only. *Would you like anything more?* or *You through?* uttered by a waiter in a restaurant convey a similar contrast in attitudes. Furthermore, a falling contour as in (59)b by a French speaker, illustrated in (60), may sound impolite to an American speaker. But it is polite in French.[22] This comparison shows that contours may be conventionalized for different uses in a language-dependent fashion, and politeness is one of the pragmatically encoded aspects of language use.

(60) Pardon Monsieur. Où est la Tour Eiffel ?
'Excuse me, Sir. Where is the Tour Eiffel?'


Pardon Monsieur Où est la Tour Eiffel ?

These examples illustrate the use of intonation in specific languages for particular means, and they all point to the same conclusion: the line between linguistic and paralinguistic use of intonation should be drawn with great care. The same words could be uttered in another way to communicate the same text, but the effects provided by the specific intonation bring in an additional component. The tonal contour used in each case contributes to the pragmatic content of the expression, which is itself part of the grammar.

[22] In Japanese, a sentence finishing at a low level 'means' the same as the sentence finishing at mid-level. However, the second pattern adds politeness (Shin Ishihara, p.c.).

As a result, a more comprehensive description of the pragmatic component of grammar may reduce the need to refer to paralinguistic uses of intonation. Only the purely physiological ones can then be called paralinguistic.

Annotated suggestions for further reading

For information structure from different perspectives, consult Féry & Ishihara (2015), a collection of chapters on various aspects of information structure.
Pierrehumbert & Hirschberg (1990) is the standard work for assigning meanings to tones. Hobbs (1990) is a response to Pierrehumbert & Hirschberg, with a simplified and systematic analysis of the meaning of tones, which is very much worth consulting. Other excellent works on the meaning of tones in English, continuing Pierrehumbert & Hirschberg's proposal, are Gulogson (2007), Bartels (1999) and Truckenbrodt (2012).
Ladd (1980, 2014) provides detailed discussions of what is paralinguistic in intonation.

Discussion points

2. Focus has been associated with special focus positions in certain languages. Discuss the relation that might exist between changes in syntax and intonational needs. You may want to consult the original sources given below.

• Hungarian places an exhaustive focus pre-verbally (É. Kiss 1998)
[Tegnap este)$_{\Phi}$]$_{\iota}$ [bemutattam Pétert)$_{\Phi}$ (Marinak)$_{\Phi}$]$_{\iota}$
yesterday evening PRT-introduced-I Peter.ACC Mary.DAT
'Yesterday evening, I introduced Peter to Mary.'
• Italian: clause-initial or clause-final (Samek-Lodovici 2005)
[L'ho incontrato a Parigi)$_{\Phi}$, (Luigi)$_{\Phi}$, (ieri)$_{\Phi}$]$_{\iota}$
him have met in Paris, Luigi, yesterday
'I met Luigi in Paris yesterday.'

2. There has been an ongoing debate about whether meanings are better assigned to contours or to individual tones. On the basis of what was said in the chapter, is there a good method for accounting for compositionality of tones, or is the association of pragmatic roles to intonation better understood in terms of contours?

3. Discuss the following example from French: is it paralinguistic or part of the grammar? It consists in the use of a low-pitched first syllable and a high-pitched second syllable, followed by a mid-level third syllable in a reduplication in French, as testified by the famous reply by Arletty in the movie *Hotel du Nord* (1938), well known by all French speakers, communicating indignation, incredulity and reproach at the same time, an instance of 'scathing' intonation.[23]

(i) Atmosphère, atmosphère, est-ce que j'ai une gueule d'atmosphère?
atmosphere, atmosphere, is-it that I have a mug of.atmosphere
'Atmosphere, atmosphere, do I look like atmosphere?'

[23] http://www.youtube.com/watch?v=6DKI0EP-RMA This is uttered as an answer to *J'ai besoin de changer d'atmosphère, et mon atmosphère, c'est toi*. 'I need a change of atmosphere and you are my atmosphere.' (by Louis Jouvet).

Chapter 7
Tone and stress at the word level

If word-stress is so hard to find, perhaps it is not there at all. (Hyman 2006:246)

7.1 Introduction

This chapter and the next one explore the typological aspects of intonation, i.e., how languages use and implement different kinds of tonal material in their grammar. This chapter focuses on tone assignment at the prosodic word (ω-word) level. It shows that tonal properties at this low level can have a variety of causes. The next chapter will explore intonational typology at the prosodic phrase and intonation phrase levels. We distinguish between different strategies in tone assignment, and discuss the grammatical system underlying the F0 contours in each case.

Until now, it has been tacitly assumed that tones are assigned in the same way across languages. This view has to be refined. In fact, which tones are implemented, and how these tones are assigned to prosodic domains may differ widely from language to language. As has already been claimed by several authors, like Hayes (1995:403), Lehiste (1970:i) and Hyman (2006:252), languages present an amazing breadth of variations in their intonational systems.

As many researchers before, we make a first distinction between stress and tone. Stress is defined in (1), see also (Liberman & Prince 1977 Halle & Vergnaud 1987, Halle & Idsardi 1995).

(1) Stress characterizes the abstract property of being prominent, and is best represented on a metrical grid.

Stress implies that one syllable, or more rarely one mora, is more prominent than the others in a specific domain, here the lexical domain. As a property of words, lexical stress is not necessarily accompanied by any invariant or intrinsic phonetic correlates. However, when it is realized individually or in a spoken sequence, it may acquire acoustic correlates of prominence, like higher pitch, larger intensity or longer duration, as compared to unstressed syllables in the same domain. Crucially, and in many cases, these phonetic correlates are facultative, and depend on the context rather than on the properties of the word itself. Thus, it is important to insist that stress is taken here as an abstract structural property, located in a ω-word by rule or idiosyncratically. The way it is realised by phonetics may be a different matter altogether. Some authors assume that all languages must have lexical stress (Hayes 1995, van der Hulst & Goedemans 2010), or that if a language has lexical stress, then obligatorily so in the entire lexicon (Hyman 2006), see chapter 3 for stress and culminativity in English. But according to the definition of stress assumed in this book, both premises are untenable. First, many tone languages do not require lexical stress, even though tones are assigned to some or most syllables in the domain of the ω-word. Moreover the existence of languages like French and Korean, without lexical stress and without lexical tone, has to be taken into account, as well. If, for some language, native speakers have no intuitions about lexical stress, or if the literature on the location of lexical stress is plainly contradictory, the language most probably does not have lexical stress, even if some phonetic properties of words resemble stress in some

contexts.[1] Second, if a language has lexical stress, it does not need to be obligatory in the entire lexicon. It will be shown below that Turkish (Levi 2005, Kamali 2011) and Japanese (Kubozono 2015), among other languages, have lexical stress in only part of their vocabulary and no lexical stress in the complementary part.

It is important to keep in mind that this chapter studies tone assignment as the consequence of a property of the ω-word. In languages lacking tone assignment *at this level*, words may end up carrying tone specifications as the consequence of tone assignment at a higher level of the prosodic hierarchy. It is thus crucial to examine word prosodies in relation to higher-level prosodies in order to be able to understand which tones are assigned at the level of the ω-word and which are assigned at higher prosodic levels. Words uttered in isolation are typically bearers of phrasal and even sentence intonation, and are thus not fit to deliver information about stress or tones. Only extensive investigation and a very good knowledge of a language can help the linguist to decide whether a language have lexical stress or not. If a word has a 'stressed' high tone on a certain syllable in a certain context, but on another syllable in another context, the high tone is probably not lexical.[2] This point is important because there have been a large number of analyses of 'lexical stresses' which turn out not to be lexical at all.[3] This is illustrated below.

Turning to tone, languages can assign tones at the level of ω-words, to Tone Bearing Units (TBU), that is syllables or moras, see the definition in (2).

(2) Lexical tones are tonal specifications (level or contour tones) assigned to syllables or to moras at the level of ω-words.

As opposed to stress, which is an abstract property of ω-words, tone has a concrete F0 component: it can be a level tone, H or L for instance, or a contour tone, rising (LH) or falling (HL), or even a more complex contour. Tone languages form a rather heterogeneous group by the number and the kind of tones they specify, by the way they associate with the syllables and also by the rules they are subject to (see M.Chen 2000 and Yip 2002 for overviews, and see section 7.6). Tone assignment at the level of the word is relatively well explored – at least in some languages –, and this chapter can build on numerous studies that take word melodies as basic for a typology of intonation and prosodic systems (see for instance the typologies in Beckman 1986, Ladd 1996/2008, Jun 2005, 2014, van der Hulst, Goedemans & van Zanten 2010 and Hyman 2006). Hyman (2006:229) proposes the following definition for a tone language: 'A language with tone is one in which an indication of pitch enters into the lexical realisation of at least some morphemes.' On the basis of this definition, the class of tone languages includes not only languages with dense tonal specifications, such as many Sino-Tibetan languages and languages spoken in sub-Saharan Africa, but also languages with sparse lexical tone, like Swedish, Norwegian and Japanese, de facto cancelling the distinction between languages assigning tones only to lexically stressed syllables and those assigning tones to unstressed syllables. In this chapter, we treat 'pitch accent languages' (or 'accentual tone languages') in a separate section from tone languages, but, like Hyman (2006), we do not assume that pitch accent languages form a separate typological category. Pitch accent languages are those languages which have both stress and tone on the same syllable, and no lexical tone assignment on unstressed syllables, see section 7.4. Hyman's definition proposes that tone

[1] Many authors have commented on the temptation for speakers of European languages to perceive pitch variations as correlates of pitch accents, see for instance Downing (2010:382) for African languages.
[2] Although this conclusion is not necessary: stress shift can be the cause of the displacement.
[3] Sometimes, the term 'post-lexical' is used instead of phrasal stress (see Gussenhoven 2004 and Jun 2005). However, it is not easy to understand what exactly a post-lexical stress means, except that it is not lexical.

is defined on morphemes, a view that is incompatible with the prosodic account of stress and tone taken for granted in this book.

In this chapter, we pass review of the different kind of stress and tone that words can have. In section 7.2, languages without tonal and accentual specification at the word level whatsoever are shortly addressed, that is languages without lexical stress and without lexical tone. Section 7.3 reviews languages with abstract lexical stress, in which stress is not accompanied by tonal specification at the ω-word level. Section 7.4 is dedicated to 'pitch accent languages': languages with lexically specified tone on the stressed syllable of ω-words. These languages resemble tone languages in specifying a contour for (stressed) syllables or moras. Section 7.5 discusses tone alignment, i.e. the way tone is aligned with the syllable it is associated to. The last group of languages, discussed in section 7.6, are tone languages, which specify tone on syllables in ω-words, stressed or not. Section 7.6.5 sums up some facts on tonogenesis. And section 7.7 contains a conclusion.

7.2 Words without lexical stress and without lexical tone

There are a number of languages for which it can be said that they have no lexical stress specification and no tonal specification at the lexical level. French is such a language. Even if the literature on French intonation disagrees on a number of issues, all researchers agree on this point: There is no minimal pair distinguishing two words on the basis of stress and there could not be such minimal pair in French. When tones are assigned to prosodic constituents, this does not happen to lexically stressed syllables. French differs in this from all other Romance languages, which do have lexical stress. French presents some properties which can be associated with stress in other languages, like a preference for disyllabic feet in some of its morphological operations, for instance in hypochoristic formation (involving nicknames and the like), see (3). Names are often shortened in the way illustrated, resulting in a reduplication of the initial, final or another syllable. The first syllable of the reduplication is open and the second one may keep the coda of the original syllable, as in (3)c. It should be noticed that this process is typical for foot formation and does not imply a lexical stress in the original word. In other words, the presence of feet does not necessarily correlate with stress (see Leben 2002 and Hyman 2006 for tonal feet). The acute accent on the transcriptions stand for a perceptually prominent syllable.

(3)	a. Dominique	→	$(\text{Dodo})_{\omega}$	[dodó]	
	b. Françoise	→	$(\text{Fafa})_{\omega}$	[fafá]	
	c. ours 'bear'	→	$(\text{nounóurs})_{\omega}$	[nunúχs]	'teddy bear' from *un ours,* 'a bear'
	d. grand-mère 'grand mother'	→	$(\text{Mémé})_{\omega}$	[memé]	'grandma'

Another property which has been sometimes advanced to motivate stress in French is the presence of schwa syllables which do not carry tones, except when they are the only vowel of a word. A word like *table* 'table' can be mono- or bisyllabic, depending on the environment, see (4)a. The same is true for a word like *ours* 'bear', see (4)b. The difference is that *table* is assumed to be underlyingly bisyllabic, whereas *ours* is supposed to be monosyllabic, due to their orthography. If a schwa is pronounced in the final position of a Φ-phrase, as often in the speech of younger speakers of Standard French (even in words ending in a vowel), it does not attract the final high tone of this Φ-phrase. In (4)c-d, by contrast, since the schwa is the only vowel of the final word, it does carry the final Φ-phrase tone. This difference shows that schwa syllables like those in (4)a-b are

epenthetic, and that they do not count as syllables' nuclei. The final schwa in (4)c-d is different in being underlying, it is thus a perfectly acceptable place for carrying a tone. In other words, schwa syllables in French cannot be used to show that some syllables are intrinsically stressed, and others are not, as proposed by Dell (1984) and Roca (1999) among others.

(4)	a. table noire	[tab nwar]/ [tablə nwar]	'black table'
	b. ours blanc	[uχs blã]/ [uχsə blã]	'white bear, white bear'
	c. prends-le	[pχã lə]	'take him'
	d. moi je	[mwa ʒə]	'me I'[4]

French assigns tonal patterns on prosodic constituents, but not to lexical words. When syllables carry tones, this is due to their position in a specific prosodic domain, and not because of a property of the word they are in. It will be shown in the next chapter that boundary tones are assigned on all prosodic phrases in French, either at the beginning or at the end, and that these boundary tones sometimes correlate with prominence. It will be proposed that the iambic pattern of hypochoristics is a consequence of the systematic assignment of boundary tones in French, and has nothing to do with lexical stress.

Other languages without lexical stress are Bella Coola (Newman 1947:132, Bagemihl 1991), some variants of Berber (Downing 2010) and a number of languages of Ethiopia and the Indian subcontinent (Hyman 1977). Moreover, native speakers of Indonesian have little or no intuition about the location of lexical stress. Van Zanten & van Heuven (1998) and van Zanten, Goedemans & Pacilly (2003) do not find phonetic correlates for Indonesian at the level of the word. Van Zanten, Stoel & Remijsen (2010) distinguish between dialects of Indonesian and observe that "Toba Batak Indonesians do have word stress, but Javanese and Jakartan Indonesians (and thus the majority of Indonesians) do not." (p.102) Aleut and West Greenlandic are also languages that lack lexical accent (see for instance Taff 1997 for Aleut and Arnhold 2014 for West Greenlandic). The only function of intonation, in these languages, is to signal phrasing. A further language without stress and without tone is Korean as described by Jun (1993, 2005).

Finnish and Hungarian can also be analysed as languages without lexical stress. In these languages, if an accent appears on a word, it is invariably assigned to the first syllable (see Suomi 2009 for Finnish). If the word is preceded by a prefix or an article, the accent is optionally located on this preceding syllable, a fact speaking against lexical stress, which is obligatorily assigned to a specific syllable.

Trubetzkoy (1939:277) and Garde (1968:98) made a distinction between demarcative and culminative stress. In one meaning of 'demarcative', stress delimits a prosodic domain, but then demarcative makes stress predictable and, as a result, it does not need to be specified in the underlying representation of words. In the other meaning of this term, stress is morphologically conditioned: it delimits stems from affixes, or separates constituents of compound form each other. In this case, words have lexical stress, and stress is often culminative. Culminative (or metrical) lexical stress is the subject of the next subsection.

7.3 Words with lexical stress and without lexical tone

In languages with obligatory lexical stress and without lexical tone, lexical stress is first of all an abstract metrical property of words. Every word has a designated lexically

[4] This is an idiomatic expression serving to signal that some persons are always speaking of themselves.

stressed syllable, which has the potential to carry a pitch accent, see for instance Bolinger (1958) for this description. In other words, lexical stress may be phonetically realized or not.[5] As in the case of languages described in the preceding section, tonal specification arises as a consequence of intonation, i.e. when tones are assigned to Φ-phrases and ι-phrases, as a consequence of syntax and semantics (see chapters 4, 6 and 8). The tonal contour of a specific word depends on the word's semantic role and on its position in the sentence as well as on the overall information structure of the sentence. The difference between languages without stress and those with stress is that some of the tones assigned at the level of the Φ-phrase and ι-phrase – the starred tones in the tone sequence model – have to be associated with the metrically strong syllable of a word, if this particular word is assigned prominence on the metrical grid. In such a constellation, prominence is realized by a pitch accent (see chapter 5, section 5.2.2). This of course does not happen if words lack lexical stress, as is the case in French.

English, Danish, Dutch and German are examples of Germanic languages with lexical stress but without tonal specification, at least in most of their dialects. Most Romance languages also belong to this class, except for French, as mentioned in the preceding section. Greek, and most Slavic languages also have lexical stress but lack tonal specification.[6]

It is often claimed that a language with lexical stress necessarily has the property of culminativity, see for instance Hayes (1995), McCarthy (2002:78), Kager (1989, 1999), Hyman (2006:231). Culminativity implies that every ω-word has a unique primary stress: to reproduce Hyman's terminology: stress is *definitional* of a ω-word. Moreover, Hyman proposes that, in languages with lexical stress, stress is obligatory in every word. This is the Obligatory Head condition. Let us first illustrate the languages in which these conditions are met before turning in the next subsection to languages in which they are not.

English is a prototypical example of a language with lexical stress according to Hyman's definition, as it has only one main stress per word (fulfilling Culminativity), and every word has a lexical stress (fulfilling Obligatory Head). Moreover, syllables are the bearers of lexical stress, another property that Hyman considers obligatory for stress. Lexical stress in English was described in section 3.5. Minimal pairs are easy to come by, as shown in (5) with words that can be alternatively nouns or verbs. If they are nouns, stress is on the first syllable; as verbs, stress is on the final syllable.

(5) pérmit$_N$ vs. permít$_V$ ábstract$_N$ vs. absráct$_V$
éxport$_N$ vs. expórt$_V$ cónstruct$_N$ vs. constrúct$_V$

Stress assignment is generally only predictable in parts of the vocabulary, like in German, Spanish and Russian, which all have partly phonological and partly morphological motivations for lexical stress location, besides unpredictable stress patterns in some words. Loanwords may behave in a different way from native lexicon, and complex words in a different way from simplex (monomorphemic) ones. See Hayes (1995), van der Hulst (1999), van der Hulst, Goedemans & van Zanten 2010) for overviews of lexical stress in a large number of languages. In many cases, more than one stressed morpheme jointly form complex words, and because of Culminativity and Obligatory Head, this

[5] In this section, only tonal aspects of word prominence are discussed, and primary stress realised as pitch accent are considered. Secondary stresses are obviously also part of the metrical representation of words (see chapter 4), and are often realised with longer duration or increased intensity or even small changes in F0.

[6] In some Serbo-Croatian dialects (Bethin 2006) and in some Dutch and German dialects (Hermans 1999, Gussenhoven & Peters 2004), lexical tonal specifications are present in a small number of words.

implies that only one primary stress survives, we will see below that this property is also found in pitch accent languages.

Consider some examples from Russian. We limit ourselves to the nominal declension, and follow Halle (1973), Idsardi (1992) and Kodzasov (1999) who provide analyses of the Russian lexical stress assignment. The following short and incomplete review only serves to illustrate the complexity of stress pattern in a language with lexical stress, as soon as idiosyncracies coming from both the stems and the affixes complicate the picture. When the entire lexicon is considered, the stress system is much more complex than what is depicted here.

Russian has three basic classes of nominal stems: *stressed* (A), *post-stressing* (B) and *unstressed* (C). The A stems have a fixed stress on one syllable of the stem, the B stems have a fixed stress on the suffix and the C stems have variable stress, often initial. According to Halle (1973:315), the distribution of nominal stems in the three classes is highly unpredictable, even though 90% of the nouns belong to class A. The same kind of predictability can be found in inflection suffixes. In (6), *-á* (nominative) is a stressed suffix, *-y* (genitive) is an unstressed suffix.[7] Stems A already have their own stressed syllable and, as a result, the stress of the inflection does not emerge on the surface. In B and C stems, this stress is the surfacing one. Stress is indicated by an acute accent on the stressed vowel. The apostrophe following consonants indicates palatalization of the consonant.

(6)	A		B		C	
I	koróv-a	‘cow’	gospoz-á	‘lady’	borod-á	‘beard’
	koróv-y	‘cow’	gospoz-ý	‘lady’	bórod-y	‘beard’
lla	goróx	‘pea’	eer'ód	‘turn’	vécer	‘evening’
IIb	zdánie	‘building’	bozestvó	‘deity’	zérkalo	‘mirror’
Ill	postél'	‘bed’	l'ubóv'	‘love’	pl'ócad'	‘horse’

Stress pattern of the suffixes demonstrate further the intricate nominal stress pattern in Russian. In table 7.1 from Idsardi (1992:113), E stands for the front yer vowel and O for the back one.

	Singular				Plural
	I	Ila	IIb	III	
Nom	á	O	o	E	y/ý/i/á
Acc	u				
Gen	ý	a		i	O
Dat	é	u		i	ám
lnst	ój	omO		Eju	ámi
Loc	é	e		i	áxO

Table 7.1. Nominal inflection suffixes in Russian (Idsardi 1992:113)

All class I endings are stressed, except for the accusative. The class II and III endings are unstressed. All oblique plural morphemes are stressed. Idsardi (1992) assumes that the oblique plural endings have a stressed thematic vowel á, sometimes unrealised. This vowel may find itself adjacent to the thematic vowel of the stem in the underlying representation, and the thematic vowel is then deleted. Stress on an ending in table 7.1 does not mean that this stress necessarily is the surfacing one.

[7] In some of the C stems, the accent can appear on a preceding preposition, as in *ná golovu* ‘on [the] head’, by clitization of the preposition to the following noun.

The nominative plural inflection suffix has several allomorphs, one of which is the unstressed y. A C stem sometimes chooses this suffix (*darý* 'gifts'), and sometimes its stressed counterpart (*zúby* 'teeth'). Interestingly, the combination of a stressed stem plus an unstressed suffix results in a form with stress on the suffix, and vice-versa. A further complication comes from the following observation: One allomorph of the locative sg. inflection not only attracts stress to itself but it also deletes an underlying stress on an A stem.

It is not the place here to go any further into the details of the intricacies of the Russian nominal stress pattern. Let us conclude with the observation that Culminativity and Obligatory Head are obligatorily fulfilled. In other words, in each word there is a single syllable bearing main stress.

7.4 Words with lexical stress and lexical tone: Pitch accent languages or accentual tone languages

The next group of languages consists of the so-called 'pitch accent languages' also called 'accentual tone languages', with both lexical stress and lexical tone falling together: only stressed syllables carry tonal contours. In these languages, lexical stress is characterized by a restricted and 'invariant tonal contour' at the level of the word (cf. Hayes 1995: 49-50).

Hyman (2006:246) calls these languages 'pick-and-choose' languages, because they choose 'among the properties that characterise prototypical tone vs. stress-accent systems'. Since tone languages vary widely in the way tones are assigned to lexemes (see section 7.6 below), some, like Japanese or Swedish, may be analysed as tone languages with very few tonal specifications (see Yip 2002, Hyman 2006 and Gussenhoven 2004 for such an analysis). Nevertheless, pitch accent languages can be categorised straightforwardly: only the lexically stressed syllable has a lexical tone specification, all other syllables in the word do not. Some pitch accent languages satisfy Obligatory Head, and some others do not. For instance, in Swedish and Norwegian, all words have a lexical stress; this is also the case in Central Franconian and Lithuanian (Blevins 1993). But in Japanese, Biskayan Basque and in Turkish, only some of them do. Let us start the presentation of pitch accent languages with the Scandinavian languages, which are Germanic languages and which present quite a number of similar properties to the languages described in the preceding section.

7.4.1 Germanic pitch accent languages

Recall from chapter 4 that Riad (2014) applied Ito & Mester's (2007, 2012) *min-max* model to the Swedish ω-words. A minimal word (ω^{min}) is characterized by obligatory and culminative *stress*, an abstract property, and a maximal word (ω^{max}) by obligatory and culminative *word accent*, a concrete property realized as a tonal specification. Furthermore minimal words are domains for syllabification (Riad 2014:118). Swedish has an obligatory lexical stress in each ω-word, and thus satisfies Culminativity and Obligatory Head. Stress is metrically determined. Its assignment has been described by Bruce (1999). Bruce (1999:556) cites a number of minimal pairs, differing only by the position of stress, like *fórmel* 'formula' vs. *formél* 'formal' or *bánan* 'the track' vs. *banán* 'banana'. In general though, syllable weight plays a role for stress position, the regularities behind stress assignment are surprisingly similar to those of German. Derivational morphology plays a major role for stress: some derivational affixes attract stress and some others do not. The facts are rather complex and cannot be discussed here.

There is a binary pitch contrast on the stressed syllable, called *accent 1* and *accent 2* (or *acute* and *grave*), each characterized by a specific pitch contour. The accents are associated with the primary stress of every word. Minimal pairs are possible – although they are quite rare, see (7) for some true and some near minimal pairs (some of the examples are from Riad 2014:183). The words in (7)b are monomorphemic and they show that the lexical contrast in the two accents appears not only in morphologically complex words, as a result of affixation, but also at the level of stems. The digits 1 and 2 placed before a syllable show that the syllable is stressed and that it has accent 1 or accent 2.

(7) a. [1]and-en ‘the duck’ [2]ande-n ‘the spirit’
[1]regl-er ‘rule-PL’ [2]regl-ar ‘mirror-PL.’
b. [1]fänrik ‘second lieutnant’ [2]sommar ‘summer’
[1]ketchup ‘ketchup’ [2]senap ‘mustard’

The distribution of both accents is predictable from the syllable structure and the location of the stress up to a certain point: In monosyllabic words or in words with final stress, accent 1 is obligatory: Accent 2 is not possible on a finally stressed syllable, because it needs a post-stress syllable in order to be realized. Moreover, the morphology also has an influence on the distribution of accents 1 and 2. Some inflectional suffixes determine the accent: for instance, inflectional plural indefinite suffix triggers accent 2. Some derivational suffixes also induce accent 2 on the stem, see (8)a. In compounds, when more than one lexical stress are coming together, the result is invariably accent 2 on the first stressed syllable, see (8)b. The diacritic ˌ indicates secondary stress not accompanied by a pitch accent.

(8) a. Derivation: [2]sjukˌdom ‘illness’, [2]underˌbar ‘wonderful’
b. Compounding: [2]sommarˌdag ‘summer day’, ba[2]nanˌskal ‘banana peel’, [2]byxˌkjol ‘pant-skirt’

Different dialects of Swedish realize accents 1 and 2 in different ways. Bruce (1977) proposed to account for the difference between accent 1 and accent 2 in Central (Stockholm) Swedish with the following rules in (9), in which both accents consist of a sequence of H and L tones, but the alignment of the tones with the syllables (pre-stress, stress, post-stress) differs (see section 7.5 for more on tone alignment).

(9) Bruce’s (1977:133) basic pitch rules for word accents
Accent 1: H in the pre-stress syllable, L in the stressed syllable.
Accent 2: H in the stress syllable, L in the post-stress syllable.

Bruce (1999) characterized accent 1 as HL* and accent 2 as H*L.

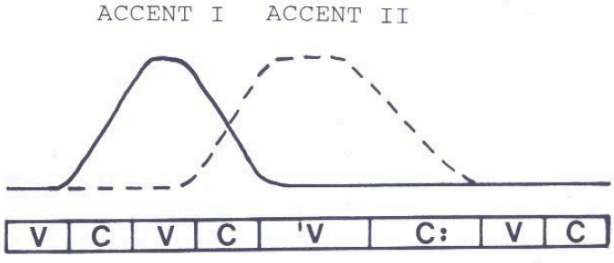


Fig.7.1 Bruce (1977:64)

Bruce found that there is a precise timing for tone association: the leading H of accent 1 appears about 120 ms. before the associated L. And the trailing L tone of accent 2 appears about 120 ms. after the associated H.

In Bruce & Gårding (1978), Bruce (1999) and Gussenhoven & Bruce (1999), not all tones characterising accents 1 and 2 are associated with segments, see also Kristoffersen (2000) for Norwegian and below. In (10), both accents have a HL contour, but in accent 1, if a tone is pre-associated, then L*, whereas in accent 2, H* is pre-associated.

(10)

a. [1]anden	b. [2]an den
|	|
HL*	H*L
'the duck'	'the ghost'
from *and-en* 'duck.DET'	from *ande-n* 'ghost.DET'

Riad's (1998a, 2014) approach of word accents is radically privative. He proposes that only accent 2 is tonally marked (with H*), and that accent 1, the unmarked variant, is tonally unspecified, see (11). The tonal specifications of accent 1 come from intonation 'pure and simple' (Riad 2014: 182). Thus, for him, the difference between the two accents is not one of timing, as in Bruce (1977), but it lies in the presence vs. absence of tonal specification. The remaining tones responsible for the difference between accent 1 and accent 2 are anchored in the phrasal intonation, see chapter 8.

(11) Riad's (1998a) privative lexical tonal opposition

accent 1:	Ø
accent 2:	H*

Other researchers have also investigated the difference in markedness or predictability of the two lexical accents, and have partly come to the same conclusion as Riad. According to Roll (2014), for instance, accent 1 on a stem has a higher predictive role than accent 2 as far as suffixation is concerned. There are fewer suffixes combining with accent 1 on the preceding stem, and thus the expectation for hearers when they are confronted with a stem bearing an accent 1 is more focused than when they are confronted with an accent 2. See also Roll, Horne and Lindgren's (2009) and Roll & Horne (2011) for some ERP experiments on the expectations induced by accents 1 and 2.

As mentioned, compounds have accent 2 on their first constituent, and no accent – or only a secondary one – on the second constituent. The association of H*L proceeds in a first step as in (12).

(12)

((σ́)ω ( )ω)ω
|
H*L

In anticipation of the next chapter, consider Swedish compounds inserted in higher prosodic domains, thus when additional boundary tones are present. Compare the minimal pairs in (13) from Bruce (1998:106, 139), also discussed by Gussenhoven & Bruce (1999) and Myrberg (2010:31). Only (13)a is a compound ([2]*mellanmålen* 'the snacks'), and (13)b is an NP with a final phrasal accent ([2]*mellan* [1]*målen* 'between the meals'). The first word *mellan* 'between' carries a H*L accent 2, but in (13)a there is a second pitch accent L*, a a phrase tone H_Φ and the final boundary tone L_ι if it is final in its clause, see section 8.3.1 for more on these tones. In (13)b, by contrast, the second word *målen* 'the meals' has the full contour of accent 1, HL*, plus the same sentence accent H_Φ and the final boundary

tone L_t as in (13)a. As a result, (13)a has two peaks and (13)b three. Bruce (1977) could demonstrate experimentally the presence of a phrasal tone, his sentence accent H (see section 8.3.1 for more on this tone).

(13) a. H* L* H_Φ
((²mellan ˌmålen)$_\omega$)$_\Phi$ 'the snacks'

b. H*L HL* H_Φ
((²mellan)$_\omega$ (¹målen)$_\omega$)$_\Phi$ 'between the meals'

Turning next to Norwegian, the facts are similar to Swedish. Following Kristoffersen (2000) and Riad's proposal, accent 1 is not specified for tone, as in Swedish, which implies that a word with accent 1 may have come from the intonational part of the melody. Again accent 2 is only specified with a H tone.[8]

(14) accent 1 accent 2

σ σ

|

H

There is always a purely intonational L*H tonal contour, as shown in (15).

(15) accent 1 accent 2

σ σ σ σ

| |

L*H H L*H

In contrast to Swedish, accent 2 in Norwegian can be associated with final syllables and thus also with monosyllabic words. For this reason, there are two possible tonal contours on a compound in Norwegian, whereas there was only one in Swedish. In (16)a since the first constituent bears accent 1, the compound also has accent 1. In (16)b, the compound has accent 2 because the first constituent bears accent 2. The examples come from Kristoffersen (2000:264), see also Gussenhoven (2004:217-219) for discussion.

(16) a. ¹feːbr + ¹nɑt → ¹feːbənɑt 'fever night'
b. ²sɔmmr + ¹nɑt → ²sɔmmənɑt 'summer night'

However, as in Swedish, the lexical high tone cannot actually be realized on a final syllable. For this reason, accent 2 on a monosyllabic word is only revealed when this word is the first part of a compound. Examples which illustrate this are scarce: *fiskesuppe* 'fish soup', from *fisk* and *suppe*, is one example, where *e* is an epenthetic or linking vowel, able to carry the H tone.

In Danish, there is no lexical tone contrast: all accents are realized alike. Thus Danish is not a pitch accent language. However, some words carry a phonation feature called *stød*, the distribution of which corresponds roughly to accent 1 in Swedish and Norwegian, and there are good reasons to assume that they have a common origin (see for instance Haugen 1967:185 for an early remark to this effect). Stød can only be realized in

[8] This view is not new in the phonology of Swedish and Norwegian. Haugen (1967:189) writes: "We shall therefore equate Accent I with STRESS alone, while Accent II will be regarded as stress plus a feature which we shall call TONE".

final stressed syllables with two sonorant moras followed by another consonant. It is a contrastive laryngalization (glottal stop or creaky voice) of the second mora of the syllable (Grønum & Basbøll 2001). It causes F0 to drop sharply at the end of the sonorants in a syllable bearing it. Because of these constraints, stød has a more restricted distribution than accent 1. There are some speculative proposals in the literature of how stød could have arisen from lexical or intonational tones (Riad 2000, Gussenhoven 2004, Kingston 2011).[9] The following minimal pairs are from Basbøll (2003, 2005:272ff) and Riad (2000:263). The apostrophe at the end of a syllable signals stød.

(17) Minimal pairs in Danish

a. Words with stød	b. Words without stød
hund [hun']'dog'	hun [hun] 'she'
vend [vɛn'] 'turn'	ven [vɛn] 'friend'
sum [sɔ́m'] 'sum'	sum [sɔ́mɐ] 'sums'

To close this short review of lexical tone contrast in Germanic languages, consider Central Franconian dialects (e.g. Hermans 1985, Schmidt 1986, Gussenhoven & van der Vliet 1999, Gussenhoven 2000a, Gussenhoven & Peters 2004) that present the same kind of contrast as the Swedish-Norwegian tones and the Danish stöd. However the contrast only appears in a small part of the vocabulary. Gussenhoven & Peters (2004) examined the so-called *Schärfung* in the German dialect of Cologne as exemplary of this pattern, and called it accent 1. As in other Central Franconian dialects, the contrast between Accent 1 (with *Schärfung*) and accent 2 (without *Schärfung*) only occurs in stressed syllables containing at least two sonorant moras. A contrast is illustrated with the word *luus* in (18). Accents 1 and 2 distinguish lexemes, as in (18). Inflectional suffixes, like plural and dative, also elicit a contrast, as was shown above for Swedish.

(18) a. Accent 1: Dat Klein es lus	[dat klɛ̯ɪn es ^{1}luus]	'The little one is clever'
b. Accent 2: Do wor en Luus	[do wo ən ^{2}luus]	'There was a louse'

The contrast between (18)a and (18)b can be described as followed: First, syllables with *Schärfung* are shorter than syllables without, and second, the melody (F0) of syllables with *Schärfung* falls steeply as compared to the other syllables, which tend to fall less steeply. Similar effects have been described for other dialects in the Central Franconian region, with slight differences between each other.

7.4.2 Japanese, Basque and Turkish

Japanese, Basque and Turkish differ from Scandinavian languages in not specifying all words for lexical stress.[10] They do not fulfil Obligatory Head: part of the monomorphemic words has lexical stress and part of it does not. These languages show an amazing similarity in their word accent systems.

In Japanese, two groups of words are often referred to as accented and unaccented words respectively (Pierrehumbert & Beckman 1988, Kubozono 1993, Poser 1984). According to Kubozono (2015), 71% of the Japanese native words are unaccented, but

[9] For instance, Riad (2000) proposed that stød is derived from a tonal contrast like the one found in Swedish and Norwegian. This contrast was lost in Danish, but left a trace in the form of a stød. In a diachronic process, glottalization appeared on the low tone of accent 1, more exactly the final boundary tone, which became grammaticalized, and eventually became the only distinctive feature between accent 1 and accent 2 in Danish.

[10] Further languages with the same peculiarity are at least Iraqw (Mous 1993), Ma'ya (Remijsen 2002) and Somali (Saeed 1999).

loanwords are accented in 93% of the cases. Only the so-called Sino-Japanese words. which are borrowed from Chinese or have been created from elements borrowed from Chinese, are equally balanced between accented and unaccented ones. Lexical pitch accent is always realized in the same way, as a falling contour, usually indicated as H*L, and it can fall on any syllable of the word, as illustrated in (19).

(19)	Initial accent	Penultimate accent	Final accent	Unaccented
	H*L	H*L	H*L	
	ínoti	kokóro	atamá	sakana
	'life'	'heart'	'head'	'fish'

As mentioned in chapter 3, Japanese is a moraic language in which the accent is assigned to a mora rather than to a syllable; the starred H* associates with the accented mora, and the following L associates with the following mora if there is one. It may also be realized at a fixed distance from the accented mora, as proposed by Pierrehumbert & Beckman (1988), but as Ladd (2008) observed, the difference in analysis is not easy to verify. Accent placement is largely unpredictable, rendering minimal pairs the rule rather than the exception. In (20), a ternary contrast between realizations of the same segmental material is illustrated. Since the vowels are monomoraic and the syllables are open in these words, each mora coincides with a syllable. In the word *hána* 'name', H* is associated with the first syllable and L with the second one. In *haná* 'flower', the L may be truncated because of lack of a tone-bearing mora following the accented one. The third variant 'nose' is unaccented and has no lexical tone. Recall from section 3.3 that, even though the accent is assigned to a mora, it is realized on a syllable (Kubozono 1999:43ff).

	H*L		H*L	
(20)	a. hána	'a name'	b. haná	'flower'

c. hana 'nose'

It was mentioned in section 3.3 that McCawley (1968) formulated a rule for the assignment of stress in loanwords in Japanese, and that Kubozono (2015) assumes that this rule is the default lexical stress rule in the entire Japanese vocabulary, see section 3.2. We return to Japanese in chapter 8, where it is shown how sentence intonation is realised. We will see there how lexical tone is part of the tonal realisation of the sentence.

Turning to Basque, the Western Lekeitio and Gernika varieties (Northern Biscayan dialects) have pitch accent properties, as described by Elordieta (1998), Hualde (1999), and Hualde, Elordieta, Gaminde & Smiljanić (2002) among others.[11] In Lekeitio Basque, most words are unaccented, but part of the vocabulary is lexically accented. Compare the unaccented *lagun* 'friend' and *léku* 'place' with an initial accented syllable. As in Japanese, there is only one tone pattern for all accented syllables: a falling pattern H*L.

Stress is a property of stems or of affixes. Complex words may have two or more accented elements. Thus, *lagún-en* 'of the friends' in (21)a is accented as a result of the preaccenting genitive plural suffix -'*en*.[12] When two accented elements come together in a word, only one of them surfaces, generally the first one, see Hualde (1999:953ff) for numerous examples of the nominal inflection in Gernika Basque. In (21)b an unaccented stem + preaccenting derivation suffix + preaccenting inflection suffix, leads to an accented stem because only the first preaccenting morpheme is active. In (21)c, the

[11] Eastern varieties of Basque do not have the distinction between accented and unaccented words.

[12] This suffix is subject to vowel harmony: it is in fact -'*an*.

accented stem surfaces as accented, even though it is suffixed with two preaccenting morphemes.

H*L
| |

(21) a. lagun ‘friend’ + -'en ‘GEN.PL’ → lagún-en ‘of the friends’
b. gixon ‘man’ + 'an ‘GEN.PL’+ 'ari ‘GEN.PL’ → gixónanari 'to the one of the men'
c. áltu ‘tall’ + -'en ‘SUPERL’ + -'ak ‘PL’ → áltuenak ‘the tallest ones’

Compounds respect the stress structure of the elements, as shown in in (22). When none of the element has a lexical stress, pitch accent is assigned to a default position: the last syllable of the first element of the compound.

(22) bélarri ‘ear’ + ondo ‘near’ → bélarriondo 'area surrounding the ear'

In some cases, unaccented words cannot be assigned a pitch accent, even when they are semantically focused. Specifically, unaccented words only have a pitch accent when they are in the final position of a Φ-phrase. In Lekeitio Basque, *neure* ‘my’ is unaccented, and it builds a Φ-phrase with the following word, in (23) with *lagunak* ‘friend’. If *neure* is focused, it has to be in preverbal position, together with the noun, because the focus is located in the preverbal position in Basque. However, since *neure* is unaccented, it cannot bear the pitch accent that marks focus. Instead, the unaccented *lagunak* becomes pitch-accented by virtue of being Φ-phrase final. Only the context makes clear that *neure* is the focus in the sentence, not the prosody (Hualde, Elordieta & Elordieta 1994:62).

(23) [Neure lagunák]$_{\Phi}$ [ekarri dau]$_{\Phi}$, [es seuriak]$_{\Phi}$.
My friend.SG.ERG brought has not yours.SG.ERG
‘My friend brought it, not yours.’

Turkish words are also divided into two classes: the lexically accented, and the unaccented ones. The lexically accented words have the same tonal pattern as in Japanese and Basque, an H*L falling contour.

Traditionally, it has been assumed that all words are lexically stressed, and that the default position of stress is the final syllable of a word (Sezer 1983, Inkelas 1999). An alternative view, first proposed by Levi 2005, and also endorsed by Kamali (2011) and Günes (2012), that is adopted here, is that all words presenting final accent are in fact unaccented. According to this view, tones assigned to the final syllables of words are boundary tones (demarcative stress).[13] Compare the place of the ‘accented’ syllable in (24), as the word becomes increasingly longer and more complex: it is always on the final syllable.

(24) [év] house ‘house’
[ev-lér] house.PL ‘houses’
[ev-ler-ín] house.PL.2ND.POSS.SG ‘your houses’
[ev-ler-in-dé] house.PL.2ND POSS.SG.LOC ‘(they are) in your houses.’
[ev-ler-in-de-kí] house-.PL.2ND.POSS.SG.LOC.PRON ‘the one in your houses’
[ev-ler-in-de-ki-lér] house.PL.2ND.POSS.SG.LOC.PRON.PL ‘the ones in your houses’
[ev-ler-in-de-ki-ler-í] house.2ND.POSS.SG.LOC.PRON.PL.ACC ‘the ones in your houses’

[13] Turkish has vowel harmony. The plural suffix alternates between *-lar* after back vowels, and *-ler* after front vowels. Suffixes with a high vowel alternate between [i], [y], [u] and [ɪ].

As for the second class of words, lexically accented stems do not loose their accent when suffixed, regardless how long the word becomes, see an example in (25). A word with lexical stress, when sufficiently long, can also have a final high boundary tone.

H*L
| |
(25) pásta-cı-lar-ımız-da
cake-DER-PL-POSS1SG-LOC
'among our cake chefs'

Examples of minimal pairs from Levi (2005) appear in (26).

(26) Mísir 'Egypt' misír 'corn'
bébek 'name of a suburb of Istanbul' bebék 'baby'
jázma 'don't write!' jazmá 'writing'

Levi (2005) compared the acoustic cues accompanying what she calls 'finally-accented' and 'non-finally accented' words, and found that the latter category of words have larger F0 excursions than the former ones, speaking for a categorical distinction between the two kinds of accents, see also Konrot (1981) for a similar result.

Suffixes can also be idiosyncratically stressed, like for instance the progressive suffix *-Íyor*, or the pre-stressing verbal negative suffix *-'mA*. Some of these suffixes are illustrated in (27), and some examples appear in (28). The capital letters stand for vowels that undergo vowel harmony. It is easy to see why these suffixes do not have a final lexical stress in the case of the disyllabic one, or why they are not themselves the bearer of stress in the case of the monosyllabic ones: all suffixes (all words, all morphemes) with a final stress would be indistinguishable from those with an unmarked final stress, and be immediately perceived as such by learners.

(27) a. idiosyncratically stressed affixes *-Ìyor* 'progr' *-índʒe* 'when' and *-érek* 'by'
b. pre-stressing suffixes: *-'me* 'neg', *-'de* 'also' *-'mA* 'neg'

(28) a. yönlendir-íyor
direct-PROG 'is directing'
b. yönlendír-me-meli
direct-NEG-NEC 'should not direct'

The combination of two pre-accenting suffixes *–'mA* (verbal negative marker) and *-'ti* (past auxiliary) is shown in (29). As in Basque, the leftmost stress survives, see (29)d.

(29) a. sinirlen-edʒék 's/she will become irritated'
b. sinirlén-me-jedʒek 's/she will not become irritated'
c. sinirlen-edʒék-ti 's/she will have become irritated'
d. **sinirlén-me-jedʒek-ti** 's/she will not have become irritated'

Compounds maintain only a single primary accent and this accent is on the leftmost accentable syllable. Thus in (30)a, the first constituent of the compound retains its idiosyncratic stress, but in (30)b, without lexical stress, the final syllable of the first word is accented. This is the same pattern as in Basque, and is prototypical for demarcative stress. Levi's (2005) suggestion that a compound is a Φ-phrase (an *Accent Domain* in her terminology) is adopted here. We will see in the next chapter that the accent pattern

illustrated in (30) is regularly found in larger domains as well. Culminativity is observed in the Turkish Φ-phrase, see section 8.3.3 for more.

(30) a. fabríka 'factory' + bad^{3}á 'chimney' → [fabríka bad^{3}a-sɪ]$_{Φ}$ 'factory chimney'
b. ajak 'foot' + kap 'cover' → [aják kab-ɪ]$_{Φ}$ 'shoe' (lit. foot cover)

To conclude this section, let us sum up the arguments for assuming a class of pitch accent languages, with different properties from both lexical stress languages and tone languages. Clearly, it is not always easy to decide whether a language is best analysed as a pitch accent language, as a tone language with an inventory of tones restricted to one, or as a language with lexical stress, especially when all these properties fall together. For this reason, some researchers prefer to classify all languages with a lexical tone contrast as tone languages. Gussenhoven (2006) exemplarily raises the question what kind of language Nubi is (a semi-extinct Arabic-based Creole of Uganda). This language presents the properties identified by Hyman as characteristic of lexical stress: culminativity and obligatory head on syllables. In Nubi, all major class words and many function words must have one syllable carrying a H tone, which can be analysed as a lexical stress. In Nubi, the lexical stress has an invariant tonal contour: it is always a high tone. Like in pitch accent languages, the sentence intonational contour depends on the placement of the lexical H tones present in it. In the classification proposed in this chapter, Nubi is a pitch accent language, combining lexical stress and tone, as Swedish and Norwegian: all words have stress. The crucial property that it shares with Japanese, Basque and Turkish in that stress has a unique (prespecified) tonal contour.

With Scandinavian languages, we saw that two different tonal contours for accents may be contrastive in a single language, as opposed to Japanese, Basque, Nubi and Turkish, where the accent has only one tonal option. Languages with two tonal options for their stressed syllables are not exceptional. According to Gomez-Imbert & Kenstowicz (2000) in Barasana (a Tucanoan language with less than 2,000 speakers from Colombia), the accent is H or HL, and the choice is lexically determined. Other languages behaving in the same way are Kaure, a nearly extinct language from Papua New Guinea (Donohue 1997) and Yucatec Maya, Mexico (Kügler, Skopeteas & Verhoeven 2007). In Yucatec Maya, only long vowels carry tones and there is a lexically determined choice from H, LH or no tone. Lithuanian has a distinction between 'acute', 'grave' and 'circumflex' accents. Acute and grave are falling melodies on short and long vowels respectively, and circumflex is a rising melody limited to long vowels (see Blevins 1993 for a phonological analysis of the Lithuanian stress pattern). The system for stress assignment in nominal classes is reminiscent of Russian, see section 7.3, and some authors assume a Balto-Slavic continuum for stress system. Finally, some dialects of Serbo-Croatian, like Neo-Štokavian, also have elements of pitch accent in their lexical stress system.

7.5 Alignment of tones

This section addresses tone 'alignment', i.e. the timing of a tone in its associated domain, especially in the syllable it is assigned to. It is necessary to distinguish between two different effects which have little in common but which have both been called alignment in the literature. The first one is related to how tone association induces meaningful phonological differences and the other one to the exact phonetic relationship between tones and segments. The first one can be seen as a categorical distinction and the second one as a gradient phonetic effect (see Ladd 2008, Gussenhoven 2004).

