Snezhina Dimitrova

Prosody in L2.
Bulgarian-Accented English

Snezhina Dimitrova

Prosody in L2. Bulgarian-Accented English

POLIS

Sofia • 2022

ISBN 978-954-796-096-1

Acknowledgments

This research was supported by the research project “Prosodic aspects of Bulgarian in comparison with other languages with lexical stress”, funded by the Bulgarian National Science Fund, project № Кп-06-Н40/11 – 12.12.2019.

Contents

1. Introduction

1.1 What is prosody?

The research presented in this book focuses on the role of prosody in learning a foreign language (henceforth L2), therefore it is worth dedicating the first couple of pages to an outline of the different views about what prosody is (or isn't). The Oxford Advanced Learner's Dictionary tells us that the first (specialised) meaning of the word refers to the (study of) the patterns of sounds and rhythms in poetry, whereas the second meaning of the word "prosody" indicates the part of phonetics that deals with stress and intonation as opposed to individual speech sounds.

The origin of the word can be traced back to the Latin *prosōdia* "accent of a syllable", from Greek *prosōidia* "song set to music". In her book "An Introduction to English Prosody" Couper-Kuhlen (1986) elaborates on the evolution of the term: in ancient Greek, the word προσωδία was originally used to denote the melodic accent of the full words in the language, but later its meaning was extended to also refer to any other features which were not indicated in the segmental string of vowels and consonants, such as vowel length. When the melodic accents of Greek disappeared and were replaced by dynamic accents, the word "prosody" itself changed its meaning yet again, and began to denote any kind of distinction based on dynamic stress. Many centuries later, this association of the term with both length and stress brought about a new meaning of the word, namely, versification.

Since the present study investigates prosody in speech, we adopt the second dictionary meaning of the word – the one related to human speech. In phonetics and phonology, the "prosodic features" of speech (Crystal and Quirk 1964, Crystal 1969), also known as "supra-

segmental features", or just "suprasegmentals" (Lehiste 1970), refer to the variations in fundamental frequency (F0, perceived as pitch), intensity (perceived as loudness), and timing or duration (perceived as length) in a spoken utterance. These variations usually (though not necessarily) extend over parts of the utterance which are longer than a single segment, hence the term "*supra*-segmental" features, that is, features which stretch above, or beyond, individual sounds. The latter term is often preferred by linguists working in the American structuralist tradition. However, "suprasegmental" phenomena, if defined in this way, could also be taken to include processes such as assimilation and sound reduction, or vowel harmony, to name but a few. Therefore, the two terms should not be taken to be completely synonymous. In the British tradition, the domain of prosody is the so-called "residue of utterance" – that part of an utterance which remains when all the segmental (vocalic and consonantal), non-linguistic (e.g., coughs and sneezes) and paralinguistic features (e.g., laughs or sobs) have been removed (Crystal 1969). Thus, the study of prosody would need to include at least the following characteristics of spoken language: loudness as a component of stress, duration as a component of rhythm and tempo, often including the study of pause distribution as well, and pitch as a component of intonation (Couper-Kuhlen 1986). Matching the multidimensional nature of human speech in general, these characteristics are also usually represented as comprising an articulatory, acoustic and auditory (perceptual) dimension (Couper-Kuhlen 1986, Grice and Baumann 2007), as shown in Table 1.1.

The periodic vibrations produced by the vocal folds when they are set in motion by the pulmonic egressive airstream coming from the lungs are responsible for the fundamental frequency (F0) in speech. Faster vibrations give rise to higher frequency, which we perceive as higher pitch, and vice versa, slower vibrations result in lower F0 which we hear as lower pitch. Fundamental frequency is measured in Hertz (Hz).

Table 1.1. The multidimensional nature of speech prosody

Articulation	**Acoustics**	**Perception**
vocal fold vibration	fundamental frequency – F0 (Hz)	pitch (high / low)
articulatory force	amplitude / intensity – I (dB)	loudness (more / less salient)
timing of articulatory gestures	length (msec)	duration (long / short)

Amplitude as a physical attribute of speech sounds depends on the amount of force or physical effort expended on the articulation of a sound or a syllable. It is related to intensity and is measured in decibels (dB), and we perceive it as loudness: bigger amplitude results in bigger intensity and the resulting sound or syllable is heard as louder, more salient than the neighbouring sounds or syllables.

Finally, all speech takes place in time. The length of time during which the articulation of a stretch of utterance takes place is perceived as its duration, which can range from several milliseconds (msec) for individual vowels and consonants to several hundred milliseconds for longer syllables, or to several seconds for longer utterances.

In this investigation, the term prosody is adopted for the study of (at least) the following aspects of human speech: stress, rhythm and tempo, and intonation. However, some of the terms just mentioned also need to be explained in more detail.

1.2 What is stress?

Stress is a notoriously difficult term to define, and scholarly definitions of the term have usually taken either an articulatory or an auditory approach to its explanation. Attempts at defining stress in articulatory terms have emphasised the greater amount of articulatory effort, or force (Jones 1976) on the part of the speaker, or the bigger muscular tension of the vocal organs involved in the articulation. Other interpretations have focused on the auditory dimension, defining stress as greater loudness (Trager and Smith 1957). For the non-linguist, a stressed syllable or word will normally be more salient or prominent, standing out from amongst its neighbouring syllables or words. But research on the acoustic correlates of such salience has unequivocally shown that perceptual prominence is usually the result of a complex interplay between at least two, and more often than not – all three of the prosodic dimensions mentioned above, so that the existing interrelation between them makes defining (and studying) linguistic stress especially challenging. In addition, segmental cues to stress, such as the aspiration of a plosive consonant in syllable-initial position in English, or the non-central nature of the vowel, can be powerful additional cues to stress which, in combination with the three prosodic parameters, further enhance prominence perception.

From a phonological point of view, stress also poses a number of challenging questions. We will briefly address two of them here, namely (i) how many degrees, or levels, of stress are distinctive, and (ii) how the position of stress is determined in words and longer stretches of utterance.

As an answer to the first of these questions, American linguists working within the structuralist tradition (e.g., Trager and Smith 1957) proposed four stress phonemes to correspond to four degrees of stress in English, namely,

primary (’),
secondary (ˆ),
tertiary (ˋ),
weak (ˇ)

For example, the first syllable of the word “entertain” will have secondary stress, the second syllable – weak stress, while the primary lexical stress will be on the final syllable:

ˆ en ˇ ter ’ tain

A pair of words such as “animate” (verb) and “animate” (adjective) will both have primary stress on the first syllable, but will be distinguished by the presence of tertiary stress (as well as full vowel quality) on the final syllable of the verb, and weak stress (and a weak vowel – schwa) on the final syllable of the adjective.

On the other hand, the majority of English pronunciation dictionaries published in the UK usually show only three degrees of stress:

primary (ˈ),
secondary (ˌ)
unstressed (which is left unmarked)

The main argument against the adoption of tertiary stress, according to this view, is the difficulty with which it is distinguished in perception. For example, the verb “entertain” will have its primary stress on the final syllable and secondary stress on the initial syllable marked as follows: ˌ enter ˈ tain or, to use the proper symbols for phonemic transcription,

/ˌentəˈteɪn/

With respect to lexical stress, in this work we will follow the British tradition of recognising primary-stressed and unstressed (or rather, weakly stressed) syllables in the citation forms of shorter words, and in addition – secondary-stressed syllables in longer words (see the example above).

1.3 What is speech tempo?

The tempo of speech is the number of speech units which are pronounced within a certain amount of time. First of all, it is important to decide what are the speech units in question. Secondly, the time span within which we are going to count their occurrence must be determined. Words emerge as candidates for such units. However, it is a well-known fact that word length, and hence their duration, can vary considerably both across languages, as well as in the same language, depending on the topic of conversation, style, individual differences between speakers, etc. For example, in scientific discourse words will usually be longer than in a friendly chat.

An alternative unit to the word for measuring the tempo of speech is the syllable, but again some languages have more complex syllable structure types (e.g., English) than others (e.g., Bulgarian) in which simpler syllable structures predominate. Compare the Bulgarian word "*сила*" /ˈsi.la/ with the English one 'strength"/streŋθ/. The Bulgarian word consists of four phonemes which are divided into two syllables, both consisting of a consonant followed by a vowel. This is a simple syllable structure, especially when compared to the English single-syllable word which comprises three consonants before and two consonants after the vowel. A sentence in which words with short, simple syllables predominate, will be heard as having a faster tempo than a sentence comprising many words like "strength". This has given rise to the very common observation that "some languages are spoken more quickly than others" (Roach 1998).