In its first meaning, alignment distinguishes between categories of accents. If speakers of a specific language can make a distinction between different tonal realisations

of pre-nuclear or nuclear tones, these tones must stand for different categories. For instance, Bruce (1977) analysed the difference between accents 1 and 2 in Stockholm Swedish as one of alignment of the L and H tones involved in both accents (cf. (9) and figure 7.1 above). Kohler (1990) related different categories of nuclear accents in German to a difference in alignment of the high tone participating in them (see section 6.4). Pierrehumbert & Steele (1989) conducted a perception and production experiment using rising contours on the expression *only a millionaire.* They produced a continuum of contours in which the rising tone was starting early to late on the accented syllable of *MILlionaire.* The speakers first listened to some of these manipulated contours, and then repeated them. The results spoke for a categorical effect. The speakers did not realize the gradient rising contours, but rather they produced either H*+L or L*+H on the word *millionaire*. Rietveld & Gussenhoven (1995) for Dutch, d´Imperio & House (1997) for Italian, Marakova (2007) for Russian and Frota (2014) for European Portuguese also investigated with success categorical perception of different nuclear tones. To sum up, in all these cases, different alignments of tones trigger different tonal categories.

The second kind of alignment of tones is discussed in more detail in the remainder of this section: is a monotonal or bitonal category realized invariantly in an intra- and/or cross-linguistic comparison? Are the precise turning points invariant or do they change according to the duration of the associated syllable, to the distance to the next accent, to the upcoming boundary? Do these properties change across languages? The *segmental anchoring hypothesis* predicts that in a specific language and in a specific phonological context, the beginning and end of tonal contours are stably anchored to the associated text, see Ladd et al. (1999). The alternative *duration anchoring hypothesis* predicts that the distance between the two parts of a tonal contour is regulated on the time dimension. The duration anchoring hypothesis predicts that the duration of a rise will be constant, regardless of the speech rate (Bruce 1977, Pierrehumbert 1980, see below).

Arvaniti, Ladd & Mennen (1998) was the first systematic study investigating alignment of tone under the anchoring perspective, using Greek. They measured the duration of prenuclear rises in long stressed syllables, like the second one in [pa.rÉM.vasi] ‘intervention’ vs. middle-long stressed syllables, like the first one in [VJÉ.nume] ‘we go out’ vs. short stressed syllables, like the fourth one in [ɣliko.LÉ.mona] ‘limes’. They found that the duration of the rise fits the duration of the accented syllable. In other words, the rise had a variable duration and a variable slope according to the length of the syllables carrying them. Both the beginning and the end of the rising contour were aligned with ‘segmental anchor points’. On the basis of these findings, they suggested that both individual tones, L and H in this precise case, might be aligned with specific points in the segmental structure, in a confirmation of the *segmental anchoring hypothesis*.

Ladd and coauthors replicated this result in several languages with lexical stress. Ladd, Mennen & Schepman (2000) showed that the alignment of prenuclear accentual rises in Dutch depends on whether the stressed vowel is phonologically tense or lax. A bimoraic tense vowel in Dutch may appear in an open syllable, but a monomoraic lax vowel needs a coda consonant because accented syllables need two moras. This implies that the following consonant belongs to the same syllable or not, depending on the preceding vowel. Even though Dutch tense [i] and lax [ɪ] are essentially identical in duration and differ only in vowel quality, and this both in stressed and unstressed syllables, a difference in alignment was found. Specifically, the L was aligned at the beginning of the accented syllable, but the H was aligned at the end of this syllable, i.e. late in the stressed vowel, when the vowel was tense, and thus phonologically long, and with the following closing consonant when the vowel was lax and thus phonologically short. This effect seems to be restricted to prenuclear accents: it could not be reproduced for nuclear accents (Schepman, Lickley and Ladd 2006).

A similar effect of syllable structure on tone alignment could also be demonstrated by T.Ishihara (2003) for Tokyo Japanese. In this language, the accentual peak occurred later relative to the end of the accented syllable when the syllable was open than when it was closed. Prieto & Torreira (2007) also found for Spanish that tone alignment is affected by syllable structure: prenuclear peaks occurred later relative to segments when the syllable was open, i.e. when the syllable duration was short. The same effect could be demonstrated when speech rate was increased. Ladd, Faulkner, Faulkner & Schepman (1999) could also confirm the segmental anchoring hypothesis for rising contours in Standard Southern British English. As in Greek, the tones were consistently aligned relative to the segmental string and the slope of the rise became steeper at faster speech rates. The initial L was aligned with the onset of the stressed syllable, as in Greek, but the H was aligned earlier, typically late in the immediately following consonant.

Finally, Atterer & Ladd (2004) compared the rise in prenuclear accents in Southern and Northern German. They found that the high tone aligns later in Southern German, see also Gibbon (1998:93) for this result. Atterer & Ladd also found that the rise takes place later in the accented syllable in German than in English. In English, the L of a prenuclear rising tone is aligned with the left periphery of the accented syllable whereas in Northern German it is aligned with the onset consonant of the accented syllable, and in Southern German with the left periphery of the accented vowel. Furthermore, they extended the comparison with Greek data, see figure 7.2. These authors assume that language-specific contrasts of alignment are at play. However, they also claim that rises or falls are just the same phonological objects in all languages. In their view, the exact phonetic alignment is a matter of phonetic detail.

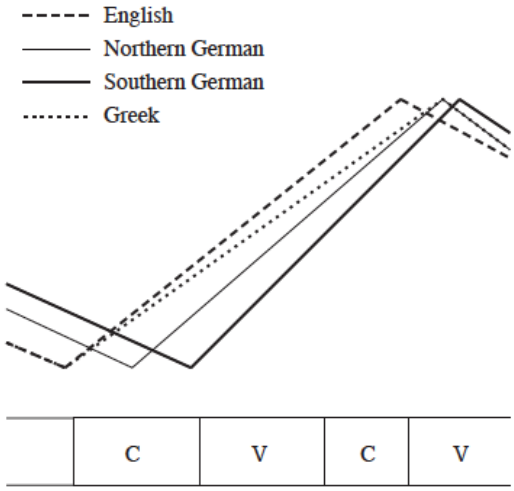


Fig.7.2 Schematic representation of alignment of H and L in prenuclear rises in different languages (from Atterer & Ladd 2004:187)

All in all, the results reported by Ladd and colleagues strongly favoured the segmental anchoring hypothesis. It should be noted, however, that all results reported so far concern prenuclear tones, and that nuclear tones behave differently, see Silverman & Pierrehumbert (1990) for English for these observations.

Moreover, as Cho (2011) observed, it may be that segmental anchoring is typical for pitch accents correlating with lexical stress. Other kinds of accents, like phrasal accents, could be governed by duration anchoring. Cho herself found that both segmental and duration anchoring are needed for a correct description of tone alignment in Korean.

In fact, duration has been found to affect the realisation of phrase accents in English, speaking for the duration anchoring hypothesis for these tones. Pierrehumbert (1980) compared the timing pattern of H* and H^- in American English. According to her,

H* is aligned at the end of the stressed syllable, except when further tones are crowded on the same syllable. As for H–, she claims that "The hypothesis suggested by this corpus of F0 contours[14] was that the H– is located at a given time interval after the L*, regardless of the stress pattern on the material following the accented syllable." (p.77) And she added: "The words in the corpus which had the longest syllable following the accented syllable were the compounds 'newsreel,' 'windmill,' 'hoosegow,' and 'headwind'. In the utterances where the L*+ H– accent fell on these words, the H– was in the first half of the second syllable, in some cases falling before the beginning of the vowel. The H– fell near or at the end of the post-accented syllable in the words where this syllable was of moderate length: 'Lieberman,' 'mothersinlaw,' 'Brobdingnag,' 'Kelloggs,' 'motherwort,' and so on. In words where the post-accented syllable was an extremely reduced CV syllable, H– fell on the next syllable after that: 'cardamon,' 'Alamo,' 'hedebo,' 'catamount,' 'rigamarole,' and so on." (p.78) To sum up, the shorter the following syllable was, the later relative to this syllable the H– was.

Northern Finnish is a further language displaying durational anchoring (Suomi 2009), although of a different kind. Here, the syllables' durations are adapted to the needs of accent. Finnish is a language in which the segmental quantity is crucial: both vowels and consonants can be short or long, independently of each other. Suomi (2009) cites the words in (31), displaying various vowel and consonant lengths.

(31) tu.li [tuli] 'fire' tuu.li [tu:li] 'wind' tul.li [tul:i] 'customs'
tu.lee [tule:] '(she/he) comes' tul.lee [tul:e:] '(she/he) may come'
tuu.lee [tu:le:] 'it blows' tuul.lee [tu:l:e:] 'it may blow'

Lexical stress is always initial in Finnish (speaking for a default demarcative stress rather than for lexical stress) and the sentence accent is usually realized as a rising-falling F0 contour on the first and second mora of the accented word, the focused one or, in the absence of a focus, the last one in the sentence (see Arnhold 2012 for the phonetic effects of focus in Finnish). The moras carrying the rise-fall are either both in the first syllable or they are distributed between the first and the second syllable, depending on their length and on the syllable structure. Suomi (2009) speaks of "tonal uniformity of the accentual F0 movement", and comments that this may help the speakers to identify the segments' quantity: If the fall-rise is fully contained in the first (open) syllable, the vowel is phonologically long, but if it is distributed between two syllables, the first vowel must be phonologically short. Furthermore, in order to accommodate the tonal fall, a moraic coda in the first syllable of an accented word is longer than its non-accented counterpart. Suomi also showed that the second syllable's length depends on whether it has to carry the falling part of the accentual contour or not. In the first case, it is much longer than in the second case.

Besides the segmental and durational anchoring hypotheses, there are quite a few obvious additional factors affecting tone alignment, like proximity of a phrase boundary, or of another tone. It is well known that tones can exert an influence on one another. For instance Xu 1998 shows how tones affect their neighbours in Mandarin, see section 5.7. Compare also Laniran & Clements (2003) for Yoruba and Féry & Kügler (2008) for German. Furthermore, in case the segmental material is too short to accommodate a full tonal contour, two different solutions are available, sometimes called *compression* and *truncation*. In case of compression, the entire contour is realized in a compressed way, with less distance between the tonal targets and/or reduced register. In case of truncation,

[14] The corpus was recorded in order to investigate the timing of H– relative to a preceding L* in a bitonal pitch accent. Two speakers were used who uttered a total of 108 sentences containing L*+H-.

the contour is not fully realized, and the tone stops midway, or it is not made at all. See Grabe (1998a-b) for the difference between English, that she takes to be a compressing language and German, assumed to be a truncating language. The difference is illustrated in figure 7.3. (from Grabe 1998b:139). In English the fall is shorter and steeper when the words are shorter. In German, the fall on the shortest word is not made.

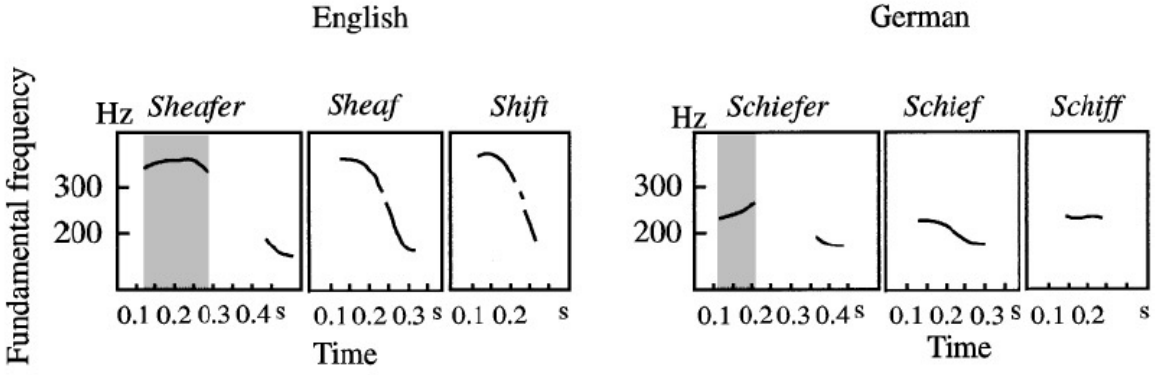


Fig.7.3 Compression in English vs. truncation in German (from Grabe 1998b:139)

7.6 Words with lexical tone: tone languages

In this section, we turn to tone specification at the level of the ω-word. It is assumed here that lexical tone is the property of a syllable or a mora to carry tonal specification independently of stress, see the definition in (2).

Important for the classification proposed here is that the tonal specification is assigned at the level of the ω-word. In this sense, pitch accent languages have tone, but lexical stress languages like English do not. Since both stress and tone are assumed to be properties of prosodic domains, Hyman's definition is not adopted here, which posits that tone is a property of morphemes, see section 7.1.

In this section, after a short overview of the main properties of lexical tones, we concentrate on Asian tone languages. Because of space limitation, the discussion can only be sketchy. For more thorough reviews of tone languages in a cross-linguistic comparison, the reader is invited to consult M. Chen (2000), Yip (2002) for Asian languages and Downing (2010) and Hyman (2011) among others for African languages.

According to Yip (2002:17), as many as 60 to 70 percent of the world's languages are tonal. She cites Africa, East and South-East Asia, the Pacific and the Americas as areas with numerous tone languages. Thus tone languages form by far the largest group of languages

Tone languages can be classified in several ways, some of which are addressed here. The first dimension along which tone languages differ concerns the bearers of tones. In most tone languages, the *TBU* is the syllable, but in some languages, it is the mora. Very often not only vowels, but also sonorant consonants carry distinctive tones. In Mandarin Chinese, even voiced fricatives can carry tones, if they are the sole segment in a syllabic rime.

The second dimension of variation among tone languages is the kind of lexical tones used: *level or contour tones*. African languages predominantly use level tones, while Asian languages generally use contour tones, for instance Mandarin Chinese, Cantonese or Vietnamese.

The third dimension of variation is the *number of tones* used distinctively, varying between two, a simple binary opposition, and ten (Sagart 1999a,b for Southeast Chinese languages). Level tones have fewer distinctions than contour tones, a consequence of the

fact that perceptually, it is more difficult to maintain a contrast between different pitch heights than between complex contour tones. Most level tone languages make a binary or ternary contrast, thus between H and L or between H, M and L, even though it is assumed that level tones may exceptionally contrast up to five different levels. Additionally to levels and contours, tones can be accompanied by non-modal phonation types like creakiness (glottalization), aspiration or breathiness, as for instance in Vietnamese, and exceptionally pharyngalization, as in Mon-Khmer.

The fourth dimension of variation is the *number of TBUs specified for tones* in ω-words. Some languages assign tones to all syllables, as for instance Mandarin (except for syllables carrying the so-called neutral tone, which can vary in form according to its environment). In this kind of language, the tonal melody of a sentence is often entirely predictable from the lexical specification of its words, modulo a restricted number of sandhi phenomena, and optionally to a small number of additional tones assigned at the level of the ι-phrase. Other languages only assign tones to few syllables, the other ones being unspecified.

A fifth dimension is the *kind of transformations* that tones show in sequences. In section 5.4, autonomy of tones was illustrated with examples from African languages. It was shown that African level tones are true autosegments by virtue of having a grammar on their own. Word melodies, spreading, tone sandhi, tonal morphemes, rules like Meussen's rule and OCP are phenomena making up tonal grammars. It was also shown that an autosegmental representation is extremely efficient for accounting for these phenomena. However, the autosegmental representation does not apply straightforwardly to Asian languages. As Hyman (2011) observes, "in Chinese it is not uncommon for a sandhi rule to replace one tonal contour by another, possibly unrelated, contour. Such is unheard of in African tone systems, where the tone clusters rarely, if ever, behave as units." Thus Asian tones are subject to tone sandhi, which are often relatively restricted, both in number and in their domains of application.

In the following, we focus on four Chinese languages. These languages are briefly described with respect of the number of tones and the invariance vs. variance of these tones in sandhi contexts. Cantonese (7.6.1) is the most invariant – or less subject to transformations – and Xiamen is the most variant language (7.6.4). Mandarin (7.6.2) and Shanghai (7.6.3) are intermediate in their tonal variation. At the end of this section, the question of tonogenesis – the origin of tones – is briefly addressed in 7.6.5.

Chinese tones are often described by reference to the four Middle Chinese (ca. 500-900 ce) categories: *ping* 'level', *shang* 'rising', *qu* (or *ʔui*) 'departing', and *ru* 'entering', see the terminology in table 7.1. *Ping* is described as long, level and low, *shang* as short, level and high, *qu* as longish, probably high and rising, and *ru* as short, with uncertain pitch and contour. The entering tones are restricted to checked syllables, i.e., to syllables with a voiceless stop or a glottal stop in their coda, and appear in the following with 'q', see M.Chen (2000) for the description of the categories. Each category can be realized in two registers, *yin* (high) and *yang* (low), giving rise to a total of eight tones. For instance, tone Ia is called *yin level* and tone IIIb *yang departing*. In many synchronic languages, it is possible to bring the lexical tones in line with the Middle Chinese categories. However, the correspondence is never exact, and we refrain to do so here. Table 7.1 illustrates the 8 tones in the Northern Wu dialect of Songjiang. Notice that the description given above for the Middle Chinese tones do not correspond to the Songjiang's tones, even though the terminology is the same.

Since, because of the complexity of their lexical tones, the Asian tones are difficult to account for in an autosegmental model such as the one proposed for African languages in section 5.4, we sometimes use the number notation of Chao (1930) in the following pages. Recall that 5 is the highest pitch, and 1 the lowest. The interested reader is referred

to Yip (2002) for an in-depth discussion about different phonological representations of tones, and to Duanmu (2007:226-28) for an in-depth criticism of Chao's notation.

Register	Tone			
	I Level (*ping*)	II Rising (*shang*)	III Departing (*qu*)	IV Entering (*ru*)
a. high (yin)	53	44	35	5q
b. low (yang)	31	22	13	3q

Table 7.1 Middle Chinese classification of tones and their values in Songjiang.

7.6.1 Invariance of lexical tone in Cantonese

Cantonese, a Yue language spoken in Hong Kong and Macao, is an extremely invariant kind of tone language. According to Lin (2002) who performed acoustic measurements of their realizations in present-day Hong Kong Cantonese, every syllable in an utterance has one of six lexical tones. Yip (2002) lists seven tones in modern Cantonese, and M.Chen (2000) nine (see Table 7.2), but both Yip and M.Chen comment that for many speakers 55 and 53 are not distinctive any more.[15] The larger number of tones proposed by M.Chen comes from the fact that he lists checked tones (CVq in Table 7.2) as independent tones.

A large number of syllables are free standing morphemes, and there is no contrast between full syllables and reduced syllables, like in Mandarin, see below. Due to these properties, the syllable has been shown to have a 'high salient psychological reality' (Wong et al. 2005, 274). The tonal pattern of a sentence is nearly entirely determined by this rich lexical tonal specification, and there is only very little room for pragmatic-driven tune changes, see section 8.4.2 for examples.

Register	Level CV(N)	Level CVq	Rising	Falling
a. high (yin)	33	55, 33	35	55/53
b. low (yang)	22	22	23	21

Table 7.2 Cantonese lexical tone system according to M.Chen (2000:16)

Examples appear in (32) from Yu (2007), see also Yip (2002:175) for a similar inventory.

(32)

Unchecked tones				Checked tones			
55 (~53)	H	si	'poetry'	55	Hq	sɪk	'to know'
33	M	si	'to try'	33	Mq	sɛk	'to kiss'
22	L	si	'affairs'	22	Lq	sɪk	'to eat'
23	LM	si	'market'				
35	MH	si	'to cause, make'				
21	ML	si	'time'				

Cantonese has a limited set of tonal changes, which are either related to tone stability (a morpheme is deleted, but the tone is re-associated, see Yu 2007 for some examples) or with tone change (or tone addition) in specific morphological contexts (see Pulleyblank 1991:448, M.Chen 2000:32ff, Yu 2007). The best-studied tone change in Cantonese is the so-called *pinjam* 'changed tone', where certain morphemes are derived from their

[15] Minor variations of some of the Cantonese tones are found in other descriptions, e.g., Hashimoto (1972), Bauer & Benedict (1997) and Flynn (2003).

semantic correlates just by changing their tone. Specifically a mid-rising tone (35 or MH) is derived from non-high level tones (33, 22, 11). Examples appear in (33) which display nominalisations of verbs (from Yu 2007).[16]

(33) a. Level tone

sou33	'to sweep'
pɔŋ22	'to weigh'
tɑn22	'to pluck'
wɑ22	'to listen'
mɔ11	'to grind'
jɐu11	'to grease'
liu11	'to provoke'
ts^hɵq11	'to hammer'
ts^hɔ11	'to plough'

b. Rising tone

sou35	'a broom'
pɔŋ35	'a scale'
tɑn35	'a missile
wɑ35	'an utterance'
mɔ35	'a grind'
jɐu35	'oil'
liu35	'a stir
ts^hɵq 35	'a hammer'
ts^hɔ35	'a plough

Except for vocative and adjective reduplication, the process is not very productive (M.Chen 2000, Yu 2007:191). Familiar and/or diminutive nouns are also often subject to *pinjam*, and there may be a difference in register between nouns subject to *pinjam* and those which are not. The former class contains more familiar nouns. However, *pinjam* is a morpholexical change, and the context of application is lexical and largely unpredictable even though it is possible to give a phonological account of the change, often using floating tones (see Yip 1980, 2002 for such an analysis).

7.6.2 Tone invariance and tone sandhi in Beijing Mandarin

Mandarin Chinese (a variant of Putonghua) as spoken in Beijing has four tones, which are largely invariant from a phonological perspective. The list of lexical tones of Beijing Mandarin is reproduced in table 7.3 from chapter 1. The Mandarin tones cannot be neatly related to the Middle Chinese tones displayed in Table 7.1.

Basic Pitch Shape	Pitch Value	Examples
T1. High Level (H)	55	mā 1 'mother'
T2. Rising (LH)	35	má 2 'hemp'
T3. Low or Falling-Rising (L)	21(4)	mǎ 3 'horse'
T4. Falling (HL)	51	mà 4 'to scold, to blame'

Table 7.3 Beijing Mandarin lexical tones

Additionally there is a so-called neutral tone that we will discuss in more detail in section 8.4.1. Here it suffices to note that the neutral tone characterises the absence of tone, see the discussion in Duanmu (2007) for arguments pointing to the toneless status of syllables with neutral tones. Neutral tone often appears in the second syllable of a disyllabic word, and takes whatever tonal specification comes from the preceding tone. Wang & Yang (2014) show that in any sequence of affix plus root, it is always the first syllable that specifies the tone of the second one, but the root always have the longer duration of the two, regardless its position relative to the affix.

A Mandarin word is mostly disyllabic, and it often has a trochaic pattern (see Duanmu 2007): the first syllable is stronger than the second one. Either both syllables

[16] Yu does not explain the notation 11 for a low level tone.

have a lexical tone, as shown in (34)a-b, or the second one loses its tone or is toneless, see (34)c.

(34) a. you-deng ‘oil lamp’ b. tʃan-hwaa ‘telephone’
35 55 51 51

c. paa-paa → paa-pə ‘dad’
51 51 51 -

‘Dual vocabulary’ is common in Mandarin. A specific word can be monosyllabic or disyllabic, as shown in (35). When it is disyllabic, the first part is often stronger than the second one.

(35) a. dian or shang-dian ‘(business)-store’
b. cai or shu-cai ‘(vegetable)-vegetable’
c. zhong or zhong-zhi ‘to plant-(plant)’

A compound can take the longer version of both words, as illustrated in (36) with words from (35). It can also take the long version of the first element and the short one of the second element, or both short versions. But short plus long in this order is not possible since it would go counter the metrical pattern strong-weak (see Duanmu 2007 and references there).

(36) shu-cai shang-dian or shu-cai dian or cai dian (but *cai shang-dian) ‘vegetable store’

Beijing Mandarin has a Tone 3 sandhi rule (T3 → T2/ __ T3), illustrated in (37). (37)a is a compound and (37)b a verb plus argument. Both form what M.Chen calls a ‘minimal rhythmic unit’ (MRU), a disyllabic unit, different from both the foot and the Φ-phrase, but which can sometimes be assimilated to the ω-word, see Duanmu (2007) for an analysis of the T3 sandhi domain based on stress and trochaic feet. T2 is mostly realized without the final rising part.

(37) a. xiao ‘small’ + gou ‘dog’ xiao gou ‘puppy’
214 214 → 35 214
b. mai ‘buy’ + ma ‘horse’ mai ma ‘to buy a horse’
214 214 → 35 214

T3 sandhi is cyclic, as shown in (38). It is regulated by morphosyntactic structure, and also depends on emphasis and speech style. In other words, the conditions of its application are rather complex.

(38) a. [lao-hu] ‘tiger’ + dan ‘gall’ lao-hu dan ‘brave’
214 214 214 → 35 35 214
b. zhi ‘paper’ + [lao-hu] ‘tiger’ zhi lao-hu ‘coward’
214 214 214 → 214 35 214

A left-branching syntactic structure usually has just one pattern, but a right-branching structure can have two (or more) applications of T3 sandhi. Thus it applies optionally and iteratively, see Duanmu (2007:256). In the following examples, tone 35 is called 2 and 214 is called 3 for simplicity.

(39) Left-branching syntactic domain: tone sandhi applies twice
mai ‘buy’ + hao ‘good’ jiu ‘wine’ [[mai hao] jiu] ‘finished buying wine’
214 214 214 → 35 35 214 (but *214 35 214, *35 214 214)

(40) Right-branching syntactic domain: tone sandhi may apply only once or twice
mai ‘buy’ + hao ‘good’ jiu ‘wine’ [mai [hao jiu]] ‘to buy good wine’
214 214 214 → 35 35 214 or 214 35 214 (but *35 214 214)

Compare the following examples in (41) illustrating the invariance of tones. At least T1, T2 and T4 are not changed in the proximity of other tones.[17] In most cases, the citation tones and the surface tones are identical, showing that no tone transformation has taken place. Compare (41) with (42) (from Duanmu 1993, also cited by Bao 2011) showing similar data in Shanghai.

(41)	san pei ‘3 cups’	san p^han ‘3 plates’	sz pei ‘4 cups’	sz p^han ‘4 plates’
Citation	55 55	55 35	51 55	51 35
Surface	55 55	55 35	51 55	51 35

Some changes in the tonal structure of Mandarin are phonetic and are motivated by co-articulation of tones leading to incomplete realizations, see Xu (2009) and section 5.7.

Mandarin assigns tone to nearly all syllables, and has a certain amount of stress, expressed in the form of trochaic feet. Foot formation is limited to some part of the vocabulary.

7.6.3 Variance of tones in Shanghai

Shanghai, a Northern Wu language of China, has the tone inventory shown in Table 7.4, from M.Chen (2000:307). As before, ‘q’ stands for the coda consonant in a checked syllable:

Basic Pitch Shape	Examples
A. 53 (HL)	tɔ ‘knife’
B. 34 (MH)	tɔ ‘island’
C. 23 (LH)	dɔ ‘peach’
D. 5q (Hq)	toq ‘sincere, earnest’
E. 12q (LMq)	doq ‘to read

Table 7.4 Shanghai lexical tones

Tones in Shanghai are much more variable than in Cantonese and in Mandarin. Citation tones are lost in Shanghai unless they occur in the initial position of a tone domain. The citation tones are split into two parts, the second one of which is shifted or spread to the following syllable, as in the following examples (from Duanmu 1993, Bao 2011). Compare (42) with (41) for Beijing Mandarin.

(42) se pe ‘3 cups’ se pø ‘3 plates’ sz pe ‘4 cups’ sz pø ‘4 plates’

[17] As discussed by Duanmu (2007:239ff), T2 is subject to some variations in a special environment. Between T1 or T2, and any tone, a middle T2 can change to T1 in connected speed. This change is generally not considered to be a phonological rule as it is dependent on speech speed.

Citation	53	53	53	53	23	53	23	23
Deletion	53	-	53	-	23	-	23	-
Surface	5	3	5	3	2	3	2	3

Duanmu (1991, 1993, 2011) and others have argued that the pattern of tone deletion and association illustrated in (42) reflects a left-prominent rhythmic organization of constituents on the metrical grid. Further examples, also cited by M.Chen (2000:308f), appear in (43) and (44). The quadrisyllabic ω-word for 'Australia' can be parsed in two tonal domains, each corresponding to a binary foot, as in the left side of (43). At the level of the ω-word, the first foot is stronger, and carries the lexical stress. As the result of a word-level stress reduction rule that conflate two stress feet into one, this word can also form a single tonal domain, as shown in the right side of the arrow in (43).

		×				×
		((×	.) (×	.))		(× . . .)
(43)		ʔo	ta	li ya	'Australia' →	ʔo ta li ya
		((34	-) (23	-))		(3 4 - -)
		((3	4) (2	3))		

A further illustration of the metrical structure of Shanghai tones is illustrated in (44). The expression in (44)a is a quadrisyllabic ω-word and is strictly separated in two feet. The structure in (44)b is a trisyllabic ω-word and causes a stress clash on the two parts of the ω-word. The clash is resolved, and as a result, the noun *ɕo zɤ* 'professor' has no tone of its own.

			×		
		(×	.)	(×	.)
(44)	a.	(ʔɤ	jã	ɕo	zɤ)
		(53	-)	(34	-)
		(5	3)	(3	4)
		Ou-yang		professor	'Professor Ou-yang'

			×			×	
		(×)	(×	.)		(× .	.)
	b.	(wã	ɕo	zɤ)	→	(wã ɕo-zɤ)	
		(23)	(34	-)		(2 3	-)
		Wang	professor			'Professor Wang'	

The tone distribution can serve to resolve structural ambiguity, as illustrated in the following examples, where the word *zã* 'bed' is the head in both instances. M.Chen treats both versions as compounds, subject to Compound stress rule, assigning main stress to the first constituent. (45)a has a complex modifier and (45)b has two modifiers. (45)a may retain the stress on *zã* (or loose it) but in (45)b, the stress clash has to be resolved. In this case only one single stress or tone domain is possible.

			×			×	Compound stress
		(×	.)	(×)		(×) (× .)	Word stress
(45)	a.	[hõ	moq]	zã	b.	hõ [moq zã]	
		23	23	23		23 23 23	Underlying tones
		(2	3)	(23)		*(23) (2 3)	Tone deletion/association
		(2	3	-)		(2 3 -)	Clash Resolution
		redwood bed				red wooden bed	

To sum up, Shanghai deletion and re-association of tones reflect a morpho-syntactically motivated prosodic constituency. Shanghai has thus both tone and stress. Tone is defined on all syllables, but stress is culminative.

7.6.4 Xiamen Min circle

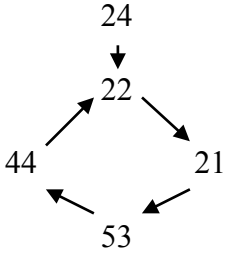
Xiamen, a South Chinese Min Nan language spoken in Taiwan, also displays tone instability and variance, albeit in a different form than Shanghai. Table 7.5 shows the basic lexical tones and their sandhi form.

Basic Pitch Shape	Sandhi form
44 (H)	22
24 (MH)	22
22 (M)	21
21 (ML)	53
53 (HM	44

Table 7.5 Xiamen lexical tones

Only the final syllable of a Φ-phrase keeps its underlying value. All other tones are subject to sandhi, in a way known as the *Southern Min Tone Circle*, see M.Chen (2000:432) and also section 4.3. Additionally to the tones appearing in (46), there are also two checked tones, i.e. tones that exist only in checked syllables: 4q (H) and 32q (ML).

(46) Southern Min Tone Circle



24: MH
22: M
21: ML
44: H
53: HM

The following examples from M.Chen (2000:433) illustrate the tone sandhi in Xiamen compounds. Only the relevant tones – those subject to variation – are indicated in (47). On the left side, the underlying version of the tone is shown, as it is final in the compound, and on the right side, the changed version, after application of tone sandhi. In this latter case, the word carrying the tone is the first element of the compound.

(47)	tsin p'ang (44)	'very fragrant'	p'ang tsui (22)	'perfume' (fragrant + water)
	p'e we (24)	'leather shoes'	we tua (22)	'shoe laces'
	wi pĩ (22)	'stomach ailment'	pĩ lang (21)	'sick person'
	k'i ts'u (21)	'build a house'	ts'u ting (53)	'roof' (house + top)
	tua hai (53)	'big ocean'	hai kĩ (44)	'ocean front'

The exact domain of tone sandhi in Xiamen has been rendered precise by M.Chen (1987, 2000) and Lin (1994), who proposed that it applies in the syntax-based Φ-phrase. Some further examples appear in (48). The first one, in (48)a, is a compound, the second one in (48)b shows a verb followed by an argument, and the third one in (48)c illustrates that the same words can be in a different syntactic configuration, here a subject plus a verb, in which case they are not subject to sandhi.

(48) a. hong - ts'e → (hong - ts'e)$_{Φ}$ 'kite'
44 44 22 44
b. pang hong - ts'e → (pang hong - ts'e)$_{Φ}$ 'to fly a kite'
21 44 44 53 22 44
c. hong ts'e hɔ ak → (hong)$_{Φ}$ (ts'e)$_{Φ}$ (hɔ)$_{Φ}$ (ak)$_{Φ}$ 'wind blows, and rain pours'
44 44 22 32q 44 44 22 32q

Compare the minimal pair in (49), where the structural ambiguity of the word sequence is resolved by forming different Φ-phrasings with different tone sandhi domains. The sequence in (49)a consists in a single Φ-phrase, a verb plus an adverb. The sequence in (49)b has an embedded clause forming a separate Φ-phrase.

(49) a. [yi [sia k'a kin]$_{VP}$]$_{S}$ 'he writes faster'
he writes more fast
44 53 53 53
(22 44 44 53)$_{Φ}$
b. [yi [sia]$_{VP}$]$_{S}$ [k'a kin]$_{VP}$]$_{S}$ 'it would be faster for him to write'
44 53 53 53
(22 53)$_{Φ}$ (44 53)$_{Φ}$

In Xiamen, the sandhi rule is a phrasal phenomenon, and thus does not correspond to any form of lexical stress.

To sum up this section so far, we have seen that lexical tone can behave differently from language to language. Facts from four Chinese languages have been used as illustrations, and it has been observed that the tonal behaviour greatly differ in term of their invariance, i.e. the way they underly tone changes and sandhi rules. Stress can be assumed in some languages but is unlikely in others. Furthermore, comparing the enormous pattern of variations of Asian tone languages with the enormous pattern of variation of African languages and American tone languages, (the latter ones were left out of consideration for lack of space), it should be clear that tone at the lexical level is not just a feature, but rather a large number of various properties (see Hyman 2006 for a clear formulation of this insight).

7.6.5 Tonogenesis

In this last section, we turn to tonogenesis, i.e. the emergence of tones. It is shown that lexical tones can have different origins, depending on the kind of tones, and their place in the intonational system.

Haudricourt (1954a,b), see also Kingston (2011), proposed that the tone contrast of Vietnamese arose in two steps. In a first step, open syllables and syllables ending with a nasal had a level tone; a final stop (written *q* in (50)) induced a rising tone, and a final voiceless fricative (written *s*) induced a falling tone on the preceding vowel. This distribution is illustrated in (50).

(50) Following C:	CVn, CV	CVq	CVs
(proto- Vietnamese)	level	rise	fall

Much later, each of these tones split into two, according to the voicing of the preceding consonants. The words beginning with a voiced stop (so-called *depressors*) developed a lower pitch than those with a voiceless stop. The resulting six-way contrast shown in (51) is the one used in North Vietnamese today. The Vietnamese names of the resulting six tones are indicated in italic.

(51) Preceding C:	pV	bV	pVq	bVq	pVs	bVs
(proto-Vietnam.)	high-level	low-level	high-rise	low-rise	high-fall	low-fall
	ngang	*huyên*	*sắc*	*nặng*	*hói*	*ngã*

According to this model, not only voicing of consonants was crucial for the emergence of tones, but also the position of the consonants in the syllable. In a nutshell, initial consonants determined the pitch height and final consonants determined contours. Although controversial, Haudricourt's model has been extremely influential (see also Hombert, Ohala & Ewan 1979 for extensions).

An alternative model of tonogenesis comes from Thurgood (2002). According to him, it is the difference in voice quality that cause the emergence of tones. Tones arise from a contrast between breathy and modal voice quality of the vowels, and this contrast is the consequence of the preceding voiced and voiceless stops. Tones are thus not directly resulting from the voicing contrast itself.

The same phenomenon arose in Chinese languages, see Haudricourt (1954a), Pulleyblank (1962), Mei (1970), Baxter (1982), and Sagart (1999a,b): lexical tone appeared from Old Chinese (500 bce) to Middle Chinese (500-900 ce). They are considered as resulting from re-interpreting initial or final consonants as pitch contrasts in the vocalic nucleus, see Table 7.1 for an overview of tones in Middle Chinese. Tone I (*ping*, level) came from open syllables or syllables closed by a sonorant. Tone II (*shang*, rising) from syllables closed by a glottal stop, and tone III (*qu*, departing) from syllables closed by s (or h). Tone IV (*ru*, entering) arose from syllables with a stop in their coda. As in the case of Vietnamese, these four tones split further under the influence of onset consonants. Later on, the voiced and voiceless stops merged together in voiceless stops but left different tones on the adjacent vowels.

Kammu, a Mon-Khmer language of Laos, has a very interesting peculiarity, see Svantesson (1989) and Svantesson & House (2006). Beside other dialects, Kammu has a tonal (Northern) and a non-tonal (Eastern) dialect, which are mutually intelligible. The tonal contrast is limited to high and low tones, which have emerged quite recently from the difference in consonant voicing, testifying in this way an on-going tonogenesis, something which has rarely been documented before. In the Northern dialect, the voicing distinction was lost after emergence of tone. However, low tones in Northern Kammu appear in syllables that begin with voiced stops and sonorants in the Eastern dialect, and high tones or rising-falling tones in Northern Kammu appear in syllables that begin with voiceless stops and sonorants in the Eastern dialect.[18]

Seoul Korean is another interesting case of potential on-going tonogenesis. It is a non-tonal language, and has a three-way distinction in stops: they can be unmarked (unaspirated, unvoiced, lax), tense [constricted glottis] or aspirated [spread glottis]. The

[18] Similar to Kammu, the Cherokee-dialect in Oklahoma has a lexical tone contrast, while the North Carolina dialect has no tone. The separation between the two dialects took place in 1838.

beginning of a Φ-phrase has been described as LHL with the first L missing when the Φ-phrase begins with a tense or an aspirated segment (Jun 1996). Silva (2006) investigated the speech of younger speakers in Seoul and found that the voicing distinction in stops is progressively being replaced by a tone distinction in vowels. Aspirated and tense segments lead to high tones ([t^{h}á oder t'á]), while the unmarked lax obstruents and the other consonants lead to low tones (Silva 2006). In other words, the duration of aspiration is no longer contrastive in younger people's speech. Aspiration in lax stops has lengthened to match that of aspirated stops. The contrast between unmarked and marked stops is maintained but has been replaced by a contrast in tone, which is gradually becoming contrastive.

Kingston (2011) remarks that "tones rarely arise *de novo* from phonation contrasts in preceding consonants, as they have done in Western Kammu dialects, Eastern Cham, Utsat, Yabem, and Korean; they are more likely to split existing tones, as they have frequently done in the Vietic, Sino-Tibetan, Hmong, Tai and Kam-Sui families."

On the basis of these observations, it seems safe to assume that laryngeal gestures associated with phonation, especially in obstruents, are the most important correlates of pitch emergence on adjacent vowels, at least in some languages. However, the cross-linguistic variation is so large that it is difficult to identify a single triggering factor for tonogenesis. For phonetic correlations between obstruent voicing and tone, the interested reader should consult Kingston (2011) and the references cited there. A primordial effect seems to be the contact of a non-tonal language to neighbouring tonal languages. This factor seems to be decisive.

In Scandinavian languages, however, tones may have emerged on a completely different basis, namely due to the emergence of new stress configurations in late Proto-Nordic, between 800 and 1200 ce (Riad 1998b, 2003, Kingston 2011), in particular from the reanalysis of the prosodic configuration of two clashing stresses, due to the syncope (elision) of medial unstressed vowels. All heavy syllables were stressed and were first separated from each other by such intervening vowels. When these vowels were lost, the heavy syllables became adjacent. Riad proposed that stress clash was resolved by destressing the second syllable through shortening rendering the affected syllable light, as shown in (52).

(52) * dǿø̩màn → dǿø.ma 'to judge'
* wúr.ðòo → wór.ðu 'words.NOM/ACC PL'.

The newly light syllable lost its stress, but retained its pitch properties, thus giving rise to a double pitch peak on a sequence of two syllables (accent 2). Only the first of these two peaks was associated with stress, thus H*. It seems that all Scandinavian dialects and languages with accent 1 and accent 2, interpreted the F0 correlates of stress as word accents in the same way, even Danish which does not have pitch accents but stød, see above.

There are more possibilities of how tones emerge: In Cheyenne (Algonquian) tone arose through vowel contraction: the long vowels of proto-Algonquian became high-toned vowels and the short vowels became low-toned. In Kickapoo, a language of the same family, a vowel with a following laryngeal fricative [h] became low, and subsequently all vowels followed by a fricative became low.

Despite these examples, it is striking that in contrast with the effect of consonants on the emergence of tones, vowels do not trigger tonogenesis very often. In his review, Kingston (2011) only cites two cases: the Angkuic language U, and Lugbara, a Moru-Madi language of the Central Sudanic branch of Nilo-Saharan. Since the effect known as intrinsic height of vowels is so pervasive (see chapter 2), the rarity of tonogenesis from

vowels is surprising. Kingston explains this with the fact that only prominent vowels have intrinsic height, non-prominent vowels do not. By contrast, consonants always act on the F0 of the neighbouring vowels.

7.7. Conclusion

This chapter has concentrated on tonal properties of ω-words and has proposed that, as far as intonation is concerned, stress and tone have to be kept apart. Stress is an abstract property identifying one syllable or one mora in a word as the most prominent one. It is best accounted for by metrical properties of ω-words, see also Hyman (2006) for such a proposal. Stress can be the consequence of a foot structure, of a particular morphological configuration, or it can be assigned by default. It can be culminative and obligatory or it can lack one or both of these characteristics. It is not characterized by phonetic correlates, and this distinguishes the present proposal from those that assume that tone and stress are distinguished by their phonetic realization, see Beckman (1986), Ladd (2008), Gussenhoven (2004) and many others. Stress is assigned at the level of the ω-word. Languages with lexical stress that do not assign tones at the level of the ω-word are intonation languages. The realisation of stress in these languages is largely dependent on sentence intonation that happens at a higher level of the prosodic hierarchy, at the level of the Φ-phrase or ι-phrase. Stress can be realised in different ways, or remain unrealised when it is in the postnuclear position. It is thus not surprising that there is a large variation as to the realisation of stress by phonetic correlates. Most Germanic, Slavic and Romance languages, as well as Greek are intonation languages, i.e. languages with stress and without tone.

In pitch accent languages, stress and tone combine and there, the realisation of stress is not dependent on sentence intonation as in pure intonation languages: it is dependent on the underlying tonal structure of the ω-word. It was shown above that Swedish and Norwegian, as well as Japanese, Basque and Turkish have both lexical stress and lexical tone. Swedish and Norwegian share culminativity and obligatoriness with the other Germanic languages. Japanese, Basque and Turkish do not have these properties. Instead they only specify a small part of their vocabulary for pitch accent.

Lexical tone is always the result of a concrete tonal pattern, it can be a level or a contour tone. Languages can contrast one or more tones. Tones can change their form in sandhi contexts or be largely invariant. Many African languages and many Asian languages have tone but no stress, although the presence of stress additionally to tone is controversial for many tone languages. Besides pitch accent languages, all other languages with tone at the ω-word level can have stress or not.

Finally some languages lack both stress and tone, like French, Berber, West Greenlandic and Bella Coola. These languages assign their tonal structure at the level of Φ-phrase and ι-phrase only, not at the level of the ω-word.

It is important to recognise that word prosodies cannot be studied in isolation. Rather they are to be integrated into phrase prosodies of different kinds to understand the import of word and the import of phrase melodies. A French word may be perceived as finally stressed, because when uttered alone, it necessarily acquires phrasal prosody. The same may be true of Hungarian, Finnish, Hindi and Georgian, to cite only a few languages with a strong phrase prosody that spills over on words. In the next chapter, we will see some of the ways how sentence intonation act on ω-word tonal pattern.

Annotated suggestions for further reading

The reader can consult several works on the typology of lexical stress.
The first large-scale study on word stress is Hayes (1995). This is a magistral work in which Hayes provides a description of many languages in a concise and uniform framework.
Van der Hulst, ed. (1999) regroups important articles on European stress systems in one large volume. The surveys on Germanic languages and on Romance languages are particularly helpful. Shorter articles examine the stress pattern of Scandinavian languages, of Basque, and stress in a diachronic perspective, among other topics. A bit problematic is the use of Halle & Vergnaud´s idiosyncratic model in most articles.
Hulst, Goedemans & van Zanten, eds (2010) extend the analysis on the world's languages, and propose an interesting and more varied analysis on language families. The reader has to become familiar with an idiosyncratic abbreviation system.
Van der Hulst ed. (2014) is a collection of papers on word stress from a theoretical point of view, all written by specialists of lexical stress. It represents an opening of perspectives relative to the preceding collections of articles by the same editor.
Hyman (2006) defends the view that stress is not obligatory in all languages. He proposes that words have suprasegmental features (or not) and is thus in agreement with the approach defended in the chapter.
As for tone, M.Chen (2000) and Yip (2002) are the standard works on tone languages. M.Chen only treats Chinese languages, and Yip covers all tone languages, although her speciality also lies on Chinese language.
Dissertations by Leben (1973), Goldsmith (1976), Odden (1981) among others provide overviews of some of the issues arising when studying lexical tones.
Tonogenesis: The article by Kingston (2011) provides a helpful overview on tonogenesis.

Discussion points

1. Hyman (2006) gives arguments against a category of pitch accent languages. Instead he pleads for the idea that languages can 'pick and choose' different properties, like ObligatoryHead and Culminativity, tone assignment and the like. After consulting his paper, discuss arguments for and against his decision.

2. Gordon (2003, 2005, 2014) shows that beside a limited number of morpholexical high-toned pitch accents in a subset of verbs that we do not address here, Chickasaw distinguishes what Gordon calls 'word stress' and 'pitch accent'. Stress is not phonemic, but rather assigned 'top down', on heavy syllables, or final syllables. Primary stress falls on the rightmost long vowel, or on the final syllable. /tala:nompaʔ/ 'telephone' has three heavy syllables, and only one long vowel. The resulting stress structure is [talá:nòmpàʔ]. Moreover, in a sequence of two open syllables with short vowels, the second syllable is rhythmically stressed and becomes half-long (and heavy): /pisalitok/ 'I looked at it' → [pisá·litòk]. /tiʃo/ 'medicine man' has final stress [tiʃó]. A high pitch accent H* appears at the end of an ι-phrase. In a simple statement, H* is simply located on the last syllable, cancelling primary stress: [talà:nòmpáʔ]. At the end of an interrogative ι-phrase, H* is generally located on the penult: [talà:nómpàʔ] and [kata:t tíʃo]? 'Who is a medicine man?' are the results. Discuss the resulting difference in accent structures depending on their position in the Φ-phrase and ι-phrase in comparison with English, where word stress is never cancelled. How is this difference accounted for? Can we really talk of 'word stress' in both cases?

Chapter 8
Sentence intonation in a typological comparison

8.1 Introduction

The preceding chapter presented a typology of tone and stress at the level of the ω-word. This chapter continues the typology and extends it to tonal structure and pitch accents at the level of the Φ-phrase and the ι-phrase, i.e. to what is called 'sentence intonation'. In the same way as distinctive tonal word specifications define a typology, tonal specifications at the level of the Φ-phrase and the ι-phrase are also typologically distinctive. In considering tonal structures at this higher level of prosodic structure, it is important to distinguish between assignment of tones on the one hand, and tonal scaling on the other. A large part of sentence intonation is the direct consequence of scaling of tones, e.g. downstep, reset and upstep. Tonal scaling is often represented in the literature with the help of tonal elements (H and L) in different combinations or with diacritics (like !, or ^ or the like), but it should be clear that tonal assignement and tonal scaling are two separate phenomena, requiring different representations (see section 6.2.3 for suggestions).

The question arises as to whether the distinctions between stress, accent and tone introduced in the preceding chapter for words help to account for the diversity of tonal structures at higher levels. For instance, it could be that if every syllable has a lexical tone of its own, it is unrealistic to expect that it will have additional tones at the sentence level. By contrast, if a language has no tonal specification at all at the level of the word, sentence intonation could be richer. In other words, there could be a trade-off between lexical tones and sentence tones. Does the presence of a dense tonal system at the level of the ω-word necessarily go together with a limited amount of sentence intonation? And, conversely, does a sparse tonal system in ω-words imply a rich tonal system at the Φ-phrase and ι-phrase levels? Although these implications may be partly true, they are not necessary. We will see that many tone languages only add a few tonal specifications at the higher prosodic levels. These languages may express pragmatic effects either by syntactic means or by register modifications, like expansion of pitch range or post-focal compression. On the other hand, languages that have either no tone specification or only few tone specifications at the level of the ω-word vary greatly in their sentence intonation. Some of them do not have a rich sentence intonation. A sparse tonal specification on ω-words does not necessarily correlate with an elaborate sentence intonation. This variation also holds for pitch accent languages. In fact, pitch accent languages display quite a wide variety of tonal effects in their sentence intonation.

Section 8.2 returns to intonation languages like English and German. In such languages, all tones – also pitch accents – are specified at the level of the higer prosodic levels; no invariant tonal contour is assigned at the level of the ω-word. The tones defining the sentence melody have their own grammar, and they associate at different points in the sentence, an important anchoring reference being the lexical stresses. As a result, sentence melodies are independent of the text carrying them; see chapter 5. It is a merit of Pierrehumbert's tone-sequence model that it is very precise about this.

By contrast, tone languages and pitch accent languages do have tone specifications at the level of the ω-word. These specifications are invariant, and sentence intonation comes additionally, in the form of tones assigned at Φ-phrase and ι-phrase. Sentence intonation has to accommodate the lexical tones, and because of this, phrasal and sentence tones are more restricted than in intonation languages. The important similarity between a

language like English and a pitch accent language is that both types of languages organize the tonal structure around the pitch accents, predefined tonally in the case of pitch accent languages, and assigned at higher levels in the case of English. Pitch accent languages are the subject of section 8.3.

Tone languages have even more predefined tones at the ω-word level, not only on the stressed syllables but also on the unstressed ones. They are thus expected to have less freedom in adding sentence intonation. Tone languages present great diversity among each other as far as sentence intonation is concerned. In section 8.4, we will concentrate on two Asian languages and one African language that use different but typical methods for implementing sentence intonation in tone languages.

The fourth type of sentence intonation includes 'phrase languages' (the first type of language addressed in the preceding chapter). They are the subject of section 8.5. These languages have tonal specifications only at the Φ-phrase and ι-phrase levels, but in contrast to Germanic languages, pitch accents are not necessarily associated with lexical stresses. Instead, tones are only assigned to Φ-phrases and ι-phrases. The result is that they can vary their position in these prosodic domains. Some of these languages, like French and Korean, do not have stress specification at the lexical level. Other phrase languages do have lexical stress specification, but these are only weakly implemented at the phrase or sentence level. This is the case with Georgian, West Greenlandic and at least some Indian languages. Section 8.6 contains a conclusion.

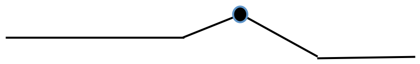
As was mentioned in chapter 6, the importance of information structure for understanding sentence intonation cannot be overrated. A large number of languages express non-standard information structure by changing the phrasing of the sentence and/or the tonal structure of a lexical stress or of the sentence intonation, in addition to using syntactic and morphological reflexes. Well-studied languages like Germanic, Slavic, Greek, and most Romance languages belong to this group. The changes resulting from information structure can take several forms. In languages with lexical stress, the tone associated with the lexical stress of a focused word is raised or otherwise changed. Or the form of a pitch accent can change under focus. Furthermore, a syllable that is not lexically accented can be rendered prominent by assigning a tone to it. Information structure can also affect intonation by changing the register in an entire prosodic domain – usually the one containing the focus. In this case, it is not only the lexically stressed syllable that changes its F0, but a larger stretch of discourse. In the same way, givenness may affect the F0 of entire prosodic domains as well, in the form of pre- or post-focal compression. In some languages, like in French, changes as a consequence of information structure can only affect Φ-phrase and ι-phrase tonal specification, in which case only the phrasing and its correlates are transformed. This happens more often in tone and phrase languages than in intonation and pitch accent languages.

8.2 Intonation languages with lexical stress

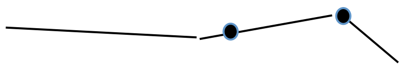
As was demonstrated in the preceding chapters for English, a prototypical intonation language, the melody of a sentence arises compositionally from different factors. We have seen that the syntactic structure is responsible for prosodic phrasing and the resulting location of pitch accents and boundary tones. The melody – the choice of tones and pitch scaling – is a consequence of the pragmatic meaning conveyed by the message, and it is the combination of tones and phrasing that expresses syntactic structure, information structure and pragmatics. In this section, the different components of intonation are put together. This is illustrated with German, another prototypical intonation language. The intonational properties of German resemble those of English, although some details of the

tonal specifications and their pragmatic interpretation differ (see among others Grabe 1998a,b, Grice & Baumann 2002, Braun 2006, Kügler 2007, Truckenbrodt 2002, 2004 for German intonation).

Examine first how a variable prosodic phrasing, expressed by changing the location of pitch accents and boundary tones, disambiguates syntactic structures (see chapters 4, 5, 6 and 9 for English examples). In (1)a, *Samuel* is a vocative. The speaker addresses somebody called Samuel and introduces her brother to him. The main clause ends with *Bruder* 'brother', and this word carries a falling pitch accent and a boundary tone. The name *Samuel* is right dislocated, and its melody is typically low and flat, as it would be in English. The postponed vocative is not accented, as was shown in chapter 6 for English. There may be a silent break between *Bruder* and *Samuel*, but this is not necessary: the tonal structure is sufficient to indicate the phrasing. In (1)b, *Samuel* is the name of the speaker's brother and Samuel is introduced to somebody else, not addressed in the sentence. The main pitch accent, the falling nucleus, is on *Samuel*, and there is a secondary rising pitch accent on *Bruder*. The two utterances have different meanings according to the placement of the boundary tone, which correlates with the syntactic structure of the sentences, and based on the location of the falling nuclear pitch accents. In German all lexical words have a lexical stress, like in English. The pitch accent on *Bruder* is realized in both variants in (1), but in (1)a, it is also the only pitch accent.



(1)a. [(Das ist mein Bruder)$_\Phi$ (Samuel)$_\Phi$]$_\iota$
this is my brother Samuel
'This is my brother, Samuel.'



b. [(Das ist mein Bruder Samuel)$_\Phi$]$_\iota$

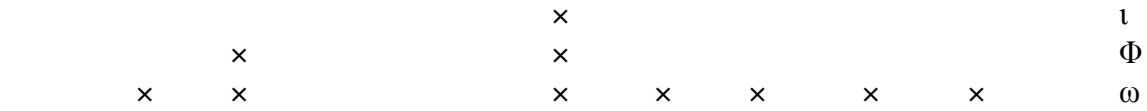
Example (2), an all-new sentence with a final verb complex, illustrates the default metrical and prosodic structure of a declarative sentence. The immediately pre-verbal constituent *Brötchen* 'rolls' carries the sentence nuclear pitch accent, and the last stressed words of each preceding Φ-phrase, *Bäckerin* 'backer' and *Früh* 'morning', also have a primary pitch accent, the head of their respective Φ-phrases. See also section 6.2.2 for pitch tracks of German sentences. Typically the nuclear accent has a falling contour and the pre-nuclear ones are rising.

									×			ι
			×				×		×			Φ
		×	×				×	×	×	×	×	ω
(2)	[(Die	junge	BÄCKERIN)$_\Phi$	(hat	in	der	FRÜH)$_\Phi$	(schon	BRÖTCHEN	backen	müssen)$_\Phi$]$_\iota$	
	the	young	baker.FEM	has	in	the	early	already	rolls	bake	must	

'The young baker already had to bake rolls early in the morning.'

If there is a narrow focus on any other word or constituent of this sentence, the metrical structure is changed (or alternatively the register of the post-focal domain is strongly

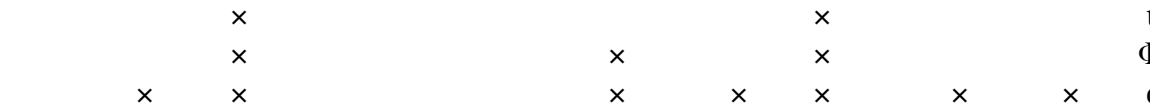
compressed, see section 6.2.3 for tonal scaling as a correlate of information structure). This is shown in (3), where the temporal adjunct *in der Früh* is narrowly focused.[1]



(3)[(Die junge BÄCKERIN)$_{\Phi}$ (hat in der FRÜH)$_{\Phi}$ (schon Brötchen backen müssen)$_{\Phi}$]$_{\iota}$

Narrow focus can also be located on other words, like *junge*, *backen, müssen* or *hat*, or on larger constituents. If a larger constituent is focused, the focal pitch accent is located on the so-called focus exponent, according to focus projection rules. If, for instance, the VP is focused, as an answer to a question like *What did the young baker do in the morning?*, the focus exponent is *Brötchen*, as in (2). In this case, the position of the nuclear stress is ambiguous among several focus structures; it is also the nuclear accent of a sentence with narrow focus on the object, as an answer to *What did the young baker already have to bake in the morning?* This ambiguity of the size of the focused constituent in the case of the default pitch accent placement was illustrated for English in section 6.2.2 with example (27). Again the similarity between English and German is conspicuous.

If the first Φ-phrase *die junge Bäckerin* is a contrastive topic, for instance in an explicit or implicit context where the subject is opposed to another one, its prominence is enhanced, and both topic and focus become equally prominent. If the highest prominence of a sentence is assigned at the level of the ι-phrase, this may imply that the topic forms its own ι-phrase, as in (4) (see Féry 2006 for this proposal).



(4)[(Die junge BÄCKERIN$_{Top}$)$_{\Phi}$]$_{\iota}$ [(hat in der FRÜH)$_{\Phi}$ (schon BRÖTCHEN$_{F}$ backen müssen)$_{\Phi}$]$_{\iota}$

Next, let's see how sentence intonation comes in. According to the tone-sequence model of intonation, it is proposed that the tonal structure is assigned to fully prosodified and metrified texts, as shown in (2), (3) and (4). Tunes in German are sequences of tones, and it is proposed in this book that tones are always associated with prosodic domains. In the case of an all-new declarative sentence, the last accent, the nuclear one, has a falling pattern, and it is analysed as a sequence of H*, assigned to the last accentable ω-word of the last Φ-phrase, i.e. the metrically strongest syllable, followed by L$_{\Phi}$, a low Φ-phrase boundary tone. The H* is associated with the metrically strongest syllable, and the L$_{\Phi}$ is associated with the Φ-phrase. This latter tone immediately follows H*, producing in this way a falling contour. A low ι-phrase boundary tone L$_{\iota}$ is located at the end of the sentence. Between the two L boundary tones, the F0 is low and flat: in (5), this contour is realized on the final verbal complex *backen müssen*. In a neutral rendition of a declarative sentence, all preceding Φ-phrases have a rising contour. Each has an L* on their last lexical accent and a high boundary tone H$_{\Phi}$ following the L*. Again the starred tones are associated with the strongest syllables of ω-words and the H$_{\Phi}$ tones are associated with the Φ-phrases. This tone spreads until the end of the Φ-phrase, inducing a high plateau. If there are metrically strong beats preceding the nuclear accent of each Φ-phrase, they also get a

[1] I set aside the question of the type of focus—informational, selective, contrastive, corrective— and the question whether a 'stronger' focus (see Féry 2011 for this concept) has larger prosodic correlates, see Braun (2006), Kügler & Gollrad (2015) and the references therein for different results in German and English.

tone, preferably H* but sometimes L* or a bitonal pitch accent, depending on their pragmatic role.

L* H_Φ L* H_Φ H* L_Φ L_ι
(5)[(Die junge BÄCKERIN)$_\Phi$ (hat in der FRÜH)$_\Phi$ (schon BRÖTCHEN backen müssen)$_\Phi$]$_\iota$

In the case of a narrow focus, as illustrated in (6) for the example (3) with focus on *Früh*, the nuclear sequence H* L_Φ is assigned to the focused word. In the following material, there is no tone to be assigned, since the nuclear tone is by definition the last one, and the result is a low and flat post-nuclear (and post-focal) intonation, as was also illustrated in chapter 6 (example (38)).

L* H_Φ H*L_Φ L_ι
(6) [(Die junge BÄCKERIN)$_\Phi$ (hat in der FRÜH)$_\Phi$ (schon Brötchen backen müssen)$_\Phi$]$_\iota$

In the case of a contrastive topic, as illustrated in (4), the metrical beat corresponding to the nuclear accent of the topic is as high as the one on the comment of the sentence. In other words, the topic is not subordinated to the last nuclear accent. As shown in (7), there is an additional H_ι at the end of the ι-phrase built on the topic, and this additional high tone induces a rise. This accounts for the fact that a topic ends higher than unmarked pre-nuclear Φ-phrases. As a result, the boundary following the topic is stronger than when the same constituent is not a topic, and this is reflected by larger phonetic correlates on the boundary tone of its prosodic domain. Thus the difference in strength between a 'normal' pre-nuclear constituent and a topic is felt on the boundary tone rather than on the low pitch accent.

L*H_Φ H_ι L*H_Φ H* L_Φ L_ι
(7)[(Die junge BÄCKERIN$_{Top}$)$_\Phi$]$_\iota$ (hat in der FRÜH)$_\Phi$ (schon BRÖTCHEN backen müssen)$_\Phi$]$_\iota$

When the sentence is a question, the nuclear accent may be changed into a rising contour L*H_Φ, followed by an H_ι. The preceding Φ-phrases do not necessarily change their tonal shape. In a question, the focus can also be changed, as in a declarative sentence, in which case the nuclear L* followed by H_Φ is located on the narrow focus. The following text remains high between H_Φ and H_ι. The tone assignment is illustrated in (8).

L*H_Φ L*H_Φ L*H_Φ H_ι
(8) [(Hat die junge BÄCKERIN)$_\Phi$ (in der FRÜH)$_\Phi$ (schon BRÖTCHEN backen müssen?)$_\Phi$]$_\iota$

Additional variations in the realization of tones are possible. The first constituent can be realized as a falling-rising pattern to enhance its topic status, for instance. The nuclear accent can be an early peak, enhancing its status as a predictable element (see Kohler 1990 and figure 6.2). As has been shown in chapter 6, tonal scaling is an important component of intonation. The default pattern of a sentence like (2) is downstep between the Φ-phrases, see section 6.2. It is best represented as downstep of the entire prosodic domains. Scaling between the individual tones can also be changed to express a changed prominence relationship between the components of the sentence (see Truckenbrodt 2002, 2004, 2015, Féry & Ishihara 2009, 2010, Féry & Kügler 2008).

In sum, intonation languages can manipulate not only the contour of pitch accents and boundary tones, but also the pitch of individual syllables by giving them more or less prominence according to their informational status. The pattern of individual tones can also

be changed for pragmatic reasons. This has been illustrated with German, but other intonation languages display the same kind of variation in their tonal components. The most important aspect is that assignment of tones is not random but is constrained by the prosodic structure and the relative prominence of prosodic constituents, as well as by their pragmatic meaning.

8.3 Pitch accent languages

Pitch accent languages resemble intonation languages in having lexical stress, but they also differ from them in that pitch accent languages are tonally specified for their lexical stress. In this section, we extend the analysis of some of the languages that were discussed in the preceding chapter to sentence intonation.

8.3.1 Stockholm Swedish

In the last chapter, we saw that Swedish has two lexical accents, accent 1 and accent 2, which were given a number of analyses: fully specified (Bruce 1977), moderately privative (Gussenhoven & Bruce 1999 and Gussenhoven 2004) or fully privative (Riad 2014); see (9) for the fully specified and the privative versions (in parentheses).

(9) a. Accent 1 b. Accent 2

[1]anden 'the duck'	[2]anden 'the ghost'
|	|
HL* (or Ø)	H*L (or H*)

In this section, it is shown that Swedish has additional tones, which are assigned at the level of the Φ-phrase and at the level of the ι-phrase.

Consider first Bruce's (1977) influential model of Swedish sentence intonation, reproduced in figure 8.1. Accent 1 melody is shown in the upper part of the figure, and accent 2 melody in the lower part. Sentences with accent 1 and sentences with accent 2 have the same melodic contour, a fall-rise-fall that Bruce decomposed into a word accent fall, a sentence accent rise and a terminal juncture fall. He proposed that the alignment between tones and segments starts later in melodies with accent 2 than in those with accent 1. Although the grammar of the contour of the entire sentence is the same in both cases, the global perceptive impression is nonetheless different because there is an additional rise on accent 2, due to the fact that the first fall starts later.

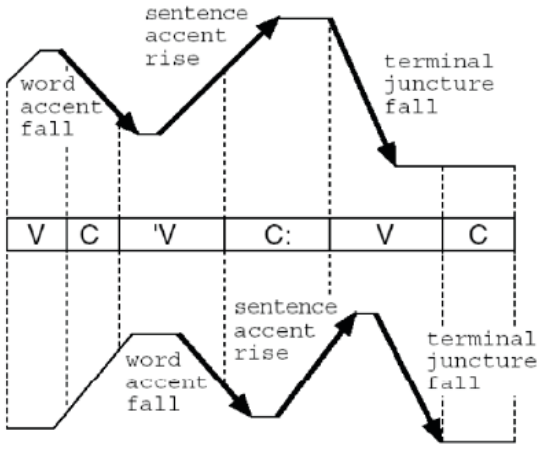


Fig. 8.1 Bruce's (1977:50) model of Swedish nuclear intonation in a declarative sentence

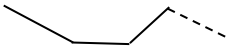
Bruce's analysis of Swedish intonation is compositional: in his approach, sentence melody is decomposed into tones coming from the lexical part of intonation and those assigned at the level of the sentence. Adapting his model to the terminology used in the present book, in addition to lexical tones (accent 1 and 2), there is a phrase-level tone (H_Φ), and a sentence-level tone, namely a final boundary tone. Traditionally, intonation research in Swedish has concentrated on word accents 1 and 2, the dialectal differences and the so-called 'focal accent'. More recently, researchers like Roll et al. (2009), Ambrazaitis (2009) and Myrberg (2010, 2013) have started to investigate sentence intonation in greater detail.

Let us first concentrate on the H_Φ tone called *sentence accent* by Bruce (1977) and *focal accent* by Bruce (1999) and other authors, see (10). The term focal accent is a misnomer since, as we will see, it is not restricted to focus, and it may appear more than once in a sentence, even if the sentence has no narrow focus.[2] Myrberg & Riad (2015) followed Riad (2014) in proposing that there are actually two kinds of accents. The first one is the word accent presented in chapter 7 (thus accent 1 and accent 2) and is called *word accent*, and the other one is what they called *focal accent*. The term *focal accent* refers to the lexical accent plus an additional phrasal tone corresponding to the sentence accent of Bruce. I follow Bruce's analysis and call his sentence accent *phrasal tone* here, written H_Φ in the following.

(10)

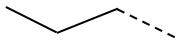
	Word accent	Focal accent
Accent 1	HL*	HL* H_Φ
Accent 2	H*L	H*L H_Φ (H*L* H_Φ in compounds)

An accented compound is realized as in (11)a, with two starred tones, following the notation of Riad (2014) plus the additional phrasal tone following the pitch accents. Compare this tone sequence with a normal word accent in (11)b, inserted into a phrase, and augmented with the phrasal tone H_Φ; L_ι is present when the word is final in the ι-phrase. It explains the final fall (terminal juncture fall in Bruce's 1977 terminology).



H* L* H_Φ (L_ι)

(11) a. [2]mellan¸målen 'the snacks'



H*L H_Φ (L_ι)

b. [2]blommorna 'the flowers'

The special realization of compounds in (11)a is explained by the fact that medial L seeks to associate with the stressed syllable of the second word in a compound, thus with the stressed syllable of *målen* in (11)a and of *hylla* in (12). In Riad's analysis, this L spreads leftwards in the compound so as to create a low contour between the two H tones.[3]

H* ⟵ L* H_Φ

(12) [2]skol-¸bok-¸hylla 'school book-shelves'

[2] Myrberg (2013:84, footnote 8) also called this tone focal accent, but she commented: '[T]here is no absolute one-to-one correlation between the presence of a focal accent and the information structural category focus. This perhaps makes the term *focal accent* slightly misleading, but I nonetheless employ it here, since it has been so widely used in the literature.'

[3] Recall that in Bruce's analysis, presented in chapter 7, the L tone is the second part of the bitonal tone on the first word.

Myrberg (2013:83) proposed that the phrasal tone H_Φ (her focal accent) appears on the last word of the Φ-phrase it is in. In (13) and (14), all words written with small capitals have a phrasal tone, in addition to their word accent. In general topics have one, and all nuclear accents and usually pre-nuclear accents also have one.

(13) A: 'What's happening?'
B: [(Maria laser en bok om DATORER)$_\Phi$]$_\iota$
Maria reads a book on computers
'Maria is reading a book about computers.'

(14) A: 'What's happened?'
B: [(Väskan med BÖCKER)$_\Phi$ (har blivit KVARGLÖMD)$_\Phi$]$_\iota$
bag.the with books has been forgotten
'The bag with books has been forgotten.'

In the case of a narrow focus, the focused word also has a phrasal tone, and the post-focal words do not. The post-focal words only have word accents, see (16). They are thus not deaccented as in English and German, since they retain their word accent. The only difference between pre- and post-nuclear realization is that in the post-nuclear domain, no additional phrasal tone is present.

Swedish also has tones at the level of the ι-phrase. All researchers agree on the presence of a final boundary tone. The following sentences, adapted from Bruce (1977) and Gussenhoven (2004:213), illustrate L_ι at the end of the ι-phrase. As shown by Ambrazaitis (2009), the final boundary tone depends on the pragmatic role of the sentence. In this point, Swedish resembles the intonation languages described in the preceding section. In (15) and (16), since the sentences are declarative, the final boundary tone is a low tone. Gussenhoven proposed an additional sentence-level initial boundary tone, which has not been adopted by Swedish intonation specialists. For this reason, it is not reproduced here.

(15) Pre-nuclear and nuclear positions

HL* HL* HL* H_Φ L_ι
a. [(man vill a^{1}namma några ^{1}längre ^{1}NUMMER)$_\Phi$]$_\iota$ accent 1
'one wants to accept some long numbers'

H*L H*L H*L H_Φ L_ι
b. [(man vill ^{2}lämna några ^{2}långa ^{2}NUNNOR)$_\Phi$]$_\iota$ accent 2
'one wants to leave some long nuns'

(16) Nuclear and post-focal positions

HL* H_Φ HL* HL* L_ι
a. [(man vill A^{1}NAMMA)$_\Phi$ några ^{1}längre ^{1}nummer]$_\iota$ accent 1

H*L H_Φ H*L H*L L_ι
b. [(man vill ^{2}LÄMNA)$_\Phi$ några ^{2}långa ^{2}nunnor]$_\iota$ accent 2

Turning to the distribution of H_Φ, it can be seen that there is one such tone at the end of the sentences in (15), but in (16), it is much earlier, at the end of the focused part of the sentences. In order to understand the distribution of this tone better, different aspects of Swedish intonation must be considered. First, Roll et al. (2009) proposed a left-edge boundary tone, written as a superscripted H, at the beginning of a main clause, which is absent at the beginning of an embedded clause. According to them, this high tone is located on the last syllable of the first prosodic word of every main clause. Thus in (17)a there are two high tones, one on the last syllable of *berättaren* 'the storyteller' and another one on *vandalerna* 'the Vandals', because these two words each introduce a main clause. That the embedded clause has main clause order can be seen from the order between *intog* 'conquered' and *inte* 'not', which is only possible in main clauses. In (17)b, the inversion of these two words signals that the embedded sentence is subordinate. In this case, there is no left-edge boundary tone H at the beginning of the embedded clause, thus on *vandalerna*.

(17) a. Berättaren^{H} menar alltså att vandalerna^{H} intog inte Gallien av en slump
the.storyteller thinks thus that the.Vandals conquered not Gaul by a chance
'The storyteller thus thinks the Vandals didn't conquer Gaul by accident'
b. Berättaren^{H} menar alltså att vandalerna^{Ø} inte intog Gallien av en slump
the.storyteller thinks thus that the.Vandals not conquered Gaul by a chance
'The storyteller thus thinks that the Vandals didn't conquer Gaul by accident'

In other words, Roll et al. proposed that the occurrence of this tone is determined on purely syntactic grounds: it is dependent on the status of the sentence as a main or subordinate clause. However, it may also be assumed that main and subordinate clauses are mapped to different prosodic domains and that the presence or absence of boundary tones in (17) correlates with prosodic domains. And indeed, Myrberg (2010) had a prosodic explanation for this tone. In her account, these accents, which she called *initiality accents*, obligatorily appear on the leftmost stressed word in a ι-phrase. She proposed distinguishing the initiality accent from what she called the focal accent because the initiality accent is aligned differently. It must be noted that if there is a focal accent early in the ι-phrase, there is no initiality accent, speaking either for complementary distribution or for identity of the two H tones.

I tentatively propose that both the *focal accent* and the initiality accent (or initial boundary tone) are the same kind of tone, namely they are both instances of the phrasal H_Φ described above. This tone is obligatory in all Φ-phrases unless the Φ-phrase is post-nuclear. This implies that they are not indicators of syntactic structure or of information structure in a direct way, but rather they are prosodic tones.

In the following illustration from Myrberg (2013:94) in (18) and figure 8.2, all H tones without diacritics are H_Φ. The sentence consists of two ι-phrases, and each ι-phrase is divided into two Φ-phrases, each with its own H_Φ. The first one in an ι-phrase associates earlier in its Φ-phrase – on the first word in the Φ-phrase – than the following ones. Moreover in the word it is realized on, it is typically aligned later than the following ones; see Myrberg (2013:85ff) for a description of the differences. There are also word accents and low boundary tones (written L% in the figure).

(18) [[De ^{2}andra skulle vara ^{2}ut¸klädda]$_{Clause}$ så [^{2}Anna ville inte vara ^{1}med]$_{Cl}$]$_{illocCl}$
the others would be dressed.up so Anna wanted not be with
'The others were getting dressed up, so Anna didn't want to join.'

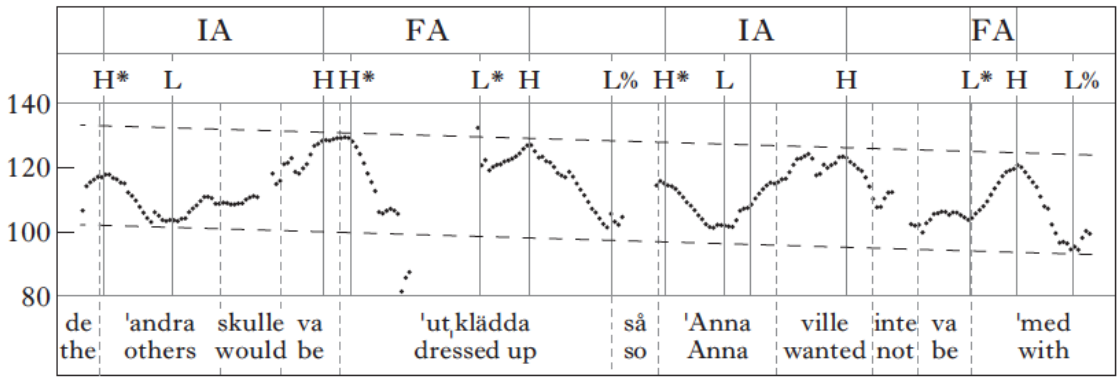


Fig. 8.2 From Myrberg (2013:99). IA stands for 'initiality accent' and FA for 'focal accent'. Both are taken to be H_Φ here.

To sum up, Swedish not only has word accents, but also phrasal tones which obligatorily appear in specific prosodic domains: an H_Φ assigned at the level of the Φ-phrase and boundary tones assigned at the level of the ι-phrase. Although the distribution and alignment of these tones remains to be clarified, it can be observed that the intonation pattern of a declarative sentence is compositional. It is fully determined by lexical, phrasal and sentence tones.

8.3.2 Tokyo Japanese

In section 3.4.2, Ito & Mester's (2012) min-max model was introduced, which illustrated the need for recursion in the prosodic structure of Japanese. It was shown there that Japanese compounds have several complex prosodic structures, all of them recursively organized. Furthermore, the unique word accent was shown in section 7.4.2 to be obligatorily realized as a falling pitch accent, thus H*L. Let us now add further tonal aspects of the Φ-phrase and ι-phrase in Japanese: first, the assignment of tones in Φ-phrases, and second, tonal scaling in Φ-phrases and ι-phrases under the influence of syntax and information structure.

An accented ω-word carries a pitch accent H*L, and it forms a Φ-phrase by itself. Moreover, Japanese assigns tones to Φ-phrases. There is an initial L_Φ in the Φ-phrase, followed by H_Φ. If there are enough free moras to carry these tones, both are realized. Compare the words in (19)a has no accented syllable and the two initial phrasal tones associate with the first and the second mora respectively; see also Gussenhoven (2004:188) and Vance (2008:142-54) for illustrations.[4] The result is a rising tone on the first two syllables and a high plateau afterwards. Likewise, (19)b with a final accented syllable starts with two unaccented moras, which leave room for the two phrasal tones. The falling pitch accent is realized on the final syllable.

L_Φ H_Φ $L_\Phi H_\Phi$ H* L

(19) a. $(garasudama)_\Phi$ 'glass beads' b. $(kake - ro)_\Phi$ 'break-off'

In a word with initial pitch accent, as in (20)a, the first two moras, here also the first two syllables, have a falling contour initiated by the pitch accent. There is no room for initial

[4] Beckman & Pierrehumbert (1986) proposed an alternative analysis: the initial rise is caused by a low boundary tone coming from the preceding phrase plus the first high tone of the Φ-phrase.

L_Φ and H_Φ. In (20)b, the second mora carries H* and L_Φ occurs on the first mora. Again the H_Φ is not realized (or it is fused with the following H*). This word is realized as a rising-falling contour.

H*L $\quad$ L_ΦH* L

(20) a. (hasi-ga)$_\Phi$ ‘chopstick-NOM’ $\quad$ b. (hasi-ga)$_\Phi$ ‘bridge-NOM’

Japanese has traditionally been analysed as having two non-recursive levels of phrasing above the ω-word and below the ι-phrase: the Minor Phrase (MiP) and the Major Phrase (MaP); see Beckman & Pierrehumbert (1986), Kubozono (1993), Selkirk & Tateishi (1991), Ishihara (2007, 2011) and many others. Ito & Mester (2012, 2013, 2015) showed that both MiP and MaP are instances of Φ-phrases, and that there is no need to distinguish between them. Following their proposal, MiP is called here ‘minimal Φ-phrase’. The minimal Φ-phrase is morpho-syntactically determined: it consists of at least one morphological word, plus possibly cliticized morphemes, thus of a maximal ω-word. It is also a culminativity domain, as it maximally contains one accent. In this sense, culminativity in Japanese is a constraining principle on the formation of minimal Φ-phrases. The examples in (21) come from Kubozono (1993), Vance (2008) and Gussenhoven (2004). First, (21)a illustrates that names in the order first name plus surname form a single minimal Φ-phrase together. In the reverse order, two minimal Φ-phrases are formed, one on each name. (21)b shows an ambiguous syntactic structure. In both versions, the NP consists of a noun modified by a preceding noun with the genitive/partitive *no* ‘the horse’s collar’, but the adjective modifies either the embedded noun ‘horse’ or the head noun ‘collar’. Even though the syntactic structure is different, the prosody is the same. Both versions consist of two minimal Φ-phrases with one accent each. Since *síro* ‘white’ and *umá* ‘horse’ are both accented, they cannot be part of the same minimal Φ-phrase.[5] And since *kubiwa* ‘collar’ is unaccented, it has to be in the same minimal Φ-phrase as the preceding accented word. As a result, the prosodic structure cannot disambiguate this ambiguous syntactic structure.

(21) a. (Kaoru Matumoto)$_\Phi$ but (Matumoto)$_\Phi$ (Kaoru)$_\Phi$

b. (siró-i)$_\Phi$ (umá-no kubiwa)$_\Phi$ = [[siró-i umá-no]$_{NP}$ kubiwa]$_{NP}$
white-INFL horse-GEN collar ‘the collar of the white horse’
[[siró-i]$_{AP}$ umá-no kubiwa]$_{NP}$
‘the white collar of the horse’

When two or more minimal Φ-phrases group together to form larger Φ-phrases, the larger domains are ‘non-minimal Φ-phrases’ (traditionally called Major Phrases or MaPs). This is the domain of downstep, sometimes called *catathesis*; see Poser (1984) and Pierrehumbert & Beckman (1988). In Japanese, downstep refers to the lowering of the pitch register following an accented mora. Only the H*L tone of accented words induces downstep; unaccented words do not. Compare the examples in (22) and their realizations in figure 8.3, from Ishihara (2015). A sequence of three accented words (AAA) appears in (22)a and the left panel of figure 8.3, and a sequence of two unaccented words plus an accented word (UUA) in (22)b and the right panel of figure 8.3. It can be seen that in the first case, each

[5] Vance (2008:156-7) showed that a word like *umá* loses its accent before *-no* and is in fact realized as if it were unaccented.

high tone is downstepped relative to the preceding one, but this does not happen in the second case. Both expressions form a Φ-phrase, but the first one contains three minimal Φ-phrases, each with the tonal structure of a Φ-phrase, thus a rising contour at the beginning. This happens only once in the second case, since there, all words form one large Φ-phrase with no internal structure.

(22) a. ((Náoya-no)$_{Φ}$ (áni-no)$_{Φ}$ (wáin-o)$_{Φ}$)$_{Φ}$
Naoya-GEN big.brother-GEN wine-ACC
'Naoya's big brother's wine'
b. (Naomi-no ane-no wáin-o)$_{Φ}$
Naomi-GEN big.sister-GEN wine-ACC
'Naomi's big sister's wine'

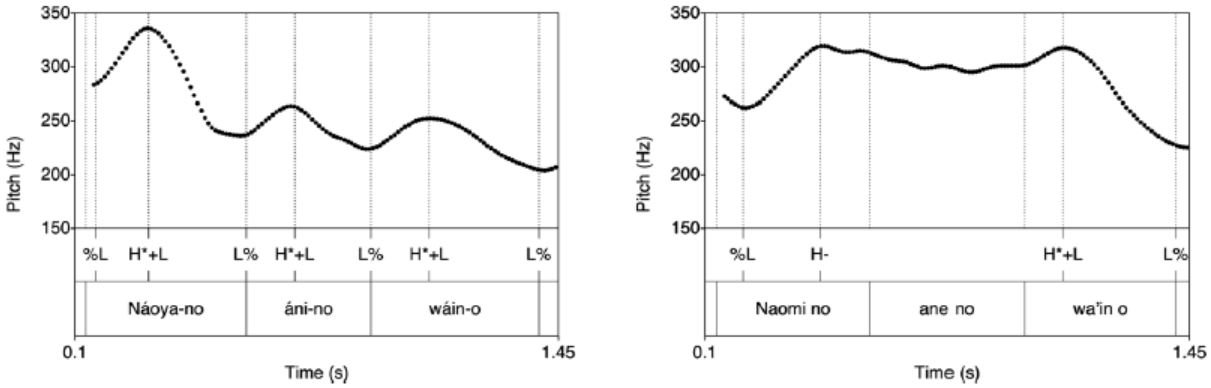


Fig. 8.3 Japanese. AAA sequence, left, and UUA sequence, right (from Ishihara 2015:572)

The effect of downstep in a sequence of accented Φ-phrases is schematized in the following graphic (figure 8.4), inspired by Ishihara (2011). Each H*L is a minimal Φ-phrase. Between two non-minimal Φ-phrases (MaPs), the F0 register is reset.

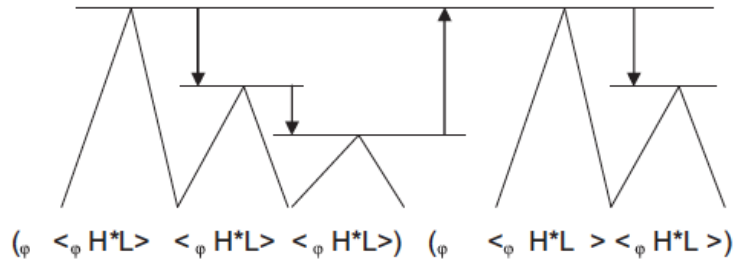


Fig. 8.4 Downstep in the Japanese Φ-phrase (from Ito & Mester 2013:25)

According to Selkirk & Tateishi (1991:529), a Φ-phrase boundary is aligned with the left boundary of a syntactic maximal projection XP. Kubozono (1989) showed that in a sequence of four accented ω-words, organized in a sequence of four minimal Φ-phrases within one non-minimal Φ-phrase, the beginning of the third minimal Φ-phrase is higher than expected by the 'flat' organization of the four minimal Φ-phrases. Kubozono calls this effect 'metrical boost'. One of his examples appears in (23).

(23) ((náoko-no)$_{Φ}$ (áni-no)$_{Φ}$ (aói)$_{Φ}$ (erímaki)$_{Φ}$)$_{Φ}$
Naoko's brother's blue scarf

The dotted line in figure 8.5 shows the expected downstep without metrical boost. The arrow pointing up shows the metrical boost on the third high tone in (23).

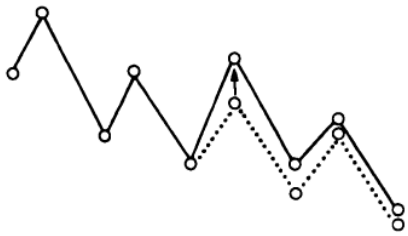


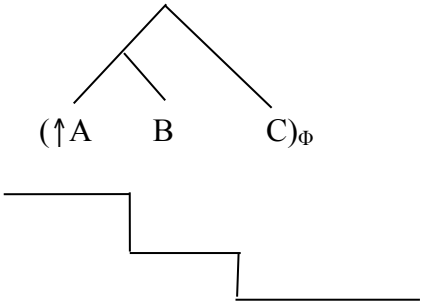
Fig. 8.5 Four minimal Φ-phrases (from Kubozono 1989:53)

Acording to Kubozono, metrical boost results from right-branching syntax-prosody mapping as in (23). He assumed a rhythmic effect as shown in (24).

(24) Right-branching syntax-prosody mapping: [[[[A] B] C] D] →
Rhythmic effect: ((A B)(C D))

The different influence that right-branching and left-branching syntactic structures exert on the corresponding prosodic structures is best illustrated with three constituents, since there the structure is truly asymmetric. A left-branching structure boosts only once, namely at the left edge of the entire structure, while a right-branching structure boosts twice, once at the left edge of the entire structure and once at the left edge of the constituent formed by B and C. As a result, the downstep on B is larger in a. than in b.

a. Left-branching [[A B] C] → (↑ABC)$_{Φ}$



b. Right-branching [A [B C]] → (↑A↑BC)$_{Φ}$

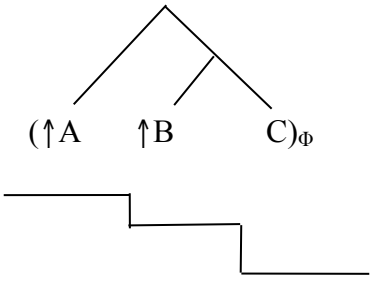


Fig. 8.6 Left- and right-branching patterns (adapted from Ishihara 2015:583)

Ito & Mester (2012) pointed out that 'minimal Φ-phrase', 'non-minimal Φ-phrase' and 'maximal Φ-phrase' are relational terms, not separate categories. Every beginning of a non-minimal Φ-phrase is at the same time the beginning of a minimal Φ-phrase. The rise induced by the phrasal tones $L_{Φ}$ and $H_{Φ}$ is always present at the beginning a Φ-phrase. It is

thus not a property of the minimal Φ-phrase (MiP) as has been assumed by Pierrehumbert & Beckman (1988), but a property of all Φ-phrases. As for the property of downstep, it is also a property of Φ-phrases in general. In minimal Φ-phrases, the effect of downstep is not apparent because they consist of either just one accented ω-word, or of several unaccented ones, which have no downstep because the downstep-inducing tonal structure is lacking.

The additional metrical boost in (23), illustrated in figure 8.5, that was difficult to explain in prosodic terms in a non-recursive model is accounted for without any problem in a recursive model: Ito & Mester (2012) claimed that the more Φ-phrases a word starts, the greater boost (and the less downstep) it has. In figure 8.7, it can be seen that each syntactic constituent is in its own Φ-phrase, not only each word but also *náoko-no áni-no* 'Naoko's brother' and *aói erímaki* 'blue scarf'. The first word *náoko-no* starts three Φ-phrases and has the largest boost, the next word *áni-no* starts only one Φ-phrase and has a minimal boost, and the third word starts two Φ-phrases and has a larger boost than the preceding and the following words, but not so large as the first one. The last word again starts only one Φ-phrase.

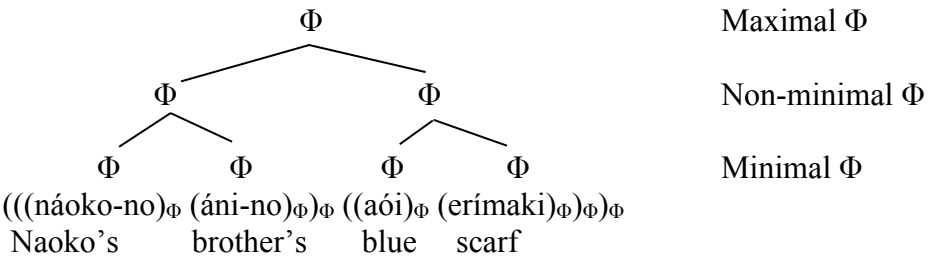


Fig. 8.7 Example (23) in the recursive model

The next level of sentence intonation is the ι-phrase. There is a low boundary tone $L_ι$ at the end of a declarative ι-phrase (see Pierrehumbert & Beckman 1988 and Gussenhoven 2004). Its effect is felt not only because of the general tendency of a declarative sentence to end low in the register, but also because of interpolation between an early but last H in the final Φ-phrase and the final $L_ι$, which causes a smoothly falling contour, see figure 8.8 for an illustration. Pierrehumbert & Beckman (1988) also assumed a low boundary tone at the beginning of the ι-phrase, but this may be unnecessary in the present analysis, as initial $L_Φ$ is already responsible for the low value observed there.

Apart from the tonal structure resulting from the assignment of tones to the prosodic domains, a further component of the sentence intonation of Japanese is what Ishihara (2004, 2007, 2011) called *Focus Intonation*. It is described in (25), slightly adapted to make it conform to the terminology of this book (reduction instead of compression is used in the original).

(25) Focus Intonation (FI) in Japanese

a. P(rosodic)-focalization: Focus is realized intonationally by a raised pitch on the focused word (or on the first word of a focus domain).

b. Post-focus compression (PFC): Post-focal material is subjected to register compression. F0 peak is lowered.

Here we use examples from Ishihara (2011) to illustrate the effect of FI. In (26)a, the sentence is all-new, which is equivalent to a wide focus on the sentence. In (26)b, only the indirect object, i.e. the second Φ-phrase of the sentence, is focused, the rest is given.

(26) a. [Aóyama-ga aníyome-ni erímaki-o ánda] $_F$
b. Aóyama-ga [aníyome-ni]$_F$ erímaki-o ánda
Aóyama-NOM sister-in-law-DAT scarf-ACC knitted
'Aoyama knitted a scarf for his sister-in-law.'

Both sentences in (26) are compared in figure 8.8. The continuous line shows the F0 contour in the all-new condition and the dashed line shows the F0 contour in the condition with narrow focus on the indirect object. A boost on the focused constituent is clearly visible, illustrating the effect of (25)a, as is the post-focal compression following the focus, accounted for by PFC in (25)b. It is noticeable that post-focally, the contour is not eliminated; there is no deaccenting as in German or English, but rather the register is compressed. The post-focal Φ-phrases may be preserved or not, the authors working on the prosodic phrasing of Japanese do not agree on this point, see below.

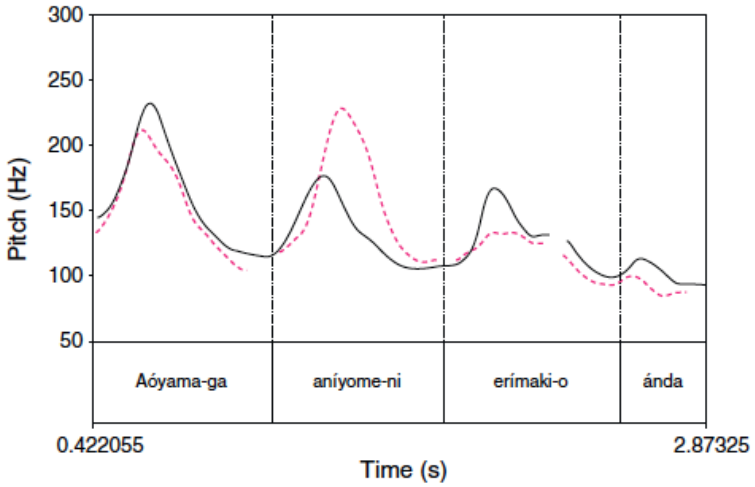


Fig. 8.8 Focus intonation in Japanese (from Ishihara 2011)

In Japanese, there is a correspondence between the scope of a wh-element and the intonation of the sentence containing it, as expressed in (27) from Ishihara.

(27) Focus Intonation – wh-Scope Correspondence
The domain of FI corresponds to the scope of a wh-question.

A declarative sentence as in (28)a shows downstep until the end of the sentence. This is illustrated in figure 8.9. Wh-questions, as in (28)b, exhibit focus intonation (FI); the narrowly focused word *nani* 'what' has a raised pitch and the post-focus material is significantly compressed by PFC, as shown in figure 8.10.

(28) a. Náoya-ga nániko-o nomíya-de nónda
Naoya-NOM something-ACC bar-LOC drank
'Naoya drank something at the bar.'
b. Náoya-ga **náni**-o nomíya-de nónda **no**?
Naoya-NOM what-ACC bar-LOC drank Q
'What did Naoya drink at the bar?'

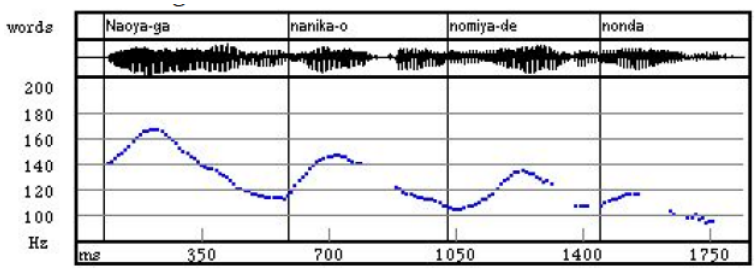


Fig. 8.9 An all-new sentence (from Ishihara 2004).

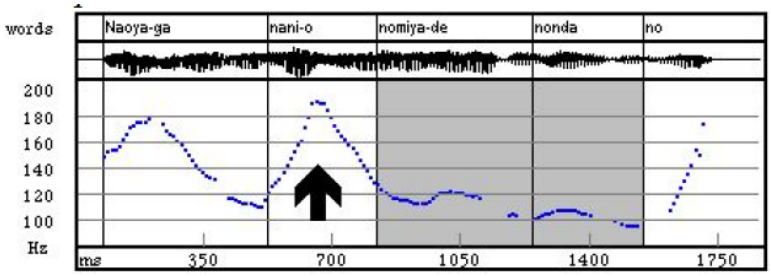


Fig. 8.10 Focus intonation in a wh-question in Japanese (from Ishihara 2004).

In standard analyses of Japanese, initiated by Pierrehumbert & Beckman (1988) (see also Nagahara 1994 and Truckenbrodt 1995), FI is obtained by inserting a MaP boundary at the left edge of the focus and deleting all the MaP boundaries thereafter. As a result of the restructuring of phrasing, downstep takes place within the newly created large MaP containing the focalized phrase and all the post-focal material.

Ishihara's (2004) alternative proposal is that focalization and PFC are independent phenomena from prosodic phrasing (see also Deguchi & Kitagawa 2002 for this view). When a wh-question has matrix scope as in (29)a, its PFC continues until the end of the matrix clause, as illustrated in figure 8.11. The crucial case is when a wh-question has embedded scope, as in (29)b. A model inserting a left boundary of a MaP, thus before *nani* 'what', and deleting all phrasing afterwards cannot explain the reset: it predicts that PFC will continue until the end of the sentence in both cases, since there, compression is just an effect of the changed phrasing. But this is not what happens. Instead, the PFC induced by the narrow focus only continues until the end of the embedded clause. After that, it stops and pitch is reset for the last part of the main clause; see figure 8.12. In this case, an ι-phrase is embedded into another ι-phrase.

(29) a. [Náoya-wa Mári-ga **náni**-o nomíya-de nónda to ímademo omótteru **no**?]$_{\iota}$
Naoya-TOP MARI-NOM what-ACC bar-LOC drank that even.now think Q
'What did Naoya still think that Mari drank at the bar?'

b. [Náoya-ga [Mári-ga **náni**-o nomíya-de nónda **ka**]$_{\iota}$ ímademo obóeteru]$_{\iota}$
Naoya-TOP MARI-NOM what-ACC bar-LOC drank Q even.now remember
'Naoya still remembers what Mari drank at the bar.'