But does faster tempo mean that more information is being exchanged? Coupé et al. (2019) who studied a sample of 17 languages from 9 different language families spread across Europe and Asia have shown that there is an interesting interplay between the language-specific structural properties and the speakers' speech rate: faster speech rate was observed in languages with simpler syllable structures and consequent smaller amount of information encoded

per syllable, whereas slower speech rates correlated with complex syllable structure and more information per syllable. As a result, all languages tended to gravitate around the same information rate – about 39 bits/sec (Coupé et al. 2019).

Phoneticians interested in studying speech tempo, however, use two measures based on syllable occurrence per second, namely, speaking rate (SR) and articulation rate (AR). Speaking /speech rate and articulation rate are both defined as the number of output units (i.e., syllables) per unit of time (usually 1 second). The difference between the two is that whereas speaking rate includes pause intervals, articulation rate does not (Jacewicz et al. 2009).

1.4 What is speech rhythm?

Rhythm is yet another prosodic phenomenon in speech which is rather difficult to define, and therefore many different definitions of the term exist. Almost all of them agree that the common feature which characterizes all kinds of rhythm is the patterning of certain speech events and their occurrence at (quasi-)regular time intervals. But scholars are divided on the issue of whether all human speech is rhythmic. Some researchers think that in spite of the occurrence of various kinds of false starts, hesitations and pauses, especially in spontaneous, unprepared speech, everything that we say is ultimately underlyingly characterized by rhythm. Others share the opinion that only some types of speech are truly rhythmic, the most obvious example being the recitation of a poem. But the big majority of authors seem to believe that all human speech is rhythmical, though maybe to different degrees.

There is more unanimity amongst researchers on the question of what constitutes the basic unit(s) of rhythmic organization in speech. The sounds of language form patterns which are called syllables, so the syllable is one possible option. Stress has been shown to affect the sound shapes of individual syllables in the flow of speech,

thus organizing them into higher order patterns called feet. Feet can also recur regularly. The recognition of the syllable and the foot as basic units of speech rhythm lies at the core of the most popular theory of speech rhythm (Pike 1945, Abercrombie 1967) which claims that all the world's languages can be divided into two groups according to the type of rhythmic organization in them. In languages like French or Spanish, the basic rhythm unit is the syllable. Speakers of such languages are said to take approximately the same amount of time in order to pronounce each syllable, and the resulting even, staccato rhythm is called syllable-timed rhythm. In languages like English and Arabic, syllables are organized by stress into feet. Each foot begins with a stressed syllable and includes all unstressed syllables that follow it, up to (but not including) the next stressed syllable. According to the theory, native speakers of such languages tend to make the intervals between consecutive stressed syllables roughly equal, and these languages are said to have stress-timed rhythm. This theory also maintains that the two types of rhythmic organization are mutually exclusive, that is, every language in the world is spoken with either stress-timed or syllable-timed rhythm. This claim has been debated for many years. The results from many experimental investigations suggest that it would be more appropriate to consider speech rhythm not in binary, but rather in scalar terms (Dauer 1983, 1987). On a scale of rhythm, English will have a place near the "stress-timed" end, French will be situated near the "syllable-timed" end, while Bulgarian may well occupy a position somewhere between them (Dimitrova 1998).

1.5 What is intonation?

Intonation is most often described as "the melody of speech", the changing pitch of the voice which can convey additional meaning or alter the meaning of what has been said. This understanding of the term seems to be especially popular amongst non-linguists.

A broader definition of "intonation" widely used in phonetics and phonology usually also includes variations in timing, loudness and sometimes voice quality as well, thus making it (almost) synonymous with the term "prosody". The present investigation adopts this second, much broader understanding of intonation. If, on the other hand, the term "melody" is used in it, it will refer specifically to the modulation of the fundamental frequency in speech.

An often-quoted definition of intonation in the research literature is the one given by Ladd: "Intonation … refers to the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings in a linguistically structured way" (Ladd 1996, p. 4). The three key points in Ladd's definition concern:

(i) the term "suprasegmental", which is used to refer to features of fundamental frequency, intensity and duration;

(ii) the fact that intonation conveys meanings at the level of the phrase or the utterance as a whole;

(iii) intonation features are grouped into categorically distinct entities and relations, and exclude any paralinguistic features.

Intonation can be used to signal a wide range of meanings. The division of speech into intonation units, each of which is characterized by a separate, complete intonation pattern, and sometimes also followed by a brief pause, can signal grammatical structure. Nolan (2006) calls intonation the "punctuation" of spoken language: it divides an utterance into chunks which often correspond to grammatical units, thus making it easier for the listener to comprehend the spoken message. For example, the difference between restrictive and non-restrictive relative clauses in English speech will normally be signaled by intonation: the non-restrictive clause in (1) below will be separated by pauses and pronounced with a complete intonation contour of its own, whereas in (2) there will be no pauses be-

fore (and probably also after) the restrictive relative clause, and its melody will be part of the preceding intonation contour:

(1) The passengers, who arrived early, boarded the train.

(2) The passengers who arrived early boarded the train.

The end of a major intonation contour usually coincides with the end of a grammatical constituent such as a short sentence or clause. But longer constituents such as a noun phrase comprising several words may be signaled by a separate intonation contour. For example, the statement (3) is likely to be pronounced with two separate intonation contours (separated by | below), a rise on the subject, followed by a fall on the predicate:

(3) The beautiful blond girl | was waving to us.

However, a speaker who uses faster speech tempo may well pronounce the whole sentence in (3) with a single falling intonation pattern. In other words, although intonation contours tend to coincide with units longer than lexical words, there is no fixed one to one correspondence between them and grammatical units in the flow of speech.

Intonation signals information structure by highlighting the important information in an utterance, and by de-accenting what is known, or old information. For example, in B's answer in (4) below, the word "wants" most probably will be highlighted, whereas all the words following it will be unaccented:

(4) A: Will Jenny be interested in a role in the school play?
B: She wants to take part in the play.

Intonation can help to distinguish between a statement and a question if the question has not been grammatically marked in any other way. For example, if sentence (5) is spoken with falling intonation, it will be interpreted as a statement, but if said with rising intonation, it will be heard as a question:

(5) They are leaving tomorrow.

Intonation can also signal the speaker's attitudes or emotional states – from friendliness, interest or surprise, to anger, shock or hostility. This so-called "attitudinal function" of intonation is primarily attributable to the kind of melody speakers choose to use in an utterance. Thus, if speaker B uses flat intonation in (6), this will most probably be interpreted by A as boredom and lack of interest, but if said with an extensive fall from high to low pitch, B's comment will indicate involvement and enthusiasm.

(6) A: Did you like the film last night?
B: It was good.

Intonation can also be used to regulate conversational behaviour. In English, if the tag question in (7) is said with falling intonation, it will be taken to express the speaker's certainty, but rising intonation on the tag will signal to the listener that the speaker is less certain and expects the listener to express an opinion:

(7) John's new car is red, isn't it?

In conversation, speakers can use intonation to also signal whether they have come to the end of what they want to say, or whether they want to go on talking without being interrupted.

Last but not least, being part of a person's accent, intonation is part of every speaker's personal or social identity.

Having outlined the importance of prosody in speech, as well as some of the key terms and concepts used for its description and scientific study, we will next discuss the role and place of prosody in L2 speech acquisition theories and research in Chapter 2. Special attention will be paid to the recently introduced L2 Intonation Learning theory, and the results from several studies of L2 prosody inspired by it will be considered.

Chapter 3 presents some of the most popular descriptions of English intonation – the British School of intonation analysis, Discourse Intonation analysis, the "levels" approach of the American structuralists, Bolinger's Pitch Accent theory, and the Autosegmental-Metrical model of intonational phonology, along with the ToBI analytical framework.

The latter approach is further considered in Chapter 4, in which it is used as a common methodological framework for a description of the intonation of American English, RP and Contemporary Standard Bulgarian.

Chapter 5 presents a comparison of the prosody of English and Bulgarian in the ToBI framework, along with some predictions about the potential difficulties which are likely to be experienced by Bulgarians who acquire the pronunciation of English.