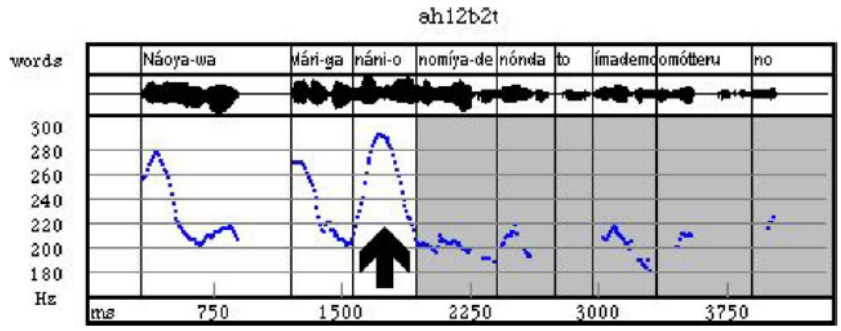


Fig. 8.11 Focus intonation in a wh-question with matrix scope (from Ishihara 2004).

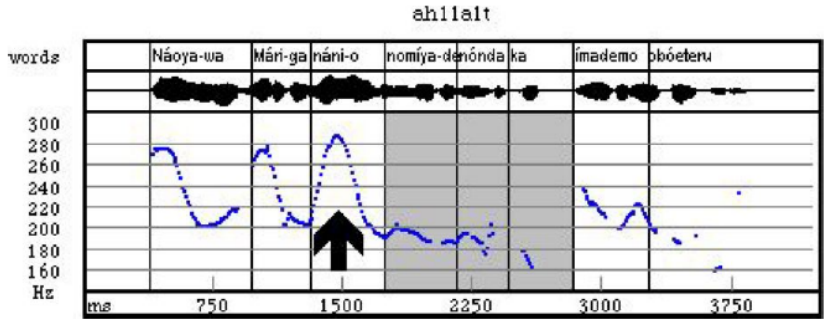


Fig. 8.12 Focus intonation in a wh-question with scope in the embedded clause (from Ishihara 2004).

To sum up this section, Japanese intonation arises compositionally from the lexical pitch accents, the phrasal tones and the boundary tones. Tonal scaling is a major component of Japanese intonation, because of downstep, pitch raising on the focus and post-focal compression.

8.3.3 Turkish

We saw in section 7.4.2 that Turkish also has accented and unaccented words, sometimes called non-finally and finally stressed words. The first class of words have a stressed non-final syllable which is invariably realized as a *falling* contour, analysed as H*L in the literature. And, like in Japanese, there is also a larger second class of unaccented words. Superficially, these words differ from Japanese unaccented words, because when uttered in isolation or in the final position of a pre-nuclear Φ-phrase, they bear a final *rising contour* aligned with their last syllable, written H_Φ. The demarcative H_Φ tone may be felt as prominent, and it has been analysed as the regular stress in the older literature, and more recently by Kan (2009) and Ipek & Jun (2013, 2014). It is analysed here as a boundary tone because, as we will see in this section, it has evident boundary properties distinguishing it from the falling pitch accent assigned to a lexically stressed syllable. Its presence is never due to the lexical properties of the word at the ω-word level, but rather to the prosodic properties of the Φ-phrase.[6]

To start with, consider sentence (30), illustrated in figure 8.13. This sentence can answer three questions, eliciting an all-new reading, a focus on the VP or a focus on the object. There is no difference in the prosodic realizations of these alternative information structures.

[6] Both Konrot (1981) and Levi (2005) pointed out several phonetic differences between the lexical falling tone H*L and the rising boundary tone H_Φ in the two classes of words, confirming the different quality of the tones.

(30) L_{Φ} H_{Φ} L_{Φ} L_{ι}
[(Yumak)$_{\Phi}$ ((MAMA-Yı)$_{\Phi}$ ye-di)$_{\Phi}$]$_{\iota}$
Yumak food-ACC eat-PAST
'Yumak ate the food.'

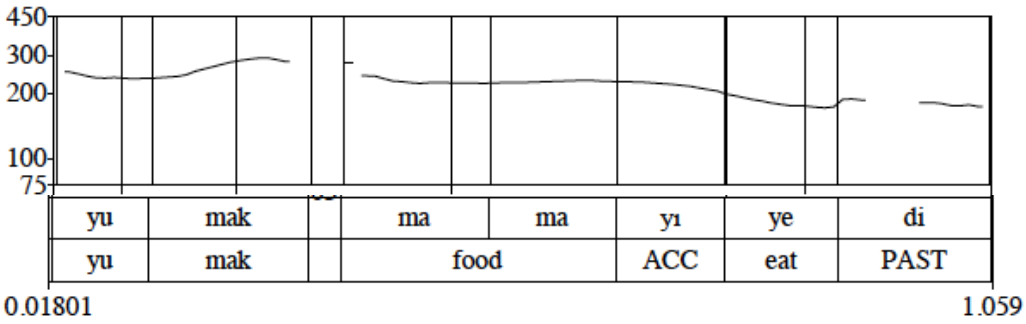


Fig. 8.13 A Turkish declarative sentence in an all-new context (from Güneş, p.c.)

It has become standard in the literature on Turkish intonation to divide an ι-phrase like (30) into three parts: pre-nucleus, nucleus and post-nucleus. The subject *Yumak* is the pre-nucleus, realized first at mid-level (written with an initial phrasal tone L_{Φ} in (30)), and then with a final high boundary tone H_{Φ} on its last syllable. The object of the sentence is the nucleus. After the initial falling contour caused by the interpolation from the preceding H_{Φ}, the contour is flat, at the mid-register of the speaker's voice. The same mid-level as in the pre-nucleus is reached, and it is again written with a phrasal tone L_{Φ}. The nucleus contains the focus of the sentence and it is felt as the sentence's most prominent position. When an utterance with canonical SOV word order is uttered in an all-new context, the object is the nucleus (Kan 2009, Kamali 2011). In other words, focus is aligned to a specific syntactic and prosodic position rather than signalled by pitch accent. Syntactically, it is pre-verbal. Prosodically, as will be shown below in more detail, it is aligned with the left edge of the final maximal Φ-phrase. The post-nucleus – the verb in (30) – is downstepped relative to the nucleus and is low and flat, except for the falling contour right at the beginning of this region, which is again an interpolation contour, this time from a low phrasal tone L_{Φ} to another low tone, the boundary tone L_{ι} at the end of the ι-phrase.[7]

The tonal analysis proposed here agrees in many points with the analyses proposed by Kamali (2011), Güneş (2013a,b) and Kamali & Truckenbrodt (2014), but it also differs from them in some aspects. First, the tones in (30), i.e. the presence of the phrasal low tone L_{Φ} at the beginning of the Φ-phrases, the high boundary tone on the pre-nucleus and the low boundary tone L_{ι} at the end of the ι-phrase are part of all the analyses. The L_{Φ} phrasal tones are responsible for the return of the F0 to a mid-low value, which remains stable until the end of both the pre-nucleus and the nucleus. Prenuclear Φ-phrases end with a high boundary tone, but in the nucleus, there is no final tonal event at all. Nevertheless, the passage from the mid-level of the nucleus to the low level of the post-nucleus is felt as the most prominent part of the ι-phrase. I propose analysing the absence of a high boundary tone H_{Φ} at the end of the nucleus as a consequence of its embedded status (see Kamali 2011 for this analysis, which is not pursued in Kamali & Truckenbrodt 2014). More specifically, the nucleus and the post-nucleus are forming a Φ-phrase together, and the

[7] Turkish has free word order. In syntax, focus has to be pre-verbal: see Güneş (2013b) who showed that the constraint on word order has a correspondent in prosody. Even though it usually appears in the pre-verbal position, as in the examples above, given material may also appear post-verbally and post-nuclearly. Alternatively given material can also be placed at the beginning of the sentence. The only prosodic constraint on word order is that the verb be part of the last Φ-phrase.

nucleus also forms an embedded Φ-phrase by virtue of being a maximal projection. It is this embedded structure that explains the absence of a high boundary tone separating the nucleus from the post-nucleus. In other words, the downstep from the nucleus to the post-nucleus is explained by recursive phrasing and tonal structure. Moreover, the L_{Φ} of the nucleus spreads to the right, but its spreading is delimited by the Φ-phrase's right edge. The unique tone of the post-nucleus is the low boundary tone L_{ι}. This tone spreads to the left until it meets the boundary separating the nucleus, indicated as | in (31). The downstep is caused by the clash between the two L tones, L_{Φ} coming from the left, and L_{ι} coming from the right; see (31) and following examples for illustration. In other words, downstep between the nucleus and post-nucleus is a purely Φ-phrase register phenomenon.

Kamali (2011) and Kamali & Truckenbrodt (2014) examined different combinations of words, both unstressed and lexically stressed; see figures 8.14 to 8.17. Sentence (31), illustrated in figure 8.14, shows that there may be more than a single pre-nuclear Φ-phrase. In such a case, both pre-nuclear Φ-phrases are at the same F0 level, and both end with a high boundary tone H_{Φ}.[8] There is no downstep between the high tones, or between the F0 register of the Φ-phrases; see Kamali & Truckenbrodt (2014) for quantitative results.

(31) L_{Φ} H_{Φ} L_{Φ} H_{Φ} L_{Φ} → | ← L_{ι}
[(Elemen-lar)$_{\Phi}$ (bunal-an-lar-ı)$_{\Phi}$ ((limonlu-ya)$_{\Phi}$ yönlendir-meli)$_{\Phi}$]$_{\iota}$
employees get.overheated-REL-PL-ACC with.lemon-DAT forward-NEC
'He should direct those who get overheated to the one with lemons.'

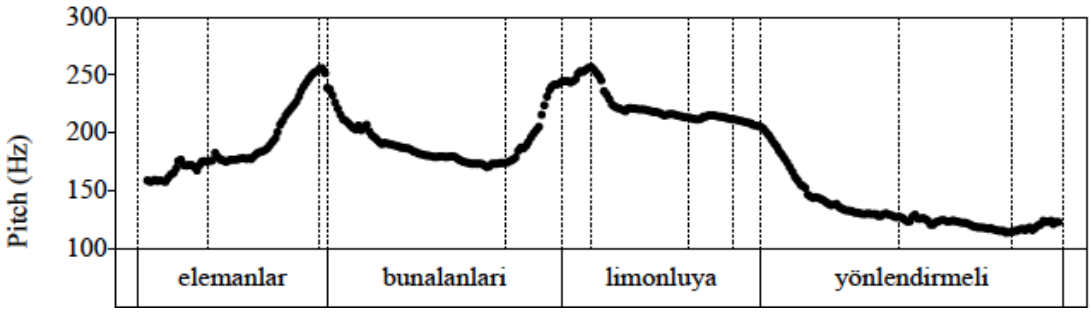


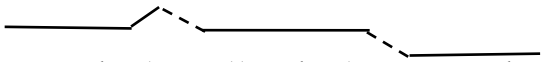
Fig. 8.14 A Turkish declarative sentence in an all-new context (from Kamali & Truckenbrodt 2014).

The focus or nucleus in Turkish does not bear an H* tone or have a particular pitch expansion aligned with it (Kamali 2011, İpek 2011, Güneş 2012, 2013a,b). The nucleus has the pitch level of the pre-nuclear non-final Φ-phrases; again see Kamali & Truckenbrodt (2014) for quantitative results. The nucleus forms a plateau which is higher than the following constituent, but on average not lower or higher than the preceding ones. By contrast, the post-nuclear verb starts lower and maintains a low pitch until the end of the ι-phrase, which is most usually near the bottom of a given speaker's pitch range.

The pattern that arises for a sentence like (30) with three unaccented ω-words is schematically illustrated in (32)a. The pre-nucleus ends with a high boundary tone. When there are more than one pre-nuclear Φ-phrase, as in (31), every pre-nucleus has the same form and the same height, as illustrated in (32)b.

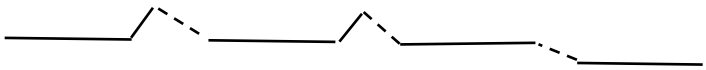
[8] See Ipek & Jun (2014) for complex pre-nucleus, containing two syntactically related nominal phrases. In their sentences, both parts of a complex pre-nucleus end at the lsame level, but the second one starts lower than the first one.

(32) L_Φ H_Φ L_Φ → | ← L_ι



a. [(Pre-nucleus)$_\Phi$ ((Nucleus)$_\Phi$ Post-nucleus)]$_\iota$

L_Φ H_Φ L_Φ H_Φ L_Φ → ← L_ι



b. [(Pre-nucleus)$_\Phi$ (Pre-nucleus)$_\Phi$ ((Nucleus)$_\Phi$ Post-nucleus)]$_\iota$

Compare next sentences with accented words, from Kamali (2011) and Kamali & Truckenbrodt (2014). The sentence in (33)a has an accented pre-nucleus H*L, the sentence in (33)b has an accented nucleus, and sentence in (33)c has an accented nucleus on the final verb. In each case, the pitch accent is assigned at the level of the ω-word.

(33) L_Φ H*L H_Φ L_Φ → | ← L_ι

a. [(NuMÁra-lar-ı)$_\Phi$ ((limonlu-ya)$_\Phi$ yönlendir-meli)$_\Phi$]$_\iota$
the numbers-PL-ACC with.lemon-DAT forward-NEC
'He should forward the numbers to the one with lemons.'

L_Φ H_Φ L_Φ H*L | ← L_ι

b. [(Bunal-an-lar-ı)$_\Phi$ ((LiMÓNlu-ya)$_\Phi$ yönlendirmeli)$_\Phi$]$_\iota$
get.overheated-REL-PL-ACC Limonlu-to forward-NEC
'He should direct those who get overheated to Limonlu.'

L_Φ H_Φ L_Φ H_Φ L_Φ H*L ← L_ι

c. [(Bunal-an-lar-ı)$_\Phi$ ((limonlu-ya)$_\Phi$ yönlenDÍR-me-meli)$_\Phi$]$_\iota$
get.overheated-REL-PL-ACC with.lemon-DAT forward-NEG-NEC
'He should not direct those who get overheated to the one with lemons.'

Figure 8.15 illustrates (33)a with a lexical accent on the pre-nuclear Φ-phrase. There is in principle no difference from a sentence without lexical stress, as in (30), except for the effect of the lexical accent itself, realized with a falling contour H*L. The presence of this pitch accent does not cancel the high tone H_Φ, which is still found at the right edge the Φ-phrase. The pre-nucleus contains two tonal peaks, one for the accented syllable and one for the boundary tone.

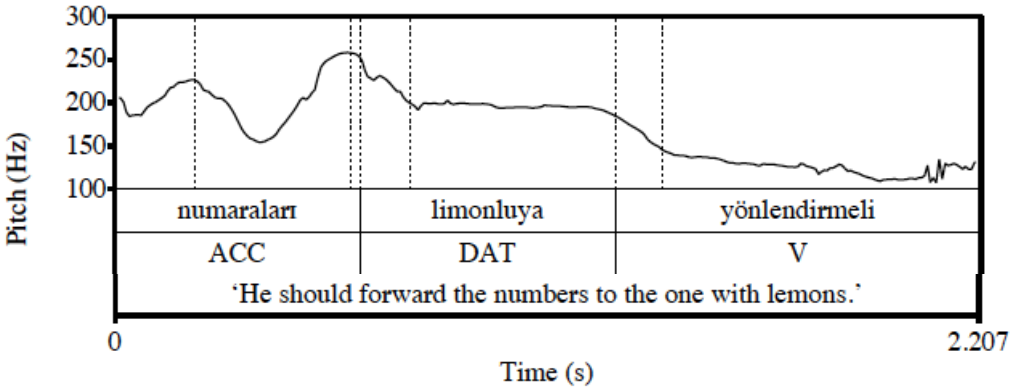


Fig. 8.15 All-new sentence with lexically accented word in the pre-nuclear position (from Kamali 2011:70).

When the nucleus has a lexical accent, as illustrated in (33)b and figure 8.16, the fall on the lexical accent is at the same time the fall into the post-nuclear Φ-phrase. There is no additional tonal marking at the right edge of the nuclear element, in particular no further step down to the post-nuclear region at the end of the nucleus. The absence of additional downstep at the juncture between nucleus and post-nucleus confirms the absence of a tone closing the nucleus. When the lexical accent H*L causes the pitch to lower in the nucleus, this lowering replaces the downstep that was observed in sentences with only unstressed words.

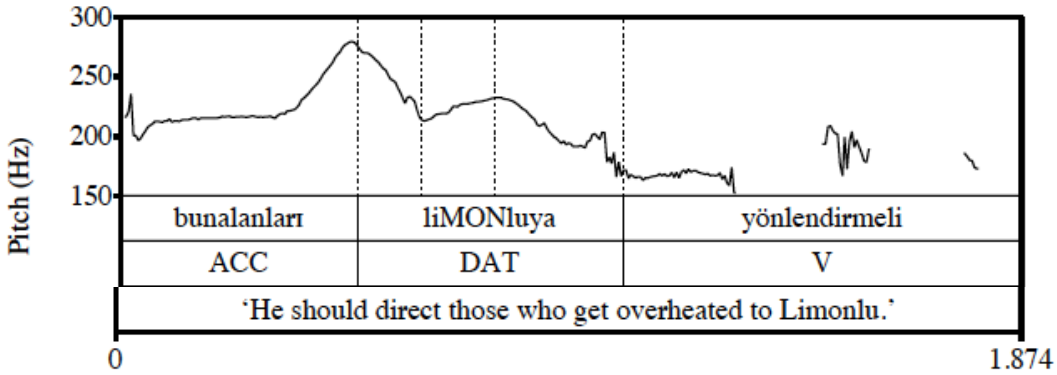


Fig. 8.16 All-new sentence with lexically accented word in the nuclear position (from Kamali 2011:73)

An accented verb can also be the nucleus, as for instance when it is negated; see (33)c. In this case, it contains the pre-stressing verbal negative suffix -'*mA*; see section 7.4.2. The H*L on the pitch accented syllable *-dir-*, preceding this suffix, induces a fall to the lowest register level, as illustrated in figure 8.17. It is conspicuous that the contour of a nuclear verb resembles the nuclear plus post-nuclear contour of unaccented constituents. Again, the nuclearity is signalled by the fact that there is no further falling contour in the sentence, and not by expansion of pitch. Also remarkable is the absence of a rise on the stressed syllable. All that is needed to signal the verb as stressed is the downstep between the first part of the word (the nucleus) and the second part (the post-nucleus).

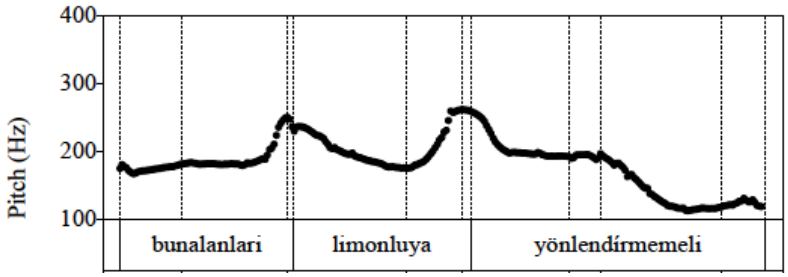
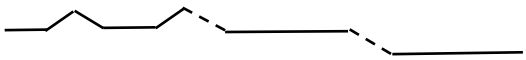


Fig. 8.17 The final verb is the nucleus (from Kamali & Truckenbrodt 2014)

The tonal patterns of sentences with accented ω-words are schematically illustrated in (34). The prosodic phrasing is unchanged in all cases, since it is the result of syntax-prosody mapping. An L_Φ always appears at the beginning of the nucleus, bringing the beginning of the nucleus to a mid-level, from where it can rise and fall again for the lexical H*L in (34)b. As shown above, when the verb is the nucleus, see (34)c, it does not realize its H*L in the same way as in the pre-nucleus or in the nucleus. Rather the L tone of the pitch accent causes a downtep at the location of the lexically stressed syllable, thus dividing the verb into a nuclear and a post-nuclear domain. It is as if the H* part of the pitch accent is not needed in this case, and is thus left unrealized. This is expressed in (34)c by putting the

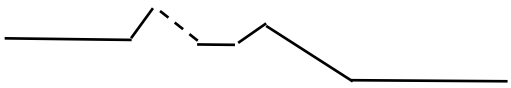
H* in parentheses. The H*L of a lexically stressed syllable is replaced by the steep fall located between the nucleus and the post-nucleus.

(34) L_Φ H*L H_Φ L_Φ → | ← L_ι



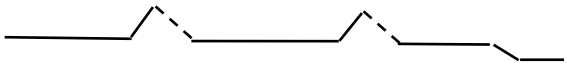
a. [(Pre-nucleus)$_\Phi$ ((Nucleus)$_\Phi$ Post-nucleus)$_\Phi$]$_\iota$

L_Φ H_Φ L_Φ H*L| ← L_ι



b. [(Pre-nucleus)$_\Phi$ ((Nucleus)$_\Phi$ Post-nucleus)$_\Phi$]$_\iota$

L_Φ H_Φ L_Φ H_Φ L_Φ (H*)L|← L_ι



c. [(Pre-nucleus)$_\Phi$ ((Pre-nucleus)$_\Phi$ Nucleus)$_\Phi$]$_\iota$

When a lexically accented verb is in the post-nuclear position, e.g. when the preceding object is the nucleus, its lexical accent is not realized. Rather it is realized flat and low. Kamali & Truckenbrodt analysed this deaccentuation as a consequence of Φ-phrase-based culminativity, resembling the pattern of deaccenting in compounds illustrated in section 7.4.2, following Levi (2002). It was shown there that in a compound, only the first accent is realized, while the following one is deleted. An alternative analysis is that the absence of pitch accent on a final verb is the consequence of compression of register. The final L_ι has the effect of completely reducing the register. The effect is that no pitch accent is possible in this domain.[9]

To sum up, Turkish has a clear intonational pattern, which arises compositionally from lexical pitch accents assigned at the ω-word level, phrasal tones assigned at the Φ-phrase level and boundary tones assigned at the ι-phrase level. Like in Japanese, all pitch accents in Turkish are lexically assigned. The presence of a nucleus is signalled both by the high boundary tone preceding it (see Ipek 2011 for some remarks on this) and the downstep following it. The effect of a medial or final focus on the immediately preceding constituent is similar to what Skopeteas, Féry & Asatiani (to appear) observed for Georgian, and Féry, Pandey & Kentner (2016) observed for Hindi. There, focus triggers as a stronger boundary tone on the Φ-phrase preceding it, clearly separating the focus from the non-focused part of the sentence.[10]

8.4 Tone languages

Tone languages form a heterogeneous group of languages, not only because of the differences in lexical tone assignment, but also as far as intonation effects are concerned.

[9] Kamali (p.c.) observed that she could not elicit unaccented verbs in nuclear position. There is no way of realizing the typical nucleus plus post-nucleus pattern on an unaccented word, showing again that pitch accents are not at play in unaccented words.

[10] Obviously, the tonal patterns of Turkish are not limited to the declarative type described in this section. Kamali & Truckenbrodt wrote, 'Similarly, pitch accents have not been documented to contribute to the pragmatics of the utterance as they do in English (Pierrehumbert and Hirschberg 1990). Nevertheless, a fair number of varying intonational pitch accents have been proposed.'

Due to the density of lexical tones, there remains little space for additional tones expressing pragmatic meanings. As a first generalization, it may be observed that all tone languages often rely on morpho-syntactic reflexes for the expression of information structure.

Since it is impossible to go into the details of the sentence intonation of tone languages because of the space limitations of this book, we examine here the consequences of a crucial difference between Asian and African languages, namely the difference in the association domains of tones. As proposed by Leben (1973), Williams (1976) and Goldsmith (1976), tones in African languages are melodies: they arise independently from syllables, and they are associated in a systematic way with prosodic domains, often ω-words, but also Φ-phrases. In their domain of association, they can be subject to all kinds of autosegmental transformations. In sections 5.3 and 7.6, we saw that tones in African languages are best analysed as forming a different phonological tier altogether; they are autosegmental and define their own grammar. Tones can also be changed for grammatical reasons, like relativization or wh-questions. In Asian languages, by contrast, tones are intrinsic to syllables, their scope is much more limited than in African languages, and they are less prone to autosegmental transformations. If tones are changed, it is a consequence of tone sandhi, but not because of grammatical roles. This difference has already been hinted at before, but it is only when considering sentence intonation that it can be fully assessed.

In this section, Beijing Mandarin and Cantonese exemplify Asian languages, and Chichewa, a Bantu language, represents the African languages. Mandarin and Cantonese have typical characteristics of Asian tone languages: tones are lexical and intrinsic to syllables. They present only few tone sandhi effects, which apply in morpho-syntactically motivated prosodic domains. Tone sandhi is not part of sentence intonation, but rather it is a ω-word or Φ-phrase effect. Tone sandhi effects do not increase or decrease under the influence of information structure, although of course, Φ-phrases can be changed under the pressure of speeded speech. Sentence intonation can be concretized in two ways in Asian languages. First by register phenomena: the high tones of specific tones can be realized with more intensity and with higher F0 when the ω-word or Φ-phrase containing it is focused. There may be post-focal compression. All in all, individual tones may be strengthened or weakened. Crucially though, the individual tones are not changed. This is illustrated with Mandarin. The second implementation of sentence intonation is illustrated with Cantonese: here additional tonal boundaries are usually ι-phrase final, and associated with a discourse particle. In Cantonese, thus, addition of tones occurs, but only at the end of ι-phrases. Here, too, individual lexical tones are not changed.

In African languages, phrasing is often changed to express information structure, and at the same time, the tonal composition of ω-words. A focused constituent may be separated from the remainder of the sentence by a phrase boundary, and phrase boundaries are often tonal or durational, as is shown with Chichewa below. The tonal correlates of phrasing affect lexical tones: high tone doubling, deletion of tones, tone insertion, spreading or dissimilation are common effects. There are no register effects, no post-focal compression, and also no or few discourse particles with variable tones, as in Cantonese.

8.4.1 Beijing Mandarin

Beijing Mandarin intonation is by far the best described of the tone languages. In principle, the tonal contour of a sentence is the result of the lexical tones and their concatenation, see section 7.6.2, plus phenomena associated with register and phrasing. There are some instances of tone assignment at the level of the Φ-phrase and the ι-phrase that we consider first. After that, phrasing and pitch range phenomena are shortly discussed as well. It will

be shown that tonal scaling of entire prosodic domains is a crucial aspect of intonation in Mandarin.

Tone assignment at the level of the Φ-phrase and ι-phrase is relatively scarce. There are only few syllables unspecified for tones that can be assigned tone at a phrasal prosodic level. These belong to two classes: first, syllables with *neutral tone*, in (35)a, and second, modal particles, in (35)b.[11] Typical examples of words in which the second (unstressed) syllable has a neutral tone include grammatical morphemes, lexical items, diminutive terms, and reduplication. Recall the notation: mā1 (Tone 1, 55, H), má2 (Tone 2, 35, R or LH), mǎ3 (Tone 3, 214, L) and mà4 (Tone 4, 51, F or HL) from section 7.6.2. In the following, the romanized notation of the tones is adopted in the examples for simplicity.

(35) a. grammatical morphemes: *làde* 'something spicy', lexical item: *bōli* 'glass', diminutive terms: *mèimei* 'sister (diminutive)', reduplication: *xiángxiang* 'to think for a little while'

b. particles: *-ba* 'agreement-soliciting', *-ma* and *-a* 'pragmatic particles', *-le* 'verbal suffix', *-zi* 'nominal suffix'

Syllables with neutral tones seem to acquire their phonetic value from the preceding syllable (Yip 1980, Shih 1997, but see Chen & Xu 2006 for a different interpretation). After tone 1 or 2, they fall to mid-level. In the case of tone 3, the fall-rise of this tone is distributed over the lexically specified syllable, thus the one carrying this tone, and the following unaccented syllable. In the same way, when a neutral tone follows, the fall of tone 4 is spread over two syllables. The behaviour of neutral tones following tones 3 and 4 is thus not compatible with tone assignment per se; rather it is just a description of how syllables without tone acquire their phonetic contour. However, the addition of a falling contour in the case of tones 1 and 2 may suggest the presence of an L_ι in declarative sentences, which can only be realized when neutral tone syllables are there to carry this boundary tone.

In the case of toneless particles, as illustrated in (35)b, things are different: particles can have an assigned pragmatic tone. Peng et al. (2005:248) cite the sentence in (36), which can end in a high or in a low tone.

(36) Tāmen bú mài yúsǎn ma
do NEG sell umbrellas PRT
'Don't they sell umbrellas?'

If the final particle bears a high tone, the speaker is asking a yes-no question. Peng et al. (2005:248) commented: '[the] boundary tone suggests a presupposition that the store should sell umbrellas. Thus, this can convey surprise, if the addressee is someone who was sent to buy an umbrella and came back empty-handed.' In the second case, they claimed that the low 'boundary tone effectively makes the utterance a statement. It might be produced by a speaker to soften an explanation of why he came back empty-handed. The English equivalent might be something like, "Well, but they don't sell umbrellas."' In other words, these boundary tones are not always present. They add pragmatic nuances to particles that already carry pragmatic meanings. Phonologically, they can be analysed as

[11] As Peng, Chan, Tseng, Huang, Lee & Beckman (2005:237) wrote, 'Containing a neutral tone syllable is one of the clearest indicators that a recurring sequence of syllables is a fully lexicalized polysyllabic word rather than a more decomposable compound or even a phrase, as in monomorphemic *dongxi* 'thing' versus the obviously compound *dongxi* 'east-west'.'

phrasal tones in need of special segmental material. The combination of particle and tone for pragmatic aims is a complex and understudied issue.

Turning now to tone scaling effects, it has been repeatedly shown in the literature that focus has an effect on the duration and on the pitch register and contour of lexical tones (Gårding 1987, Y.Chen 2010, Jin 1996, Shih 1988, Xu 1999, 2005, Y.Chen & Gussenhoven 2008). Xu (1999) used SVO sentences consisting of five syllables and three words with different lexical tones, inserted in four different focus structures (see section 5.7). The result for a sentence consisting of a sequence of Tones 1, 3, 2, 1, 1 (noted H, L, R, H, H by Xu) is shown in figure 8.18. It is clearly visible that, in this tonal configuration, an initial focus raised the F0 of the focused word and compressed the F0 of the post-focal domain (dashed line), a medial focus raised the high part of the focused word with a rising tone, and also compressed the post-focal domain, but it had no effect on the pre-focal material (continuous line), and a final focus had only a mild raising effect on the focused constituent, and no effect on the pre-focal material (the dotted line). It had nearly the same contour as the sentence uttered in an all-new context (the line called 'none'). This latter effect may be due to the fact that the last tone never changed: in all conditions (not only in figure 8.18), the last tone was a high tone.

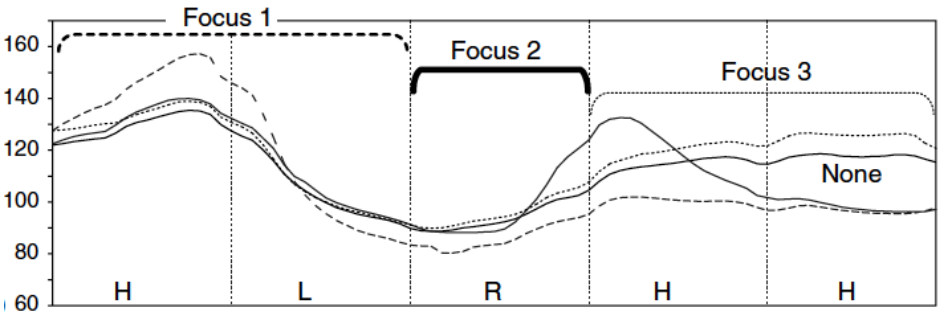


Fig. 8.18 The effect of focus in Mandarin in a sentence consisting of words with different tones (from Xu 2005:232)

Obviously, the lexical tones are decisive for the tonal pattern of the whole contour. There is no additional tone at the phrase level or at the sentence level for the sake of expressing focus. There is also no pitch accent and, in this particular case, no change in the global phrasing. The phonetic changes induced by focus are located in the lexical tones and in the way they influence each other. Depending on which tones are involved, the effect of focus changes considerably. It is most visible in a sentence consisting of high tones only, but is insignificant when the last but one tone is low and a focused rising-falling configuration precedes. The reader is referred to Xu's seminal study to assess the differences. His results confirm Shih's (1988) study on contextual tonal variations in Mandarin due to downstep. Shih found that the amount of F0 lowering in a tone due to downstep differed depending on which of the four tones preceded it. The F0 was lowered most when the tone was preceded by the L tone, but only moderately so when preceded by the R and F tones. There was no lowering after an H tone.

On the basis of their own results, Y.Chen & Gussenhoven (2008) partly confirmed Xu's results and partly rejected his analysis. They tested three focus conditions called NoEmphasis, Emphasis and MoreEmphasis (roughly equivalent to all-new, narrow focus and corrective focus). They found a robust and gradual increase in syllable duration from the first to the third of these conditions, as well as a non-gradual expansion of F0 going from NoEmphasis to Emphasis, but no such effect going from Emphasis to MoreEmphasis.

They also found a robust effect of tones, with a falling tone expanding more than rising and high tones. They wrote:

> 'Together with the specific details of the F0 range expansion and syllable duration increase, a number of further phonetic adjustments under emphasis make it clear that lexical tones are produced with enhanced distinctiveness of their F0 contours. Specifically, the wider F0 range was mainly manifested by raised F0 maxima, and sometimes by lowered minima, depending on the identity of the tone. Moreover, the low F0 of an L tone was sometimes cued by creakiness. In the case of LH and HL tones, they showed clear rising or falling F0 contours despite the significant increase of syllable duration. Moreover, the rise for the LH tone was strongly delayed, while the HL tone had a delayed as well as raised F0 peak. As an *ensemble*, these adjustments enhance the distinctiveness of the contrasts among the lexical tones' (Y. Chen & Gussenhoven: 2008744).

They concluded from their results that focus cannot be considered as a simple function of pitch range manipulation, in the same way as in intonation languages, as proposed by Xu (2005). Recall from section 5.7 that for Xu, focus is understood as a phonetic instruction to the articulators. According to Y.Chen & Gussenhoven, the phonetic correlates of focus in Mandarin are compatible with hyperarticulated phonemic features.[12] They compare tonal realization under emphasis with vowel articulation under emphasis: the greater articulatory force applies equally to segments and tones of the focused syllables. In Xu & Xu (2005), sentences are assumed to always have the same partition: a pre-focus phase, in which accents are fully intonated, the focus itself, with a boost on the stressed syllable, and a post-focus phase with optional post-focal compression. According to Shih (1988) and Y.Chen & Gussenhoven (2008), the effects of focus are tone-dependent and concern all phonetic correlates, not only F0. The question arises of how these effects are phonologically represented. Y.Chen & Gussenhoven assume mediation by an abstract metrical component responsible for a higher prominence of the focused constituent as compared to the other constituents.

8.4.2 Hong Kong Cantonese

Recall from section 7.6.1 that Cantonese has 6 lexical tones, and that nearly every syllable is specified for tone. Cantonese intonation has been described a number of times, see for instance Lin (2002), Flynn (2003) and Wong et al. (2005). According to Flynn, who provided a phonetic analysis of a live radio interview, an overall declination is observed in Cantonese intonation for all sentence types both at the ι-phrase level and at the level of the Φ-phrase (called 'intonation group' by Flynn). This study showed that lexical tones are able to maintain their contrast in spite of the intonation-driven declination, because, even at the end of the prosodic domain, room is still available for generating pitch differences with relative heights within the shrinking pitch range in the intonation group, a phenomenon that was also observed above for Swedish, Japanese and Mandarin Chinese. At the beginning of a new Φ-phrase, F0 is often reset.

Further, both Flynn (2003:45-46) and Lin (2002:89) noted that while stress does not occur in Cantonese, pragmatic contrast can be achieved through prominence. The duration of a selected syllable is lengthened, and this is often accompanied by expansion of pitch range, in a way similar to what was described for Mandarin Chinese in the preceding section.

[12] In questions, the F0 of an entire sentence is raised (Haan 2002, Ho 1977, Shen 1990). Pitch raising of an entire sentence for pragmatic purposes is often observed in tone languages.

Law (1990:107) suggested that Cantonese has boundary tones which can be added after the last lexical tone of an intonation phrase; see also Yip (2002) and Sybesma & Li (2007) for confirmation. These boundary tones elicit pragmatic meanings: the high-tone variant adds a tentative pragmatic meaning to the sentence, the low tone is more assertive, and the mid tone can be interpreted as neutral. The boundary tones can exceptionally be realized by themselves, at the end of the sentence, as an additional tone on the last syllable, but generally they need a discourse particle functioning as a TBU. An example from Law appears in (37): the particle ɛ marks the utterance to which it is attached 'as a suggestion made by the speaker to the hearer'. Depending on the tone on the final particle (55=high, 11=low, 33=mid), the sentence can be translated in different ways. In (37), part of the glossing and translations are from Sybesma & Li (2007), and the rest is from Law, i.e. the tones and some of the translations.[13]

(37) ngo-dei jat cai heoi tai-hei ɛ
we together go see-movie PART
'How about we go see a movie $ɛ^{55}$?' [what do you think?]
'How about we go see a movie $ɛ^{11}$' [would be fun]
'Let's go see a movie $ɛ^{33}$!' [I won't take 'no' for an answer]

Independently of their intonation, the roles of Cantonese discourse particles are extremely difficult to pin down, and appear to be highly variable. It is often said that there are about 40 of them, and that they all differ in their pragmatic import. A significant contribution of Cantonese discourse particles is their influence on intonation when utterances are produced with the conventionalized pitch of the particles; see Law (1990) and Sybesma & Li (2007) among others for in-depth discussion of a large number of particles.[14]

8.4.3 Chichewa

The large group of Bantu languages, exemplified by Chichewa in this section, displays different prosodic correlates of phrasing, but also similarities among sub-groups. Common to many Bantu languages is the underlying two-tone system (H, L), with the high tone as the active one (Kisseberth & Odden 2003). Tone and duration are generally manipulated in Bantu phrasing, although the exact phonetic properties of this manipulation can vary from one language to another (Cole 1955, Doke 1954, Zerbian 2007). For instance in Chimwiini (Kisseberth & Abasheikh 1974), a Bantu language with distinctive vowel length, underlying long vowels are shortened at the end of phrases. Kimatuumbi (Odden 1987) also shortens vowels, but within phrases rather than at phrase ends. As for the tones, Kisseberth & Odden (2003:62) claimed that mobility of high tones is one of the most fundamental phenomena of Bantu tonology. The end of a Φ-phrase is the location of different effects. In Kinyambo (Bickmore 1990), a high tone is deleted before another word within the same phrase. In Chaga (McHugh 1999), Φ-phrase-final high tones are raised to superhigh; high tones can be inserted in this context. By contrast, in Northern Sotho, high tones are blocked at phrase boundaries. It is a general feature of Bantu

[13] Law commented on these examples in the following terms: 'The difference among the three utterances seems to be that (37)c is a neutral suggestion; whereas (37)a, which ends with '$ɛ^{55}$', sounds more tentative, so that the hearer is free to propose other things; and (37)b, compared with (37)c, appears to be stronger in force, having the connotation that the speaker is more determined in finding agreement from the hearer. […] This would be predicted if we assume that '$ɛ^{55}$' is derived from associating 'ɛ' with the weakener, '$ɛ^{11}$' from linking 'ɛ' to the strengthener, and that '$ɛ^{33}$' results from the default tone associated with an inherently toneless particle.'

[14] See also Ding & Féry (2014) for examples related to discontinuous nominal phrases in Cantonese.

languages that the last syllable of a Φ-phrase shows tonal extrametricality. This is then signalled by two different and often mutually exclusive effects: either the final syllables are not associated with a high tone, as in Chichewa, see below, or they cannot act as a target for tonal rules that spread or shift high tones, as in Zulu (Zerbian 2007).

The prosody of Chichewa has been described extensively by Kanerva (1990), who showed that prosodic phrasing originates by mapping Φ-phrases and ι-phrase to syntactic constituents. He also showed that information structure plays a role in prosodic phrasing. Like in other Bantu languages, the phenomena associated with phrasing can be classified in two groups: changing vowel duration and changing tone. The phonological cues associated with the final edge of the Φ-phrase in Chichewa are summed up in (38).

(38) a. Penultimate lengthening: A vowel in the penultimate syllable of a Φ-phrase is lengthened.

b. Tone retraction: An underlying H tone is retracted from the final mora of a Φ-phrase to the penultimate mora.

c. Non-final doubling high: An H tone is doubled (spread one mora to the right) as long as the target of spreading is not in the domain-final disyllabic foot.

Penultimate lengthening is the process by which the last but one syllable of a Φ-phrase is lengthened. A word with a lengthened syllable has sometimes been analysed as a stressed word (see for instance Doke 1927/1992, Büring 2010 and Downing & Pompino-Marschall 2014). Problematic for this analysis is that discourse-given constituents are lengthened exactly like new or focused ones: there is no difference between prominent and non-prominent Φ-phrases for the duration of the penultimate syllable. Moreover in Chichewa and in Zulu, lengthening only occurs at phrase boundaries, not at word boundaries, so that lengthening can safely be understood as phrase-correlated. Thus interpreting a purely phrasal effect as an indicator of prominence assumes that prominence can be unrelated to lexical stress or information structure. It simply implies that correlates of phrasing are intrinsically prominent. Tone retraction and non-final doubling high are illustrations of tonal mobility in the phrasal domain. In tone retraction, a high tone underlyingly associated with a final syllable cannot appear at this location and shifts to the preceding syllable. Non-final doubling high illustrates high tone spreading to the following syllables, but again as long it does not provide a final syllable with a high tone. Both tonal rules are means to the same aim: avoidance of a Φ-phrase final high tone, and presence of a defaut low tone instead. Thus every Φ-phrase ends with a lowering of the F0.

As an example of the application of these processes, consider (39)a, /ⁿjiⁿgá/ ‘bicycle’ with an underlying H tone on the final mora, which is pronounced without any change when it is not Φ-phrase-final. The penultimate vowel is lengthened when the word is phrase-final. Moreover the underlying final tone is retracted to the penultimate syllable because of tone retraction. In (39)b, the word /kugúla/ ‘to buy’ is changed because of non-final doubling. This process spreads the underlying non-final H tone to the following mora. Non-final doubling is blocked, however, when the target of spreading is within the last two syllables of a Φ-phrase.

(39) a. /ⁿjiⁿgá/ ‘bicycle’ in (*ⁿjiⁿgá yábwi:no*)$_{Φ}$ ‘good bicycle’ but (…ⁿjí:ⁿga)$_{Φ}$

b. /kugúla/ ‘to buy’ → [kugúlá] in (*kugúlá nya:ma*)$_{Φ}$ ‘to buy meat’

Let us examine some of the syntax-driven phrasing in Chichewa. Some of Kanerva’s original examples appear in (40) to (42). Under sentence-wide focus, Chichewa’s pre-verbal subjects are only optionally wrapped in a Φ-phrase of their own followed by a Φ-

phrase for the entire verbal phrase, as shown in (40)a and (42) (Kanerva 1990:157,159, also Truckenbrodt 1999 and Samek-Lodovici 2005:732). (40)b-c and (41) illustrate cases where a subject is phrased in a separate Φ-phrase from that of the VP. When the subject is focused, as in (41), or when it is a (contrastive) topic, it is in a separate Φ-phrase (Cheng & Downing 2011, Downing & Pompino-Marschall 2013). In short, a focused subject and a topic are obligatorily aligned with the right edge of a Φ-phrase. Other Bantu languages behaving like Chichewa are Tsonga (Kisseberth 1994) and Chimwiini (Kisseberth & Abasheikh 1974).[15]

(40) a. (tinabá kálú:lu)$_{Φ}$
we-stole hare
'We stole the hare.'
b. (fí:si)$_{Φ}$ (anadyá ḿká:^{n}go)$_{Φ}$
hyena ate lion
'The hyena ate the lion.'
c. (kagá:lu)$_{Φ}$ (ka:fa)$_{Φ}$
small.dog died
'The small dog died.'

(41) {Who wrote a letter to the woman?}
[(Malú:me$_{F}$)$_{Φ}$ (a-ná-lémbera mkází kála:tax)$_{Φ}$]$_{ι}$
1.uncle 1SUBJ-TAM-write.to 1.woman 9.letter
'The uncle wrote the woman a letter.'

(42) [(wa-á-pátsa bambo chi-yá:ni)$_{Φ}$]$_{ι}$
1SUBJ-TAM-give 1.father 7-what
'What has s/he given to father?'

The next examples in (43) and (44) show that the entire VP is contained in a single Φ-phrase when the sentence is all-new. This behaviour is not specific to Chichewa among the Bantu languages but has also been reported for Northern Sotho (Zerbian 2007), Kinyambo (Bickmore 1990) and Haya (Byarushengo, Hyman & Tenenbaum 1976). In general, a conjoint phrasing of verb and following object is very common across Bantu. It is attested in Chichewa, Chimwiini (Kisseberth & Abasheikh 1974), Tsonga (Kisseberth 1994) and Kinyambo (Bickmore 1990).[16]

(43) (M-fúumu)$_{Φ}$ (i-ná-pátsa mw-aná zóóváala)$_{Φ}$
9-chief SIMPLE.PAST-give 1-child 10.clothes
'The chief gave the child clothes.'

(44) a. {What happened?/What did he do?}
(([anaménya nyumbá ndí mwáála]$_{F}$)$_{Φ}$)$_{ι}$
3SBJ-TAM-hit 9.house with 3.rock
'He hit the house with a rock.'

[15] Northern Sotho deviates from this pattern by phrasing a left-dislocated element in a separate Φ-phrase (dislocation being clearly indicated by clitic doubling). Other Bantu languages behave like Northern Sotho and incorporate their left-dislocated elements into the following Φ-phrase, like Haya (Byarushengo, Hyman & Tenenbaum 1976) and Chizigula for adverbs (Kenstowicz & Kisseberth 1990).
[16] Other Bantu languages like Chimwiini (Kisseberth & Abasheikh 1974) and Kimatuumbi (Odden 1987) show separate phrasing of two post-verbal objects.

The effect of focus in Chichewa amounts to the following: focus targets specific constituents or syntactic heads, which occur in a Φ-phrase of their own, as do all following constituents. If the focused constituent is final, no special phrasing is required. The behaviour of post-focal given constituents is best analysed as right-dislocation, as acknowledged by Bresnan & Mchombo (1987).[17]

(45) a. {What did he hit with the rock?}
((anaménya [nyuúmba]$_F$)$_Φ$ (ndí mwáála)$_Φ$)$_ι$

b. {What did he do to the house with the rock?}
(([anaméenya]$_F$)$_Φ$ (nyuúmba)$_Φ$ (ndí mwáála)$_Φ$)$_ι$

To sum up, all Bantu languages considered have in common that the verb is phrased together with the following object in one prosodic constituent, and that right-dislocated given objects are phrased separately. Left-dislocated constituents may or may not be phrased together with the main clause, depending on the language. But these languages differ from each other in the phrasing of subjects and further constituents.

It is not an accident that when investigating the effect of focus on grammar in Bantu languages, researchers have insisted on the prosodic phrasing and the morpho-syntactic changes, but they have largely ignored tonal changes. As Zerbian (2006) has shown for Northern Sotho, there is most often no tonal effect at all, except for those accompanying phrasing. In the same way, Downing & Pompino-Marschall (2014) did not find any tonal effect of focus in Chichewa. It can be that the there is truly no sentence intonation besides the changes due to the phrasing and summed up in (38).

It has been repeatedly shown that tones can also accompany purely syntactic changes, and the literature abundantly illustrates this. Here, only one example is shown for Chichewa. Mchombo (2001) shows that if a verb with a low tone is in a relative clause, the low tone is changed into a high tone, compare (46)a with (46)b-c with the verb uku-sáká 'it is hunting'. The tone change happens both in case the relative marker u-méné is present and in case it is deleted or dropped. The tone marking functions as a phonological cue of the different constructions.

(46) a. mkángo **u-ku**-sáká mbûzi
3-lion 3SM-pres-hunt 10-goats
'The lion is hunting goats'
b. mkángó u-méné **ú-kú**-sáká mbûzi
3-lion 3SM-relpro 3SM-pres-hunt 10-goats
'The lion which is hunting goats.'
c. mkángó **ú-kú**-sáká mbûzi
3-lion 3SM-pres-hunt 10-goats
'The lion which is hunting goats.'

Other African tone languages have special intonational properties. According to Rialland & Embanga Aborobongui (2015), Embosi, a Gur language, has no Φ-phrase. This is a two-tone language without downdrift, e.g. the high tones are not downstepped when a low tone separates them. Moreover, complex sentences have no segmental or tonal specification whatsoever that could be interpreted as indicators of a phrasing between two parts of a clause. Both inside of the ι-phrase and across two clauses of a complex sentence, hiatus is

[17] According to Downing & Pompino-Marschall (2013), focus does not necessarily cause separate Φ-phrases.

avoided by vowel deletion or coalescence. Genzel (2013) shows that Akan, a Kwa language, has no tonal correlate of information structure. The only change in intonation in this language is a raising of register and a final low tone at the end of interrogative sentences, a result that she interprets as supporting Rialland's (2009) description of lax question prosody.

The survey of effects of sentence intonation in tone languages is very sketchy. The literature is much richer that what can possibly be summed up here, and the reader is encouraged to consult it extensively.

8.5 Phrase languages

Most typologies of word prosodies predict that languages unambiguously belong to a single prosodic system, where the prosodic systems are organized in cells in a 2 x 2 table, and the systems have one value each for exactly two binary features, like [± stress accent], [± tone] defined at the lexical or morphemic level (see Beckman 1986, Ladd 1996 and Jun 2005 among others, but see Hyman 2006 for another perspective). For instance, English is [+stress accent, –tone] and both Swedish and Mandarin Chinese are [+stress accent, +tone]. The features are often defined in terms of phonetic correlates like pitch, duration or intensity. If a language has neither stress nor accent nor tone at the lexical level, it is defined entirely by negative values. However, all languages have tonal excursions on at least some words, and the question is where the excursions come from. As should have become clear from the preceding sections, more often than not, tonal excursions are not assigned to ω-words directly, but at another level of the prosodic hierarchy, at the Φ-phrase or ι-phrase level, and sometimes as reflection of other parts of grammar (syntax or pragmatics).

In this section, a 'new' group of languages is discussed, called here 'phrase languages', as already proposed in Féry (2010). Phrase languages differ from intonation languages in that their tonal specifications are mostly assigned at the level of Φ-phrases and ι-phrases, and do not correlate with lexical stress. Specifications at the level of the word are thus sparse, absent or only weakly implemented. In phrase languages, most tones are non-lexical (or 'post-lexical'). Phrasal tones can be peripheral (boundary tones) or non-peripheral. French and Korean are characterized by an absence of tone and stress at the ω-word level and are thus typical phrase languages. Other languages with similar properties are Georgian, Hungarian, Finnish (Arnhold 2014a), West Greenlandic (Arnhold 2014b), as well as a number of Indian languages (Patil et al. 2007, Féry, Pandey & Kentner 2016). Some of these languages may have lexical stress, like Georgian and Hindi, but the phrasal tones are predominant, and cancel or partially overwrite lexical pitch accents, as is the case in Hindi (see Féry, Pandey & Kentner 2014) or Georgian (Skopeteas, Féry & Asatiani 2009, Skopeteas & Féry 2014). Turkish was categorized as a pitch accent language in section 8.3, because this language has tonal specifications at the level of the ω-word in part of its vocabulary, but it must be observed that a large part of its tonal pattern arises at the Φ-phrase level, and it is because of this predominance of the phrasal tonal structure that Turkish may be called a phrase language (see Güneş 2012, 2015 for this suggestion). In the following, French and Hindi are discussed as examples of phrase languages.

8.5.1 French

French is a typical phrase language. It has neither stress nor tone at the ω-word level, and the tonal patterns of this language arise at the Φ-phrase level. Interestingly, French is nevertheless very rich in intonational patterns. Hirst & Di Cristo (1996) and Di Cristo (1998) claimed that Φ-phrases in French (which they called Tonal Units) are delimited by

an ‘early rise’ and a ‘late rise’. Jun & Fougeron (2000, 2002) also found an initial rise at the beginning and a final rise at the end of the Φ-phrase (which they called Accentual Phrase or AP). They analysed the first rise as a phrasal boundary and the final one as a pitch accent assigned to the Φ-phrase.[18] The definition of Jun & Fougeron’s AP is rather vague: ‘one or more Content Words (Wc), optionally preceded by one or more Function Words (Wf), and demarcated by the primary stress’ (2002:149-150).

Welby (2006) experimentally investigated the patterns used in declarative sentences containing a focus, concentrating on how the low and high tones of the two rising contours proposed by the authors mentioned above are associated and phonetically aligned to the text. She used lists of two or three coordinated nouns like *le randonneur et le mélomane* ‘the hiker and the music lover’ in the subject position of a sentence, as instances of APs. She found that the peripheral tones (L in the case of the initial rise, and H in the case of the late rise) are anchored in a stable way, but the internal tones (H in the initial rise, and L in the final rise) vary greatly in their association behaviour. She also found an important difference between the two rises in duration data. The last H is accompanied by syllable lengthening, while the first one is not. Results of articulatory studies also reveal that the first H is articulated more weakly than the last one (see Rolland & Loevenbruck 2002 for similar results). Due to variation in alignment, the rises in French do not behave like typical pitch accents since they cannot be claimed to be associated with a metrically strong syllable. On the basis of her results, Welby tentatively suggested that both the initial and the final contours, not only the initial one, mark boundaries, emphasizing the difference between French and English regarding the role played by pitch accents.

In the following, all tones are analysed as phrasal tones, which does not mean that these tones do not express prominence. In fact they often do. The notion of phrase language suggests that the assignment of tones does not happen at the ω-word, but at the Φ-phrase and ι-phrase levels. In the example mentioned above from Welby, *le randonneur et le mélomane*, both nouns form a Φ-phrase each, and together they form a larger Φ-phrase (see (47)a). It is assumed here that the phrasing is based on the morpho-syntax and that it is recursive; see also Féry (2014) and Destruel & Féry (2015) for detailed discussions of prosodic phrasing in French.

D’Imperio & Michelas (2014) found for the subject in their example (47)b that there is a smaller downstep between the high tone of the first word (on the last syllable of *mamie*) and the high tone of the last word (last syllable of *Rémy*) than between the first word and the second word (last syllable of *amis*), as shown in figure 8.19. They suggest that this difference in the scaling of the high tones requires an additional level in the prosodic structure, and they call it *intermediate phrase*, following Beckman & Pierrehumbert (1986) for English. But in fact, assuming a recursive prosodic structure, as shown in (47)c, explains the downstep and reset pattern of the complex noun phrase better, see also Féry 2014. The high boundary tone of the entire Φ-phrase (on *Rémy*) is higher than the high tone on *amis*, because it ends a maximal Φ-phrase. D´Imperio & Michelas find that it nearly or completely reaches the high reference line of the entire Φ-phrase. Assuming that each Φ-phrase defines its reference line, the last word returns to this height

[18] Delais-Roussarie (1995), Delais-Roussarie et al. (2002), Post (2000, 2002) and Gussenhoven (2004:253) proposed that all rising tones should be treated as pitch accents. A number of studies on French have distinguished between the so-called *accent logique* or *accent grammatical* for the final tonal contour and the *accent d’insistance* for the initial one (see, for example, Grammont 1933, Malmberg 1969, Léon 1972, Martin 1975, 1980, Rossi 1980, Dell 1984, and Mertens 1990), although some authors have made a distinction between the initial rising contour and the *accent d’insistance* (see, for example, Di Cristo 1998:198). We won’t be concerned with the difference here, but see Féry (2001) for some remarks on the *accent d’insistance*.

or near it. The high tones found on each content word are indicators of the following left boundary in the complex Φ-phrase.

(47) a. ((le randonneur)$_{Φ}$ (et le mélomane)$_{Φ}$)$_{Φ}$
the hiker and the music.lover
b. ((la mamie)$_{AP}$ (des amis)$_{AP}$ (de Rémy)$_{AP}$)$_{ip}$ (demandait l´ institutrice)$_{ip}$
b. $_{Φ}$(la mamie $_{Φ}$(des amis $_{Φ}$(de Rémy))) (demandait l´ institutrice)$_{Φ}$
the granny of.the friends of Rémy asked the teacher
'The granny of Rémy's friends asked for the teacher.'

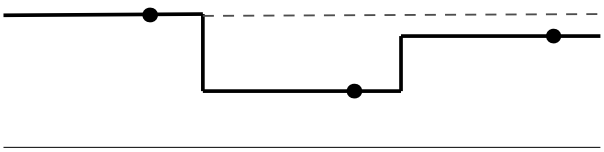


Fig. 8.19 Scaling of tones in a complex Φ-phrase in French

Consider the example in (48), from Féry, Hörnig & Pahaut (2010), for an illustration of a similar downstep relationship, but this time the first Φ-phrase is not embedded into a larger one. Figure 8.20 illustrates the sentence. As before, each Φ-phrase is separated from the following Φ-phrase by a rising tonal pattern, written $L_{Φ}H_{Φ}$. $L_{Φ}$ is aligned with the beginning, and $H_{Φ}$ with the end of each non-final Φ-phrase. The final Φ-phrase contains just an initial $H_{Φ}$ and there is a low boundary tone $L_{ι}$ at the very end of the ι-phrase. The scaling relation between the high tones of each word is again dependent on the syntactic and prosodic structure. This sentence was realized in an experimental set-up using spontaneous speech.

$L_{Φ}$ $H_{Φ}$ $L_{Φ}$ $H_{Φ}$ $L_{Φ}$ $H_{Φ}$ $H_{Φ}$ $L_{ι}$
(48) [(A ta droite)$_{Φ}$ (donc à la gauche (du gorille)$_{Φ}$)$_{Φ}$ (un cheval)$_{Φ}$]$_{ι}$
To your right thus to the left of.the gorilla a horse
'To your right, thus to the left of the gorilla, there is a horse.'

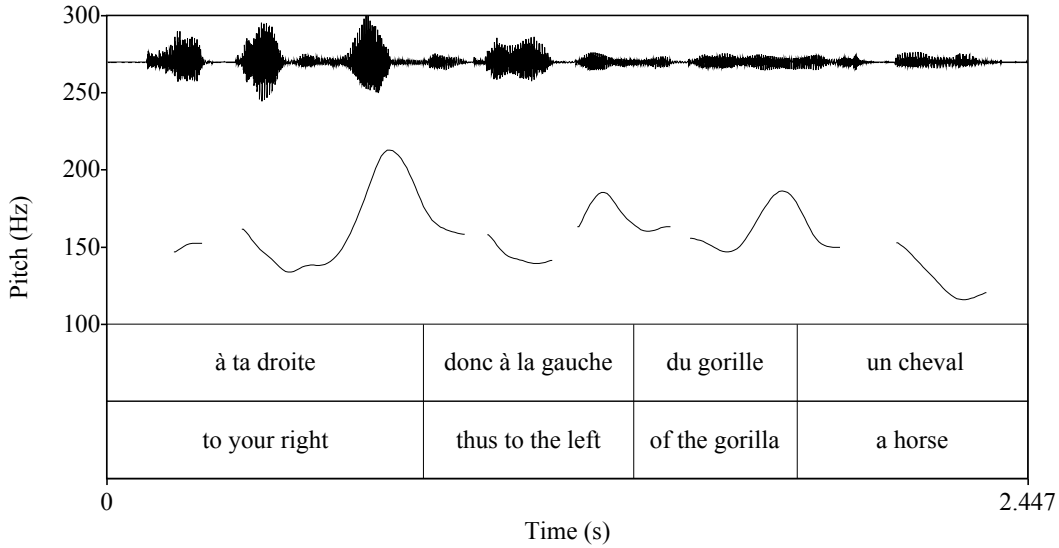


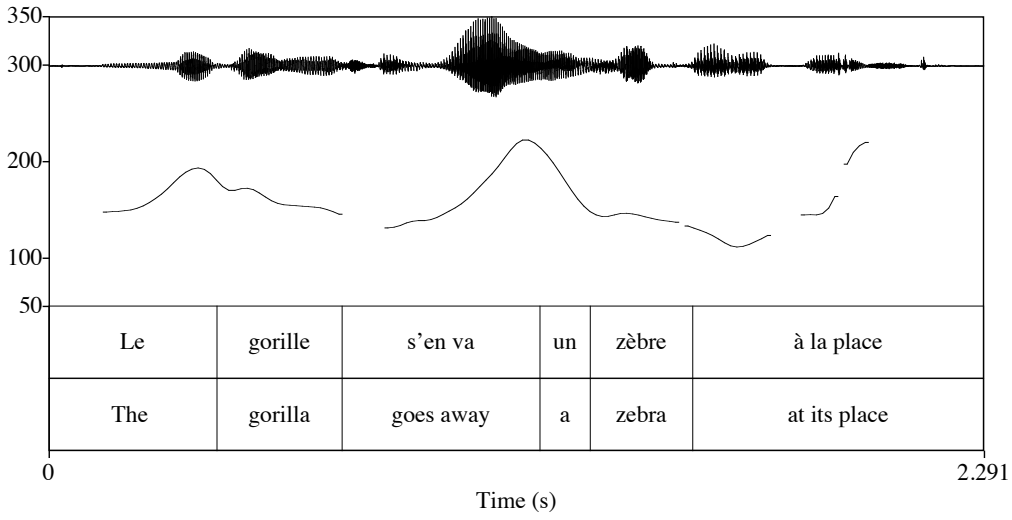
Fig. 8.20 Scaling of high tones in French

There are non-random variations in the tonal pattern of French that are explained by the pragmatic roles of phrasal tones. Consider the example in (49), illustrated in figure 8.21,

from Féry, Hörnig & Pahaut (2010). The first word, an article, has a rising pattern, and the following noun is falling, thus the inversed pattern from the one taken to be the default one in French. The following verb is rising, ends high, and is also lengthened durationally, expressing a continuation pattern. The following clause has exactly the same pattern, except for the fact that the initial rise is not there. Instead the article starts high from the high level reached by the preceding high tone and the noun *zebra* has a falling pattern. The adverb *à la place* has the rising contour of the verb in the preceding clause. This is a typical list or continuation pattern in French, which applies regardless of the function of the individual words carrying the tones. In this pattern, the phrasal tones associate with Φ-phrases, though not necessarily at their edges. This kind of patterns show very clearly that tones in French cannot be straightforwardly associated with prominence (see also Ladd 2008:62 for a similar remark).

	$L_\Phi H_\Phi$	L_Φ	H_Φ	H_Φ	L_Φ	L_Φ	H_I
(49)	[((Le	gorille)$_\Phi$	s'en va)$_\Phi$	(un zèbre)$_\Phi$	(à	la	place)$_\Phi$]$_\iota$
	the	gorilla	away goes	a zebra	at	the	place

'The gorilla has gone away; there is a zebra instead.'



In

Fig. 8.21 List pattern in French

There are other tonal patterns expressing pragmatic roles in French; see Ladd's (1980) examples in section 6.4 and discussion point 4 in chapter 6.

Let us briefly examine the effect of information structure on French prosody. Lambrecht (1994), Destruel (2013), Hamlaoui (2009) and Féry (2014) among many others have proposed that French is a language that changes word order and syntactic structure for the sake of information structure: among other syntactic strategies, focused subjects readily form cleft-sentences in spontaneous discourse, and givenness and topic are frequently marked by left- or right-dislocation. The realization of given final constituents has been discussed by Rossi (1985), Touati (1987), Di Cristo (1998), Jun & Fougeron (2000), Delais-Roussarie et al. (2002), Beyssade, Hemforth and Marandin & Portes (2009) among others: a low plateau, a succession of downstep and a continuous fall are all indicative of givenness. But givenness can also have no effect at all: In a production experiment, Hamlaoui, Corridun & Féry (2012) found that nominal phrases consisting of a noun and an adjective, like *moineau orange* 'orange sparrow' do not vary their F0 value in a significant way according to the place of focus, whether on the noun, the adjective or on the entire nominal phrase. Givenness does not decrease the prominence of the final adjective in such

a constituent. Féry (2014) formulated the hypothesis that post-focal compression is Φ-phrase-based in French: the absence of post-focal compression in the noun phrase is due to the fact that only entire Φ-phrases can be compressed, and that adjectives, being part of the Φ-phrase of the noun, cannot be prosodically compressed as a consequence of givenness. Assuming that a sequence of a verb plus argument forms a single (recursive) Φ-phrase and that a sequence of a verb plus adjunct forms two separate Φ-phrases, the latter pattern should be subject to post-focal compression much more than the former one. Experimental results with laboratory speech by Destruel & Féry (2015), with 11 speakers and 4 sentences in 14 conditions each, confirmed this hypothesis. Compare instances of the sentences in (50), (51) and (52). The first one illustrates a sentence containing a verb plus an argument, illustrated in figure 8.22. There was not much difference between a realization of this sentence with a focused argument and with a given one. Post-focal compression was non-significant in this case. In these experimental sentences as well, every Φ-phrase is separated from what precedes it by a high Φ-phrase boundary; see above.

(50) {What happened?/Who did Jeremy recognize?}
[(Jérémy)$_{\Phi}$ (a reconnu (le voleur (de caramel)$_{\Phi}$)$_{\Phi}$)$_{\Phi}$]$_{\iota}$
Jeremy has recognized the thief of caramel

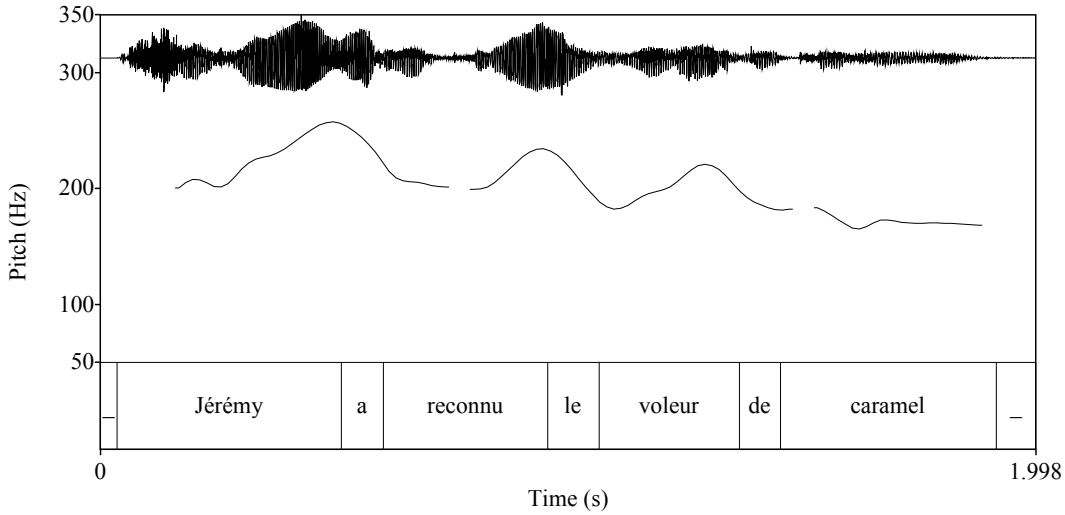


Fig. 8.22 Verb plus argument in a French sentence

The next sentences illustrate a sequence of a verb plus adjunct in an all-new French sentence in (51) and when the adjunct is given in (52). The figures 8.23 and 8.24 illustrate the difference between the two. In the first case, the adjunct is typically separate from the verb, downstep is less than in the case of an argument, and there may by a short break between verb and adjunct. In the second case, post-focal compression is obvious, and regular, although it is not obligatorily realized in this extreme way.

(51) {What happened with the caramel thief?}
Jérémy l' a reconnu à la sortie du village
Jeremy CL.3SG has recognized at the exit of.the village

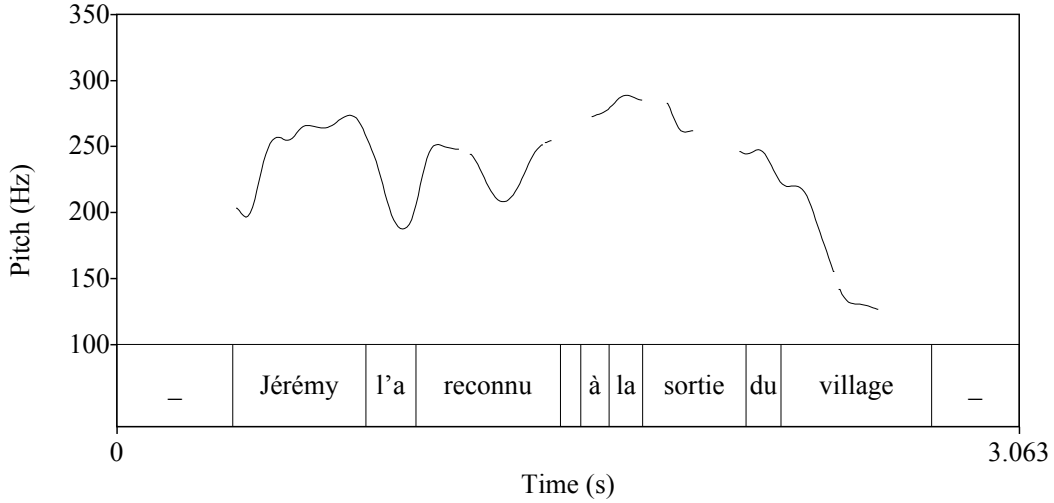


Fig. 8.23 Verb plus adjunct in an all-new French sentence

(52) {What happened at the exit of the village?}
Jérémy l' a reconnu, [à la sortie du village]$_G$
Jeremy CL.3SG has recognized at the exit of.the village

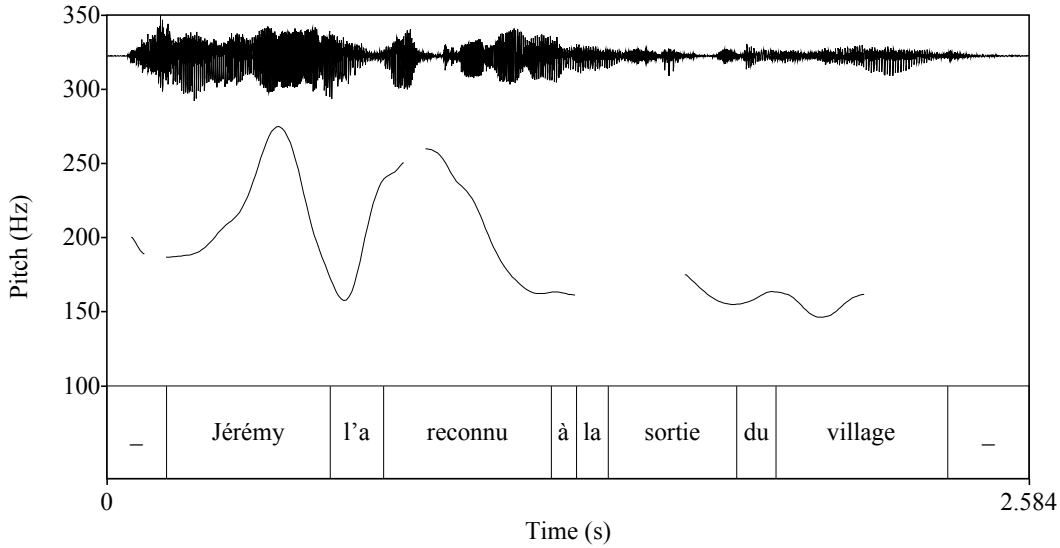


Fig. 8.24 Verb plus given adjunct in French

It must be clear that the realization shown in figure 8.24 is not obligatory in French. About half of the experimental sentences with an argument showed no compression at all, in the sentence in figure 8.23 for an argument. It is not to be excluded that compression is only possible in the case of right-dislocation, which is optional in (52).[19]

In short, French is a phrase language in the sense that tonal assignment, tonal patterns and tonal scaling are phrase-based. It might be that the high degree of variation in the tonal structure of French correlates with this property. Phrasal tones are freer to associate with any syllable in a phrase, as was illustrated in figure 8.21. Moreover, tonal

[19] In this case, right-dislocation is indistinguishable from 'normal' syntax, since there is no possible clitic doubling.

contours (rather than pitch accents forms) are bearers of meaning to a larger degree than in English.

8.5.2 Hindi

In this section, we turn to Hindi, an Indo-Aryan language from India. This is a language with lexical stress, but lexical stresses are not always implemented in the phrasal tonal structure (see Pandey 1989, to appear, for a formal proposal for word stress in Hindi and Puri 2013 for a thorough review of the literature on lexical stress in Hindi). Hayes (1995:163), following Pandey, used the term "non-phonemic lexical accent".[20] Lexical stress is realized in isolation, when the words are pronounced out of context, but it is often only weakly, if at all, perceivable in spontaneous speech. The reason for this latter property is the absence of a systematic pitch accent corresponding to lexical stress. When accounting for the absence of the Nuclear Stress Rule in Hindi, Bansal (1976:27) writes the following: "The main differences [between Hindi and English] were found to be in tonicity. Except where the location of the tonic syllable is determined by contextual emphasis or contrast, English places the 'tonic' on the last important word in the tone-group, but no such rule operates in Hindi." Intonational markers in the form of phrasal tones, are often more prominently realized than lexical stress.

Moore (1965) showed that every non-final prosodic phrase in Hindi, which he called 'foot', starts with a low tone. The prosodic phrase is defined as "one to several syllables in length, which normally is uttered with a continuously rising pitch from beginning to end." Harnsberger & Judge (1996) analysed the low part of the rising contour as a low pitch accent, annotated as L* in the tone sequence model, and the high part of the rising contour as a boundary tone H%, or H-.

As for the phonetic correlates of focus, Moore (1965), Harnsberger & Judge (1996) and also Dyrud (2001) found that it has the effect of inserting a boundary tone at the left edge of the focused phrase. They find that the rising pitch pattern may also show a higher excursion, a greater intensity and longer duration. Post-focally the pitch range may be compressed or even be completely flat and deaccented, although rising pitch accents are still realized in compressed pitch range (Harnsberger & Judge 1996). Similarly, Puri (2013) finds a larger F0 excursion in narrow focus than in wide focus and post-focal compression.

Jyothi, Cole, Hasegawa-Johnson & Puri (2014) asked ten non-expert speakers of Hindi and one expert (the paper's last author) to identify prominent words and prosodic boundaries in 10 excerpts from spoken short narratives. They found that the non-experts do not agree with each other and with the expert as far as prominent words are concerned (κ = 0.15). The agreement is better with prosodic boundaries (κ = 0.41). For English speakers, Yoon, Chavarría, Cole & Hasegawa-Johnson (2004) found agreement rates at 89% for boundaries and 86% for pitch accent. Mo, Cole, and Lee (2008) found a mean kappa (κ) of 0.582 for non-expert speakers for prominence and boundaries.

Patil et al. (2008) conducted a simple production experiment with 20 native speakers to investigate the effects of word order and focus in controlled lab data using scripted speech. There were three conditions for focus (wide focus vs. subject focus vs. object focus), and two conditions for word order (SOV vs. OSV). They found that each non-final Φ-phrase started with a low tone and ended with a high tone but the final verb, had a falling tone. When the first constituent was focused, it showed an earlier and slightly higher pitch rise, and the pitch contour was compressed in the post-focal region. When the

[20] Pandey (1989) emphasized the fact that Hindi is spoken by a very large population, of whom a large part are native speakers of other Indian languages, Indo-Aryan or others, which may constitute an explanation for the difficult state of lexical stress in Hindi.

second constituent was focused, there was no focus marking on it. All in all, the most robust effect in F0 was indirect: focus elicited compression of the postfocal material. In all other conditions, there was no effect on the height of the high tone of the focused constituent.

Féry, Pandey & Kentner (to appear) conducted an experiment with spontaneous data, in which 16 Hindi speakers realized sentences with a focused subject or a focused object. The F0 realization was again a low tone at the beginning and a high tone at the end of every non-final Φ-phrase, and a falling pattern on the final Φ-phrase, see (53) and (54) for pitch tracks. There was also small increase of F0 and intensity as a result of focus, as well as a small amount of post-focal compression in pitch and intensity accompanying givenness, in line with the results of Patil et al. (2008).

Correlates of phrasing at the level of the focused constituent were indicative of an effect of phrasing for focus: a boundary at the left edge of the focused constituent was realized by different means which were not all prosodic: First, there were more occurrences of the optional article *e:k* 'a' in case the constituent was in focus. Second this article was often encliticised to the preceding constituent although, morphologically, the article belonged to the following constituent. And third, glottal stop insertion before the noun and before the article occurred much more often in case the constituent was focused than in case it was not. These three strategies have the same result: they enhance the separation of the focused constituent from the rest of the sentence and, in this way, increase its perceptive prominence.

(53) Object is focused {Is the woman hitting a tree or a flower?}
vəhã: ʤəhã: ni:la: a:ka:ʃ hæ: ərət [e:k pe:ɽ ko:]$_F$ ma:r rəhi hæ:
there where blue sky be-3S woman one tree DAT hit AUX-PROG-F be-3S
'There where the sky is blue, the woman is hitting [a tree]$_F$.'

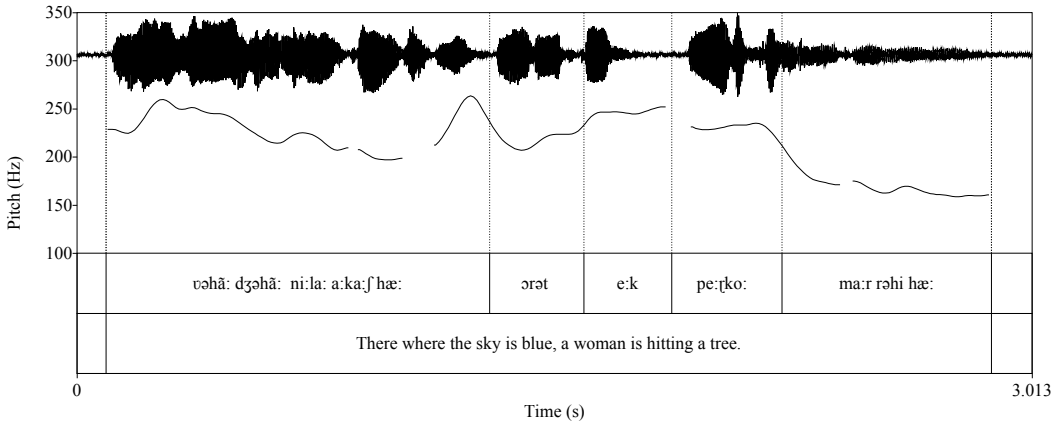


Fig.8.25 Object focus in Hindi: Encliticisation of *e:k* to the preceding constituent

(54) Subject is focused {Is a man or a woman hitting the man?}
ʤəhã: ni:la: a:ka:ʃ hæ: ʊəhã: [ʔe:k ʔərət]$_F$ a:dmi: ko: ma:r rəhi: hæ:
where blue sky be-3S there one woman man DAT hit AUX-PROG-F be-3S
'Where the sky is blue, there, [a woman]$_F$ is hitting a man.'

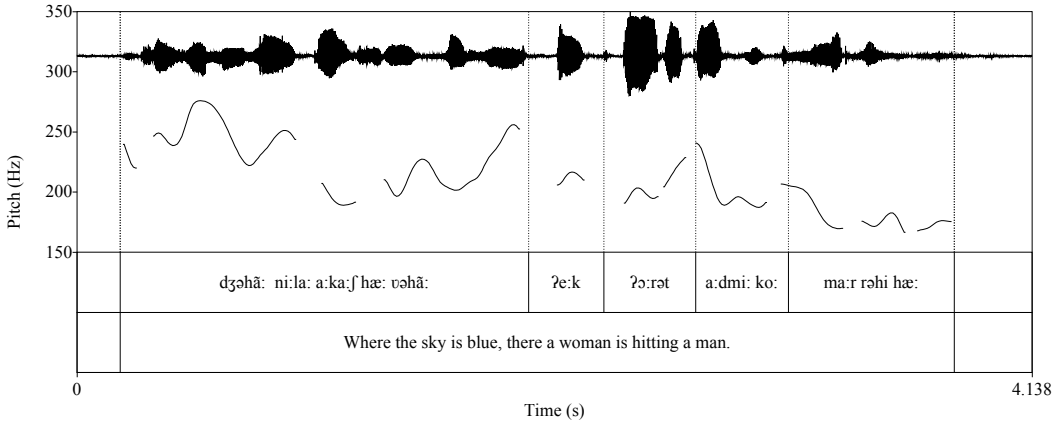


Fig.8.26 Subject focus in Hindi: Glottal stop insertion before the focus

Even if the prosodic correlates used to mark focus can be said to enhance the prominence of the focused constituent, and this both in French and Hindi, these correlates are not the same as the ones usually discussed for intonation and pitch accent languages. More specifically, there is no increase in F0 on a specific pitch-accented syllable. Rather the entire Φ-phrase containing the focus is rendered more prominent, with addition of high or low tones at the periphery of the Φ-phrase, as in French, or with correlates increasing the separation of the focused Φ-phrase from the adjacent non-focused ones, as in Hindi. Moreover, other phonetic correlates are also present, as has been shown with Hindi. Post-focal compression applies in both languages, although here, too, there is a difference with the Germanic languages. In French, it seems that only entire Φ-phrases are subject to post-focal compression, and in Hindi, compression is present but less than in other languages.[21]

Not only French and Hindi, but also other Indian languages (see Féry 2010, Khan 2014 for Bangladeshi Bengali) and West Greenlandic (Inuit-Inupiaq branch of Eskimoan; see Rischel 1974 and Arnhold 2014) and Korean (Jun 1998, 2005) are phrase languages. They associate tones at the level of the Φ-phrase and the ι-phrase. Some languages, like Hungarian and Finnish, assign phrasal tones in a purely demarcative way, and may be called phrase languages as well. This remains to be investigated.

To sum up, phrase languages differ from 'classic' intonation languages in organizing the tonal structure of sentences on the basis of Φ-phrases. English and other intonation languages organize their tonal structure around pitch accents, which generally coincide with lexical stresses. These differences explain a number of properties of these languages.

8.6 Conclusion

This chapter has passed review of different categories of intonation and has proposed four large classes of languages: intonation languages, pitch accent languages, tone languages and phrase languages. As a motivation for these four classe, it has been proposed that tones can arise at different levels of prosodic struscture and that all levels have to be considered in order to decide on the categorization of a specific language. In other words, languages cannot be classified solely according on the tonal structure of the words, as has been done in the past. Sentence intonation has to be considered as well. In chapter 7, it was shown that only lexical tones and stress are assigned at the level of ω-word, and that further tones are assigned at Φ-phrases and ι-phrases. Languages differ greatly in the tones they assign at these levels, and it is a task for the future to understand the typological tonal grammar

[21] This could also be shown for Indian English by Féry, Pandey & Kentner (to appear).

better. The in-depth compositional tonal meaning has been restricted to English so far, but this part of prosody has to acquire much more importance if intonation has to find its way in formal grammar.

Annotated suggestions for further reading

Jun (2005) and (2014) are useful short papers on the intonation in different languages. All the papers use a ToBI annotation.

The Special Issue of Phonology 32 (2015), edited by Lisa Selkirk and Seunghun Lee, contains papers on the tonal phonology and syntax-prosody interface of several languages.

Discussion points

Find native speakers of Swedish, German and French and have them utter the following sentences. Provide a prosodic and tonal analysis for each sentence.

(1) a. Jag SKILDRAR [1]anden. b. Jag SKILDRAR [2]anden. (from Gussenhoven 2004)
'I am depicting the duck.' 'I am depicting the ghost.'

(2) a. Während der NACHT fahren KRANkenwagen meistens SCHNELL.
during the night drive ambulances most.of.the.time fast
'At night ambulances usually drive fast.'
b. Während der NACHT fahren Krankenwagen meistens schnell.
c. KRANkenwagen fahren meistens schnell während der Nacht.

(3)a. Emilie préfère des bonbons.
Emilie prefers INDEF-ART sweets
'Emilie prefers sweets.'
b. Emilie, elle préfère des bonbons.
c. C'est des bonbons qu' elle préfère, Emilie.
it is INDEF-ART sweets that she prefers Emilie

Chapter 9
The processing of prosody

9.1. Introduction

Imagine you have to give the solution to the following arithmetical problem: 3 x 4 + 2. Since the linear arrangement of the operations is no guarantee of their order, and since the final solution depends on the order of the operations, it is necessary to know whether 3 x 4 or 4 + 2 is to be processed first. An oral version of this problem is likely to be disambiguated. This is because prosodic grouping is expressed by tonal scaling, boundary tones and durational cues, produced by the speaker and decoded by the hearer. This chapter investigates the role played by prosody in the comprehension and production of speech from a psycholinguistic perspective. Lexical, syntactic, contextual, semantic and pragmatic information, as well as information structure, are the linguistic elements expected to be processed partly through the help of prosody. The chapter examines whether and how intonation can facilitate and/or disrupt processing of an utterance, especially in the case of structural ambiguity, like the one mentioned above. Because the beginning of a sentence usually does not provide much information on how it will be continued, linguistic structures are ambiguous most of the time. If a sentence begins with *Anna saw that*…, it can be continued with 'nice shirt' or 'her mother was angry' or 'covered in snow', to cite just a few possibilities. In other words, temporary – or local – ambiguities arise frequently, although this does not seem to cause much of a problem in speech processing, especially when the context provides sufficient cues for comprehension. Global ambiguities, on the other hand, can cause more problems downstream if they are not resolved by any further lexical or syntactic information in the utterance (Cutler, Dahan & van Donselaar, 1997) or by the context. Global ambiguities refer to ambiguities that are not resolved at the end of a sentence. We will see numerous examples of both kinds of ambiguities below.

Although a supporting role of prosody is largely acknowledged, some authors, like Allbritton et al. (1996), Watt & Murray (1996) and Stirling & Wales (1996), have claimed that prosody has very little effect on structural disambiguation, if at all. One of the tasks of this chapter is to examine the arguments for and against assigning a central role of prosody to speech processing, especially of ambiguous structures. It goes without saying that in examining prosody for speech comprehension, we do not ignore the speaker. For a comprehender to be able to process prosody in discourse, it is necessary to have a speaker who has produced the prosody in the first place.

Independently of its role in disambiguating structures, Frazier, Carlson & Clifton (2006) explained why prosody is helpful in processing speech and why it is worth investigating its role in detail. They observed that remembering a string of digits is easier when it is parsed in smaller units, such as 174 31 86 as compared to 1743186 presented in one block (Reeves et al. 2000, Sturges & Martin 1974): it is the structuring which facilitates the exercise of remembering. And in fact, as has been shown throughout this book, one of the central tasks of prosody is to provide a structure to a continuous string of segments, syllables and words. The prosodic structure organizes a sentence, or a sequence of sentences, in smaller units, relating them to each other in significant ways. In many languages, syllables at the end of Φ-phrases and ι-phrases are lengthened and carry special boundary tones that help the listener to identify them as final. Nursery rhymes and long epic poems are often organized in stereotyped and rhythmic patterns, helpful for memorization. Because of these rhythmic properties, Frazier et al. considered prosody as a kind of pivot around which comprehension takes place. It should be added that the

accenting of some words and deaccenting of others is used to convey the information structure of a sentence, and that the choice of tonal contours provides the listeners with elements allowing them to understand the speakers' attitude regarding what they say.

As detailed in chapter 6, prosody is the sole means for marking many distinctions in meaning at the lexical, semantic, pragmatic, information structural and interpersonal levels in numerous languages, and the processing of prosody is crucial for this reason. However, as a large part of the existing literature on the processing of prosody has focused on its role in disambiguating syntactic ambiguities, the present chapter will follow the tradition and concentrate on this aspect.

Useful summaries of the topic of this chapter can be found in Cutler, Dahan & van Donselaar (1996), Shattuck-Hufnagel & Turk (1996), Frazier et al. (2006), Speer & Blodgett (2006), Wagner & Watson (2010) and Breen (2014), all of which provide background information on this chapter's topic available at different periods of time. The present chapter can only do justice to part of the large literature on the subject.

In reviewing the results achieved in this branch of research, we will encounter different methods for investigating the processing of prosody: offline methods like naming tasks, forced tasks or multiple choice tasks as well as online methods like reading time, brain imaging, ERPs, and eye-tracking (as in the visual-world paradigm) are some of the methods used to investigate processing issues. No clear separation between these methods will be aimed at in this chapter, due to lack of space but also because they yield largely convergent and comparable results in the case of prosody research.

This chapter examines several components of prosody in turn, the first and best investigated one being the phrasing of utterances, in particular how prosody disambiguates syntactic structures by assigning them different prosodic phrasings. The role of phrasing for prosody processing is discussed in section 9.2. Further components of prosody that are addressed in this chapter are the role of stress and rhythm in section 9.3. Section 9.4 is dedicated to 'implicit prosody' (Fodor 1998, 2002a,b, Bader 1998). Implicit prosody studies the role of prosody in silent reading. Since ambiguous structures and ambiguous attachment that typically involve prosodic mismatching affect silent reading more than other ambiguities, a natural inference is that prosody also guides comprehension in silent reading. This is especially interesting because prosody is only poorly marked in written language.

9.2 Prosodic phrasing and speech comprehension

Often cited examples of local ambiguity from Speer, Kjelgaard & Dobroth (1996) and from Kjelgaard & Speer (1999) appear in (1) and (2).[1] Let us concentrate on example (1) for an illustration of the distinction between early and late closure, upon which syntactic proposals of resolving ambiguity are often based. When encountering *the door* after the transitive verb *checks*, readers have a tendency to attach it as the object of the verb, and this works fine if the sentence continues like (1)a. In this case, the embedded clause finishes late (after *door*), hence the term *late closure*. If, however, the sentence continues like (1)b, where the verb *is* is heard or read, the comprehender may have already integrated *door* as an object of *checks* and may have to go back in order to interpret this sentence in the intended way. In (1)b, *checks* is used intransitively and *door* is the subject of the main clause. The embedded clause finishes early (before *door*): in this case one speaks of *early closure*. The b. versions of these examples are so-called *garden-path*

[1] The examples were inspired by Slowiaczek's (1981) and Carroll & Slowiaczek's (1987) examples (i):
(i) a. *Because her grandmother knitted pullovers / Kathy kept warm in the wintertime.* (late closure)
b. *Because her grandmother knitted / pullovers kept Kathy warm in the wintertime*. (early closure)

sentences ('to lead someone down the garden path' means to mislead somebody; see Bever 1970), where comprehenders are forced to revise their initial parse.

(1) a. Late closure: [Whenever the guard checks the door]$_\iota$ [it's locked]$_\iota$
b. Early closure: [Whenever the guard checks]$_\iota$ [the door is locked]$_\iota$

(2) a. Late closure: [When Roger leaves the house]$_\iota$ [it's dark]$_\iota$
b. Early closure: [When Roger leaves]$_\iota$ [the house is dark]$_\iota$

Frazier & Fodor (1978) and Frazier & Raynier (1987) explained the preference for the first versions of these sentences in syntactic terms. Late closure was formulated as a syntactic principle; see (3).

(3) Late Closure (Frazier 1978:49)
When possible, attach incoming material into the clause or phrase currently being parsed.

Further principles posited that syntax processes the structure incrementally and economically: *Minimal attachment* (Choose the syntactic structure requiring the smallest number of syntactic nodes) and *Active filler strategy*/*Minimal chain condition* (Construct as few and as short movement chains as possible). All these syntactic principles have the same aim: every time that a constituent is added, it is attached as 'minimally' to the structure built up so far as possible. Creating a new node or a new clause is more costly. Similarly to (1) and (2), when encountering *the answer*, the structure (4)a is preferred over (4)b.[2]

(4) a. John IP[VP[V[knew NP[the answer] very well]]].
b. John IP[VP[V[knew CP[IP[NP[the answer VP[was wrong]]]]]]].

Warren, Grabe & Nolan (1995), Speer et al. (1996), as well as Kjelgaard & Speer (1999) added a prosodic component to these preferences, and showed that the perception of an early prosodic boundary, like the one after *checks* or *leaves* in (1) or (2), reduces the bias in favour of late closure. They claimed that if the edge of an ι-phrase conflicts with the preferred syntactic structure (the one leading to late closure), the sentence is reprocessed at the point where the conflict arises. Speer et al. (1996) and Kjelgaard & Speer (1999) conducted a series of experiments using speeded phonosyntactic grammaticality judgement, end-of-sentence comprehension, and cross-modal naming tasks. They could demonstrate that prosody has a facilitatory role in sentences with cooperating prosody, i.e. in which the prosody corresponds to the syntactic structure. It has a negative effect in the case of conflicting prosody, in which prosodic boundaries are placed at misleading points in the syntactic structure. Processing took more time in the latter case. They could also show that the garden-path effect demonstrated previously in reading experiments was reduced when the sentences were presented with a felicitous prosodic structure in listening experiments. They argued that their results are not compatible with models suggesting that prosodic information is processed later than syntactic information, like those of Pynte & Prieur (1996) and Marcus & Hindle (1990). In such models, prosodic information contributes only to the reanalysis of previous syntactic commitments, but the models cannot predict any online effect of prosody on processing. More specifically, these latter authors proposed that initially syntax determines the parsing of a sentence, and

[2] See below for cases going against this tendency.

that prosodic information contributes to the final structuring of the sentence, but that it is not available during initial processing.

Most studies on speech processing in the realm of prosody have investigated the influence of phrasing on the comprehension of sentences. In a nutshell, the result of this intensive effort is that prosodic boundaries help the speaker to organize the syntactic structure. In this literature, the effect of increased duration and/or pauses at the end of Φ-phrases and ι-phrases has been studied (Cooper & Paccia-Cooper 1980, Lehiste 1973, Marslen-Wilson, Tyler, Warren, Grenier & Lee, 1992, Snedeker & Trueswell 2003, Watson & Gibson, 2004, Ferreira 1993, Schafer, Speer, Warren, & White 2000, Wagner 2005, Breen, Watson & Gibson 2011 and many others). All these authors concentrated on temporal correlates and neglected tonal correlates (see section 9.3 for effects of pitch accents and F0 in general).

The sentences in (5)a and (6)a are additional examples of garden-path sentences (Bever 1970). The beginning of the garden-path sentence induces a preferred reading, which has to be corrected and re-interpreted on the basis of following material. In (5) and (6), *raced* and *dropped* are preferrably read as the finite verb of the main clause, as in the b. versions, which show the preferred parsing, without parenthetical or non-at-issue information. The first continuation in each pair, the a. version, induces a reading in which *raced* or *dropped* is the verb of an embedded reduced relative clause, because of the presence of the following finite verb. Similarly to the examples (1) and (2), the garden-path effect can be eliminated or attenuated if the prosody strongly supports the less preferred reading (see also Bader 1996, 1998 for a comparable result in German, discussed in section 9.4).

(5) a. Garden-path: The horse raced past the barn fell.
b. The horse raced past the barn and fell.

(6) a. Garden-path: The package dropped from the plane contained important news.
b. The package dropped from the plane. It contained important news.

Another type of structural ambiguity that can be resolved with adequate prosodic boundaries appears in (7). In this case, the scope of the negation differs along with the continuation of the main clause. Koizumi & Bradley (2007) showed that this ambiguous sentence has a preferred reading, the one illustrated in (7)a, namely that Jane did not purchase the blouse, and that the reason for her decision was that the blouse was made of silk: *because* has scope over the negation (BEC>NOT). The dispreferred reading is the one in which she did purchase the blouse, and the reason for her decision was not the fabric but something else; see (7)b. In this case, the negation has scope over the embedded causal clause (NOT>BEC).

(7) a. [Jane didn't purchase the white blouse]$_\iota$ [because it was silk]$_\iota$ (and silk is difficult to wash).
b. [Jane didn't purchase the white blouse [because it was silk]$_\iota$]$_\iota$ (but because it was white).

Koizumi & Bradley (2007) speculated that one of the reasons for the preference for the first reading over the second one is to be found in prosody.[3] They proposed that a

[3] Frazier & Clifton (1996:55) propose a purely syntactic explanation for the preferred interpretation, along the following lines and 'unless the sentence is presented in a context where it is already presupposed that *John left*': The negation is integrated into the structure and semantically interpreted as soon as it can. This provides an advantag for negating the VP in the main sentence.

prosodic explanation 'must refer crucially to the (default) two-phrase contour that is compatible only with BEC>NOT, versus the (marked) single-phrase contour that NOT>BEC seems to require.' However, they have no explanation of why a two-phrase contour should be preferred over a single-phrase contour. Indeed this preference goes against syntactic explanations like Late Closure and Minimal Attachment, which require the material to be attached to the on-going clause. But in fact, we have seen the reason for the preference in chapter 4, when we saw that syntax and prosody are both recursive. The constraint MATCH CLAUSE requires that sentence (7)a consist of two ι-phrases and as we saw above, this is an unproblematic prosodic structure. In the dispreferred reading of (7)b, by contrast, the negation has to be part of the same ι-phrase as the causal clause in order to express the intended scope, and this involves embedding of an ι-phrase into a larger one, a dispreferred prosodic structure according to NON-RECURSIVITY. The marked prosodic structure is itself a consequence of the semantic and syntactic relationship between the parts of the sentence.

The phonetic correlates of prosodic boundaries are highly variable, and it is also important to account for this variability. This aspect is called the 'non-invariance' of the prosodic structure and is the source of many difficulties for attempts to treat prosody as purely phonetic. According to Clifton, Carlson & Frazier's (2002) 'rational speaker hypothesis', the speakers are rational, and employ intonation in a manner consistent with the intended message. Carlson, Clifton & Frazier (2002) also formulated the 'informative prosodic boundaries hypothesis' stating that the interpretation of a prosodic boundary is compared and processed relative to other boundaries in the same sentence. If, other things being equal, a prosodic boundary is inserted just before a constituent, and this boundary is not obligatorily required by grammar, the speaker shows in this way that the constituent is not part of the preceding phrase. Take the example in (8), in which *after dinner* can be attached to *learned* or to *telephoned*. It is not the size of the boundary that is decisive for the interpretation, but rather whether it is smaller or bigger than the other one in the same sentence. In (8)a, the second boundary is of the size of a Φ-phrase, or is even non-existent (marked by 0), and is thus smaller than the first one, which has the size of an ι-phrase. In this case, it is likely that *after dinner* is attached to *telephoned*. In (8)b and c, the relation between the boundaries is reversed; the boundary is larger at the second position, and *after dinner* should preferably be attached to *learned.*

(8) Susie learned | that Bill telephoned | after dinner.

a.	ι	0, Φ
b.	0	Φ, ι
c.	Φ	ι

Clifton et al. (2002) and Frazier et al. (2006) proposed that the same principle applies to other cases of attachment ambiguities, for instance attachment of relative clauses, like in (9); see also the next section for discussion. Their hypothesis was supported in several experiments with different realizations of sentences containing attachment ambiguities.[4]

(9) I met the daughter | of the colonel | who was on the balcony.