Finally, Chapter 6 describes some recent experimental work conducted within the Autosegmental-Metrical approach to intonation analysis, and Chapter 7 presents an outlook on future research as well as the teaching of the prosody of Bulgarian-accented English.

2. Prosody in L2

2.1 The place of prosody in theories of Second Language Acquisition (SLA)

It is a well-known fact that L2 learners often experience difficulties with the perception and the production of certain aspects of the sound system of the foreign language. There are a number of theories which have been developed in order to try to account for these difficulties. A notable characteristic feature of the big majority of these theories is that they tend to focus almost exclusively on the acquisition of the segmental sounds of the target language.

The suprasegmental, or prosodic, characteristics of L2 speech have for a long time been ignored by educators and researchers alike. L2 teachers have tended to focus on the vowels and the consonants of the foreign language, on the assumption that mastering the individual sounds is crucial, if not sufficient, for efficient communication in the L2. Researchers have also long ignored L2 prosody, not least because of the lack of consistent methodology for comparing the prosodic features of two or more languages, and for making predictions about learner problems in the acquisition of the suprasegmental features of the target language. Some of the most popular L2 learning models, such as the Speech Learning Model (Flege 1995, 1997, 2007), the Native Language Magnet model (Kuhl 1991, 1992, 2000), and the Perceptual Assimilation Model (Best 1995, Best and Tyler 2007) focus almost exclusively on the segmental level. Most of the early predictions made by the original models have been based on research carried out with data from learners who acquire the foreign language in a predominantly L2 environment where the target language is the official language used in daily interaction. Later

studies (e.g., Piske 2007, Tyler 2019) have also tried to explore the models' implications for L2 students learning the language through formal instruction in the foreign language classroom in the environment of the learners' mother tongue.

Some of the most popular models and their significance for the teaching and learning of the suprasegmental characteristics of a second / foreign language are discussed below. No distinction will be made between "second language (L2)" and "foreign language" (FL), in particular "English as a foreign language (EFL)", and the terms will be used synonymously in the following brief outline of theories which have dominated the field of second language acquisition (SLA) and the discussion of the place of prosody in them.

The Contrastive Analysis (CA) hypothesis (Lado 1957) claims that the patterns which are likely to cause difficulty in learning a second or foreign language can be reliably predicted and described by comparing the target (L2) language system with the system of the learner's native language (L1). The main prediction is that those elements in the linguistic system of the target language which are similar to elements in the learner's mother tongue will be easy to acquire, while elements which are different will be difficult to learn. In the original version of Contrastive Analysis, the greatest pronunciation challenge to the learner was predicted to be presented by cases in which two speech sounds which were allophones in the L1 had to be re-assigned to two distinct phonemes in the L2. Contrastive Analysis attributed all learners' errors in the L2 to interference from the L1.

In an extension of CA, Stockwell and Bowen (1965) posited eight degrees of difficulty in the acquisition of the sound system of the foreign language, depending on the status of the target sounds in the L1, that is, whether they were separate phonemes, allophones of the same phoneme, or entirely absent from the phonological system of the learner's native language. Of these, the acquisition of L2

allophones which were absent from the L1 sound system was predicted to pose the greatest challenge of all.

Empirical observations to support such predictions can be drawn from the practice of teaching segmental features of English pronunciation to Bulgarians. For example, the acquisition of the aspirated allophones of the voiceless plosives /p t k/ is a well-known problem, even at a fairly advanced level. However, research has also demonstrated that certain errors predicted by Contrastive Analysis are equally well explainable in other ways: for example, learners of Swedish from diverse L1 backgrounds have been shown by Johansson (1973) to experience problems with the acquisition of the same phonemes which Swedish children have been shown to acquire later in childhood: such difficulties can be accounted for in terms of ease of articulation. Research has also demonstrated that there are numerous errors which, although predicted by the Contrastive Analysis hypothesis, have never been attested in the language of the foreign learners.

In spite of the limitations of Contrastive Analysis, it has palyed, and still continues to play, an important role in second language teaching and learning, since it makes it possible to predict and explain in a rather straightforward way some of the main problems of L2 learners, including pronunciation problems at the segmental level. Where speech prosody is concerned, however, another major problem emerges which has prevented the efficient application of Contrastive Analysis to the teaching and learning of L2 prosody, namely, the unavailability of methodologically comparable decsriptions of the prosody of the mother tongue and the target language. The lack of a universally agreed framework for the description of language prosody precludes the application of Contrastive Analysis to an analysis of the similarities and differences between the learner's L1 and L2, and a prediction of learner difficulties in the field of L2 stress, rhythm and intonation.

The Markedness Differential hypothesis (Eckman 1977) puts forward the idea that on the basis of the markedness relations in universal grammar, a systematic comparison of L1 and L2 can predict the areas of difficulty that a language learner is likely to experience. In phonetics and phonology, this means that unmarked sounds like /p/, /d/, /a/, etc. which can be found in most of the world's languages, will be relatively easy for the foreign learner to acquire, whereas marked sounds such as /θ/ and /ʒ/ will be far more difficult to learn.

The Interlanguage hypothesis first introduced the idea that, while learning a foreign language, learners create an intermediate language variant called "interlanguage" which is "a separate linguistic system based on the observable output which results from a learner's attempted production of a target language (TL) norm" (Selinker 1972). A learner's interlanguage is created through L1 transfer and L2 input. It may preserve some features of the mother tongue, and may at the same time overgeneralize some of the rules of the foreign language. Interlanguage theory changed attitudes towards learners' errors, viewing them as dynamic features which may be absent from both the L1 and the L2 of the speaker. Fossilization is the final stage of interlanguage, and is reached when a given linguistic form stops evolving and "freezes", irrespective of whether or not it is correct. Investigations into learners' interlanguage characteristics, however, have shown that some interlanguage rules, including phonological rules, may result from neither L1 transfer nor L2 input.

The problem with both the Markedness Differential and the Interlanguage hypotheses was the same as the one mentioned in connection with Contrastive Analysis, namely, the lack of comparable descriptions of the prosodic systems of the L1 and L2 which would allow analyses and predictions of the difficulties which L2 learners are likely to be faced with. The above SLA hypotheses were suc-

ceeded in the last decade of the 20th century by several influential models of L2 speech production and perception, of which we will outline briefly the Native Language Magnet Theory, the Perceptual Assimilation Model, the Speech Learning Model, the Perceptual Assimilation Model-L2 and the Revised Speech Learning Model.

The Native Language Magnet model (NLM) (Kuhl 1991, 1992, 2000) focuses on early speech perception, suggesting that infants categorize the speech sounds of the mother tongue by creating in their brains a "sound map" – a complex network, or filter, which may then interfere with the acquisition of the phonemes of an L2. The prototype sounds in the sound map then act like magnets and tend to attract similar sounds, so that "initial learning" (of the L1 sounds) "can alter future learning" (that of the L2 sounds) (Kuhl 2000, p. 11855).

The Perceptual Assimilation Model of L2 speech learning (PAM/PAM-2) was put forward to deal primarily with non-native speech perception in the L2 and how it is shaped by the L1 sound system of the learner (Best 1995, Best and Tyler 2007). It claims that a learner's success at acquiring phonological contrasts which exist in the L2 but are absent in the L1 ultimately depends on the way in which the L2 phonemes have assimilated to the phonological system of the learner's L1. In other words, the system of the mother tongue influences the perception of the foreign language, and unless phonetic differences which signal phonological contrast in the L2 have been assimilated so as to preserve the dissimilarity, perceptual learning is required in order for the learner to acquire the new L2 phonological category.

In its current version, the Perceptual Assimilation Model puts forward six possible patterns of assimilation or association between L2 speech sounds and L1 phonological categories. For example, it predicts that if two L2 sounds which are contrastive in the foreign language are judged to be equally good (or bad) matches of a single

L1 phonological category, then it will be very hard to learn to discriminate between them. On the other hand, if two contrastive L2 sounds are assimilated to two different L1 phonological categories, it will be much easier for the learner to discriminate between them. If the two contrastive L2 sounds are matched to the same L1 category, but one of them is a better fit than the other, or if two L2 sounds cannot be matched to any L1 type in spite of the fact that both fall within the phonological space of the mother tongue, then perception will depend on the degree of difference between the two L2 sounds. In other words, PAM presents a much more complex account of the way in which the sounds of the foreign language are perceived.