Watson & Gibson (2004) mentioned additional factors playing a role in the probability of choosing between possible interpretations. They showed that prosodic phrasing is also affected by the length (in words or syllables) of syntactic phrases, as well as by their complexity. Holding other factors constant, the probability of a prosodic boundary

[4] See, however, Snedecker & Casserly (2010) for the claim that both relative and absolute boundaries are active in processing.

occurring at some position is an increasing function of the length of the preceding and following constituents. As a consequence, increased length of surrounding phrases should reduce the informativeness of a prosodic boundary because listeners need not assume that the syntactic structure of the sentence motivated the boundary; see for instance Hirose (2003) for this result in Japanese (and section 9.4 for more).

Clifton, Carlson & Frazier (2006) confirmed the effect of length and complexity on prosodic boundaries also for different data: when two short phrases (short proper names) are conjoined, as in (10)a-b, the placement of the initial prosodic phrase boundary has a substantial effect on the interpretation. In these sentences, the conjoined constituents are so short that no prosodic boundary is necessary. If the speaker chooses to realize boundaries, there must be a reason, like grouping of constituents, as illustrated in (10). When long phrases, as in (10)c-d, replace these short phrases, the effect of prosodic boundary placement becomes smaller. In this case, boundaries are not only induced by disambiguation of grouping, but also by the length of the constituents.

(10) a. [Pat]$_{\Phi}$ [or Jake and Lee]$_{\Phi}$ convinced the bank president to extend the mortgage.
b. [Pat or Jake]$_{\Phi}$ [and Lee]$_{\Phi}$ convinced the bank president to extend the mortgage.
c. [The plantation owner]$_{\Phi}$ [or the tenant farmer and the new caretaker]$_{\Phi}$ convinced the bank president to extend the mortgage.
d. [The plantation owner or the tenant farmer]$_{\Phi}$ [and the new caretaker]$_{\Phi}$ convinced the bank president to extend the mortgage.

Clifton et al.'s results are theoretically relevant. They argue against the localist view, which posits that the phonetic cues accompanying prosodic boundaries are defined as a function of the strength of the syntactic boundaries, as in Nespor & Vogel (1986) or Selkirk's (1995) approaches. Calhoun (2010) also proposed that a more flexible understanding of the role of prosody is needed, in the sense that there is no one-to-one relation between syntactic and prosodic constituents, but that the prosodic boundaries depend on the other prosodic boundaries in the same sentence.

A large number of studies have used ERPs (event-related potential, located in the brain, an electrophysiological response to a stimulus) and confirmed the presence of components such as the N400, a negative deflection around 400 ms after a target, which correlates with semantic processing (see Kutas & Hillyard 1980), and the P600, a positive deflection around 600 ms correlating with syntactic processing (see Osterhout & Holcomb 1992). Steinhauer, Alter & Friederici (1999) used ERPs to demonstrate an immediate effect of a prosodic boundary in a study of German sentence processing. They used examples like those in (11). In (11)a, *Anna* is the indirect object of *verspricht* 'promises' and in (11)b, it is the direct object of the complement clause verb *entlasten* 'support'. The different syntactic structures are accompanied by different prosodic boundaries. By cross-splicing the prosodic and the syntactic realizations, they obtained syntax-prosody mismatches. When hearing a boundary tone at *verspricht* in (11)a, the participants erroneously interpreted *Anna* as the object of *arbeiten*, which does not make any sense in German. In other words, the presence of cues associated with the right edge of a prosodic phrase prevented inclusion of *Anna* within the first clause.

(11) a. [(Peter verspricht Anna)$_{\Phi}$ (zu arbeiten)$_{\Phi}$]$_{\iota}$ und…
Peter promises Anna to work and…
'Peter promises to work Anna and...'
b. [(Peter verspricht)$_{\Phi}$ (Anna zu entlasten)$_{\Phi}$]$_{\iota}$ und…
Peter promises Anna to support and…
'Peter promises to support Anna and...'

The evidence for the detection of the mistake and reprocessing of the syntactic structure was an increased negative effect N400, a semantics-driven effect (difficulties in lexical integration of the verb), followed by a positive effect P600, a syntax-driven effect (structural revision) at the verb *arbeiten*. Furthermore, they discovered a positive shift in the ERPs that they called the 'Closure Positive Shift' (CPS), which follows the phrase-final words.

Roll, Horne & Lindgren (2009) and Roll & Horne (2011) showed with ERP experiments that the effect of prosody on syntactic processing is immediate in Swedish as well. They compared subordinated clauses with embedded main clauses. The difference between the two is visible from the different word order between a sentence adverb – a negation in their material – and the verb; see section 8.3.1 for a discussion of their data. Moreover only (embedded) main clauses are preferably accompanied by an initial high tone. They found a positive effect P600 in the case of an embedded main clause, a less preferred pattern, which was reduced when an initial high tone was present. This effect was absent in embedded subordinated clauses. They concluded that left-edge boundary tones in Swedish play a role in syntactic processing by facilitating the processing of the structure they delimit.

A further aspect, which has been debated in the literature, concerns the question of 'audience design'. The point of departure of this question was an observation by Watt & Murray (1996) and Allbritton, McKoon & Ratcliff (1996), among others, who found that trained speakers can disambiguate syntactic structure with the help of prosody better than untrained speakers. They concluded that the utterances used in experiments are not representative of speech in general because trained speakers are more likely to use prosody to disambiguate syntax. One example of the ambiguous sentences used by Allbritton et al. is illustrated in (12).

(12) When you learn gradually you worry more.

The question that audience design tries to answer is the following: do speakers prosodically disambiguate structures for the hearer's (audience's sake)?[5] In order to investigate this question, Schafer et al. (2000) designed an experiment in which participants – a driver and a slider – played a game using a game board and moves of game pieces. They used a pre-specified set of oral instructions to communicate moves, like those in (13). In (13)a, *the square* is the object of the first sentence and in (13)b, it is the subject of the next sentence.

(13) a. When that moves the square, it should land in a good spot.
b. Good choice. When that moves, the square will encounter a cookie.

Schafer et al. found that participants systematically used prosodic cues, like final lengthening, to disambiguate the syntactic role of constituents like *the square* in (13).

Snedeker & Trueswell (2003) and Kraljic & Brennan (2005) also used dialogues in the form of instructions between participants to investigate audience design. They used the visual-world paradigm, and found that prosody plays an early role for comprehension, as early as lexical and other phonological information. In this paradigm, participants look at pictures while hearing sentences containing elements that tend to direct their attention to parts of the picture. The eye movements are recorded. As these authors observed, visual-world eye-tracking is well suited for psycholinguistics because eye movements are very closely time-locked to the speech stream, and thus allow researchers to efficiently

[5] The alternatives would be that speakers do not disambiguate prosodically, or that they always disambiguate. In this latter case, they disambiguate 'for themselves' or 'for the speaker's sake'.

investigate participants' interpretations of speech in real time. Snedeker & Trueswell (2003) used sentences like in (14)a, in which the listener is instructed by another person to either tap a frog that is carrying a flower, or to tap a frog with an instrument in the form of a flower. Thus 'with a flower' can be an instrument or a modifier. Kraljic & Brennan (2005) used sentences like in (14)b, in which the listener is instructed to put the dog into a basket located on top of a star, or to take the basket containing the dog and put it on the star. In this case, there is an ambiguity between a goal and a modifier.

(14) a. Tap the frog with the flower.
b. Put the dog in the basket on the star.

Snedeker & Trueswell (2003) found that speakers only use disambiguating prosody when they are aware that they need to disambiguate structures for the audience, e.g. when there are two frogs in the visual display, one with a flower and one without. They suggest that the speaker must be aware of the listener's needs and must be able to meet these needs. By contrast, Kraljic & Brennan report that speakers always disambiguate, independently of the presence of an ambiguity in the design of the layout. According to them, speakers do not change their prosody to adapt to the needs of an audience. In this line of explanation, prosodic correlates like prosodic boundaries, prominence, tone scaling and the like are by-products of the syntax and of the phonetics. The prosodic structure is not a disambiguating device for the sake of the listeners but rather prosody arises automatically.

A few remarks are in order to shed light on the differences between the experimental results. First, the kind of ambiguities used in the different experiments differed. Kraljic & Brennan (2005) used more complex structures than Snedeker & Trueswell (2003), since one of the two constituents was always complex, whereas the sentences in Snedeker & Trueswell were always simple. In other words, complexity in Kraljic & Brennan's (2005) examples increased the probability of speakers inserting prosodic boundaries that might have helped listeners to disambiguate. And structural simplicity in Snedeker & Trueswell's examples prevented speakers from using phrasing cues unless they were forced to disambiguate (for the hearer's sake).

Moreover, as Kraljic & Brennan themselves observed, it could have been that the experimental design was more often ambiguous for the speakers than they assumed at the start of the experiment. If this observation is correct, the design required a disambiguation for the hearers in more cases than originally planned in this experiment, and the experiment did not deliver an answer to the question as to whether the speakers use prosody for themselves (speaker's sake) or for the hearers (audience's sake).[6]

As has been proposed in chapter 4, prosody is part of the grammar, and as such is present as soon as speech is produced. Nespor & Vogel (1986) attributed the difference in the success of intonation in disambiguating structural ambiguities to prosodic structure itself. According to them, the syntax is not directly responsible for the insertion of prosodic boundaries, but rather the prosodic structure accompanying the syntax, as has been shown throughout this book. In some cases, the prosodic structure does not clearly distinguish two readings of an ambiguous syntactic structure, as in (14)a.

An important aspect of the syntax-prosody interface is the direction of branchingness, which has only rarely been investigated from a psycholinguistic perspective. As an illustration, consider the pair of sentences in (15). When the embedded sentence is located after the main clause, as in (15)a, it does not induce the same kind of

[6] In Schafer et al.'s experiment, the ambiguities concerned the role of a constituent across two clauses. This kind of ambiguity concerns ι-phrases and it is to be expected that ι-phrases are more often – and more clearly – separate from each other prosodically than Φ-phrases, since the level of constituency is larger in ι-phrases.

prosodic disrupture as an embedded clause at the beginning of the sentence, as in (15)b. In (15)a, assuming recursive prosody, there is no right ι-phrase boundary separating the main clause from the embedded clause, whereas there is one in (15)b.

(15) a. [Jonathan doesn't notice [the girl is smiling at him]$_{\iota}$]$_{\iota}$
b. [When leaving forever]$_{\iota}$ [Jane didn't look back]$_{\iota}$

Kentner & Féry (2013) conducted production experiments with groupings of three and four names in German and came to the conclusion that the direction of branchingness is crucial. In a left-branching structure as in (16)a, the Φ-phrase boundary after *Sam* induces a larger reset than the same boundary after the first constituent in a right-branching structure, as in (16)b, thus the one after *Steve* in this case. If the prosody is recursive in a similar way to syntax, the difference between (16)a and (16)b must be attributed to their prosodic structure; see chapter 4 for recursion in prosodic structure.

(16) a. (Steve and Sam)$_{\Phi}$ (or Bob)$_{\Phi}$
b. (Steve)$_{\Phi}$ (or Sam and Bob)$_{\Phi}$

9.3 The role of pitch accents and rhythm

This section emphasizes the importance of taking the role of pitch accents and rhythm as seriously as the impact of phrasing discussed in the preceding section. This has not always been done in the literature, as testified by the following example. Recall that Clifton et al. (2002) and Frazier et al. (2006) investigated the effect of prosodic boundaries on relative clause attachment. But in fact, as can be observed from their figure 9.1 illustrating sentence (9) with two different prosodic boundary structures, the two realizations differ not only in the prosodic boundaries. The noun *daughter* is accented in the first case and unaccented (or compressed) in the second, even though they assign it an H*. Moreover pitch accents and boundaries can be realized by rising or falling contours, which fulfil different semantic roles (see chapter 6). Clifton et al. (2002) and Frazier et al. (2006) mention the differences in the tonal contours and pitch accents, and they illustrate and comment each pitch track. But they do not consider them further in their discussion of the results. Nor do they address the implications of the difference in tonal contours.

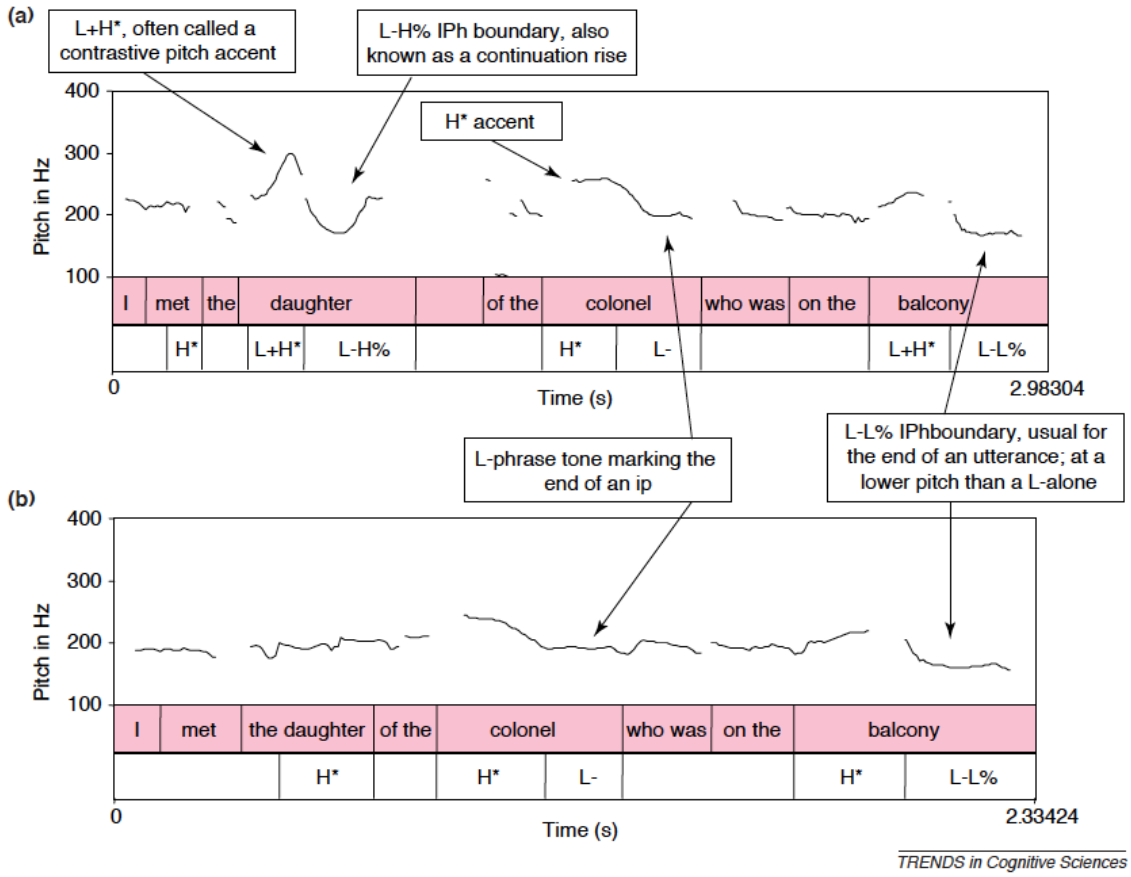


Fig. 9.1 Two readings of a sentence with a global ambiguity from Frazier et al. (2006)

Frazier et al.'s (2006) example is illustrative of a wider phenomenon. The role played by pitch accents and by tones is often dismissed as irrelevant in the literature on speech comprehension. It is perceived as a mere side effect of phrasing, and, as a result, it has been much less investigated than phrasing. Lehiste (1973) claimed that speakers mainly use manipulations on the time dimension to disambiguate meanings, intonation (i.e. pitch variation) playing only a subordinate role. She stipulated that since intonation already has so many different roles, it is probably not available to fulfil this additional one. In the same vein, Price, Ostendorf, Shattuck-Huffnagel, & Fong (1991) claimed that pitch accents only play a supporting role in disambiguation. The assumption of a restricted role of pitch accents has been maintained in the experimental literature so far in different guises, and a large number of experiments, some of which have been summed up or mentioned in the preceding section, have only manipulated prosodic boundaries to investigate the influence of prosodic structure on comprehension without considering that variation in pitch accents can have an effect on processing as well.

This concentration on temporal factors and the neglect of tonal correlates is a problem in the evaluation of the literature discussed in this chapter. This book has demonstrated that pitch is important for speech, and that it plays a crucial role in meaning. If tones are not carefully controlled for, they may induce differences in meaning and/or information structure all by themselves. An additional problem is that the poor understanding of the role of pitch accents has often led researchers to assume deaccenting when only the kind or the size of pitch accents was changed (see for instance Dahan et al. 2002 for such a case, discussed below). In the remainder of this section, the role played by accents in particular and by tones in general is addressed. First the presence vs. absence of pitch accents will be examined, called accent 'quantity'. Second we will discuss the role played by the form of the pitch accents, called accent 'quality'.

Speech comprehension is facilitated when words are accented. Birch & Garnsey (1995) demonstrated experimentally that the surface form of a word is kept in memory a longer time if the word is in focus. Moreover, several researchers have investigated the facilitation effect of adjusting the accent pattern of a sentence along with its information structural pattern. Bock & Mazzella (1983) found that response time in a simple comprehension task is shorter when new information is accented and given information is not, compared to conditions in which the accent pattern is reversed. Gussenhoven (1983) investigated the role of pitch accents in sentences inserted in different information-structural contexts and found that wrongly assigned post-nuclear accents are perceived as deviant. Moreover, he was also interested in the role of pre-nuclear accents. He compared mini-dialogues such as in (17), in which the answer consisted of a verb and an argument, with dialogues as in (18), which contained a verb followed by an adjunct. In the a-versions, the whole VP was focused, and in the b-versions, only the direct object or the adjunct was focused, the difference being elicited by the preceding question.

(17) Verb and argument

a. {Do you live by yourself?}
I [share a flat]$_F$ (the whole VP is focused)

b. {I hate sharing things, don't you?}
I share [a flat]$_F$ (the argument NP is focused)

(18) Verb and adjunct

a. {Where will you be in January?}
We will be [skiing in Scotland]$_F$ (the whole VP is focused)

b. {Where will you be skiing?}
We will be skiing [in Scotland]$_F$ (the adjunct PP is focused)

The sentences were spoken by naïve speakers. Gussenhoven cross-spliced questions and answers, so as to obtain both answers in both contexts. He then presented the two accent structures to listeners who had to decide which of two answers was the most appropriate as a response to a preceding question. He found that the presence of an accent on the verb in addition to the expected accent on the object in (17) did not change the acceptability of the pitch accent structure, even for (17)b, where the verb was given. The speakers did not do better than the chance level when required to choose between the two answers on the basis of such an accent pattern. By contrast, the absence of a stress on the verb in (18)a was an indicator that the verb had to be given (and thus not focused), so that the listeners disambiguated the two readings better than in the predicate-argument condition in the same task. When only the verb was accented, and nothing else, acceptability dropped considerably (see also Birch & Clifton 1995, 2002 for reproduction and confirmation of Gussenhoven's results).

Dahan, Tanenhaus, & Chambers (2002), Nooteboom & Kruyt (1987) and Terken & Nooteboom (1987) showed that comprehension time increased when accents were placed in an inappropriate way relative to information structure requirements. Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, & Carlson (1995) showed experimentally that a contrastive accent on an adjective allowed the listener to immediately select the one member of a set of items contrasting on exactly that adjective with the other items. Finally, Féry & Stoel (2002) demonstrated with a grammaticality judgement task that listeners of German expected post-nuclear and post-focal given material to be deaccented. They also found that added pre-nuclear accents are not perceived as deviant.

However, as was shown in chapters 6 and 8, focus is not automatically realized with the highest pitch in the sentence and givenness does not automatically lead to

deaccenting. In some circumstances, the expectation of a focus on a certain item has a facilitating effect on its perception, even if no F0 correlates of focus are present; see for instance Cutler & Fodor (1979) for such a result. Bishop (2014) found that listeners evaluated the exact same pitch accent as more prominent when they expected it to correspond to a narrow focus than to a wide focus. Terken & Hirschberg (1994) showed experimentally that prior mention does not necessarily lead to deaccentuation of a given entity. Deaccentuation is more likely if the repeated entity occupies the same grammatical role as in a previous mention. They explain the results in terms of predictability of reference of the entities (see below for more on the effect of predictability and different kinds of givenness on production and comprehension of different pitch accents, and see Wagner 2005 for a similar result with a semantic interpretation). Likewise, Bard & Aylett (1999) could show with a corpus analysis that repeated mention is not necessarily sufficient to trigger deaccenting of a referent.

In some cases, a lexical ambiguity is resolved by the presence vs. absence of pitch accents. Some German examples of discourse particles have been discussed in the literature: *wieder* 'again', *selbst* 'self', *auch* 'also' and *schon* 'already' have different meanings according to their status as accented or unaccented particles (see Féry 2012 for an analysis of this difference in terms of information structure); see for instance the difference in meaning of *schon* in (19). In (19)a, unaccented *schon* is a temporal adverb meaning *already*. In (19)b, accented *schon* is a modal particle focusing on the affirmative and concessive meaning of the remainder of the sentence.

(19) a. Anna war schon DA, als Maria kam.
Anna was already there when Maria came
'Anna was already there when Maria came.'
b. Anna kann SCHON nett sein.
Anna can already nice be
'Anna can certainly be nice (but…).'

The same holds in English and German for wh-words like *what*, *who* or *when*. Schafer, Carlson, Clifton, & Frazier (2000) conducted experiments on the ambiguity of *who* and *when* in their role as interrogatives or as elements introducing a relative clause as in (20)b or a temporal clause as in (21)b; see also Cooper & Sorensen (1977) and Carlson, Clifton & Frazier (2001). Schafer et al. (2000) found a bias in favour of the relative clause reading in (20). However, a pitch accent on the wh-word significantly increased the number of readings as an embedded question. They interpreted their result in terms of focus: an interrogative word is intrinsically focused and a pitch accent reinforced the focus reading.

(20) a. I asked the pretty little girl WHO is cold.
b. I asked the pretty little girl who is COLD.

(21) a. Joshua began to wonder WHEN his girlfriend got a tattoo.
b. Joshua began to wonder when his girlfriend got a TATTOO.

A different perspective on the role of pitch accents in speech processing comes from the kind of accents, their contour, thus from the 'accent quality'. A large number of experiments have investigated the effect of different forms of accents on speech processing, to which we now turn (see chapter 6 for the meaning of individual tones according to Pierrehumbert & Hirschberg 1990). The best-investigated question addresses the distinction between new information accent and contrastive accent.

Schafer, Carter, Clifton, & Frazier (1996) used a certain type of sentence, illustrated in (22), to show that syntactic ambiguity resolution can be affected by pitch accents. In an offline auditory comprehension study, the listeners had to choose between attachment of the relative clause *that the mechanic was carefully examining* to *propeller* (N1) or to *plane* (N2). An accent on Nl or N2 increased the probability that the relative clause would be understood as modifying the accented noun. Moreover, the listeners were even more likely to choose the more deeply embedded noun (*plane*) when it received a pitch accent LH*, a contrastive accent, than when the accent was H*. They called this preference the *Focus Attraction Hypothesis*. When accented, a noun is more prominent than when unaccented, and this explains the preference for attaching the relative clause to the accented noun.

(22) a. ... the propeller of the PLANE that the mechanic was carefully examining.
b. ... the PROPELLER of the plane that the mechanic was carefully examining.

Turning to online methods, the visual-world paradigm is an ideal tool for investigating the role of prosody in speech processing, since, as mentioned above, eye fixations are sensitive to the acoustic-phonetic signal and react in a fine-grained fashion. Dahan, Tanenhaus & Chambers (2002) used lexical competitors with identical stressed initial segments, like *candy* and *candle* (also called 'cohorts') in experiments in which listeners had to process information as quickly as possible. The task consisted in moving pictures of objects on a computer screen to different locations. Examples appear in (23). The authors manipulated the new/given status of the referents of the target words, as well as the form of the accent assigned to the referents. In one condition, the referent had a relatively high H* and in the other condition, the high tone was downstepped relatively to a preceding one, thus !H*. They qualified the second realization as 'deaccented'. They found a preference for interpreting 'deaccented' nouns as anaphoric and accented nouns as non-anaphoric. Moreover the listeners interpreted the less accented noun as referring to the most prominent previously mentioned entity and a more accented noun as referring to a new, previously unmentioned referent, or to a previously mentioned entity carrying a different thematic role. When hearing a non-downstepped H* pitch accent on the critical part of the target word, the listeners fixated the new referent more often than the given one. The differences appeared very early after onset of the target word, ca. 300 ms later, showing once more that pitch accents are detected rapidly and are also rapidly integrated into the discourse model.

(23) a. Put the candle/candy below the triangle.
b. Now put the CANDLE above the square.
c. Now put the candle ABOVE THE SQUARE.

Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard & Carlson (1995) and Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995) showed that listeners can use contrastive accents to facilitate disambiguation of referents. Ito & Speer (2006, 2008) investigated processing of adjectives and nouns in felicitous and infelicitous contexts. Participants heard instructions for decorating small Christmas trees and had to choose ornaments out of a set. Ito & Speer found that the placement of LH* on the colour adjective of a noun phrase (e.g. *orange drum* in (24)b) created a stronger bias towards contrastive information than the non-contrastive H*.

(24) a. Hang the green drum.
b. Next hang the orange drum.

Watson, Tannenhaus & Gunlogson (2008) also used the visual-world paradigm to compare the perception of H* and LH* in English. Their hypothesis was that LH* would elicit more contrastive interpretations and H* only new information. Their definition of contrastive and new information is illustrated in the three-step instruction in (25). The second mention of *camel* in (25)c is contrastive information, while the first mention of *candle* there is new information.

(25) a. Click on the camel and the dog.
b. Move the dog to the right of the square.
c. Now, move the CAMEL/CANDLE below the triangle.

They found that even though LH* had a bias in favour of contrastivity, the interpretation of such accents may overlap and H* is interpreted both as new information and as contrast. This overlapping role of pitch accents may compromise Pierrehumbert & Hirschberg's (1990) claim that different accents map one-to-one to information structural roles (see also Chen, den Os & de Ruiter 2007 for related work on British English, using both natural and synthetic speech in a visual-world paradigm, who came to a similar conclusion to Watson et al. 2008).

The literature on German has been effective in showing a relationship between the kind of accent and the pragmatic role, probably because there is only little variability in the kind of pitch accents that German declarative sentences can have, as compared to English at least. In an experiment using the visual-world paradigm and continuation of sentences with different kinds of pitch accents, Féry, Kaiser, Hörnig & Weskott (2009) used Kaiser & Trueswell's (2004) material to investigate whether listeners make use of information regarding the direction of a pitch accent, i.e. falling or rising, to predict post-nuclear deaccenting, and thus reference to a given constituent. The material consisted of pictures with three referents, in the example described below a nurse, a doctor and a patient. Féry et al. (2004) used sentences such as (26)c inserted in a short narrative as in (26)a-b. The target sentences were spoken in two ways. In the first one, the verb *befragt* 'asks/questions' had a rising intonation, which in German signals that a falling accent is still needed for a following focus. In this case, the following adverb *gleich* 'soon' had a high tone. In the second case, the verb was uttered with a falling contour, signalling that nothing new or focused should follow, and *gleich* had a low tone. In both cases, the subject was realized with a rising pre-nuclear accent. The pictures and the sentences were presented to listeners up to the adverb *gleich*. The listeners were asked to end the sentence in the way they found most appropriate. In the first case, a listener was expected to complete the sentence with the yet unmentioned referent *die Patientin*. In the second case, the already mentioned *nurse* was expected to be chosen to complete the sentence.

(26) a. A doctor and a nurse are leaning on the hospital reception desk.
b. It is almost two o'clock.
c. Der Arzt befragt gleich die Krankenschwester / die Patientin
the doctor.NOM questions soon the.ACC nurse / the.NOM patient
'The doctor will question the nurse / the patient soon.'

The results showed a significant difference between the continuations to the contours with a high tone on *gleich* and those with a low tone on this word. Discourse-old referents were chosen more often when a low tone occurred, confirming an anticipatory effect of prosody. Not only a contrastive vs. an informational pitch accent can trigger anticipatory effects as to intended referents. In this experiment it was the tonal contour of the verb and

post-verbal adverb which gave information to the listeners about the tonal structure and thus about the expected information status of the following object, speaking for top-down and non-local processing of prosody.

9.4 Implicit prosody

Implicit prosody is based on the hypothesis that speakers project a default prosody on what they read silently. In principle, listeners prefer to interpret sentences in their default prosodic pattern, corresponding to wide focus, and they need good reasons to depart from this preference. Moreover the implicit prosody hypothesis implies that the silently projected prosody may have an effect on the assumed syntactic structure. In other words, the prosodic representation at the phrase and sentence level plays a causal role in syntactic processing. The idea that readers hear an inner voice when they read is not new, and in the realm of psycholinguistics and intonation research, Chafe (1988:397) remarked that many people hear an inner voice when reading silently, with 'specific intonations, accents, pauses, rhythms, and voice qualities, even if the writing itself may show these features poorly, if at all.' It is this property of silent reading, together with the increasing evidence accumulated since the eighties showing that prosody is integrated extremely rapidly when processing speech, that led Fodor (1998, 2002a,b) and Bader (1996, 1998) to investigate the effect of prosody on silently read material.

Bader (1996, 1998) was one of the first to experiment with the role of implicit prosody. He reflected that when encountering a garden-path structure, a reader must be led to the wrong reading by prosody at the same time as by syntax, even when reading silently. He formulated the Prosodic Constraint on Reanalysis, expressed in (27), which claims that it is more difficult to reanalyse a syntactic structure when prosody has to be revised at the same time.

(27) Prosodic Constraint on Reanalysis (PCR, Bader 1998:8)
Revising a syntactic structure is difficult if it necessitates a concomitant reanalysis of the associated prosodic structure.

In other words, his prosodic hypothesis postulates that the prosodic structure is erected from the syntactic structure, but that it does not automatically distinguish different syntactic structures. Only the syntactic structures that also differ in their prosodic structures can be disambiguated by prosodic means. This is also the view expressed by Lehiste (1973) and Nespor & Vogel (1986:258).

Bader (1998) experimented with sentences like (28). The sequence *ihr Geld* can be analysed in two different ways. The preferred reading is a noun phrase 'her money' in which *ihr* is a possessive determiner. In (28)a, the preferred reading is available, since the verb *beschlagnahmt* 'confiscated' only requires a direct object. In (28)b, this reading is not available at the point where the verb *anvertraut* 'entrusted' is met, since this verb requires an indirect object, which is lacking if the sequence *ihr Geld* is interpreted as the direct object. The readers have to correct their analysis and assign the other possible reading to the sequence *ihr Geld*, namely a dative pronoun for *ihr* 'to her' and a bare accusative noun for *Geld.*

(28) a. Zu mir hat Maria gesagt, dass man ihr Geld beschlagnahmt hat.
to me has Maria said that one her money confiscated has
'Maria said to me that someone confiscated her money.'
b. Zu mir hat Maria gesagt, dass man ihr Geld anvertraut hat.
to me has Maria said that one her money entrusted has

'Maria said to me that someone entrusted money to her.'

In both readings of the sentence (28), *ihr* is a function word (determiner or pronoun) and function words are usually not accented. This means that both readings involve a preference for accenting *Geld* and leaving *ihr* unaccented. But when *ihr* is the indirect object, it is more stressed than when it is a possessive determiner. In online experiments using eye-tracking and silent reading, Bader found longer reading times on the verb – the disambiguating region – in the reading (28)b than in (28)a. He explains this result as the consequence that, upon encountering *anvertraut* 'entrusted', the reanalysis required a simultaneous shift in phrasal stress from *Geld* 'money', a bare noun in this case, to *ihr* 'to her', a dative pronoun, and it is this reanalysis which takes time.

In order to disentangle the role of syntax and the role of prosody in processing such sentences, Bader added a focus particle, as in (29). In this case, the default accent structure of the sentence is radically changed as compared to (28)b. The presence of a focus particle in the dative pronoun version renders the stress on the associated element *ihr* compulsory and more prominent than the stress on *Geld*. The result was an increase in reading time on the disambiguating region due to the additional cost of updating not only the syntactic but also the implicit prosodic representation.

(29) … dass man sogar ihr Geld anvertraut hat.
… that one even her money entrusted has
'that someone entrusted money even to her.'

Fodor (1998, 2002a,b) took a more general perspective on the issue of implicit prosody. The puzzle she was confronted with is the following: if prosody in reading must be projected on the basis of the lexical string and the syntactic structure assigned to it, as proposed by the syntactic hypothesis (30) (see for instance Slowiaczek 1981, Cooper & Paccia-Cooper 1980 for this view), how could prosody ever influence the assignment of syntactic structure to the lexical string? In such a view, prosody should only redundantly interpret the established meaning, since its only function is to represent sentence structure.

(30) Syntactic Hypothesis
The prosodic structure of a sentence is immediately determined by the syntactic structure. Different prosodic structures emerge automatically when the syntactic structure differs.

Fodor proposed that sentence comprehension involves both a syntactic and a prosodic parser operating in parallel. At the same time, she went a step further than Bader in proposing that the prosodic parser is always active, also when it is not necessary to disambiguate a syntactic structure with prosody. The role of the prosodic parser is more general. The result of the syntactic parser (which she called 'the sausage machine' in Frazier & Fodor 1978) is provided with prosodic phrases. This prosodic parser is active in silent reading, even though no real prosody can be heard. Instead, a default prosodic contour is projected onto the stimulus, which may influence syntactic ambiguity resolution. 'Prosody is mentally projected by readers onto the written or printed word string. And – the crucial point – it is then treated as if it were part of the input, so it can affect syntactic ambiguity resolution in the same way as overt prosody in speech does' (2002b:1).

The Implicit Prosody Hypothesis is formulated in (31).

(31) The Implicit Prosody Hypothesis ((IPH, Fodor 2002a:113)
In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction.

The prosody parser favours prosodic phrases roughly balanced in length, a property called 'the same-size sister constraint' by Fodor (1998:302, see also Gee & Grosjean 1983 for English and French, Kubozono 1989 for Japanese, and Myrberg 2013 for Swedish). A famous consequence of the same-size sister constraint is that it may have an influence on attachment preferences, as illustrated in the set of data in (32) with complex NPs from Fodor (2002a: 306-307). In (32)a, the divorced person is indifferently the bishop or the daughter, as the whole noun phrase forms a single Φ-phrase. In (32)b, the daughter is likely to be divorced, as a balanced phrasing separates 'the recently divorced' from 'bishop's daughter'. In (32)c, the preference is for the bishop to be the divorced one, because the entire noun phrase consists of two Φ-phrases: 'the recently divorced bishop's' and 'daughter-in-law'.

(32) a. (The divorced bishop's daughter)$_{\Phi}$
b. (The recently divorced)$_{\Phi}$ (bishop's daughter)$_{\Phi}$
c. (The recently divorced bishop's)$_{\Phi}$ (daughter-in-law)$_{\Phi}$

Compare the prediction of the IPH with a syntactic account. Late Closure in (3) and Minimal Attachment predict that it is always the bishop who is divorced.

In the same way, the Late Closure principle predicts an advantage for low attachment of relative clauses, as the relative clause should preferably attach to the phrase that is currently available. When two readings of an ambiguous sentence do not differ in the number of syntactic nodes, the preferred structure is the one where new items are attached to the clause or phrase which is currently being processed, thus to *actress* in (33).

(33) Someone called [the servant$_{NP1}$ of the actress$_{NP2}$] who was on the balcony.

This preference has been reported for English (for instance in Cuetos & Mitchell 1988). However, empirical findings indicate that Late Closure is not able to account for all languages. In a seminal study comparing data from English and Spanish, Cuetos & Mitchell (1988) found that Spanish prefers to attach a relative clause high in this structure. This difference between languages was futher confirmed by a series of studies showing that languages differ as to their preference to attach high or low. A prosodic break between *actress* and the relative clause increases the probability of high attachment. If a specific language generally assigns a left boundary at the beginning of a relative clause, thus for reasons other than for disambiguation, then the preference will be for high attachment in general. This is because in many languages, the presence of a prosodic break in this position correlates with a high attachment preference; see Augursky (2006) for a discussion of attachment preferences in many languages, and for an overview of the huge literature on high vs. low attachment.

As was shown above, see the discussion around (10), Watson & Gibson (2004) and Clifton, Carlson & Frazier (2006) showed that the length of constituents influences the insertion of prosodic boundaries. Speakers are more likely to place boundaries after longer constituents than after shorter ones. Hirose (2003) explored the effect of phonological length on prosodic phrasing and relative clause attachment in Japanese in silent reading. She could show that the length of proper names affects the resolution of

prosodic ambiguities because, upon encountering a long name, speakers tend to insert a prosodic boundary before continuing the sentence. Augursky (2006) also showed for German that complex NPs have an effect on attachment preferences in silent reading.

In the same way as a prosodic boundary may be implicit in silent reading, so can an accent be implicit. Not only at the lexical level can words have a default or preferred metrical pattern, but a series of studies have also confirmed that implicit phrasal or sentential stress can influence syntactic preferences. Ashby & Clifton (2005) could show that quadrisyllabic words with two stressed syllables, like *ÙLtiMÁtum, RÀdiÁtion* or *SÌtuÁtion*, take more time to read than quadrisyllabic words with only one stressed syllable, like *inSÁnity, inTÉNsity* or *auTHÓrity*, everything else being equal (number of letters and word frequency). They also showed that the sheer number of syllables does not matter. This indicates that speakers assign lexical stress to words when reading silently and that stressed syllables take more time to process than unstressed syllables.

In an online experiment, Breen & Clifton (2011, 2013) provided evidence for the existence of implicit metrical representations. They used pairs of homographs for verbs and nouns in garden-path sentences. Nouns have a trochaic pattern (strong-weak) and verbs have an iambic pattern (weak-strong). An example of the sentences they used appears in (34). They demonstrated longer reading times for the last part of the sentence in the verb reading of these words than in the noun reading. They interpreted their results as evidence for the need to shift a default accent (the nominal one, thus the trochaic default foot) to a marked one (the verbal one, an iambic marked foot).

(34) a. The brilliant abstráct$_V$ the best ideas from the things they read.
b. The brilliant repórt$_V$ the best ideas from the things they read.
c. The brilliant ábstract$_N$ was accepted at the prestigious conference.
d. The brilliant réport$_N$ was accepted at the prestigious conference.

Finally, some studies address the role of stress clashes in the processing of read material, suggesting that listeners make predictions about temporal patterns, and that they prefer a regular alternating rhythmic pattern without arhythmic disruptions.

Warren, Grabe & Nolan (1995) used sentences like those in (35). Listeners heard the sentences up to *problems* and then had to choose (i.e. read from a screen) the word *are* or a distractor.

(35) a. Whenever parliament discusses *Hong Kóng, próblems* are solved instantly.
b. Whenever parliament discusses *Hóng Kong próblems*, they are solved instantly.

Their results showed that the position of the accent in words like *Hong Kong* influenced the listeners' expectations about whether the words would be followed by a prosodic boundary. When there was a stress shift, participants were more likely to expect a following initially stressed word responsible for the stress shift (*próblem*) in the same prosodic domain. As a result, they were less likely to anticipate a following prosodic boundary. Warren et al. could thus demonstrate that stress patterns may influence the syntactic analysis of temporarily ambiguous sentences.

Kentner (2012:3) speculated that '[u]nder the assumption of a speech-like prosodic-phonological representation in reading, and given their immediate availability, stress and linguistic rhythm should exert their influence from the very beginning of the parsing process.' He tested this hypothesis with experiments using material like that illustrated in (36), where the sequence *nicht mehr* 'not more/not anymore' is ambiguous between a comparative and a temporal reading. When *mehr* is a comparative, it is generally accented, whereas as a temporal adverbial, *nicht mehr* is accented on *nicht*.

Furthermore, a following verb was either initially stressed, as in *nachweisen* ‘to prove’ or had a stress on the second syllable, as in *ermitteln* ‘to determine’.

(36) Der Polizist sagte, dass man…
‘The policeman said that one…
a. … nicht mehr NACHweisen/erMITteln kann, wer der Täter war.
… couldn’t prove/determine anymore who the culprit was.’
b. … nicht MEHR NACHweisen/erMITteln kann, als die Tatzeit.
… couldn’t prove/determine more than the date of the crime.’

Kentner found that unprepared readers of these sentences stressed *mehr* less often before *nachweisen* than before *ermitteln*. Kentner attributed this result to avoidance of stress clash. He could reproduce this effect in a silent reading task. Reading times in the disambiguating region were longer when the verb *nachweisen* was used in a comparative context, a reading that necessitates a prosodic representation with a stress clash between *mehr* and the first syllable of the verb. He concluded that the implicit rhythmic structure of *nicht mehr* led to an initial interpretation as a temporal adverbial, but that this interpretation had to be reanalysed as a negated comparative when the following sequence was uninterpretable. Kentner concluded that lexical stress immediately contributes to parsing decisions. His study adds to the growing body of literature pointing to the ease with which lexical-prosodic features are computed in reading and he does not exclude that on the basis of his work and similar future studies, we will be in a position to show that syntax can be influenced by prosody.

In closing, the growing number of studies investigating the role of prosody in speech production and comprehension are extremely useful for understanding intonation better. In the past, linguistic studies have too often been conducted with written material, and it is high time that spoken language is studied. What is still missing is a typological comparison: tone languages and phrase languages are nearly unexplored as far as psycholinguistics is concerned.

Annotated suggestions for further reading

Cutler, Dahan & van Donselaar (1996), Shattuck-Hufnagel & Turk (1996), Frazier et al. (2006), Speer & Blodgett (2006), Wagner & Watson (2010) and Breen (2014) are useful summaries of the research on processing of prosody, especially the first one that gives a larger view on the themes, even though it is the oldest study.

Chapter 10
Summary and conclusion

In this concluding chapter, let's sum up the essential results of this book.

First, intonational phenomena like lexical tones, stress, pitch accents, boundary tones and phrase tones are always assigned to ω-words, Φ-phrases or ι-phrases, thus at specific levels of the prosodic hierarchy. In the indirect reference approach to prosody taken in this book, syntax does not interact directly with tones. As mentioned above, this happens only through the presence of morpho-syntactically motivated prosodic domains. These prosodic domains result from the mapping between morpho-syntactic constituents and prosodic domains, which can be modified by minimality and maximality requirements, as well as by other constraints on well-formedness and the balanced structure of these domains. All languages map morpho-syntactic constituents to higher-level prosodic domains, although the syntax-prosody mapping displays variation in a cross-linguistic comparison. And the phonetic and phonological correlates of the prosodic domains also differ greatly from each other, only part of them being tonally defined. More research on the mapping between morpho-syntactic constituents and prosodic domains is clearly needed, especially in non-Indoeuropean languages.[1] It has been taken for granted in this book that the prosodic domains are recursive, where recursivity is defined as the possibility of a specific domain to be part of a larger one at the same level. Recursivity avoids proliferation of prosodic domains, and in this way simplifies the prosodic system. It also predicts a straightforward mapping between syntax and prosody. However, since recursivity is a controversial issue, it will be necessary in the future to verify case by case whether recursivity is also motivated by phonological factors.

Second, a compositional conception of intonation, in Laniran & Clements' (2003) understanding of this term, has emerged from the preceding chapters: melodies arise by concatenation of tones and the way they interact. Languages differ from each other in the way their intonation arises compositionally, and, crucially, tones and tonal patterns that are superficially similar may arise from very different underlying tonal patterns. This is because the phonetics realizing the phonological structure is limited. Tones can be lexical tones assigned at the level of prosodic words (ω-words), phrase tones assigned at the level of prosodic phrases (Φ-phrases) and boundary tones assigned at the level of Φ-phrases and intonation phrases (ι-phrases). Very often, tones are not simply concatenated but they also interact in specific ways, and the melody of a sentence is the result of the way they combine with each other. In some languages, most intonational tunes arise from only one kind of tone, as is the case with tone languages with densely specified lexical tones, or, at the other end of the spectrum, there are languages with only phrase tones. In other languages, melodies arise from different kinds of tones. The task of the phonologist is to provide a phonological analysis of the underlying components of the melodies, a mere sequence of tones cannot be enough.

African languages typically have only a few distinctive lexical tones, but these tones interact in different ways. African tones are often described by means of an autosegmental tonal grammar, which can be accounted for independently of the

[1] Some prosodic domains may be absent in specific languages: Hyman (1985) cited Gokana as a language without the syllable (but see Hyman 2011b), and Rialland & Aborobongui (2010) suggested that Embosi, a Bantu language, has no Φ-phrase, for example.

syllables the tones are associated with. According to some descriptions, tones in Ewe, for instance, may be assigned to ω-words. This is not true of many Asian tone languages, like Mandarin, where tones are inalienable parts of the syllables they are associated with, and do not spread or change so easily. In these languages, the number of distinctive tones may be much larger. Another dimension of variation of tone interaction is how phrase tones interact with pitch accents, like in English. If a pitch accent is final in its domain, it may fall together with a boundary tone, as has been proposed for intonation languages. In languages without pitch accents, like in French, phrase tones may interact with each other to express pragmatic meanings in a freer way than in languages with pitch accents.

The third important insight is that tones may vary in form as a function of information structure (pitch accents differ with the informational role of the words carrying them) and pragmatic content (an accent is often realized in a different way in a question and in a declarative sentence). The relation of intonation to information structure and pragmatic content is pervasive. Changing the tonal structure of an expression adds a pragmatic nuance, and in certain cases, changes its meaning. Pragmatic roles of intonation are especially conspicuous in languages which can accommodate a large variety of different contours, thus in languages which do not have lexical specifications for tones. In intonation languages and in phrase languages, tones typically correlate with pragmatic meanings in the way described by Pierrehumbert & Hirschberg (1990). Melodies of sentences or parts of sentences are paired with meanings as a result of the compositional role of the individual tones in the sentence melody. Notice that this use of the term 'compositional' differs from the first one. Here it is restricted to the semantic component of intonation: the meaning of a sequence of tones as a result of the individual tones in the sequence and the way they combine, especially pitch accents and boundary tones. The research in this field has been sparce, probably because the relation between tones and meaning is not obvious, is difficult to pin down, and is inconsistent, variable and versatile, especially as a grammar unrelated to the text. The fact is that individual tones acquire their meaning in relation to the text they are associated with and in relation to each other. As far as meaning is concerned, phrase tones may play a similar role to the one played by lexical tones, especially in phrase languages like French.

Fourth, the model of intonation that has been used in this book originated in the tone-sequence model of Bruce (1977) and Pierrehumbert (1980). In this model, tones define their own grammar, which means that a specific sequence of tones may or may not be well formed. In Pierrehumbert's approach and in much subsequent work, tonal phenomena have been related to prominence, because the best-studied languages are Germanic languages, where pitch accents are central to intonation. Prominence has been expressed by means of a metrical representation which allows the pairing of metrically prominent positions with starred tones. A metrically prominent position is associated with a pitch accent. In this way, the fact that focused or otherwise important constituents are signalled by higher (or lower) pitch is a direct consequence of the metrical representation. This relation between prominence and pitch height takes place at the level of the individual accents but is also expressed by tonal scaling of entire prosodic phrases. As an example of the relation between prominence and scaling of registers, take the large number of languages with post-focal or post-nuclear compression. However, the effects of register and F0 scaling, like downstep, reset and the like, are not easy to account for in a purely linear account of tones. Many different assumptions about tonal scaling have been made: for instance, Pierrehumbert (1980) proposed that specific sequences of tones have an effect on register, and Ladd (1983)

assumed a special feature (still present in ToBI conventions). Perhaps tonal scaling is not really expressible by individual tones, and we need a better representation of scaling effects. Moreover, in recent years, it has become more and more obvious that tonal scaling is not necessarily a direct translation of prominence. First, changes in register can be automatic in sequences of prosodic phrases with similar prominence, as already acknowledged in Beckman & Pierrehumbert (1986), and second, changes in the F0 height of individual tones of prosodic domains can be used for the expression of meaning. Tone languages can expand or compress the register of domains for questions for instance, and parentheticals or non-restrictive relative clauses are also often signalled by changes in the register. In other words, languages can use scaling of tones for reasons other than prominence, and the various uses of pitch changes beyond prominence are still waiting to be studied in-depth.

I hope that by addressing a number of issues in the realm of intonation and prosody, this book has contributed to define the place of intonation in grammar, namely at the interface between different linguistic modules, phonology, phonetics, syntax, semantics and pragmatics being the most important ones. Some important aspects have not been touched upon at all, such as acquisition of intonation, both in first language acquisition (L1) and in second language acquisition (L2), or attrition. Intonation of pidgins and creoles has also not been addressed. The work done in speech synthesis and text-to-speech models has also not been dealt with in this book. Intonation is still a young discipline of linguistics. However, due to the advances in speech technologies, it is now possible for all linguists to investigate intonation, and the progress made in recent decades gives good grounds to expect a rapid development in the field in the years to come.

References

Ahoua, Firmin. 1996. *Prosodic Aspects of Baule*. Cologne: Køppe.

Allbritton, David W., Gail Mc Koon & Roger Ratcliff. 1996. Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22. 714-735.

Ambrazaitis, Gilbert. 2009. *Nuclear intonation in Swedish – Evidence from experimental-phonetic studies and a comparison with German.* Travaux de l'institut de linguistique de Lund 49. Lund: Lund University.

Anttila, Arto, Adams B. Bodomo. 2000. Tonal polarity in Dagaare. In: Vicky Carstens & S. Parkinsons (eds.). *Trends in African Linguistics* 4*: Advances in African linguistics*. Trenton, NJ: African World Press. 119-134.

Arnhold, Anja. 2012. *Finnish Prosdy: Studies in Intonation and Phrasing*. Dissertation. Goethe-Universität Frankfurt.

Arnhold, Anja. 2014. Prosodic structure and focus realization in West Greenlandic. In: Jun, Sun-Ah (ed.). *Prosodic Typology* II. *The phonology of intonation and phrasing.* Oxford: Oxford University Press. 216-251.

Arregui, Karlos. 2015. Focus projection theories. In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Arvaniti, Amalia. 2011. The Representation of Intonation. In: Oostendorp, Marc van, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.). *The Blackwell Companion to Phonology Vol. 2*. Blackwell Publishing. 757-780.

Arvaniti, Amalia, Robert D. Ladd & Ineke Mennen. 1998. Stability of Tonal Alignment: the case of Greek Prenuclear Accents. *Journal of Phonetics* 26. 3-25.

Arvaniti, Amalia, Robert D. Ladd & Ineke Mennen. 2000. What is a starred tone? Evidence from Greek. In: Broe, Michael B. & Janet B. Pierrehumbert (eds.). *Papers in Laboratory Phonology* V. Cambridge: Cambridge University Press. 119–131.

Ashby, Jane & Charles Clifton. 2005. The prosodic property of lexical stress affects eye movements during silent reading. *Cognition* 96. B89-B100.

Atterer, Michaela & Robert D. Ladd. 2004. On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics* 32. 177-197.

Augurzky, Petra. 2006. *Attaching relative clauses in German: The role of implicit and*

explicit prosody in sentence processing. Unpublished Ph.D. thesis. MPI for Human Cognitive and Brain Sciences, Leipzig.

Bader, Markus. 1996. *Sprachverstehen. Syntax und Prosodie beim Lesen*. Opladen: Westdeutscher Verlag.

Bader, Markus. 1998. Prosodic influences on reading syntactically ambiguous sentences. In: Fodor, Janet Dean & Fernanda Ferreira (eds.). *Reanalysis in sentence processing*. Dordrecht: Kluwer. 1-46.

Bagemihl, Bruce. 1991. Syllable structure in Bella Coola. *Linguistic Inquiry* 22. 589–646.

Bansal, Ram Krishna. 1976. *The Intelligibility of Indian English*. 2nd edn. Monograph 4. Hyderabad: CIEFL.

Bao, Zhiming. 2011. *Chinese Tone Sandhi*. In: Oostendorp, Marc van, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.). *The Blackwell Companion to Phonology Vol. 5*. Blackwell Publishing. 2561-2585

Bard, Ellen Gurman & Matthew P. Aylett. 1999. The dissociation of de-accenting, givenness, and syntactic role in spontaneous speech. In: *Proceedings of the International Conference on Phonetic Sciences*. San Francisco. 1753–1756.

Bartels, Christine. 1997. *Towards a compositional interpretation of English statement and question intonation*. Unpublished Ph.D. dissertation. Amherst: University of Massachusetts.

Bartels, Christine. 1999. *The Intonation of English statements and questions. A Compositional Interpretation.* New York: Garland.

Basbøll, Hans. 2003. Prosody, productivity and word structure: the stød pattern of Modern Danish. *Nordic Journal of Linguistics* 26. 5–44.

Basbøll, Hans. 2005. *The phonology of Danish*. Oxford: Oxford University Press.

Bauer, Robert S. & Paul K. Benedict. 1997. *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In: Elordieta, Gorka & Pilar Prieto (eds.). *Prosody and meaning*. Berlin: De Gruyter Mouton. 119-161.

Baxter, William H. 1982. *A handbook of Old Chinese phonology*. Berlin & New York: Mouton de Gruyter.

Beaver, David, Brady Zack Clark, Edward Flemming, Florian T. Jaeger & Maria Wolters. 2007. When semantics meets phonetics: Acoustical studies of second-

occurrence focus. *Language* 82. 245-276.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3. 255–309.

Berg, Rob van de, Carlos Gussenhoven & Toni Rietveld. 1992. Downstep in Dutch: Implications for a model. In: Docherty, Gerhard & Robert D. Ladd (eds.). *Papers in Laboratory Phonology* II: gesture, segment, prosody: Cambridge: Cambridge University Press. 335-367.

Berman, Arlene & Michael Szamosi. 1972. Observations on sentential stress. *Language* 48. 304-325.

Bethin, Christina Y. 2006. Stress and tone in East Slavic dialects. *Phonology* 23. 125-156.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In: Hayes, J. R. (ed.). *Cognition and the development of language*. New York: John Wiley and Sons. 279-362

Beyssade, Claire, Barbara Hemforth, Jean-Marie Marandin & Cristel Portes. 2009. Prosodic Markings of Information Focus in French. In: Yoo, Hi-Yon & Elisabeth Delais-Roussarie (eds.). *Proceedings of the conference "Interface Discourse and Prosody"*. Paris. 109-122.

Bing, Janet M. 1979. Aspects of English Prosody. Unpublished PhD Dissertation. University of Massachusetts at Amherst.

Birch, Stacy L., & Garnsey, Susan M. 1995. The effect of focus on memory for words in sentences. *Journal of Memory and Language* 34. 232-267.

Bishop, Christopher M. 2014. In: Gorka Elordieta & Pilar Prieto (eds.). Prosody and meaning. Berlin: De Gruyter Mouton. 239-269.

Blevins, Juliette. 1993. A tonal analysis of Lithuanian nominal accent. *Language* 69. 237-273.

Blevins, Juliette. 1995. The syllable in phonological theory. In: Goldsmith, John A. (ed.). *The handbook of phonological theory*. Cambridge, MA & Oxford: Blackwell. 206–244.

Bock, Kathryn J. & Joanne R. Mazella. 1983. Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition* 11. 64–76.

Boersma, Paul & David Weenink. 1999-2014. *Praat: doing phonetics by computer [Computer program]*. Version 5.3.70, retrieved 2 April 2014 from http://www.praat.org/

Bolinger, Dwight L. 1958. A theory of pitch accent in English. *Word* 14. 109-149.

Bolinger, Dwight L. 1964. Around the edge of language: Intonation. *Harvard Educational Review* 34. 282–293. Reprinted in Bolinger. 1972. 19–29.

Bolinger, Dwight L. 1965. Pitch accent and sentence rhythm. In: Abe, I. & T. Kanekiyo (eds.). *Forms of English: Accent, Morpheme, Rhythm*. Tokyo: Hokuou. 139–180.

Bolinger, Dwight L. 1972. *Intonation: Selected Readings*. Harmondsworth: Penguin.

Bolinger, Dwight L. 1978. Intonation across languages. In: Greenberg, J. (ed.). *Universals of Human Language*, Volume 2 (Phonology). Stanford: Stanford University Press. 471–524.

Bolinger, Dwight L. 1982. Intonation and its parts. Language 58, 505-533.

Bolinger, Dwight L. 1986. *Intonation and its Parts: The Melody of Language*. Stanford, CA: Stanford University Press.

Brazil, David. 1975. *Discourse intonation*. Birmingham: English Language Research, University of Birmingham.

Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49. 451-493. doi: 10.1177/00238309060490040201

Breen, Mara. 2014. Empirical Investigations of the Role of Implicit Prosody in Sentence Processing. *Language and Linguistics Compass* 8. 37-50.

Breen, Mara & Charles Jr. Clifton. 2011. Stress matters: effects of anticipated lexical stress on silent reading. *Journal of Memory and Language* 64. 153–70.

Breen, Mara & Charles Jr. Clifton. 2013. Stress matters revisited: a boundary change experiment. *Quarterly Journal of Experimental Psychology.* 1896-1909.

Breen, Mara, Duane Watson & Edward Gibson. 2011. Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes* 26. 1532–62.

Bresnan, Joan. 1971. Sentence stress and syntactic transformations. *Language* 47. 257-281.

Bresnan, Joan. 1972. Stress and syntax. *Language* 48. 326-342.

Bruce, Gösta. 1977. Swedish word accent in sentence perspective. *Travaux de l'Institut de Linguistique de Lund* 12. Gleerup: Lund.

Bruce, Gösta. 1998. Allmän och svensk prosodi. Lund: Lund University.

Bruce, Gösta. 1999. Swedish. In: Hulst, Harry van der (ed.). 1999. *Word prosodic system in the languages of Europe*. Berlin & New York: Mouton de Gruyter. 554-567.

Bruce, Gösta. 2005. Intonational prominence in varieties of Swedish revisited. In: Jun, Sun-Ah (ed.). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 410–429.

Bruce, Gösta & Eva Gårding. 1978. A prosodic typology for Swedish dialects. In: Gårding, Eva, Gösta Bruce & Robert Bannert (eds.). *Nordic prosody*. Lund: Gleerup. 219–228.

Büring, Daniel. 1997. *The Meaning of Topic and Focus - The 59th Street Bridge Accent*. London: Routledge.

Büring, Daniel. 2003. On D-Trees, Beans, and B-Accents. *Linguistics & Philosophy* 26. 511-545.

Büring, Daniel. 2015. In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Calhoun, Sasha. 2010. The centrality of metrical structure in signaling information structure: a probabilistic perspective. *Language* 86. 1-42.

Carroll, Patrick J. & Maria L. Slowiaczek. 1987. Models and modules: Multiple pathways to the language processor. In: Garfield, J. (ed*.). Modularity in knowledge representation and natural language understanding*. New York: Academic Press. 221-247.

Catford, J.C. 1988. *A Practical Introduction to Phonetics*. London: Clarendon Press.

Chafe, Wallace, L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: Li, Charles N. (ed.). *Subject and topic*. New York: Academic Press. 25-55.

Chafe, Wallace L. 1988. Punctuation and the Prosody of Written Language. *Written Communication* 5. 395-426.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time*. Chicago/London: University of Chicago Press.

Chen, Aoju, Carlos Gussenhoven & Toni Rietveld. 2004. Language-Specificity in the Perception of Paralinguistic Intonational Meaning. *Language and Speech* 47. 311-48.

Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4. 109-150.

Chen, Matthew Y. 2000. *Tone sandhi*. Cambridge: Cambridge University Press.

Chen, Yiya. 2010. Post-focus f0 compression - Now you see it, now you don't. *Journal of Phonetics* 38. 517-525.

Chen, Yiya & Carlos Gussenhoven. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36. 724-746.

Chen, Yiya, Els den Os & Jan Peter de Ruiter. 2007. Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review* 24. 317–344

Chen, Yiya & Yi Xu. 2006. Production of weak elements in speech - evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica* 63. 47–75.

Cho, Hyesun. 2011. The timing of phrase-initial tones in Seoul Korean: a weighted-constraint model. *Phonology* 28. 293–330.

Chomsky, Noam. 1971. Deep structure, surface structure, and semantic interpretation. In: Steinberg, D. & L. Jakobovits (eds.). *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge: Cambridge University Press. 183-216.

Chomsky, Noam. 1995. *A Minimalist Program for Linguistic Theory*. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24. 239-297.

Clark, Herbert H. & Susan E. Haviland. 1977. Comprehension and the given-new contract. In: Freedle, R. O. (ed.). *Discourse Processes: Advances in Research and Theory, Vol.* 1. Discourse production and comprehension. Norwood, NJ: Ablex. 1-40

Clark, Herbert H., and C. R. Marshall (1981). Definite reference and mutual knowledge. In Aravind K. Joshi, B. Webber, and I. A. Sag, (eds.). *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Clements, George N. 1979. The description of terraced-level tone languages. *Language* 55. 536–558.

Clements, George N. 1984. Principles of Tone Assignment in Kikuyu. In: G. N. Clements & John Goldsmith (eds). *Autosegmental Studies in Bantu Tone*. Dordrecht: Foris Publications. 281-340.

Clements, George N. & K. C. Ford. 1980. On the phonological status of downstep in Kikuyu. In: Goyvaerts, D. (ed.). *Phonology in the 80s*. Ghent: Story Scientia. 309-57.

Clements, George, N. & Annie Rialland. 2007. Africa as a phonological area. In: Heine, B. & D. Nurse (eds.). *The Linguistic Geography of Africa*. Cambridge: Cambridge University Press. 36–85.

Clifton, Charles Jr., Katy Carlson & Lyn Frazier. 2002. Informative prosodic boundaries. *Language and Speech* 45. 87-114.

Clifton, Charles Jr., Katy Carlson & Lyn Frazier. 2006. Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review* 13. 854–861.

Cohn A., C. Fougeron & M. Huffman. 2012. Oxford Handbook of Laboratory Phonology. Oxford: Oxford University Press.

Constant, Noah. 2014. Contrastive topic: Meanings and realizations. Unpublished Ph.D. thesis. Amherst, Massachusetts: University of Massachusetts

Cooper, William E. & Jeanne Paccia-Cooper. 1980. *Syntax and speech, volume* 3. Cambridge, Massachusetts: Harvard University Press.

Cooper William, E., Stephen J. Eady & Pamela R. Mueller. *1985*. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77. 2142–2156.

Cooper, William E. & J. Sorensen. 1981. *Fundamental frequency in sentence production*. Heidelberg: Springer.

Cruttenden, Alan. 1986/1997. *Intonation*. Cambridge University Press, Cambridge.

Crystal, David. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Crystal, David & Randolph Quirk. 1964. *Systems of prosodic and paralinguistic features in English*. The Hague: Mouton.

Cuetos, Fernando & Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30. 73-105.

Cutler, Anne, Delphine Dahan & Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40. 141-201.

Cutler, Anne & Fodor, Jerry A. 1979. Semantic focus and sentence comprehension. *Cognition* 7, 49-59.

Dahan, Delphine, Michael K. Tanenhaus & Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47. 292–314.

Davis, Stuart. 2005. "Capitalistic" vs. "Militaristic": The Paradigm Uniformity Effect Reconsidered. In: Downing, Laura, Tracy A. Hall & Renate Raffelsieffen (eds.). *Paradigms in Phonological Theory.* Oxford: Oxford University Press. 107-121.

Deguchi, Masanori & Yoshihisa Kitagawa. 2002. Prosody and wh-questions. In: Hirotani, Masako (ed.). Proceedings of the North Eastern Linguistics Society, volume 1. *NELS* 32. Amherst, Massachusetts: GLSA, University of Massachusetts. 73–92.

Delais-Roussarie, Elisabeth. 1995. *Pour une approche parallèle de la structure prosodique: Etude de l'organisation prosodique et rhythmique de la phrase française.* Unpublished Ph.D. thesis. Université de Toulouse: Le Mirail.

Delais-Roussarie, Elisabeth, Annie Rialland, Jenny Doetjes & Jean-Marie Marandin. 2002. The Prosody of post-focus sequences in French. In: Bel, Bernard & I. Marlien (eds.). *Proceedings of Speech Prosody. 2002*. Aix en Provence. 239-242.

Delattre, Pierre. 1966. Les dix intonations de base en français. *The French Review* 40. 1-14.

Dell, François. 1984. L'accentuation dans les phrases en français. In: Dell, François, Daniel Hirst & Jean-Roger Vergnaud (eds.). *Forme sonore du language.* Paris: Hermann. 65-122.

Destruel-Johnson, Emilie. 2013. *An empirical investigation of the meaning and use of the French* c'est-*cleft*. Unpublished Ph.D. dissertation. Austin, Texas: University of Texas at Austin.

Destruel, Emilie & Féry, Caroline. 2015. *Compression in post-verbal sequences in French*. Submitted.

Dilley, Laura, Robert D. Ladd & Astrid Schepman. 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics* 33. 115–119.

D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. *Proceedings of Eurospeech* (Rhodes). 251-4.

D'Imperio, Mariapaola & Amandine Michelas. 2014. Pitch scaling and the internal structuring of the Intonation Phrase in French. *Phonology* 31. 95 - 122

Ding, Picus Sizhi & Caroline Féry. 2014. Word Order, Information Structure and Intonation of Discontinuous Nominal constructions in Cantonese Information Structure. *Cahiers de Linguistique Asie Orientale* 43. 1-40.

Dolphyne, Florence Abena (1988) The Akan (Fante-Twi) language. Its sound system and tonal structure. Accra: Ghana Universities Press.

Donohue, Mark. 1997. Tone systems in New Guinea. *Linguistic Typology* 1. 347–386.

Downing, Laura. 2010. Accent in African languages. In: Hulst, Harry van der, Goedemans, Rob & Zanten, Ellen van (eds.). *A survey of word accentual patterns in the languages of the world*. Berlin & New York: Mouton de Gruyter. 381-427.

Downing, Laura J. & Bernd Pompino-Marschall. 2013. The focus prosody of Chichewa and the Stress-Focus constraint: a response to Samek-Lodovici 2005. *Natural Language and Linguistic Theory* 31.

Downing, Laura J. & Maxwell Kadenge. 2014. *Motivating prosodic stems in Shona*. Handout of the workshop 'On the prosodic hierarchy in a typological perspective'. Stockholm University, March 2014.

Dretske, Fred I. 1972. Contrastive Satements. *Philosophical Review* 81. 411-437.

Duanmu, San. 1991. Stress and syntax-phonology mismatches: tonal domains in Danyang and Shanghai. In: Bates, D. (ed.). *Proceedings of the Tenth West Coast Conference on Formal Linguistics*. Stanford University Association. 127-137.

Duanmu, San. 1993. Rime length, stress, and association domains. *Journal of East Asian Linguistics* 2. 1-44.

Duanmu, San. 2007. *The Phonology of Standard Chinese* (Second edition).Oxford: Oxford University Press.

Duanmu, San. 2011. Chinese syllable structure. In: Oostendorp, Marc van, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.). *The Blackwell Companion to Phonology*, volume 5. Malden, MA & Oxford: Wiley-Blackwell. 2151-2777.

Dyrud, Lars O. 2001. *Hindi-Urdu: Stress Accent or Non-Stress Accent?* Ph.D. thesis, University of North Dakota.

Eady, Stephen J. & William. E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80. 402–415.

É.Kiss, Katalin. 1998. Identificational focus versus information focus. Language 74. 245-273.

Elfner, Emily 2015. Recursion in prosodic phrasing: Evidence from Connemara Irish. *Natural Language and Linguistic Theor*y.DOI 10.1007/s11049-014-9281-5

Eliasson, Stig. 2010. German and Swedish quantity: A comparison. In: Dammel, Antje, Sebastian Kürschner & Damaris Nübling (eds.). *Kontrastive germanistische Linguistik* 1. Hildesheim, Zürich & New York: Georg Olms Verlag. 5-55.

Eliasson, Stig & Nancy LaPelle. 1973. Generativa regler för svenskans kvantitet. *Arkiv för nordisk filologi* 88. 133-148.

Elordieta, Gorka. 1998. Intonation in a pitch-accent variety of Basque. Annuario del Seminario de Filología Vasca Julio de Urquijo. *International Journal of Basque Linguistics and Philology* 32. 511-569.

Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure.* Cambridge: Cambridge University Press.

Fanselow, Gisbert & Denisa Lenertová. 2010. Left peripheral focus: Mismatches between syntax and information structure. *Natural Language and Linguistic Theory* 29.169-209.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague, Netherlands: Mouton, 2nd edition. 1970.

Feldhausen, Ingo. 2010. *Sentential Form and Prosodic Structure of Catalan*. Amsterdam: John Benjamins Publishing Company.

Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100. 233–53.

Féry, Caroline. 1998. German Word Stress in Optimality Theory. *Journal of Comparative Germanic Linguistics* 2.2. 101-142.

Féry, Caroline. 2001. Focus in French. In: Féry, Caroline & Wolfgang Sternefeld (eds.). *Audiatur Vox Sapientiae*. A Festschrift for Arnim von Stechow. Berlin: Akademie-Verlag. 153-181.

Féry, Caroline. 2006. The prosodic basis of topicalization. In: Schwabe, Kerstin & Susanne Winkler. *On Information Structure, Meaning and Form*. Amsterdam/ Philadelphia: Benjamins. 69-86.

Féry, Caroline. 2010. The intonation of Indian languages: an areal phenomenon. In: Hasnain, Imtiaz & Shreesh Chaudhury (eds.). *Problematizing Language Studies. Festschrift for Ramakant Agnihotri*. New Delhi: Aakar Books. 288-312.

Féry, Caroline. 2011. German sentence accents and embedded prosodic phrases. *Lingua 121*. 1906-1922.

Féry, Caroline. 2012. Prosody and information structure of the German particles selbst, wieder and auch. In: Borowsky, Toni, Shigeto Kawahara, Takahito Shinya & Mariko Sugahara (eds.). *Prosody Matters: Essays in Honor of Elisabeth O. Selkirk*. London: Equinox Press. 441-468.

Féry, Caroline. 2013. *Focus as prosodic alignment. Natural Language and Linguistic Theory 31*. 683–734.

Féry, Caroline. 2014. Final Compression in French as a Phrasal Phenomenon. In Katz Bourns, Stacey & Lindsy L. Myer (eds.). *Perspectives on Linguistic Structure and Context. Studies in Honor of Knud Lambrecht*. Amsterdam: Benjamins. 133-156.

Féry, Caroline. 2015. Extraposition and Prosodic Monsters in German. In: Frazier, Lyn & T.Gibson (eds.). *Explicit and Implicit Prosody in Sentence Processing*. Amsterdam: John Benjamins Publishing Company.

Féry, Caroline, Robin Hörnig & Serge Pahaut. 2010. Phrasing in French and German: an experiment with semi-spontaneous speech In: Gabriel, Christoph & Conxita Lleó (eds.). *Intonational Phrasing at the Interfaces*: *Cross-Linguistic and Bilingual Studies in Romance and Germanic.* Amsterdam: Benjamins. 11-41.

Féry, Caroline & Shinichiro Ishihara. 2009. The Phonology of second occurrence focus. *Journal of Linguistics 45*. 285-313.

Féry, Caroline & Shinichiro Ishihara. 2010. How Focus and Givenness Shape Prosody. In: Zimmermann, Malte & Caroline Féry (eds.). *Information Structure*. Theoretical, Typological, and Experimental Perspectives. Oxford: Oxford University Press. 36-63.

Féry, Caroline & Shinichiro Ishihara. 2015. Introduction. In: Féry, Caroline & Shinichiro Ishihara (eds.). *Handbook of information structure*. Cambridge: Cambridge University Press.

Féry, Caroline, Elsi Kaiser, Robin Hörnig, Thomas Weskott & Reinhold Kliegl. 2009. Perception of intonational contours on given and new referents: A completion study and an eye-movement experiment. In: Boersma, Paul & Silke Hamann. *Perception in Phonology*. Berlin: Mouton De Gruyter. 235-266.

Féry, Caroline & Shinichiro Ishihara. (eds.). 2015. *Handbook of information structure*. Oxford: Oxford University Press.

Féry, Caroline, Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36. 680-703.

Féry, Caroline, Pramod Pandey & Gerrit Kentner. To appear. *The prosody of Focus and Givenness in Hindi and Indian English*. Studies in Language.

Féry, Caroline & Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82. 131-150.

Féry, Caroline & Fabian Schubö. 2010. Hierarchical prosodic structures in the intonation of center-embedded relative clauses. *The Linguistic Review* 27. 289–313.

Féry, Caroline & Hubert Truckenbrodt. 2005. *Sisterhood and Tonal Scaling. Studia Linguistica. Special Issue "Boundaries in intonational phonology" 59.* 223-243.

Fiedler, Ines, Katharina Hartmann, Brigitte Reineke, Anne Schwarz & Malte Zimmermann. 2009. Subject Focus in West African Languages. In: Zimmermann, Malte & Caroline Féry (eds.). *Information Structure from Different Perspectives*. Oxford: Oxford University Press. 234-257.

Flynn, Choi-Yeung-Chang. 2003. *Intonation in Cantonese*. München: Lincom.

Fodor, Janet Dean. 1998. Learning to Parse? *Journal of Psycholinguistic Research* 27. 285-319.

Fodor, Janet Dean. 2002a. Prosodic Disambiguation in Silent Reading. In: Hirotani, Masako (ed.). *Proceedings of the Thirty-second Annual Meeting of the North-Eastern Linguistic Society*. 113-137.

Fodor, Janet Dean. 2002b. Psycholinguistics Cannot Escape Prosody. *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence. 83-88.

Frascarelli, Mara. 2000. *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian.* Dordrecht. Kluwer.

Frazier, Lyn, Katy Carlson & Charles Jr Clifton. 2006. Prosodic phrasing is central to

language comprehension. *Trends in Cognitive Sciences* 10. 244-249.

Frazier, Lyn & Janet Dean Fodor. 1978. The Sausage Machine: A new two-stage parsing model. *Cognition* 6. 291-326.

Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14. 176-210.

Frota, Sonia. 2000. *Prosody and Focus in European Portuguese*. New York. Garland Publishing.

Frota, Sónia. 2014. A focus intonational morpheme in European Portuguese: Production and perception. In: Elordieta, Gorka & Pilar Prieto (eds.). *Prosody and meaning*. Berlin: De Gruyter Mouton. 163-196.

Fry, D.B. 1955. Duration and Intensity as Physical Correlates of Linguistic Stress. *Journal of the Acoustical Society of America* 27. 765–768.

Garde, Paul. 1968. *L'accent.* Paris: Presses Universitaires de France.

Gårding, Eva. 1983. A generative model of intonation. In: Cutler, Anne & Robert Ladd (eds.). *Prosody: Models and measurements*. Heidelberg: Springer. 11-25.

Gårding, Eva. 1987. Speech act and tonal pattern in Standard Chinese. *Phonetica* 44. 13-29.

Gee, James P. & François Grosjean. 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15. 411–458.

Genzel, Susanne (2013) Lexical and post-lexical tones in Akan. PhD dissertation. University of Potsdam.

Ghini, Mirco. 1993. Ø-formation in Italian: A new proposal. In: Dyck, Carrie (ed.). *Toronto Working Papers in Linguistics* 12. Department of Linguistics, University of Toronto. 41-78.

Gibbon, Dafydd. 1998. Intonation in German. In: Hirst, Daniel & Albert Di Cristo (eds.) *Intonation Systems: Survey of Twenty Languages*. Cambridge. Cambridge University Press. 1998. 78-95.

Giegerich, Heinz J. 1992. *English Phonology*. Cambridge: Cambridge University Press.

Goldsmith, John A. 1976. *Autosegmental phonology*. Ph.D. thesis. MIT. Distributed by IULC 1976. New York: Garland Press. 1979.

Gomez-Imbert, Elsa & Michael Kenstowicz. 2000. Barasana tone and accent. *International Journal of American Linguistics* 66. 419–463.

Gordon, Matthew. 2003. The phonology of pitch accents in Chickasaw. *Phonology* 20. 173-218.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language & Linguistic Theory* 23. 595–653.

Gordon, Matthew. 2014. Disentangling stress and pitch-accent: a typology of prominence at different prosodic levels. In: Hulst, Harry van der, (ed.). *Word stress*. *Theoretical and typological issues*. Cambridge: Cambridge University Press. 83-118.

Grabe, Esther. 1998a. *Comparative intonational phonology*. English and German, Ph.D. thesis. University of Nijmegen: Max Planck Institute Series in Psycholinguistics.

Grabe, Esther. 1998b. Pitch realization in English and German. *Journal of Phonetics* 26. 129–143.

Grammont, Maurice. 1933. *Traité de phonétique.* Paris: Librairie Delagrave.

Grice, Martine & Stefan Baumann. 2002. Deutsche Intonation und GToBI. Linguistische Berichte 191. 267–298.

Grønnum, Nina & Hans Basbøll. 2001. Consonant Length, Stød and Morae in Standard Danish. Phonetica 58. 230-253.

Gryllia, Stella, Caroline Féry, & Frank Kügler. 2014. *Pitch Accents and Relative Clause Attachment in German*. Potsdam: Manuscript.

Gundel, Jeanette, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274-307.

Güneş, Güliz. 2012. *Prosodic parentheticals and prosodic integration: A case study in Turkish.* Ms. University of Groningen.

Güneş, Güliz. 2013a. Limits on Syntax-Prosody Mapping in Turkish Prosody of Finite and Non-finite Clausal Parentheticals. In: Erguvanlı Taylan, Eser (ed.). *Journal of Linguistics Research* (Dilbilim Araştırmaları Dergisi), special issue "Updates in Turkish Phonology". Istanbul: Boğaziçi University Press.

Güneş, Güliz. 2013b. On the Role of Prosodic Constituency in Turkish. In: Özge, Umut (ed.). *Proceedings of WAFL8 MIT Working Papers in Linguistics*. Cambridge, MA. 115-128

Güneş, Güliz. 2015. Deriving prosodic structures. PhD Thesis. Groningen. LOT 606.

Gunlogson, Christine. 2003. *True to Form: Rising and Falling Declaratives as Questions in English*. New York: Routledge.

Gussenhoven, Carlos. 1983a. *A semantic analysis of the nuclear tones of English*. Bloomington, Indiana: Indiana University Linguistics Club (IULC). Reprinted in Carlos Gussenhoven 1984.

Gussenhoven, Carlos. 1983b. Stress shift and the nucleus. *Linguistics* 21. 303–39.

Reprinted in Carlos Gussenhoven 1984.

Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, Carlos. 1991. The English rhythm rule as an accent deletion rule. *Phonology* 8. 1–35.

Gussenhoven, Carlos 1992. Sentence accents and argument structure. In: Roca, Iggy (ed.). *Thematic Structure. Its role in grammar*. Berlin: Foris. 79-106.

Gussenhoven, Carlos. 2000. The boundary tones are coming: on the non-peripheral natural of boundary tones. In: Michael B. Broe and Janet B. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press. 132-151.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: University Press.

Gussenhoven, Carlos. 2006. The prosody of Nubi between stress and tone. *Phonology* 23. 193–223.

Gussenhoven, Carlos & Gösta Bruce. 1999. Word prosody and intonation. In: Hulst, Harry van der (ed.). *Word Prosodic systems in the languages of Europe*. Berlin: Mouton de Gruyter. 233–271.

Gussenhoven, Carlos & Jörg Peters. 2004. A tonal analysis of Cologne Schärfung. *Phonology* 21. 251–285.

Gussenhoven, Carlos & Peter van der Vliet. 1999. The phonology of tone and intonation in the Dutch dialect of Venlo. *Journal of Linguistics* 35. 99–135.

Hale, Kenneth L. 1973. Deep-surface canonical disparities in relation to analysis and change: an Australian example. In: Sebeok, A. (ed). *Linguistics in Oceania*. *Current Trends in Linguistics* 11. The Hague: Mouton. 401-458.

Hale, Kenneth & Elisabeth Selkirk. 1987. Government and Tonal Phrasing in Papago. *Phonology Yearbook* 4. 151-183

Halle, Morris. 1973. The accentuation of Russian words. Language 49. 312-348

Halle, Morris & William Idsardi. 1995. General properties of stress and metrical structure. In: Goldsmith, John A. (ed.). *The handbook of phonological theory*. Oxford: Blackwell. 403-484.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge: MIT Press.

Halliday, Michael A.K. 1967. Notes on transitivities and theme in English. *Journal of*

linguistics 3.Part 1: 37-81. Part 2: 199-244.

Halliday, Michael A.K. 1970. Language structure and language function. In: Lyons, J. (ed.). *New horizons in linguistics*. Harmondsworth: Penguin. 140-65.

Hamblin, Charles L. 1973. Questions in Montague Grammar. *Foundations of Language* 10. 41-53.

Hamlaoui, Fatima. 2009. Focus, contrast and the syntax-phonology interface: the case of French cleft sentences. *Current issues in unity and diversity of languages: Collection of the papers selected from the 18th International Congress of Linguistics 2008*. Seoul: The Linguistic Society of Korea.

Hamlaoui, Fatima, Sascha Coridun & Caroline Féry. 2012. Expression prosodique du focus et du donné au sein des groupes nominaux [N A] du français. *Proceedings of Congrès mondial de linguistique française 12*. Lyon.

Harnsberger, James D. and Jasmeet Judge. 1996. Pitch range and focus in Hindi. The Journal of theAcoustical Society of America 99. 2493.

Hashimoto, O.K. Yue. 1972. *Studies in Yue Dialects 1: Phonology of Cantonese*, Cambridge: Cambridge University Press.

Haudricourt, André-George. 1954a. Comment reconstruire le chinois archaique. *Word* 10. 351–364.

Haudricourt, André-George. 1954b. De l'origine des tons en vietnamien. *Journal Asiatique* 242. 69–82.

Haugen, Einar. 1967. On the rules of Norwegian tonality. *Language* 43. 185–202.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?. *Science* 298. 2769-79.

Hayes, Bruce. 1980. *A metrical theory of stress rules*. Unpublished Ph.D. Dissertation. Cambridge: MIT.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253-306.

Hayes, Bruce. 1990. Precompiled phrasal phonology. In: Inkelas, Sharon & Draga Zec (eds.). *Phonology-Syntax-Interface*. Chicago: University of Chicago Press. 85-108.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce & Aditi Lahiri (1991) Bengali intonational phonology. *Natural Language and Linguistic Theory* 9. 47-99.

Hedberg, Nancy, Juan M. Sosa and Emrah Görgülü. To appear. The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory*. De Gruyter.

Hermans, Ben. 1985. Het Limburgs en het Litouws als metrisch gebonden toontalen. *Spektator* 14. 48–70.

Hermans, Ben. 1999. In: Harry van der Hulst (ed.). Word prosodic system in the languages of Europe. Berlin: Mouton de Gruyter. 633-658.

Hirose, Yuki. 2003. Recycling Prosodic Boundaries. *Journal of Psycholinguistic Research* 32. 167-195.

Hobbs, Jerry R. 1990. The Pierrehumbert-Hirschberg theory of intonational meaning made simple: Comments on Pierrehumbert and Hirschberg. In: Cohen, P. R., J. Morgan & M. E. Pollack (eds.). *Intentions in Communication*. Cambridge, MA: The MIT Press. 313–323.

Höhle, Tilmann N. 1992. Über Verum-Fokus im Deutschen. In: Jacobs, Joachim (ed.) *Informationsstruktur und Grammatik* (= *Linguistische Berichte*, Sonderheft 4). 112-141.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37-58.

Horne, Merle. 1990. Empirical evidence for a deletion analysis of the rhythm rule in English. *Linguistics* 28. 959–81.

Hualde, José Ignacio. 1999. *Basque accentuation.* In: Hulst, Harry van der (ed.). *Word prosodic systems in the languages of Europe*. Berlin: Mouton de Gruyter. 947-993.

Hualde, José Ignacio, Gorka Elordieta & Arantzazu Elordieta. 1994. *The Basque dialect of Lekeitio*. Bilbao: Euskal Herriko Unibertsitatea.

Hualde, José Ignacio, I. Gaminde Elordieta & Rajka Smiljanić. 2002. From pitch-accent to stress-accent in Basque. In: Gussenhoven, Carlos & Natasha Warner (eds.). 2002. *Laboratory Phonology 7*. Berlin and New York: Mouton de Gruyter. 547-84.

Hulst, Harry van der. (ed.). 1999. *Word prosodic system in the languages of Europe.* Berlin & New York: Mouton de Gruyter.

Hulst, Harry van der. (ed.). 2014. *Word stress*. Theoretical and typological issues. Cambridge: Cambridge University Press.

Hulst, Harry van der, Rob Goedemans & Ellen van Zanten. 2010. *A survey of word accentual patterns in the languages of the world*. Berlin & New York: Mouton de Gruyter.

Hyman, Larry M. 1977. On the nature of linguistic stress. In: Larry M. Hyman, (ed.). *Studies in stress and accent.* Los Angeles: Department of Linguistics, University of Southern California. 37–82.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris. Reprinted 2003 with an introduction, Stanford: CSLI.

Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23. 225-257.

Hyman, Larry M. 2011a. The Representation of Tone. In: Oostendorp, Marc van, Colin J Ewen, Elizabeth Hume & Keren Rice (eds.). *The Blackwell Companion to Phonology Vol 4*. Blackwell Publishing. 1078-1102.

Hyman, Larry M. 2011b. Does Gokana really have no syllable? Or: what's so great about being universal? Phonology 28.55-85.

Idsardi, William James. 1992. *The computation of prosody*. Unpublished PhD dissertation. Cambridge: MIT.

Inkelas, Sharon. 1999. Exceptional stress-attracting suffixes in Turkish: representations vs. The grammar. In: Kager, René, Harry van der Hulst & Wim Zonneveld (eds.). *The Prosody-Morphology Interface*. Cambridge: Cambridge University Press. 134–187.

Inkelas, Sharon & Zec, Draga. (eds.). 1990. *Phonology-Syntax-Interface*. Chicago: University of Chicago Press.

Ipek, Canan. 2011. Phonetic realization of focus with no on-focus pitch range expansion in Turkish. *International Conference of Phonetic Science* XVII, Hong-Kong.

Ipek, Canan & Sun-Ah Jun. 2013. Towards a model of intonational phonology of Turkish: Neutral intonation. In: *Proceedings of Meeting on Acoustics* 19. Acoustical Society of America.

İpek, Canan & Sun-Ah Jun. 2014. Distinguishing phrase-final and phrase-medial high tone on finally stressed words in Turkish. In: *Proceedings of the 7th Speech Prosody International Conference*. Dublin.

Ishihara, Shinichiro. 2004. *Prosody by Phase: Evidence from Focus. Intonation–Wh-scope Correspondence in Japanese*. In: Ishihara, Shinishiro, M. Schmitz & A. Schwarz (eds.). Interdisciplinary Studies in Information Structures 1. *Working Papers of the SFB* 632. Potsdam. 77-119.

Ishihara, Shinichiro. 2007. Major phrase, focus intonation and multiple spell-out (MaP, FI, MSO). *The Linguistic Review* 24. 137–167.

Ishihara, Shinichiro. 2011. Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121. 1870-1889.

Ishihara, Shinichiro. 2015. Syntax–phonology interface. In: Kubozono, Haruo (ed.). *Handbook of Japanese Phonetics and Phonology*. Berlin: Mouton de Gruyter. 569-620

Ishihara, Takeshi. 2003. A phonological effect on tonal alignment in Tokyo Japanese. *Proceedings of 15th ICPhS, Barcelona* 1. 615-18.

Ito, Junko & Armin Mester. 2007. Prosodic adjunction in Japanese compounds. In: Miyamoto, Y.& M. Ochi (eds.). Formal approaches to Japanese linguistics: *Proceedings of FAJL* 4. Cambridge, Massachusetts: MIT Department of Linguistics and Philosophy. 97–111.

Ito, Junko & Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In: Borowsky, T., S. Kawahara, T. Shinya & M. Sugahara (eds.). *Prosody matters: Essays in honor of Elisabeth Selkirk*. London: Equinox Publishers. 280-303.

Ito, Junko & Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124. 20–40.

Ito, Junko & Armin Mester. 2015. Word formation and phonological processes. In: Kubozono, H (ed.). *Handbook of Japanese Phonetics and Phonology.* Berlin. Mouton de Gruyter. 363-395.

Ito, Kiwako & Shari R. Speer. 2006. Using interactive tasks to elicit natural dialogue. In: Augurzky, P. & D. Lenertova (eds.). *Methods in empirical prosody research*. Berlin. Mouton de Gruyter. 229–257.

Ito, Kiwako & Shari R. Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58. 541–573

Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar.* Cambridge, Massschusetts: MIT Press.

Jacobs, Joachim. 2001. The dimensions of topic-comment. In: *Linguistics* 39. 641–681.

Jin, S. 1996. *An acoustic study of sentence stress in Mandarin Chinese*. Unpublished Ph.D. thesis. Ohio State University, Columbus.

Joshi Aravind K. 1982. The role of mutual beliefs in question-answer systems. In: Smith Neil. (ed.). *Mutual knowledge.* New York: Academic Press. 181–197.

Jun, Sun-Ah. 1993. *The Phonetics and Phonology of Korean*, Unpublished Ph.D. thesis. Ohio: Ohio State University.

Jun, Sun-Ah. 1996. Influence of microprosody on macroprosody: A case of phrase

initial strengthening. *University of California Los Angeles Working Papers in Phonetics* 92. 97–116.

Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15. 189–226.

Jun, Sun-Ah. 2002. Syntax over Focus. In: Hansen, John H. L. & B. Pellom (eds.). Proceedings of ICSLP Denver, CO. 2281-2284.

Jun, Sun-Ah. 2005. Korean intonational phonology and prosodic transcription. In Jun, Sun-Ah (ed.) *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2014. *Prosodic Typology.* Oxford: Oxford University Press.

Jun, Sun-Ah & Cécile Fougeron. 2000. A phonological model of French intonation. In: Botinis, A. (ed.). *Intonation: Analysis*. Dordrecht: Kluwer. 209–242.

Jun, Sun-Ah & Cécile Fougeron. 2002. The Realizations of the Accentual Phrase in French Intonation. *Probus* 14. 147-172.

Jun Sun-Ah & Hee-Sun Kim. 2007. VP focus and narrow focus in Korean. Proceedings of International Conference of Phenetic Science XVI, Saarbrücken. 1276-1279.

Jyothi, Preethi, Jennifer Cole, Mark Hasegawa-Johnson & Vandana Puri. 2014. An Investigation of Prosody in Hindi Narrative Speech. Proceedings of Speech Prosody.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Kager, René. 1999. *Optimality Theory. A Textbook*. Cambridge. Cambridge University Press.

Kahnemuyipour, Arsalan. 2009. *The syntax of sentential stress*. Oxford: Oxford University Press.

Kaiser, Elsi & John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94. 113–147.

Kamali, Beste. 2011. *Topics at the PF Interface of Turkish*. Unpublished Ph.D. thesis. Harvard University.

Kamali, Beste & Hubert Truckenbrodt (2014). *Elements of intonation in Turkish.* Submitted.

Kan, Seda. 2009. Prosodic domains and the syntax-prosody mapping in Turkish. Unpublished master's thesis, Bogaziçi University.

Kanerva, Jonni M. 1990. Focusing on phonological phrases in Chichewa. In: Inkelas, Sharon & Draga Zec (eds.). *Phonology-syntax-interface*. Chicago: University of Chicago Press. 145-161.

Kenstowicz, Michael & Mairo Kidda. 1987. The obligatory contour principle and Tangale phonology. In: Odden, David (ed.). *Current approaches to African linguistics* 4. Dordrecht: Foris. 223–238.

Kentner, Gerrit. 2012. Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition* 123. 1–20.

Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. *The Linguistic Review 2013* 30. 277 – 311.

Khan, Sameer. 2007. *Phrasing and Focus in Bengali.* Proceedings of the 16thInternational Congress of Phonetic Sciences (ICPhS).

Kingston, John. 2011. *Tonogenesis*. *The Blackwell Companion to Phonology*. In: Oostendorp, Marc van, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.). Oxford: Blackwell Publishing. 2304-2333.

Kiparsky, Paul. 1982. *Lexical Phonology and Morphology. Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co.

Kisseberth, Charles & Mohammad Imam Abasheikh. 1974. Vowel length in Chi-Mwi:ni : a case study of the role of grammar in phonology. In: Bruck, Anthony, Robert A. Fox & Michael W. La Galy (eds.). *Papers from the parasession on natural phonology*. Chicago: Chicago Linguistic Society. 193–209.

Kisseberth, Charles & David Odden. 2003. Tone. In: Nurse, Derek & Gérard Philippson (eds.). *The Bantu languages*. London & New York: Routledge. 59–70.

Kjelgaard, Margaret M. & Shari R. Speer. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40. 153-194.

Klein, Wolfgang & Cristiane von Stutterheim. 1987. Quaestio und referentielle Bewegung in Erzählungen. Linguistische Berichte 109. 163-183

Kodzasov, Sandro. 1999. Russian. In: Hulst, Harry van der (ed.). *Word prosodic system in the languages of Europe*. Berlin & New York: Mouton de Gruyter. 852-869.

Kohler, Klaus J. 1987. Categorical pitch perception. *ICPhS* II (Tallinn) 5. 331-3.

Kohler, Klaus J. 1990. Macro and micro F0 in the synthesis of intonation. In: Kingston, John & Mary E. Beckman (eds.). *Papers in Laboratory Phonology* I*:*

Between the Grammar and Physics of Speech. Cambridge: Cambridge University Press. 115–138.

Koizumi, Y. & D. Bradley. 2007. Non-syntactic factors in proecssing the 'Not-Because' ambiguity in English. *Proceedings of the 8th Tokyo Conference on Psycholinguistics (TCP2007)*. Tokyo, Hituzi Syobo. 187-212.

Konrot, Ahmet. 1981. Physical correlates of linguistic stress in Turkish. In University of Essex Language Centre occasional papers. University of Essex Language Centre. 26-53.

Kraljic, Tanya & Susan E. Brennan. 2005. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology* 50. 194-231.

Kratzer, Angelika & Elisabeth Selkirk. 2007. Phase theory and prosodic spellout: The case of verbs. *The Linguistic Review* 24. 93-135.

Krifka, Manfred. 1998. Scope inversion under the rise-fall contour in German. *Linguistic Inquiry* 29. 75-112.

Krifka, Manfred. 2008. Basic Notions of information Structure. *Acta Linguistica Hungarica* 55. 243–276.

Kristoffersen, Gjert. 2000. *The Phonology of Norwegian*. Oxford: Oxford University Press.

Kubozono, Haruo. 1989a. The mora and syllable structure in Japanese: evidence from speech errors. *Language and Speech,* 32. 249 -278.

Kubozono, Haruo. 1989b. Syntactic and Rhythmic Effects on Downstep in Japanese. *Phonology* 6. 39-67.

Kubozono, Haruo. 1993. *The Organization of Japanese Prosody*. Tokyo: Kurosio.

Kubozono, Haruo. 1999. Mora and syllable. In: Tsujimura, Natsuko (ed.). *The handbook of Japanese linguistics*. Oxford: Blackwell. 31–61.

Kubozono, Haruo. 2008. Japanese Accent. In: Miyagawa, Shigeru (ed.). *The Oxford Handbook of Japanese Linguistics*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780195307344.013.0007

Kubozono, Haruo. 2015. Japanese Accent. In: Kubozono Haruo (ed.). *Handbook of Japanese Phonetics and Phonology*. Berlin: Mouton de Gruyter.

Kubozono, Haruo. 2016.

Kügler, Frank. 2007. The intonational phonology of Swabian and Upper Saxon. Tübingen: Niemeyer.

Kügler, Frank. 2015. Phonological phrasing and ATR vowel harmony in Akan. *Phonology* 32.

Kügler, Frank & Caroline Féry. 2015. Postfocal downstep in German. Potsdam. Ms.

Kügler, Frank & Anja Gollrad. 2015. Production and Perception of Contrast: The case of the fall-rise contour in German. *Frontiers in Psychology* 6-1254. doi:10.3389/fpsyg.2015.01254

Kügler, Frank, Stavros Skopeteas & Elisabeth Verhoeven. 2007. Encoding Information Structure in Yucatec Maya: *On the Interplay of Prosody and Syntax Interdisciplinary Studies on Information Structure* 8. Potsdam. 187-208.

Kuno, Susumo. 1972. Functional sentence perspective: *A case study from Japanese and English. Linguistic Inquiry* 3. 269-320.

Kutas, Marta & Steven A. Hillyard. 1980. Reading senseless sentences: brain potentials reflect semantic anomaly. *Science* 34. 203-205.

Ladd, D. Robert. 1980. *The Structure of Intonational Meaning*: *Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. Robert. 1983a. Phonological features of intonational peaks. *Language,* 59. 721–59.

Ladd, D. Robert. 1983b. Peak features and overall slope. In: Cutler, Anne. & Ladd, D. Robert. (eds.). *Prosody: models and measurements.* Heidelberg: Springer. 39–52.

Ladd, D. Robert. 1986. Intonational phrasing: the case for recursive prosodic structure. Phonology *Yearbook* 3. 311-340.

Ladd, D. Robert. 1988. Declination 'reset' and the hierarchical organization of utterances. *Journal of Acoustics Society of America* (JASA) 84. 530-544.

Ladd, D. Robert. 1990. Metrical representation of pitch register. In: John Kingston & Mary E. Beckman (eds.). *Papers in laboratory phonology* 1: *Between the Grammar and the Physics of Speech*. Cambridge: Cambridge University Press, 35–57.