The Speech Learning Model (SLM) claims that the accuracy with which L2 segments are perceived is a major determinant of the accuracy with which those segments will subsequently be produced by the foreign learner. Flege (1997) maintains that an L2 sound will be classified as identical with an L1 phoneme if it meets three criteria, namely, the two are transcribed with the same IPA symbol, their acoustic characteristics are similar, and they are perceptually similar as well. An L2 sound will be classified as similar to an L1 phoneme if the two are represented by the same transcription symbol but they are acoustically and perceptually different. An L2 sound will be classified as new if it is different from the L1 phoneme in terms of all three characteristics mentioned above. Because the sound categories of L1 and L2 co-exist in a common phonetic space, in the first of the above cases the learner will classify the new L2 sound within an already existing L1 category without any need to modify it. In the second case, the learner will modify an already existing category by adding the L2 sound to it. In the last case, the learner will set a new category for the new L2 sound.

The Speech Learning Model also claims that cross-language phonetic interference tends to be bi-directional, affecting L2 speech

production, but also exerting an influence on certain aspects of L1 speech, especially that of learners who have been exposed to the foreign language from an early age. In other words, because the elements of the phonetic systems of L1 and L2 co-exist in a common phonological space, they constantly affect and mutually influence each other. Unlike the "Critical Period Hypothesis" (Lenneberg 1969), the Speech Learning Model maintains that the ability to form new phonetic categories remains active throughout an individual's life span, but in L2 learning it takes time, and correlates with the amount and nature of the input received by the learner. All of the above assumptions and hypotheses of the original SLM model again crucially concern the learning of individual speech sounds (segments).

The Speech Learning Model and the Perceptual Assimilation Model both predict that the most difficult L2 sound to acquire will be the ones which are "similar" because they will be classified in the same L2 category, whereas the "different" L2 sounds will be the easiest ones to learn. These predictions are different from the ones made by theories such as the Markedness Differential Hypothesis. Unlike PAM, SLM does not address the issue of the link between L2 speech production and perception. Best and Tyler (2007) developed an extension of the original model – PAM-L2 which also takes into consideration the role of the learner's knowledge level, as well as the importance of gestures, the influence of L1 and L2 phonetics and phonology and their interaction in the perception of L2 sounds.

The recently published Revised Speech Learning Model (SLM-r, Flege and Bohn 2021) presents eleven key aspects of foreign language speech learning. Some of the most important claims of SLM-r are that L2 learners' experience is different from that of monolingual speakers of the target language, therefore the production and perception of learners and monolinguals will never match; produc-

tion and perception of the L2 sounds "co-evolve" without one of the two taking precedence over the other; L2 category formation is independent of the learner's age of exposure to the L2; learners have access to features of the target language which are not part of the L1 – the so-called "full access" feature hypothesis. Last but not least, SLM-r considers in detail inter-subject (rather than inter-group) phonetic variability, and attempts to offer a more complex account of L2 phonetic and phonological acquisition.

Although the main assumptions of the models briefly reviewed above target almost exclusively the acquisition of the segmental system of the respective L2, they were reviewed because at least some of the predictions which they make could eventually be applied to the learning of suprasegmental features as well.

As a result of the preoccupation with segmental acquisition, both L2 teaching and L2 research have suffered considerably from the lack of a comprehensive theory and model of L2 suprasegmental, or prosodic, learning. Studies of the prosodic difficulties of L2 students which have been carried out roughly up until the 1990s are considerably fewer than those dedicated to the investigation of segmental errors. Besides, the big majority of those studies tend to focus on the problems experienced by L2 learners of English from various L1 backgrounds (see, e.g., Gut 2009, Mennen 2007, 2015 for a review). Many of these studies point to the existence of transfer or interference from the speakers' L1 which plays an important role in the production of L2 intonation patterns. However, the diversity of methodological approaches on which such studies are based makes it difficult to compare and interpret the results reported in them. Mennen (2007, 2015) highlights the importance of having a speech model which accounts for the prosodic and intonational aspects of L2 speech learning by providing a framework for cross-language comparisons and allows for making predictions about the degree of difficulty presented by different aspects of L2

intonation for learners from diverse L1 backgrounds. The Autosegmental-Metrical model of intonational phonology on which Mennen's L2 Language Intonation Learning Theory (LILt) is largely based has been able to fill this big gap.

2.2 The L2 Intonation Learning Theory

Since the last decades of the 20th century, the Autosegmental-Metrical (AM) model of intonational phonology (e.g., Pierrehumbert 1980, Pierrehumbert & Beckman 1988, Ladd 1996/2008) has emerged as a strong contender for a general framework of intonational analysis, as proved by the abundance of contrastive research (including work on L2 prosody) which has been carried out since the start of the millennium.

In the context of SLA, an important consequence of the emergence of the AM model and a major step towards the development of a comprehensive model of L2 prosody acquisition is the L2 Intonation Learning Theory (LILt) (Mennen 2007, 2015). Largely inspired by Ladd's (1996/2008) dimensions of cross-language variation, the theory attempts to offer an extensive account of the most frequently observed prosodic problems experienced by L2 learners, especially those in the area of intonation. Some of the basic tenets of LILt are outlined below.

First of all, in line with the now largely popular Autosegmental-Metrical framework, Mennen's theory also draws an important distinction between phonological representation and phonetic implementation. Mennen hypothesises that L2 learners first acquire the phonological patterns in the foreign language, and only afterwards try to master the phonetic implementations of those patterns. She therefore insists that, due to this, "a perceptually similar error may in fact have different underlying causes, which can be either difficulties with the phonological structure of the L2 or with its phonetic realisation … it is important for teaching purposes to distinguish be-

tween phonological and phonetic errors, so that the source of the problem can be addressed in teaching" (Mennen 2007, p. 71).

Mennen's LILt theory distinguishes four major dimensions along which L2 intonation may deviate. The first of these – the systemic dimension – deals with the inventory of structural prosodic elements and their distribution. The categorical elements can be pitch accents, accentual units of different size (prosodic words, accentual phrases, intonation phrases, etc.) or boundary phenomena, as outlined in Pierrehumbert's (1980) original version of the Autosegmental-Metrical theory. This dimension also involves the ways in which structural elements such as pitch accents combine with one another – for example, what combinations of High (H) and Low (L) pitch targets are admissible in a given language. In addition, it also looks at tune – text association (Ladd 1996, p. 119), that is, the way the tune is mapped onto the segmental string.

The second dimension of the LILt model – the realisational, or phonetic, dimension – is concerned with the phonetic implementation of the categorical elements of the system: this may involve tonal alignment – the actual lining up of the pitch accents with the segmental string of the utterance, tonal scaling (i.e., the relative height of a pitch accent), and the shape, or slope of a pitch accent, e.g., shallow vs. steep rises or falls.

The third dimension in Mennen's LILt model is the semantic one: it deals with the "functionality" of the categories comprising the phonological system, that is, the ways in which the systemic elements are used to signal intonational meanings such as broad vs. narrow focus or interrogativity.

The fourth and final dimension of LILt – the frequency dimension takes into account the "frequency of use", or how often the structural elements are used.

Mennen (2015) also considers the possibility to extend the predictions of some of the segmental SLA models, most notably those

of SLM and PAM-L2, to L2 intonation learning, based on the already accumulated amount of knowledge about L2 speech learning and on the theoretical assumptions of existing segmental models. She therefore proceeds to put forward several hypotheses:

(i) Both SLM and PAM-L2 assume that production problems in L2 are perceptually motivated, since the perception of the segments in the target language is "filtered through" the automatic perceptual strategies of the L1, as a result of which the L2 segments are recognised as instances of L1 categories, and thus L1 interference occurs. Like SLM and PAM-2, LILt hypothesises that learners' problems with the production of L2 intonation will have a perceptual basis, and will be attributable to poor perception of L2 intonational cues which are either not present, or are different in the L1. Besides, "it is more difficult to determine the existence and perception of categories for intonation than it is for segments, because of the close intertwining of gradient and categorical variations in intonational form, each of which convey both linguistic and paralinguistic meaning" (Mennen 2015). Therefore, when trying to predict the relative difficulty of L2 intonation categories for the learner, both their form (that is, the realisational dimension) and their meaning (that is, the semantic dimension) should be taken into account. Finally, although LILt is in agreement with both SLM and PAM-L2 that perception lies at the core of many learner difficulties, it also recognises that sometimes other explanations of the observed deviations are possible, which may have to do with articulation or with acoustic memory storage. This first hypothesis / set of assumptions of LILt regarding the role of perception as predictor of the intonation difficulties of foreign learners remains a largely unexplored area as far as Bulgarian learners of English are concerned.

(ii) Another important assumption made by the SLM and the PAM-L2 which the LILt model also shares concerns the interrelation between the level of phonological contrast and that of phonetic

implementation. LILt explicitly claims that both similarities as well as dissimilarities between the intonation of the mother tongue and the target language can be both systemic and realisational, and that the second kind of (dis)similarities may have serious consequences for the correct discrimination, categorization and production of a L2 contrast. Besides, the contexts and positions in which a contrast occurs should always be taken into account as well.

(iii) Both SLM and PAM-L2 acknowledge the role of age of arrival (AOA) and age of learning (AOL) as important predictors of learners' successful acquisition of the L2 segmental system. LILt hypothesises that the same is valid for intonation. Research on the importance of these factors for acquiring L2 intonation carried out so far, albeit relatively limited in scope, suggests that although it may be true that "the earlier, the better", the influence is not the same for each intonational dimension.

(iv) Another theoretical assumption maintained by the SLM and the PAM-L2 concerns the fact that virtually the same perceptual learning abilities are at the disposal of both children learning L1 or L2 and to adults learning a L2, therefore learners can constantly improve the language-specific phonetic properties of the target language in the course of learning it, and can ultimately approximate, or even reach, the target language norms in their production. LILt claims that there is no principled reason why the process of acquiring L2 intonation should be different; therefore, it can be expected that with experience, the production of L2 intonation will move closer to the L2 norms.

(v) The SLM and the PAM-L2 both maintain that because L1 and L2 categories share a common phonological space, languages interact; moreover, the nature of the interaction is bi-directional: there may be cross-linguistic assimilation which is the result of merging properties of the two languages, so that learners produce segmental sounds whose properties are intermediate between the L1 and

the L2. Mennen (2015) hypothesises that such intermediate values between the L1 and L2 are to be expected due to merging effects in intonation as well. Interaction at the segmental level has also been shown to take the form of dissimilation, or polarization: LILt proposes that the same can also be expected for intonation as well.

From the point of view of the most recently proposed SML-r model, the first of the above hypotheses can in some instances be questioned, since SLM-r claims that an L2 sound can in some cases be accurately produced even if it has not been accurately perceived. The third and fourth hypothesis can similarly be questioned, since SLM-r maintains that younger age does not necessarily contribute to better acquisition, neither can it be expected that L2 learners will ever attain perfect mastery of the target language's system.

In conclusion, the L2 Intonation Learning theory (LILt) proposed by Mennen (2015) has gained popularity and is beginning to be more widely used as a tool for the comparison of the similarities and differences between L1 and L2 intonation, and for the formulation and testing of research hypotheses regarding the difficulties experienced by foreign learners when they acquire the prosody of L2. A number of related questions remain to be addressed, such as

- the extent to which L2 intonation acquisition depends on the acquisition of the segmental system and of other prosodic properties of the L2, such as prosodic length and prosodic structure;
- the role of universal constraints in L2 intonation learning;
- the similarities and differences between learners from different L1 backgrounds, or between learners with the same L1 background who acquire different L2s, etc.

But in spite of the many questions which still seek an answer, LILt remains the most thorough and well-developed theory of L2 intonation learning to date.

2.3 Some recent applications

Some recent studies of L2 intonation which use a LILt-based theoretical approach to the exploration of L2 prosody include Andreeva's (2017) investigation of the prosodic encoding of phrasal prominence and information structure in German and Bulgarian and Pešková's (2020) analysis of the prosody of Czech and German learners of Spanish and Czech learners of Italian. As both studies involve another Germanic language – German, and so the experimental methods and results could inform similar investigations into L2 English prosody, they are discussed briefly below.

On the basis of an extensive review of existing research, and following the dimensions of Ladd's (1996) original model, Andreeva (2017) draws a number of conclusions about the similarities and differences between German and Bulgarian. She lists the following similarities:

- in the systemic dimension: the same tonal inventory;
- in the semantic dimension: given information is not always de-accented, the same focus type can be expressed with different pitch accents;
- in the semantic dimension: broad focus is expressed with early peak accents with a falling onglide;
- in the semantic / realisational dimension: givenness lowers pre-nuclear pitch accents and cancels post-nuclear ones.

The differences between German and Bulgarian which Andreeva notes are:

- in the systemic dimension: given information is de-accented much less frequently in Bulgarian than in German;
- in the realisational dimension: the high target of H* is aligned early in the syllable in Bulgarian and in the middle of the syllable in German;

- in the realisational dimension: contrastive and non-contrastive focus are distinguished in Bulgarian by reducing the prosodic prominence of pre-nuclear accents in contrastive focus, and in German – by late upstepped realisations of the nuclear accents;
- in the semantic dimension: in contrastive focus, Bulgarian uses mainly H*, while German uses mostly L+H*.

In addition, Andreeva (2017) also follows Flege's Speech Learning Model in order to investigate the influence of L1 on the production of German focus prosody by advanced Bulgarian learners of the language. Bulgarian and German use identical pitch accents with falling onglides in broad focus, similar pitch accent types in narrow contrastive focus, similar phonetic realisation of the default H* pitch accent, similar de-accentuation patterns of given information, and different implementation of global and local cues in contrastive vs. non-contrastive focus. On this basis, Andreeva hypothesises that

(i) Bulgarian speakers of German will produce more pitch accents than native German speakers;

(ii) Bulgarian speakers of German will produce fewer L+H* pitch accents than native German speakers;

(iii) Bulgarian speakers of German will transfer the early alignment of the peak of H* from their L1 (Bulgarian) in L2 (German);

(iv) Bulgarian speakers of German will produce the nuclear accent with later alignment and higher scaling in contrastive compared with non-contrastive focus in the L2.

Andreeva's analysis of data from five Bulgarian speakers of German and six native German speakers revealed that although Bulgarians produced fewer pre-nuclear accents in the L2 than in their native

language, the pre-nuclear pitch accents in Bulgarian-accented German were nevertheless considerably more (89% in non-contrastive and 86% in contrastive focus) than those in the readings of the group of native speakers of German (64% and 59%, respectively). This supports hypothesis (i). The second hypothesis was also confirmed by the experimental results, according to which the Bulgarian speakers used fewer L+H* and their distribution depended on position in the sentence (the pitch accent was more often used sentence-initially). No significant difference between native and L2 speakers of German was found with respect to the alignment of the H* peak in broad focus. In narrow focus, native German speakers produced very few H* pitch accents, while Bulgarian L2 speakers were inconsistent and the peak was reached anywhere within the accented syllable. Andreeva explains this finding with the differences in vowel length between the vowel systems of the two languages, and points to "the causal link between the acquisition of the phonetic implementation of intonational categories and the acquisition of segmental phonology" (Andreeva 2017, p. 78). The last hypothesis about the later alignment and higher scaling of the nuclear pitch accent in contrastive focus was likewise confirmed by the experimental data.

Andreeva's (2017) investigation, to our knowledge, is the first systematic application of Ladd's (1996) (and by extension – also of Mennen's) model to the analysis of the prosody of Bulgarian-accented L2 speech.

Pešková (2020) studied the prosody of Czech and German learners of Spanish and Czech learners of Italian. She showed that, in spite of L1 transfer, L2 intonation can be learned, and that L1 transfer cannot account for all the non-native-like patterns and tonal events in L2s. On the basis of her findings, and in line with Mennen's (2015) L2 Intonation Learning theory (LILt), Pešková (2020, 2021) proposes a nine-point Developmental L2 Intonation hypothesis which claims that:

"(1) Phonological features of intonation are acquired earlier than phonetic features of intonation;

(2) Pragmatically unmarked structures are acquired earlier than marked structures;

(3) Patterns with a heavy semantic weight are acquired earlier than patterns with no changes in meaning;

(4) Patterns that exist in both L1 and L2 are acquired earlier than new patterns, provided that they convey the same meaning;

(5) Patterns that do not lead to changes in the semantic dimension fossilize faster;

(6) Patterns that are phonetically similar in the learners' L1 and the target language fossilize faster than phonetically different patterns;

(7) Patterns in functionally weaker positions fossilize faster than patterns in functionally stronger positions;

(8) New but frequent and perceptually prominent patterns tend to be subject to overgeneralization;

(9) Rising boundary tones (being unmarked or "universal" forms) tend to be overgeneralized in all types of questions."