Ladd, D. Robert. 1996/2008. *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D . Robert. 2014. *Simultaneous Structure in Phonology.* Oxford: Oxford University Press.

Ladd, D. Robert, D. Faulkner, H. Faulkner & Astrid Schepman. 1999. Constant segmental anchoring of f0 movements under changes in speech rate. *Journal of the American Acoustic Association* 106. 1543–54.

Ladd, D. Robert, Ineke Mennen & Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the American Acoustic Association* 107. 2685–96.

Ladd, D. Robert & Alex Monaghan. 1987. Modelling rhythmic and syntactic effects on accent in long noun phrases. In: Laver, J. & M. Jack (eds.). *Proceedings of the European Conference on Speech Technology* 2. Edinburgh: CEP. 29–32.

Ladefoged, Peter & Keith Johnson. 2005/2011. *A Course in Phonetics*. Wadsworth Cengage Learning.

Lakoff, George. 1971. Presupposition and relative well-formedness. In: Steinberg, D.D. & L.A. Jakobovits (eds.). *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology.* Cambridge: Cambridge University Press. 329-340.

Laniran, Yetunde O. 1992. *Intonation in tone languages: the phonetic implementation of tones in Yoruha.* Unpublished PhD thesis, Cornell University.

Laniran, Yetunde O. & G. N. Clements. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31. 203-250.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge. Cambridge University Press.

Law, Sam-po. 1990. *The Syntax and Phonology of Cantonese Sentence-final Particles*. Unpublished Ph.D. thesis. Boston University, Boston.

Leben, William R. 1973. *Suprasegmental phonology*. Unpublished Ph.D. thesis. Cambridge: MIT.

Leben, William R. 1978. The representation of tone. In: Fromkin, V. A. (ed.). *Tone: A Linguistic Survey*. New York: Academic Press. 177–219.

Leben, William R. 2002. Tonal feet. In: Gut, Ulrike & Dafydd Gibbon (eds.). *Typology of African prosodic systems*. Bielefeld: University of Bielefeld. 27–40.

Lehiste, Ilse. 1970. *Suprasegmentals.* Cambridge, Massachusetts: MIT Press.

Lehiste, Ilse. 1973. Phonetic disambiguation of syntactic ambiguity. *Glossa* 7. 107–122.

Léon, Pierre R. 1972. *Prononciation du français standard. Aide mémoire d'orthoépie à l'usage des étudiants étrangers.* Paris: Didier.

Levi, Susannah. 2002. *Intonation in Turkish: the realization of noun compounds and*

genitive possessive NPs. Ms. University of Washington.

Levi, Susannah. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35. 73–97.

Levin, Juliette. 1985. *A metrical theory of syllabicity*. Unpublished Ph. D. thesis. Cambridge, MA: MIT.

Liberman, Mark Y. 1978. *The intonational system of English*. Bloomington: Indiana University Linguistic Club.

Liberman, Mark Y. & Ivan Sag. 1974. Prosodic form and discourse function. *Papers from the 10th regional meeting. Chicago Linguistc Society*. 416-427.

Liberman, Mark Y. & Janet B. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In: Aronoff, M. & R.T. Oehrle (eds.). *Language sound structure.* Cambridge, MA: MIT Press. 157-233.

Liberman, Mark Y. & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249-336.

Lieberman, Philip & Sheldon Michaels. 1962. Some Aspects of Fundamental Frequency and Envelope Amplitude as Belated to the Emotional Content of Speech. *Journal of the Acoustical Society of America* 34. 922-927.

Lieberman, Philip. 1967. *Intonation, Perception, and Language.* Cambridge, Massachusetts: MIT Press.

Lin, Jianping. 2002. *Xianggang Yueyu Judiao Yanjiu*. A Study on Sentence Intonation in Hong Kong Cantonese. Ph.D. dissertation. Hong Kong: Chinese University of Hong Kong.

Lin, Jowang. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11. 237-76.

Lohnstein, Horst. 2015. Verum focus. In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. & Alan S. Prince. 1986. *Prosodic phonology*. Amherst: Ms University of Massachusetts & Waltham: Brandeis University.

McCarthy, John J. & Alan S.Prince. 1993a. *Prosodic Morphology* I: Constraint interaction and satisfaction. Amherst: Ms. University of Massachusetts & New Brunswick, New York: Rutgers University.

McCarthy, John J. & Alan S. Prince. 1993b. Generalized Alignment. In: Booij, Geert &

Jaap van Marle (eds.). *Yearbook of Morphology 1993*. Dordrecht: Kluwer. 79-153.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McCawley, James J. 1978. What is a tone language? In: Fromkin, Victoria (ed.). *Tone: A Linguistic Survey*. New York: Academic Press. 113-131.

Makarova, Veronika. 2007. The effect of pitch peak alignment on sentence type identification in Russian. Language and Speech, 50. 385-422.

Malmberg, Bertil. 1969. *Phonétique française.* Lund: Liber Läromedel.

Marcus, Mitshell & Donald Hindle. 1995. Description theory and intonation boundaries. In: Altmann, G. (ed.). *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press. 483-512

Marslen-Wilson, William D., Lorraine K. Tyler, Paul Warren, P. Grenier & C.S. Lee, 1992. Prosodic Effects in Minimal Attachment. *The Quarterly Journal of Experimental Psychology* 45. 73-87.

Martin, Philippe. 1975. Analyse phonologique de la phrase française. *Linguistics* 146. 35–67.

Martin, Philippe. 1980. Sur les principes d'une théorie syntaxique de l'intonation. In: Rossi, Léon (ed.) *Problèmes de Prosodie* 1*(Studia Phonetica* 17*)*. Paris: Didier. 91–101.

Marty, Anton. 1918. *Gesammelte Schriften*. Vol II. Halle: Max Niemeyer Verlag.

Meeussen, Achilles 1970. Tone typologies for West African languages. *African Language Studies* 11, 266–271.

Mei, Tsu-lin. 1970. Tones and prosody in Middle Chinese and the origin of the Rising tone. *Harvard Journal of Asian Studies* 30. 86–100.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27. 661–738.

Mertens, Pierre. 1990. Intonation. In: Blanche-Benveniste, Claire, Mireille Bilger, Christine Rouget & Karel van den Eynde (eds.). *Recherches Le Français parlé.* Paris: CNRS Éditions. 159–176.

Mo, Y., Cole, J., & Lee, E-K. 2008. Naïve listeners' prominence and boundary perception. *Fourth Speech Prosody*. 735-736.

Mohanan, K.P. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.

Moore, Robert R. 1965. *A study of Hindi intonation*. Ph.D. thesis, University of Michigan.

Mous, Maarten. 1993. *A grammar of Iraqw*. Hamburg: Buske.

Myrberg, Sara. 2010. The Intonational Phonology of Stockholm Swedish. *Acta Unversitatis Stockholmiensis*. Stockholm Series in Scandinavian Philology. New Series 53.

Myrberg, Sara. 2013. Sisterhood in prosodic branching. *Phonology* 30. 73-124.

Myrberg, Sara & Tomas Riad 2015. On the expression of focus in the metrical grid and in the prosodic hierarchy. In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Nagahara, Hiroyuki. 1994. Phonological phrasing in Japanese. Unpubished Ph.D. dissertation. University of California, Los Angeles.

Nash, Jay A. 1992–94. Underlying low tones in Ruwund. *Studies in African Linguistics* 23. 223–278.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris. Reprinted 2007 by Mouton de Gruyter.

Newman, Stanley. 1946. On the stress system of English. *Word* 2. 171-87.

Newman, Stanley. 1947. Bella Coola I: phonology. *IJAL* 13. 129–134.

Nooteboom, S. G. & J. G. Kruyt. 1987. Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America* 82. 1512–1524.

O'Connor, J. D. & G. F. Arnold. 1961. *Intonation of Colloquial English.* London: Longman.

O'Connor, J. D. & G. F. Arnold. 1973. *Intonation of Colloquial English*. London: Longman, Second edition.

Odden, David. 1981. *Problems in Tone Assignment in Shon*a. Unpublished Ph.D. dissertation. University of Illinois.

Odden, David. 1982. Tonal phenomena in KiShambaa. *Studies in African Linguistics* 13. 177–208.

Odden, David. 1984. Stem Tone Assignment in Shona. In: Clements, George N. & J. Goldsmith (eds.). *Autosegmental studies in Bantu tone*. Berlin: De Gruyter. 255-280.

Odden, David. 1995. *Tone: African languages*. In: John A. Goldsmith (ed.). *Handbook of phonological theory*. 444–475.

Odden, David. 2005, *Introducing Phonology*. Cambridge: Cambridge University Press.

Ohala, John. 1983. Cross-language use of pitch: An ethological view. *Phonetica* 40. 1-18.

Osser, Henry A. 1964. *A Distinctive Features Analysis of the Vocal Communication of Emotion*. Unpublished Ph.D. thesis. Cornell University.

Osterhout, Lee & Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 3. 785-806.

Palmer, Harold Edward. 1922. *English Intonation with Systematic Exercises*. Cambridge: Heffer.

Pandey, Pramod. 1989. Word accentuation in Hindi. *Lingua* 77. 37-73.

Patil, Umesh, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry & Shravan Vasishth. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics* 1. 53-70.

Peng Shu-hui, K.M. Marjorie, Chiu-yu Tseng Chan, Tsan Huang, Ok Joo Lee & Mary E. Beckman. 2005. Towards a Pan-Mandarin System for Prosodic Transcription. In: Jun, Sun-Ah (ed.). *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 230-70.

Perlmutter, David. 1995. Phonological quantity and multiple association. In: Goldsmith, John A. (ed.). *The handbook of phonological theory*. Oxford: Blackwell. 307-317.

Philippson, Gérard. 1998. Tone reduction vs. metrical attraction in the evolution of Eastern Bantu tone systems. In: Hyman, Larry M. & Charles W. Kisseberth (eds.). *Theoretical aspects of Bantu tone.* Stanford: CSLI. 315-29.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis. Cambridge, Massachusetts, MIT. Published by New York: Garland Press. 1990.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, Massachusetts: MIT Press.

Pierrehumbert, Janet B. & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P., J. Morgan & M. Pollock (eds.). *Intentions in communications*. Cambridge, MA: MIT Press. 271-311.

Pierrehumbert, Janet B. & Shirley Steele. 1989. Categories of tonal alignment in

English. *Phonetica* 46. 181–96.

Pike, Kenneth L. 1945. *The intonation of American English*. University of Michigan Press.

Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Unpublished Ph.D. thesis. Cambridge, MA: MIT.

Post, Brechtje. 2000. *Tonal and Phrasal Structures in French Intonation*. The Hague: Thesus (Subsidiary of Holland Academic Graphic).

Potts, Christopher 2005. *The logic of conventional implicatures* (Oxford Studies in Theoretical Linguistics). Oxford: Oxford University Press.

Price, Patti J., S. Ostendorf, Stefanie Shattuck-Hufnagel & C. Fong. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 9. 2956–70.

Prieto, Prieto, Chilin Shih & Holly Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* 24. 445-473.

Prieto, Pilar & Francisco Torreira. 2007. The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35. 473-500.

Prince, Alan & Paul Smolensky. 1993/2004. Optimality Theory: Constraint Interaction in Generative Grammar. Malden, MA, and Oxford: Blackwell.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In: Cole, Peter (ed.). *Radical Pragmatics*. New York: Academic Press. 223-256.

Pulleyblank, Edwin G. 1962. The consonantal system of Old Chinese. *Asia Major, Series* 2. 58–144, 206–265.

Puri, Vandana. 2013. *Intonation in Indian English and Hindi late and simultaneous bilinguals.* Ph.D. dissertation, University of Illinois at Urbana- Champaign.

Pynte, Joel & B. Prieur. 1996. Prosodic breaks and attachment decisions in sentence processing. *Language and Cognitive Processes* 11. 165–192.

Reetz, Henning & Allard Jongman. 2009. *Phonetics: Transcription, Acoustics, and Perception*. Oxford: Wiley-Blackwell.

Reeves, Carolyn et al. 2000. Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory & Cognition* 26. 1638–1654.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53-94.

Remijsen, Bert. 2002. Lexically contrastive stress accent and lexical tone in Ma'ya. In: Gussenhoven, Carlos & Natasha Warner (eds.). 2002. *Laboratory Phonology* 7, Berlin & New York: Mouton de Gruyter. 585–614.

Riad, Tomas. 1998a. *Towards a Scandinavian tone accent typology*. In: Kehrein, W. & R. Wiese (eds.). Phonology and morphology of the Germanic languages. *Linguistische Arbeiten* 386. Tübingen: Niemeyer. 77–109.

Riad, Tomas. 1998b. The origin of Scandinavian tone accents. *Diachronica* 15. 63–98.

Riad, Tomas. 2000. The origin of the Danish stød. In: Lahiri, Aditi (ed.). *Analogy, leveling, markedness: Principles of change in phonology and morphology*. Berlin & New York: Mouton de Gruyter. 261–300.

Riad, Tomas. 2003. Diachrony of the Scandinavian accent typology. In: Fikkert, Paula & Haike Jacobs (eds.). *Development in prosodic systems*. Berlin & New York: Mouton de Gruyter. 91–144.

Riad, Thomas 2014. *The Phonology of Swedish*. Oxford: Oxford University Press.

Rialland, Annie. 2007. Question prosody: an African perspective. In: Gussenhoven, Carlos & T. Riad (eds.). *Tones and tunes* I: *Typological studies in word and sentence prosody*. Berlin: Mouton de Gruyter: 35-62.

Rialland, Annie. 2009. The African lax question prosody: Its realisation and geographical distribution. *Lingua* 119. 928–949.

Rialland, Annie & Martial Embanga Aborobongui. 2015. *How intonations interact with tones in Embosi (Bantu C25), a two-tone language without downdrift*. Paris. Manuscript.

Rietveld. A. C. M. & Carlos Gussenhoven 1995. Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* 23. 375-85.

Rischel, Jørgen. 1974. *Topics in West Greenlandic phonology. Regularities underlying the phonetic appearance of word forms in a polysynthetic language*. Copenhagen: Akademisk Forlag.

Roberts, Craige. 1996. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. In: Yoon, J.H. & Andreas Kathol (eds.). OSU *Working Papers in Linguistics* 49: Papers in Semantics. 91-136.

Roca, Iggy. 1999. Stress in the Romance Languages. In: Hulst, Harry van der (ed.). 1999. *Word prosodic system in the languages of Europe*. Berlin & New York: Mouton de Gruyter. 659-811.

Roll, Mikael. 2014. *Brain network for tones cuing grammatical structure*: Talk given at

TIE 2.Utrecht.

Roll, Mikael & Merle Horne. 2011. Interaction of right-and left-edge prosodic boundaries in syntactic parsing. *Brain research* 1402. 93-100.

Roll, Mikael, Merle Horne & Magnus Lindgren. 2009. Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses – An ERP study. *Journal of Neurolinguistics* 22. 55–73.

Rolland, Guillaume and Hélène Loevenbruck. 2002. Characteristics of the accentual phrase in French: An acoustic, articulatory and perceptual study. In: Bel, Bernard & Isabelle Marlien. *Proceedings of the speech prosody 2002 conference.* Aix-en-Provence. 611–614.

Rooth, Mats. 1985. *Associations with focus*. Unpublished Ph.D. thesis. Amherst: University of Massachusetts.

Rooth, Mats. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1. 75-116.

Rooth, Mats. 2015. Alternative semantics. In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Rossi, Mario. 1980. Le français, langue sans accent? In: Fónagy, Ivan & Pierre P. Léon. *L'accent en français contemporain.* (*Studia Phonetica* 15). Montréal: Didier. 13–51.

Saeed, John. 1999. Somali. Amsterdam: Benjamins.

Sag, Ivan & Mark Y. Liberman. 1975. *The intonational disambiguation of indirect speech acts*. Papers from the 11th regional meeting, Chicago Linguistic Society, 487-497.

Sagart, Laurent. 1999a. The origin of Chinese tones. In: Kaji, Shigeki (ed.). *Cross-linguistic studies of tonal phenomena: Tonogenesis, typology, and related topics*. Tokyo: ILCAA. 91–103.

Sagart, Laurent. 1999b. The Roots of Old Chinese. *Current Issues in Linguistic Theory* 184. Amsterdam: John Benjamins.

Samek-Lodovici, Vieri. 2005. Prosody-Syntax Interaction in the Expression of Focus. *Natural Language and Linguistic Theory* 23. 687-755.

Schafer, Amy, Katy Carlson, Charles Jr. Clifton & Lyn Frazier. 2000. Focus and the Interpretation of Pitch Accent: Disambiguating Embedded Questions. *Language and Speech* 43. 75 –105.

Schafer, Amy, Juli Carter, Charles Jr Clifton & Lyn Frazier. 1996. Focus in Relative Clause Construal. *Language and Cognitive Processes* 11. 135–163

Schafer, Amy, Shari R. Speer, Paul Warren & S. David White. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29. 169-182.

Schepman, Astrid, Robin Lickley & D. Robert Ladd. 2006. Effects of vowel length and right context on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34. 1-28.

Scherer, Klaus R. 2000. A cross-cultural investigation of emotion inferences from voice and speech: implications for speech technology. In: Yuan, B., T. Huang & X. Tang (eds.). *Proceedings of the Sixth International Conference on Spoken Language Processing* 2. Beijing: China Military Friendship Publish. 379-82.

Scherer, Klaus R., Tom Johnstone & Gudrun Klasmeyer. 2003. Vocal expression of emotioin. In: Davidson, R.J., H. Goldsmith & K.H. Scherer (eds.). *Handbook of the affective sciences*. Oxford: Oxford University Press. 443-56.

Schmerling, Susan. 1976. *Aspects of English sentence stress*. Austin: University of Texas Press.

Schmidt, Jürgen Erich. 1986. *Die Mittelfränkische Tonakzente* (Rheinische Akzentuierung). Stuttgart: Franz Steiner.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and Other Constraints on the Placement of Accent. *Natural Language Semantics* 7. 141-177.

Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11. 563-605.

Selkirk, Elisabeth O. 1984. *Phonology and syntax*. The relation between sound and structure. Cambridge, Massachusetts: MIT Press.

Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371-405.

Selkirk, Elisabeth O. 1995. Sentence prosody: intonation, stress, and phrasing. In: Goldsmith, John A. (ed.). *The handbook of phonological theory*. Oxford: Blackwell. 550–569.

Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In: Morgan, James L. & Katherine Demuth (eds.). *Signal to syntax. Bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates. 187-213.

Selkirk, Elisabeth O. 2000. The Interaction of Constraints on Prosodic Phrasing. In: Horne, Merle (ed.). *Prosody: Theory and Experiment.* Dordrecht. Kluwer. 231-261.

Selkirk, Elisabeth O. 2011. The syntax–phonology interface. In: Goldsmith, John A., J. Riggle & A. Yu (eds.). *The handbook of phonological theory* 2. Oxford: Blackwell. 435-484.

Selkirk, Elizabeth. O. & Tong Shen. 1990. Prosodic domains in Shanghai Chinese. In: Inkelas, Sharon & Draga Zec (eds.). *Phonology-Syntax-Interface*. Chicago: University of Chicago Press. 313-337.

Selkirk, Elisabeth O. & Koichi Tateishi. 1991. Syntax and downstep in Japanese. In: Georgopoulos, Carol & Roberta Ishihara (eds.). *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda*. Dordrecht: Kluwer Academic Publishers. 519–543.

Sezer, Engin. 1983. On non-final stress in Turkish. *Journal of Turkish Studies* 5. 61–69.

Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence comprehension. *Journal of Psycholinguistic Research* 25. 193-247.

Shih, Chilin, 1988. Tone and intonation in Mandarin. Working Papers, Cornell. *Phonetics Laboratory* 3. 83–109.

Shih, Chilin. 1997. Mandarin third tone sandhi and prosodic structure. Wang, Jialing & Norval Smith (eds.) .*Studies in Chinese Phonology*. Berlin: Walter de Gruyter. 81-123.

Sievers, Eduard. 1901. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprache, 5th edition*. (1st edition.1876). Leipzig: Breitkopf and Härtel.

Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23. 287–308.

Silverman, Kim E., Mary E. Beckman, John Pitrelli, Mori Ostendorf, Colin Wightman, Patti Price, Janet B. Pierrehumbert & Julia Hirschberg. 1992. *To BI: a standard for labeling English prosody*. Proceedings of the 2nd International Conference on the Processing of Spoken Language. 867–870.

Silverman, Kim E. & Janet B. Pierrehumbert. 1990. The timing of pre-nuclear high accents in English. In: Kingston, John & Mary. E. Beckman (eds.). 1990. *Papers*

in Laboratory Phonology I: Between the Grammar and Physics of Speech. Cambridge: Cambridge University Press. 72–106.

Skopeteas, Stavros & Caroline Féry. 2015. Focus and intonation in Georgian. Submitted

Skopeteas, Stavros, Caroline Féry & Rusudan Asatiani. 2009. Word order and intonation in Georgian. *Lingua* 119. 102–127.

Skopeteas, Stavros, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry, & Manfred Krifka. 2006. Questionnaire on Information Structure (ISIS Vol. 4). Potsdam: Universitätsverlag Potsdam.

Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *JASA* 100.4. 2471–2485

Slowiaczek, Maria L. 1981. *Prosodic units as processing units*. Unpublished Ph.D. dissertation, Amherst: University of Massachusetts.

Snedeker, Jesse & John Trueswell. 2003. Using prosody to avoid ambiguity? *Journal of Memory and Language* 48. 103–130.

Snider, Keith. 1999. *The Geometry and Features of Tone*. Arlington: Summer Institute of Linguistics & The University of Texas at Arlington.

Speer, Shari R. & Allison Blodgett 2006. Prosody. In: Traxler, Matthew & Morton Gernsbacher (eds.). *Handbook of Psycholinguistics*. Oxford: Elsevier. 505-538

Speer, Shari R., Margaret Kjelgaard & Kathryn M. Dobroth. 1996. The influence of prosodic strucuture on the resolution of temporary syntactic ambiguities. *Journal of Psycholinguistic Research* 25. 249-271.

Stalnaker, Robert 1974. *Pragmatic presuppositions*. In: Munitz, Milton & Peter Unger (eds.). Semantics and philosophy. New York: New York University Press. 197–213.

Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy* 25. 701-721.

Steedman, Mark. 2000. Information Structure and the Syntax-Phonology Interface. Linguistic Inquiry 31. 649-689.

Steedman, Mark. 2014. The surface-compositional semantics of english intonation. *Language* 90. 2–57.

Steinhauer, Karsten, Kai Alter & Angela D. Friederici. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature* 2. 191-196.

Stewart, John M. 1966. The typology of the Twi tone system. *Bulletin of the Institute of African Studies* 1.

Stirling, Lesley & Roger Wales. 1996. *Does prosody support or direct sentence processing? Language and cognitive processes* 11. 1996. 193-212.

Sudhoff, Stefan, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.). 2006. *Methods in Empirical Prosody Research*. Berlin. Mouton de Gruyter.

Suomi, Kari 2009. Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics* 37. 397–416.

Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in northern Mon-Khmer. *Phonetica* 46. 60–79.

Svantesson, Jan-Olof & David House (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology* 23. 309-333.

Sybesma, Rint & Boya Li. 2007. The dissection and structural mapping of Cantonese sentence final particles. *Lingua* 117. 1739-1783.

Taff, Alice. 1997. Intonation patterns in Unangan. In: Botinis, A., G. Kouroupetroglou & G. Carayannis (eds.). 1997. *Intonation: Theory, Models and Applications*. Proceedings of an ESCAWorkshop, Athens (Greece). ESCA and University of Athens, Department of Informatics. 301–304.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen. M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–1634.

Terken, Jaques & Julia Hirschberg. 1994. Deaccentuation of words representing given information: effects of persistence of grammatical function and surface position. *Language and Speech* 37. 125–145.

Terken, Jaques & S. G. Nooteboom. 1987. Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information. *Language and Cognitive Processes* 2. 145–163.

't Hart, J., R. Collier & A. Cohen 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press. 141–189.

Thurgood, Graham. 2002. Vietnamese and tonogensis. *Diachronica* 19. 333–363.

Tomioka, Satoshi 2010. Contrastive Topics Operate on Speech Acts. In: Zimmermann, Malte & Caroline Féry (eds.). *Information Structur*e. Oxford: Oxford University Press. 115–138.

Topintzi, Nina 2010. *Onsets: suprasegmental and prosodic behaviour*. Cambridge: Cambridge University Press.

Touati, Paul 1987. *Structures Prosodiques du Suédois et du Français*. Lund University Press.

Trager, George L. & Henry Lee Jr. Smith. 1951. *An Outline of English Structure. Norman*, OK: Battenburg Press. Reprinted by the American Council of Learned Societies, Washington, 1957.

Trubetzkoy, Nikolai Sergejewitsch. 1939. *Grundzüge der Phonologie*. Prag (=TCLP 7) (Published again 1958 by Vandenhoeck & Ruprecht, Göttingen).

Truckenbrodt, Hubert. 1995. *Phonological Phrases: Their Relation to Syntax, Focus and Prominence*. Unpublished Ph.D thesis. Cambridge, Massachusetts, MIT.

Truckenbrodt, Hubert. 1999. On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30. 219-255.

Truckenbrodt, Hubert. 2002. Upstep and embedded register levels. *Phonology 19*. 77–120.

Truckenbrodt, Hubert. 2004. Final lowering in non-final position. *Journal of Phonetics* 32. 313-348.

Truckenbrodt, Hubert. 2005. A short report on intonation phrase boundaries in German. *Linguistische Berichte* 203. 273-296.

Truckenbrodt, Hubert. 2006. Phrasal stress. In: Brown, Keith (ed.). The Encyclopedia of Languages and Linguistics, 2nd Edition. Oxford. *Elsevier*. 572-579.

Truckenbrodt, Hubert. 2007. Upstep on edge tones and on nuclear accents. In: Gussenhoven, Carlos & Tomas Riad (eds.). *Tones and tunes: Studies in word and sentence prosody* 2. Berlin: Mouton De Gruyter. 349-386.

Truckenbrodt, Hubert. 2012. Semantics of intonation. In: Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics. An international handbook of natural language meaning. Vol. 3*, 2039-2969. Berlin: de Gruyter.

Truckenbrodt, Hubert. 2015. Focus, intonation and tonal height In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford Handbook of information structure*. Oxford: Oxford University Press.

Truckenbrodt, Hubert & Féry, Caroline. 2015. Hierarchical organization and tonal scaling. *Phonology* 32. 1-29.

Uldall, Elizabeth. 1964. Dimensions of Meaning in Intonation. In: David, Abercrombie, D. B. Fry, P. A. D. McCarthy, N. C. Scott & J. L.M.Trim (eds.). *In honour of Daniel Jones*. London: Longmans. 271-279.

Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.

Verhoeven, Elisabeth & Kügler, Frank (2015) Accentual preferences and predictability. An acceptability study on split intransitivity in German. Lingua 165(B), 298–315. http://dx.doi.org/10.1016/j.lingua.2014.09.013]

Vicente, Luis 2006. Negative short replies in Spanish. *Linguistics in the* Netherlands 23. Amsterdam: John Benjamins. 199–211.

Vigário, Marina (2014) Distinguishing prosodic constituents from recursive domains. Talk presented at the Workshop "The Prosodic Hierarchy in a Typological Perspective", Stockholm University, March 14-15.

Vander Klok Jozina, Michael Wagner & Heather Goad. 2014. The Prosodic Marking of Focus in English, Québec French, and European French. McGill University: Submitted.

Wagner, Michael. 2005. *Prosody and recursion*. Unpublished Ph.D thesis. Cambridge, Massachusetts: MIT.

Wagner, Michael Watson, Duane 2010. Experimental and theoretical advances in prosody: A review. Language and Cognitive Processes 25. 905-945.

Wang, B. & Nuoyi Yang. 2014. A preliminary study of tonal influence on weak syllables in Mandarin. *4th International Symposium on Tonal Aspects of Languages* (TAL-2014). http://www.isca-speech.org/archive.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61. 747-776.

Warren, Paul, Esther Grabe & Francis Nolan. 1995. Prosody, Phonology and Parsing in closure ambiguities. *Language and Cognitive Processes* 10. 457-486.

Watson, Duane & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19. 713–755.

Watson, Duane G., Michael K. Tanenhaus & Christine A. Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* vs. L+ H*. (Methode: visual

world eye-tracking) *Cognitive Science* 32. 1232-1244.

Watt Sheila M. & Wayne S. Murray. 1996. Prosodic form and parsing commitments. *Journal of Psycholinguistic Research* 25. 291-318.

Welby Pauline. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34. 343–371.

Welmers, William E. 1959. Tonemics, morphotonemics, and tonal morphemes. *General Linguistics* 4. 1–9.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price, 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *JASA* 91.3. 1707–1717

Wilkinson, Karina. 1988. Prosodic structure and Lardil phonology. *Linguistic Inquiry* 19. 25-34.

Williams, Edwin. 1976. Underlying tone in Margi and Igbo. *Linguistc Inquiry* 7. 463-484.

Winkler, Susanne. 2015. Ellipsis and information structure. In: Féry, Caroline & Shinichiro Ishihara (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Wong, Wai Yi, Marjorie Chan & Mary E. Beckman. 2005. An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In: Jun, Sun-Ah (ed.). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 271-300.

Xu, Yi. 1998. Fundamental frequency peak delay in Mandarin. *Phonetica* 58. 26–52.

Xu, Yi 1999. Effects of Tone and Focus on the Formation and Alignment of F0 Contours. *Journal of Phonetics* 27. 55-105.

Xu, Yi 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46. 220–251.

Xu, Yi 2009. Timing and coordination in tone and intonation - An articulatory-functional perspective. *Lingua* 119. 906–927.

Xu, Yi & Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33. 159–197.

Yip, Moira. 1980. *The tonal phonology of Chinese*. Unpublished Ph.D. thesis. Cambridge, Massachusetts: MIT.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yoon, T-J., Chavarría, S., Cole, J., & Hasegawa-Johnson, M. (2004). Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. *Interspeech 2004*. 2729-2732.

Yu, Alan C. L. 2007. Understanding near mergers: the case of morphological tone in Cantonese. *Phonology* 24. 187-214

Zanten, Ellen van, Rob Goedemans & Jos Pacilly. 2003. The status of word stress in Indonesian. In: Weijer, Jeroen van de, Vincent J. van Heuven & Harry van der Hulst (eds.). *The phonological spectrum* 2. Suprasegmental structure. Amsterdam & Philadelphia: Benjamins. 151–175.

Zanten, Ellen van & Vincent J. van Heuven. 1998. Word stress in Indonesian: its communicative relevance. *Bijdragen tot de Taal-, Land- en Volkenkunde* 154. 129–149.

Zanten, Ellen van, Ruben Stoel & Bert Remijsen. 2010. Stress types in Austronesian languages. In: Hulst, Harry van der, Rob Goedemans & Ellen van Zanten (eds.). *A survey of word accentual patterns in the languages of the world*. Berlin & New York: De Gruyter. 87-112

Zerbian, Sabine. 2006. *Expression of Information Structure in the Bantu Language Northern Sotho*. PhD dissertation, Humboldt University of Berlin.

Zerbian, Sabine. 2007a. Phonological Phrasing in Northern Sotho. *The Linguistic Review* 24. 233-262.

Zerbian, Sabine. 2007b. The Subject/Object-Asymmetry in Northern Sotho. In: Schwabe, K. & S. Winkler (eds.). *Information Structure and the Architecture of Grammar: A Typological Perspective.* Amsterdam: John Benjamins. 323–345.

Zimmermann, Malte. 2008. Contrastive focus and emphasis. *Acta Linguistica Hungarica* 55. 347–360.

Zubizaretta, Maria Luisa. 1998. *Prosody, Focus and Word Order*. Cambridge, Massachusetts: The MIT Press.

Glossary

Accent: Realized prominence at the level of the word, the prosodic phrase or the intonation phrase. It is to be distinguished from stress, which is an abstract property.

Alignment: In Optimality Theory, alignment requires that the left or right edge of some constituent coincide with the right or left edge of some other constituent.
Alternative meaning: Alignment also characterizes the relative timing of events in the F0 contour relative to the segmental string and the syllable.

Amphibrach: A foot of the form weak-strong-weak.

Ambisyllabicity: An ambisyllabic segment belongs to two syllables: It is typically coda of one syllable and onset of the following one.

Anapest: A foot of the form weak-weak-strong.

Appendix: Additional segments following the core syllable, in English always coronal obstruents.

Association: In autosegmental phonology, association refers to the abstract phonological property of belonging together in some way. It is visually displayed by so-called association lines.

Audience design: A psycholinguistic notion to the effect that speakers prosodically disambiguate structures for the hearer's (audience's sake).

Autosegmental-metrical model: Another name for the tone-sequence model, coined by Ladd (1996).

Bottom line: The lowest value of a speaker's register.

Boundary tone: A tone associated with the edge of a prosodic constituent.

British school: A school in intonation research characterized by a holistic approach of nuclear tones, and a tool for the learning of English intonation by foreign students.

Catathesis: see downstep.

Categorical change: A change that causes a different category in phonology, phonetics, or another part of grammar.

Categorical sentence: Categorical sentences consist of a topic and a comment.

Coda: The part of the syllable's rime that consists of the segments following the syllable's nucleus.

Comment: In a sentence with a topic, the remainder of the sentence is the comment, the counterpart of the topic, which also contains the focus of the sentence.

Compositionality: Tones assigned to syllables, feet, prosodic words, prosodic phrases and intonation phrases all contribute compositionally to the final melody of a sentence. Besides the tones themselves, interactions between tones also play a role.

Compound Stress Rule: The Compound Stress Rule (CSR), from Chomsky & Halle (1968) assigns stress to the first element of a binary compound.

Contour tone: A tonal unit (i.e. a lexical tone) that consists of more than one level tone, and that is realized as a transition from high to low or low to high, or even as more complex tonal movements.

Core syllable: Syllable containing maximally a certain number of segments, moras and positions.

Culminativity: The principle of culminativity requires all content words to have a stressed syllable and all prosodic constituents to have a unique head.

Dactylus: A foot of the form strong-weak-weak.

Declination: The continuous decline of F0 in the course of a higher-level prosodic constituent, mostly understood as a physiological process. It is the result of decreasing subglottal pressure, causing a concomitant gradual fall in F0.

Delayed Peak: A high pitch accent is realized later than the stressed syllable, for instance on the following syllable.

Direct reference hypothesis: In the direct approach to intonation, pitch accents and boundary tones are directly mapped to syntactic or semantic constituents and/or roles.
Directionality: Feet are formed starting from the beginning or the end of the prosodic domain.
Dominance: Feet are first or end headed and the head determines the strong side. In binary feet, this parameter distinguishes first-headed trochees and end-headed iambs. The dominant side is labelled *s* and the recessive side *w*.
Downdrift (or automatic downstep): In a sequence of two H tones, the second one is lower than the first one when they are separated by a low tone.
Downstep (or catathesis): In a sequence of two H tones, the second one is lower than the first one.
Downtrends: Downtrends are the sum of the decreasing F0 tendencies over the course of an utterance: declination, downstep and downdrift.
Duration: Duration refers to the temporal organization of tones, segments, syllables, words, and sentences. It also includes small pauses between chunks of discourse.
Early peak: A high pitch accent is realized earlier than the stressed syllable, for instance on the preceding syllable.
Edge-based model: In this group of approaches to the syntax-phonology interface, syntactic constituents are mapped to prosodic constituents at one edge only.
Excursion size: See *span*.
Extrametricality: Feet assignment sometimes behaves as if a word-final segment, syllable, or even foot were not there, they are extrametrical.
F0 (fundamental frequency): F0 is the raw substance of the voice, measured in Hertz (Hz), produced by the vibration of the vocal cords. In the study of speech, it is called pitch. The course of F0 is perceived as pitch, the melody of speech.
Faithfulness constraint: In Optimality Theory, faithfulness constraints require identity between input and output.
Final lengthening: Final lengthening denotes the slowing down of speech at the end of prosodic constituents. Prosodic constituents typically take more time in final positions.
Final Lowering: The last high tone in a series of downstepped high tones is lower than predicted by the algorithm calculating the non-final downstepped values.
Focus: Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions.
Foot: The foot is a metrical unit that plays a crucial role for stress assignment. In the prosodic hierarchy, it is one of the 'rhythmic' categories. It is located between the syllable and the prosodic word. Feet also play a major role in morphological processes.
Garden-path sentences: In such sentences, comprehenders wrongly assume a certain prominent or usual syntactic structure, but are forced to revise their initial parse on the basis of incoming words.
Geminate consonant: A long consonant, taking two positions at the syllable's level, and usually belonging to two different syllables.
Givenness: A given constituent is entailed by the preceding discourse, or it is salient in the context.
Global ambiguity: Global ambiguities refer to ambiguities that are not resolved at the end of a sentence.
Gradient change: A change in the phonetic dimension fails to cause a different category in phonology, phonetics, or somewhere else.
Head: In some languages, each prosodic constituent has a strong element, called the head. As an example, the head determines the strong side of a foot.
Alternative meaning from the British School: In a melodic sequence, the stretch of speech containing the nucleus.

Homomorphism: Homomorphism is a structure-preserving map between two categories that does not assume a one-to-one correspondence. Some information contained in one category may be lost in the other.
Iamb: A foot of the form weak-strong.
Implicit prosody: Implicit prosody predicts that in silent reading, a default prosodic contour is projected onto the stimulus that may influence syntactic ambiguity resolution.
Indirect reference hypothesis: In the indirect approach to intonation, phonological rules do not access syntactic structure directly but rather through the intermediary of prosodic domain formation.
Information structure: On the one hand the extra-linguistic, cognitive or mental states of referents, actions, locations and temporality, and on the other hand the formal and communicative aspects of language for the expression of information structural roles, especially topic, focus and givenness.
Intensity: Intensity or amplitude denotes the loudness of speech.
Intermediate Phrase: The intermediate phrase is assumed by Beckman & Pierrehumbert (1986) as a tonal domain below the intonation phrase (ι-phrase). It is also used in ToBI.
Interpolation: The melodic contour between two syllables specified for tones smoothly goes from the first tonal value to the second one, across all intermediate syllables that are unspecified for tone, i.e. not associated with any tone.
Intonation languages: These languages are characterized by a large variation in tonal contours according to pragmatic meanings. Melodic contours are assigned at the higher-level prosodic domains.
Intonation phrase (ι-phrase): A higher-level prosodic domain more or less equivalent to a clause, although it can also be smaller.
Isomorphism: Isomorphism in syntax-prosody interface denotes a one-to-one bijective correspondence between syntax and prosody.
Junctural accent: In Japanese binary compounds, there is only one accent for the entire compound, the so-called 'junctural accent' often located on the first syllable of the second word.
Juncture: Phonetic correlates of a break between morphemes, words or prosodic constituents.
Leading tone: In the tone-sequence model, a leading tone precedes a starred tone. The sequence of these two tones defines a bitonal pitch accent.
Left/right-branching: The direction of branching in syntax is important for assignment of heads in prosody. Left-branching syntax calls for initial heads, and right-branching syntax for final ones.
Level tone: Level tones have only one value, high or low, sometimes also mid. In languages with only level tones, contours arise by concatenation.
Lexical stress: Stress assigned to prosodic words. Lexical stress is an abstract property of words.
Lexical tone: Tones associated with or assigned to ω-words are called lexical tones.
Liaison: Characterizes the process of pronouncing an otherwise mute final consonant when a following vowel follows, so that the consonant becomes the onset of the next syllable (*enchaînement*).
Local ambiguity: Local ambiguities refer to ambiguities that are only local in the sentence and that are resolved at the end of a sentence.
Markedness constraint: In Optimality Theory, markedness constraints require unmarked patterns.
Match: In Optimality Theory, Match constraints require that syntactic and prosodic constituents correspond one-to-one to each other.
Metrical Grid: Represents relative prominence in prosodic constituents with the help of columns of beats (written ×).

Metrical Tree: Represents relative prominence in prosodic constituents with the help of binary trees labelled *strong* and *weak*.
Microprosody: Some individual speech sounds, like voiceless and voiced obstruents, have intrinsic properties that result in perturbations or distortions of the acoustic signal.
Modal phonation: Plain phonation, without creakiness, aspiration or breathiness.
Mora: Constituent of the syllable. In quantity-sensitive languages, every segment from the rime on contributes with one mora to syllable weight, or only segments contributing to the nucleus do. Moras determine stress: heavy syllables attract stress.
Non-modal phonation: Phonation with creakiness or breathiness, as for instance in Vietnamese.
Nuclear stress: The nuclear stress is the strongest position in its domain, prosodic phrase or intonation phrase.
Nuclear stress rule: The Nuclear Stress Rule (NSR, Chomsky & Halle 1968) assigns stress to the last full vowel of a domain.
Nucleus: Syllable peak, usually a vowel.
Offline methods: Measure language processing in delayed time. Examples of offline methods are naming tasks, grammaticality judgments and forced decisions.
Online methods: Measure language processing in real time. Examples of online methods are reading time, brain imaging, ERPs and eye tracking.
Onset: The part of the syllable that comes before the syllable's rime, it consists of one or more consonants.
Onset maximisation principle: In syllable formation algorithms, the still unsyllabified consonants are made part of the onset as far as possible. The remaining consonants are then made part of the coda of the preceding syllable.
Optimality-theoretic model of intonation: Optimality Theory assumes that speech (output) is based on underlying representations (input) and that the relationship between the two is determined by a language-specific ranking of universal constraints.
OT constraint: Universally valid principle of Optimality Theory, expressed in form of a rankable principle.
Paralinguistics: The part of intonation that is not part of grammar, but that is driven by emotion and/or physiology.
PENTA (Parallel ENcoding and Target Approximation): Xu's (2005) model that directly relates the phonetic component of speech to their grammatical functions.
Phrase language: A new type of language that emphasizes the role of phrasal tones in the tonal system of some languages: Sentence intonation is primarily the product of higher-level specifications.
Phrase tone: Tone assigned at the level of a prosodic phrase.
Pitch: Perceived F0 or level of F0 at a certain moment or changing in time.
Pitch accent: An accent realized by pitch on a metrically prominent syllable.
Pitch accent languages: Languages in which tones define lexical distinctions only on stressed syllables.
Praat: A speech analysis software by Boersma & Weeninck, freely downloadable at www.praat.org.
Pre-head: Intonational notion from the British School: the stretch of speech preceding the head.
Primary stress: Most important stress in a prosodic word, a prosodic phrase or an intonation phrase.
Prominence: A syllable, a word, or a larger prosodic constituent is prominent when its role in a sentence is highlighted by phonetic cues, e.g. higher F0, larger intensity, longer duration.
Prosodic hierarchy: A hierarchy of prosodic constituents, from mora to utterance. Lower-level prosodic categories are rhythmic categories and higher-level ones are interface categories.
Prosodic phrase (Φ-phrase): A prosodic constituent which is more or less isomorphic to syntactic phrases, as for instance nominal phrases, prepositional phrases or verbal phrases.

Prosodic word (ω-word): A prosodic constituent more or less isomorphic to a grammatical word.
Proximity/Anti-proximity: Principle stating that "constituents in one grouping are prosodically proximate and those separated by a syntactic boundary are prosodically distant."
Quantity-insensitivity: In quantity-insensitive languages, each syllable has the same status: they do not differ in weight.
Quantity-sensitivity: In quantity-sensitive languages, moras determine whether a syllable is heavy or not.
Raddoppiamento sintattico 'syntactic doubling': an initial consonant is geminated in some Italian dialects.
Range/pitch range: An interval in the F0 used by a particular speaker in general, or in particular sentences. This is the F0 domain that their voice covers. Often means the same as *pitch register*.
Recursivity/recursion: 'The basic operation that allows the generation of a potentially infinite array of discrete expressions out of a finite set of elements.' (Hauser, Chomsky & Fitch 2002)
Reduplication: A morphological process that consists in repeating a grammatical element of the size of a syllable or a foot.
Reference line: The bottom line and the top line of individual speakers' register. The lowest value for a given speaker S is also called the bottom line and the highest value the top line.
Register/pitch register: The interval on the vertical F0 scale used in a stretch of discourse, with H tones fixed to the top of the interval, and L tones fixed to the bottom of the interval. Often means the same as *pitch range*.
Relation-based model: In this model, left or right morpho-syntactic boundaries are first translated into phonological boundaries, and rules apply in the resulting phonological domains.
Rendaku 'sequential voicing': A dissimilation rule of Japanese, which voices the first voiceless obstruent of a non-initial word of a compound except if this word already contains a voiced obstruent.
Reset: Refers to the return of the F0 to a reference line when a prosodic domain of the size of a prosodic phrase or an intonation phrase ends and a new one begins. Reset can be total or partial.
Resyllabification: Process of syllabification across morphemes overwriting the syllabification already assigned.
Rhyme: That part of the syllable that participates in poetic verses' identical rhymes.
Rhythm Rule: Stress shift to avoid a stress clash in two adjacent words.
Rime: The syllable's constituent consisting of the nucleus plus all consonants following the nucleus, thus the coda (and the appendix if present). The rime determines the syllable's weight.
Sandhi (rule): A segmental phonological process at the juncture of two morphemes, or two ω-words, or two even larger constituents.
Secondary stress: A stress or pitch accent that is not the primary stress of its domain.
Semi-syllable: Final segments after a final core syllable that can be felt as an additional (incomplete) syllable, as for instance [r] in *fire*.
Similarity: A principle calling for levelling in terms of their metrical strength constituents that are at the same syntactic level.
Sonority hierarchy: Defines how segments are organized in terms of their (abstract) sonority or loudness: vowels are the most sonorous and obstruents the less sonorous segments.
Sonority sequencing principle: The sonority sequencing principle requires segments not to increase in sonority from the nucleus to the margins (Sievers 1901).
Span/pitch span: Pitch span is used for concrete tonal events, and refers to the difference between two concrete values, a high one and a low one. It is also called excursion size.
Stress: Stress characterizes the abstract property of being prominent, and is best represented on a metrical grid.

Stress-XP: An optimality-theoretic constraint from Truckenbrodt (1995): Each syntactic phrase (XP) must contain a phrasal stress (where 'phrasal stress' stands for the head of a prosodic phrase).
Stress bearing unit (SBU): Syllables and moras can be stress-bearing units.
Stress Subordination Convention: The Stress Subordination Convention (SSC) reduces by one step the stresses already obtained by previous application(s) of the Nuclear Stress Rule (NSR), Chomsky & Halle (1968).
Strict Layer Hypothesis: Principle regulating the organisation of the prosodic constituents in a hierarchy.
Syllable: Prosodic constituent located between the mora and the foot. It consists of a syllable peak and optional margins organized along the sonority sequencing principle.
Tail: A concept from the British School: The stretch of speech before the nucleus.
Text-tone association: Mechanism of how tones are associated with specific syllables in the text to deliver specific phonological tone sequences.
Thetic sentence: A sentence without topic, denoting a single-event.
ToBI (**To**ne and **B**reak **I**ndices): An annotation convention for intonation originally based on Pierrehumbert's (1980) tone-sequence system.
Tone: An abstract phonological (discrete) category, which has a concrete F0 component: it can be a level tone, H or L for instance, or a contour tone, rising (LH) or falling (HL), or even a more complex contour.
Tone alignment: Tone alignment is the timing of a tone in its associated domain, especially in the syllable it is assigned to.
Tone bearing unit (TBU): Syllables and moras are tone-bearing units.
Tone language: Language in which tones define lexical distinctions, not only on stressed syllables.
Tone sandhi: A tonal phonological process at the juncture of two morphemes or two ω-words, or even larger prosodic constituents.
Tone-sequence model of intonation: also called Autosegmental-Metrical model of intonation. It refers to the separation of tones and text to different planes of the phonological representation.
Tone spreading: Tone spreading is a method for assigning a melody to syllables unspecified for tones. The result of tone spreading is that these syllables are tonally specified. Alternatively a tone can also spread to tonally specified syllables and in the process, change the tonal value of these syllables.
Tonogenesis: The emergence or origin of tones.
Top line: The highest value of individual speakers' register.
Topic: A topic is a denotation of a referential expression about which the remainder of the sentence expresses a proposition.
Trailing tone: In the tone-sequence model, a trailing tone may follow a starred tone. The sequence of these two tones defines a bitonal pitch accent.
Trochee: A foot of the form strong-weak.
Universal feet inventory: Singles out syllabic trochees, moraic trochees and iambs as the only possible feet for lexical stress (Hayes 1995).
Upstep: A tonal value is higher than would be expected.
Alternative definition: a tone returns to the top line (the reference line) after a sequence of downstepped tones.
Visual world paradigm: Online method used for the study of language processing based on pictures (Visual World Paradigm) and eye-tracking measurements.
Weight-by-position: The effect of this principle is that a consonant acquires a mora by virtue of being in a coda.
Xiamen Min circle: A series of tone sandhi rules in Xiamen that can be visualized as forming a circle.

Languages
Basque
Beijing Mandarin
Hong Kong Cantonese
Chichewa
Dutch
English
French
Hindi
Italian
Georgian
German
Italian
Tokyo Japanese
Shanghai
Stockholm Swedish
Turkish
Xiamen