To return to the original LILt model, in addition to the outline of the model, Mennen (2015) also presents an overview and gives many illustrations, mostly from research on the major spoken varieties of English and a range of (mostly European) languages. In her review of L2 intonation studies, she given ample evidence for the existence of deviations attested along all four dimensions, though those in the second one – the realisational dimension – seem to be the most numerous of all, mostly concerning the alignment of pitch accents, their timing and scaling. She also admits that the LILt model is far from being unproblematic. For example, the dimensions can interact with one another, or it may be difficult to uniquely ascribe

a certain instance of L2 intonational deviance to one of the four dimensions.

However, in spite of its limitations, the L2 Intonation Learning theory constitutes the first model aiming to provide a systematic and detailed account of L2 prosodic learning. It will be thus extremely useful to be able to utilize for the first time a model specially developed for characterizing L2 intonation in order to analyse the prosody in the speech of Bulgarian learners of English, albeit with the important caveat that LILt should be treated "as an evolving or 'working' model, which is subject to change when more data are published" (Mennen 2015, p. 17).

3. English intonation models

Analyses of the intonation of the mainstream standard varieties of English have provided a sound basis for the development of influential models of intonation, as well as for the design of materials for the teaching of English intonation to foreign learners. The different traditions on the two sides of the Atlantic and the different goals behind the analyses have given rise to a number of popular descriptions which will be reviewed briefly below.

Descriptions of the intonation of English can be classified in terms of several criteria (for a review, see Lecumberri 1997 and Dimitrov 2017, among others). First, they can be subdivided into phonetic and phonological approaches. Of course, neither a purely phonetic, nor an exclusively phonological account of intonational variation would be fully satisfactory, therefore it is hardly surprising that existing descriptions have incorporated both phonetic and phonological aspects of intonation in varying degrees.

A second possible classification – one into holistic versus compositional accounts – would subdivide current views of English intonation into models which treat the pitch contour as an inseparable whole, as opposed to views which decompose it into a finite number of separate constituents. The holistic view has a long tradition in British English intonation analysis, beginning with Palmer (1922) and continuing into the latter decades of the 20th century in the work of Kingdon (1958), Crystal (1969), Halliday (1967), O'Connor and Arnold (1973), Gimson (1970), Cruttenden (1986) and more recently – of Wells (2006) and Roach (2009). It is also evident in the research of British authors working within the framework of Discourse Analysis, such as Brazil et al. (1980), Bradford (1988) and Brazil (1994). Descriptions within the American Structuralist tradition such as Pike

(1945) and Trager and Smith (1951) also largely adhere to the holistic approach.

Compositional accounts have come to the fore after the publication of Pierrehumbert's pioneering work "The phonology and phonetics of English intonation" (1980). Unlike the holistic view in which the meaning is attributed to the intonational contour as a whole, the compositional view treats the meaning of the tune as the result of the individual contribution of all constituents which comprise it (Pierrehumbert and Hirschberg 1990).

Despite some similarities suggested by the holistic approach to the analysis of intonation contours, descriptions of British English and American English intonation up until the last couple of decades of the 20th century have developed along very different lines. The major distinction between the two comes down to the well-known "levels vs. configurations" debate in intonational analysis. Therefore, they will be discussed separately below.

3.1 The British tradition of intonation analysis

Ever since Kingdon's (1958) and O'Connor and Arnold's (1973) descriptions of English intonation became popular in the field of English language teaching, holistic accounts have dominated the area. In spite of this fact, terminological confusion has also been abundant. To begin with, stress, loudness and prominence have often been used to refer to the same phenomenon. The terms tone unit, intonation group, intonational phrase, etc., have all been used to name the main unit of intonational analysis. The same holds for terms such as pitch contour, tune, and so on. In the following discussion, we will preserve the original terminology as put forward by the respective author, but will provide equivalent terms if and where necessary.

The British School of intonation analysis has always considered pitch contours to be unitary. Armstrong and Ward (1926) are

amongst the first to consider whole sentences as being characterized by 'tunes' as functional units. Since the nineteen-thirties, tunes have usually been divided into several parts: a (optional) "head" – the section preceding the main "sentence stress", the "nucleus" – the syllable which carries the sentence stress, and (optionally) the "tail" – the syllable(s) after the nucleus. Some accounts explain the functions of a tune with reference to the combination of head and nucleus, while others treat tunes as truly unitary. All descriptions, however, pay special attention to the "nuclear tone" – the pitch movement which begins on the nuclear syllable and continues to the end of the tone unit.

As aptly discussed by Ladd (1980), the two major traditions of prosodic analysis of English speech are adequately illustrated by reference to the work of Trager and Smith (for American English) and Kindgon (for standard Southern British English). We begin by looking at the system proposed by Kingdon (1958) and further developed during the second half of the 20th century by Crystal (1969), Halliday (1967), O'Connor and Arnold (1973), Gimson (1970), Cruttenden (1986), with Wells (2006), Cruttenden (2008) and Roach (2009) being amongst the most recent examples.

Kingdon's (1958) system is a manifestation of developments aiming at the creation of a comprehensive system of prosodic notation suitable for the purpose of teaching English intonation to foreign learners. Kingdon's "tonetic" system comprises two static and three kinetic tones. The static tones are

H – high level, and

L – low level.

The kinetic (moving) tones in Kingdon's system are given numbers as follows:

I (rising), with two variants: Ih (high rising) and Il (low rising);

II (falling), also with two variants which, however, are not distinctive;

III (falling-rising), with two variants: undivided and divided (in the latter, the rise starts on a secondary stressed syllable).

The system also includes the following complex tones:

IV (rising-falling): a complex tone which is a modification of the falling tone:

V (rising-falling-rising): another complex tone, a modification of the falling-rising one.

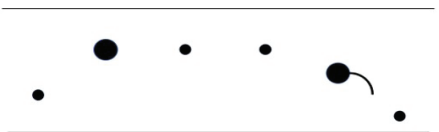
Kingdon is the pioneer who proposed the use of the "tonetic stress marks" in order to transcribe the tones in running text, e.g.,

Jo
\ John h (high fall)
n

a
/ Maria Ma (low rise)
ri

The British School views intonation as comprising a complex set of features from different prosodic systems. The most central of these features are tone, pitch range and loudness, while rhythmicality and tempo are closely related to them. In other words, rather than restricting the definition of intonation to pitch variation alone, a wider definition is adopted, which results in a more complex formal description of the phenomenon, but allows for "an ultimately less involved semantic statement" (Crystal 1969). Priority is given to those features which involve pitch movement, namely, tone and pitch range, since pitch contrasts lie at the centre of intonation as a prosodic phenomenon. Stress is viewed as word-level potential for prominence, or as prominence realised by means other than pitch, while the term prominence itself is reserved for salience achieved with the help of pitch change.

For the British School, the maximal functional unit to which meaning can be assigned in intonation analysis is the tone unit. The boundaries between tone units in the flow of speech can be marked by pause, other junctural phenomena (such as final lengthening), or signaled by a perceivable change of pitch (step-up or step-down). The only compulsory element in a tone unit is the nucleus, or tonic syllable, which is the carrier of the nuclear tone – a gliding pitch movement of a particular kind. The nuclear syllable is the last accented syllable in the tone unit; syllables which follow it are unaccented and constitute the tail of the tone unit. If there are stressed (and also usually pitch prominent) syllables preceding the nucleus, the first of them marks the beginning of the head of the tone unit. The head stretches up to, but does not include the tonic. Any unstressed syllables preceding the first stressed (or the nuclear) syllable constitute the pre-head. The pre-head, head and tail are all optional constituents.



We went to the sta- -tion.

Figure 3.1. A "tadpole" representation of the intonation in the statement "We went to the station."

Figure 3.1 gives an example of the "tadpole notation" type of representation of the tune which is typical for the British School. It consists of small dots which correspond to unstressed syllables, larger dots which indicate the stressed (and usually also pitch prominent) syllables, and a tail, or curve which is attached to the large dot that stands for the nucleus: the curve shows the direction of the pitch glide. In the example in Figure 3.1, "we" constitutes the

pre-head, “went to the” – the head, “sta-“ is the tonic syllable, and “-tion” makes up the tail of the tone unit.

Every tone unit has a single tonic syllable, or nucleus, on which one out of a small number of pitch glides, or tones, must occur. As noted by Crystal (1969), this placement of the nuclear tone has often been referred to as “sentence stress”. However, tone units are seldom co-extensive with sentences or clauses, therefore the term “sentence stress” is rather misleading.

Nuclear tones are of three main types – simple, compound and complex. Simple tones are unidirectional: rising (transcribed ˊ or ↗), falling (transcribed ˋ or ↘) or level (→). (We will be using the arrow transcriptions in the rest of the review.) There are high and low varieties of the rising and falling tone. They are treated as a matter of pitch range, and a separate complex system of pitch range contrasts is postulated. It is claimed that in English there is only a single contrastive degree of widening or narrowing of each tone. The combination of pitch direction and pitch range, or height, accounts for the meaning of the nucleus. Thus, the combination of a falling tone with wider pitch range will be interpreted as more emphatic, categorical, etc. than the default combination with normal pitch range.

Complex tones involve nuclei in which there is a change of the direction of the pitch movement on the tonic syllable. The most frequently occurring English tones within this category are the fall-rise ↘↗ and the rise-fall ↗↘, but occurrences of the rise-fall-rise ↗↘↗ and fall-rise-fall ↘↗↘ have also been attested. The first element of the fall-rise and the rise-fall is usually phonetically more prominent than the second element. With respect to pitch range, either the whole tone, or only one element of it may be widened or narrowed.

The compound tones, also called correlative or binuclear tones (Crystal 1969), are combinations of two kinetic tones, namely ↘+↗

and ↗+↘ which function as a single unit over a longer stretch of utterance. A comparison with the respective complex tones shows that the differences are semantically contrastive, as in the following pair of examples from Crystal (1969), where prominent syllables are transcribed in block capitals:

(1) compound ↘+↗ tone: I ↘ THOUGHT it would ↗ RAIN (and it did) vs.

(2) complex ↘+↗ tone: I thought it would ↘↗ RAIN (but it snowed)

Crystal's (1969) data on the occurrence of the basic tone types in English reveal that the simple fall is the most frequently occurring tone – it was found in 51.2% of the tone units in Crystal's corpus, followed by the simple rise – 20.8%, the complex fall-rise – 8.5%, the compound fall + rise – 7.7%, the complex rise-fall – 5.2%, the level tone – 4.9%, and the compound rise + fall – 1.7%.

Some authors working within the British tradition of intonation analysis, however, do not recognise the level as a separate nuclear tone – this is the stance of Kingdon and Halliday, among others. Other authors, such as O'Connor and Arnold and Gimson, distinguish two types of each simple kinetic tone, e.g., a high fall which starts above the middle of a speaker's pitch range, and a low fall which starts at or below mid range. The system of 7 nuclear tones proposed by O'Connor and Arnold (1973) for the purpose of teaching English intonation to foreign learners of the language is among the best-known descriptions of English intonation. The tones they recognise are as follows: low fall, high fall, rise-fall, low rise, high rise, fall-rise, mid level. Figure 3.2 shows diagrammatically with the help of the tadpole notation the pitch movement in the low fall (left) and the high fall (right) over the two syllables of a bi-syllabic word with lexical stress on the first syllable, such as the word "twenty".

Figure 3.2. O'Connor and Arnold's low fall (left) and high fall (right) realised on a bi-syllabic word with initial stress, e.g., "twenty"

Halliday (1967) views English intonation contrasts as purely grammatical. He distinguishes four hierarchically ordered phonological units – tone group, foot, syllable and phoneme. A tone group comprises two elements of structure – pre-tonic and tonic. The primary tone contrasts are carried by the tonic. For Halliday, there are three meaningful sets of choices in English intonation – "the three Ts". The choice of tone is only one of the so-called "three Ts" of the British School of intonation analysis. The first choice has to do with the way in which the flow of speech should be broken down into chunks – tone groups. This kind of choice is called **T**onality. The second "T" refers to the choice of word(s) on which the speaker wants to focus the listener's attention by highlighting their stressed syllable(s). This choice is known as **T**onicity. The last one is the choice of nuclear pitch movement and is simply called **T**one.

The same nuclear tone system as that proposed by O'Connor and Arnold is also adopted by Cruttenden (2008). Wells (2006) – one of the most recent additions to the series of English intonation courses for foreign learners, proceeds from a basic choice between falling versus non-falling tones, with default options as follows: a fall for statements, exclamations, wh-questions and commands, and a rise for yes-no questions. For utterances with two intonation phrases, the default choice on the main part is the fall, and on the subordinate or dependent part – a non-fall. The three falling tones in Wells' system are the high fall, low fall and rise-fall. The various

kinds of non-falling tone include the high rise, low rise, mid level and fall-rise, and it is often necessary for a further distinction to be recognised between rises and fall-rises.

Wells describes the basic meaning of a fall as indication that the information is complete, and that of a rise or fall-rise – as signaling that there is something more to come; the fall is presented as the default choice of tone for statements, wh-questions, exclamations and commands, the rise is the default tone for yes-no questions, and the fall-rise signals particular implications.

In his (mainly theoretical) account of the intonational system of British English RP, Roach (2009) operates with a simplified system of only three simple tones (fall, rise and level) and two complex tones (fall-rise and rise-fall), although he leaves open the possibility to utilize pitch height as a further choice so as to distinguish between high and low varieties of the simple tones. In Roach's account, the main functions of the nuclear tones are:

Fall – a "neutral" statement; gives the impression of finality and definiteness;

Rise – signals that there is something more to follow; in a conversation – an invitation for the other speaker to continue; encouraging; the default tone in general questions and lists;

Fall-rise – signals givenness as well as limited agreement, response with reservation; hesitation, doubt;

Rise-fall – conveys emphasis, strong approval or disapproval, surprise;

Level – routine or uninteresting statement, may suggest boredom.

The author warns that each of the five tones can have many more meanings, depending on the context in which it is used, but that certainly not every English tone could be used in any context.

3.2 Discourse Intonation analysis

Although the British School model of English intonation described above has probably been the most influential approach, and continues to dominate the field of English pronunciation teaching for foreign learners, it would be wrong to assume that it is the only one. An important offspring of the research carried out in the late 1970s and the early 1980s within the framework of the Birmingham School of Discourse Analysis was the model of Discourse Intonation put forward by Brazil et al. (1980) and elaborated upon in a number of textbooks specifically designed for teaching discourse intonation to the foreign learner (Bradford 1988, Brazil 1994, Cauldwell 2002). As stated by Brazil et al. (1980), "… a view of language as discourse and communication, where utterance value depends crucially on interactive function within the discourse, needs a system of intonational analysis which is distinct from that appropriate to a syntactic or semantic view of language." Thus, Discourse Intonation analysis makes no attempt to relate intonation to grammar or to speakers' attitudes. The focus is on the ability to use language as a means of communication, and the systematic choices which the speaker makes in order to convey meaning concern the division into tone units, the choice of prominence and tonic syllable, and three subsystems relating to pitch, namely, tone, key and termination.

Tone choice is limited to a set of four tones: fall (the telling, or proclaiming tone), rise, fall-rise (the referring tones), and level. In addition, speakers have the choice of placing prominent syllables low, mid, or high in relation to the previous prominence. The key sub-system operates on the onset prominence and comprises high, mid and low key: mid key is the unmarked option, high key signals contrast, whereas low key signals equivalence. The termination subsystem operates on the tonic prominence: mid termination is unmarked, high termination adds the meaning "I expect your judgement on this", and low termination signals "This is discourse-final".

Although Discourse Intonation analysis was quite popular in English-as-a-foreign-language classrooms at the end of the twentieth century (or at least in the classrooms of those teachers who could afford to spend time teaching English intonation), the bulk of recent English intonation teaching materials seem to attest a return to the traditional British School approach.

3.3 The American approach: levels vs. configurations

The "levels vs. configurations" debate in the study of intonation refers to the way in which pitch patterns should be analysed. Scholars working within the British tradition in general tend to adopt the "configurations" approach, as discussed above. On the other hand, the American structuralist tradition of intonation analysis posits a number of phonemic pitch levels (usually four) and describes pitch movements in terms of contours which constitute sequences of pitch level phonemes.

K. Pike's "The Intonation of American English" (1945) is the first comprehensive treatment of the topic in the American structuralist tradition. His work is notable for the fact that it considers a combination of suprasegmental factors such as rhythm, pause, length and stress and their contribution to speech prosody. But the most notable contribution to intonation theory and analysis which Pike makes is undoubtedly the introduction of the "levels" approach. In this approach, the speaker's pitch range is divided into several relative heights of pitch. Each height is considered to be phonemic. As explained by Bolinger (1972), these levels "bear the same relationship to intonational configurations as such phonemic entities as vowels and consonants bear to words". Although configurations are referred to as "contours", what is essential is not the movement of the pitch from one level to the next, but rather the sequence of levels. Pitch movement is only a means for getting from one phonemic level to another.

Pike distinguishes four "relative but significant levels (pitch phonemes)" for (American) English which constitute the basic building blocks for intonation contours: he labels them extra-high, high, mid and low, and gives them numbers from 1 (extra-high) to 4 (low). Note that most of Pike's successors number the levels in reverse, that is, level 1 is low and level 4 is extra-high.

The meaning is carried by the contour as a whole, but the levels, being the building blocks, contribute to the contour and therefore to the meaning. According to Pike, intonation meanings should not be confused with the syntactic uses to which they are put, and it is not correct to speak about "statement intonation" or "question intonation". However, some generalisations can nevertheless be made regarding the meaning of levels. Thus, contours which include pitch level 1 tend to signal surprise or unexpectedness, pitch level two is the default for stressed syllables, while level 4 is the default for unstressed syllables at the end of a fall. It must be noted, however, that numerous exceptions exist, therefore it might be more fruitful to draw generalisations on the basis of grouping together contours with related form and meaning.

As for the transcription of contours, a stressed syllable which begins a primary contour is marked by [°]; any syllable whose stress is lexically determined is potentially the starting point of a primary contour. If there are any unstressed syllables immediately before the stressed syllable of a primary contour, and if they are pronounced together with it, they constitute a precontour. A precontour (if there is one) and a primary contour make up a total contour, e.g.,

He 'said so.
3- °2- -4
'Tom did it.
°2- -4 (The examples are from Pike 1945)

Trager and Smith (1951, 1957) further developed the levels approach to the analysis of intonation, their second revision largely

in response to the reactions of scholars involved in the preparation of English-teaching materials. As pointed out by Ladd (1980), theirs was the best-known application of Bloomfieldian principles of phonemic analysis to suprasegmentals. Trager and Smith posit four stress phonemes: primary corresponds to "sentence stress" or "nuclear stress", while weak corresponds to unstressed. Stress is manifested by loudness and is strictly a separate element of the system from that of pitch. Like Pike (1945), they also work with four pitch phonemes: low (^{1}), middle (^{2}) , high (^{3}) and extra high (^{4}) – note that the numbering is the reverse of that used by Pike.

In Trager and Smith's model, intonation patterns in American English typically consist of three pitches and a terminal contour. The initial pitch is typically / ^{2}/, the central one is most often / ^{3}/ in statements, questions, etc., but can be / ^{4}/ if there is emphasis. The final pitch is most often / ^{1}/ at the end of a statement, / ^{2}/ at the end of a sentence non-final clause, and / ^{3}/ at the end of some types of questions. The final pitch is additionally modified by the terminal contour which can be level / | / (usually in non-final clauses), rising / ‖ / (in many non-final clauses as well as in yes-no questions) or falling / # / (usually in statements and wh-questions). In addition to these three terminal junctures, Trager and Smith also postulate one "internal open juncture" (transcribed / + /) which distinguishes minimal pairs such as "night rate" and "nitrate". Some examples of short English utterances analysed and annotated with the help of Trager and Smith's system follow:

^{3} John ^{1} #

^{2} Why are you ^{3} going ^{1} #

In the first example (a statement), the fall begins at level 3 and ends at level 1, followed by a terminal falling contour. It is realised on the single-syllable utterance "John". In the second example (a wh-question), the utterance begins on the default initial pitch at level 2,

rises to level 3 – the typical central pitch, and falls to level 1, again followed by a terminal fall.

3.4 Bolinger

Another notable American contribution to the description of suprasegmentals is that of Bolinger (1958, 1986, 1989). An important distinction first made by Bolinger as part of his theory of rhythm is that between two distinct vowel categories in English: full and reduced vowels. Another crucial contrast suggested by him as part of his theory of pitch accent in English is the one between stress as a lexical property expounded by loudness, and accent as its manifestation by means of a pitch movement in an utterance.

Although Bolinger's theory of pitch accent for American English has certain similarities to the configurations approach of the British School, since he claims that pitch configurations are more meaningful than pitch level sequences, he allots much more importance to stress, or prominence. Bolinger's pitch accents are at the same time markers of prominence and building blocks of the intonation contour. In Bolinger's theory, pitch and stress are interdependent – pitch prominence is the main cue to stress, and stress changes affect intonation contours. Therefore, intonational morphemes should be defined in terms of both pitch and stress, hence the term "pitch accent". Other cues such as length, loudness, rhythm and vowel quality also contribute to prominence, but the role of pitch is primary. Bolinger distinguishes three major accent types in American English:

Accent A – the most frequent shape: relatively high pitch followed by a jump down, "an intonational configuration whose distinguishing feature is an abrupt fall in or from the syllable that is made to stand out by the fall" (1989, p. 3);

Accent B – almost as frequent as A, is characterized by an upward movement, a jump up to the syllable which stands out, while

any following unaccented syllables either continue the gradual rise, stay level, or slightly fall;

Accent C – the least frequent of the three, is "a kind of anti-accent A" (1958, p. 143), which is marked by "down to" rather than "down from" pitch movement.

To sum up, Bolinger's theory differs from both the levels and the configurations approach in that it insists on the interdependence of pitch and stress, thus anticipating the Autosegmental-Metrical model of intonational phonology.

3.5 The Autosegmental-Metrical model and ToBI

The Autosegmental – Metrical (AM) model of intonational phonology which was first put forward by J. Pierrehumbert (Pierrehumbert 1980, Pierrehumbert and Beckman 1988, Ladd 1996/2008) has offered a new perspective on the investigation of suprasegmental features, and on comparative intonation research. As a consequence of the application of the Autosegmental-Metrical approach to the analysis of the intonation in a range of languages (Jun 2005, 2014), it has also exerted its influence on the teaching of (English) intonation to foreign learners. To give just one example, Estebas – Vilaplana (2015, 2018) compared the prosodic features in the speech of two groups of Spanish students of English phonetics who were trained using different methodology: one group was taught in the tradition of the British School of intonational analysis, whereas the other was trained using what the author calls TL_ToBI – a version of Pierrehumbert's (1980) system adapted to the needs of teaching English intonation in a distance learning setting. Estebas-Vilaplana found that "students instructed with TL_ToBI produced more native-like intonation patterns" than those instructed with the British School model, "suggesting that a system based on tonal targets and their association to the metrical structure is more helpful", especially for self-tuition purposes (Estebas – Vilaplana 2015, p. 42).

As pointed out by Beckman et al. (2005), the Autosegmental-Metrical framework and the related ToBI system are soundly based on many years of research into the prosody of English, including experimental evidence from both speech production and speech perception. The term "autosegmental-metrical" itself was proposed by Ladd (1996) to refer to the two interrelated sub-systems of (i) speech melody – the autosegmental tier, and (ii) prominence and phrasing – the metrical structure. The Autosegmental-Metrical model distinguishes between the phonology of intonation and the phonetic realisation, or implementation of phonological categories. It analyses the continuous pitch contour as comprising a number of constituent primitives, in other words, it recognises the compositionality of the contour. Intonation is analysed as a sequence of Low (L) and High (H) tones and various combinations of the two. The tones are defined in relative terms: H stands for tones which are high in comparison with other tones in the contour and with respect to the speaker's pitch range, and L represents tones which are low with respect to the particular speaker's range and in comparison with other tones in the same contour. All tones have pragmatic meaning and contribute to the meaning of the utterance (Pierrehumbert and Hirschberg 1990).

Tones are autosegments, that is, they are independent of the segmental string of vowels and consonants, but are associated to certain structural positions. These positions are determined with reference to the metrical representation of an utterance. The structural positions to which tones associate are of two kinds: constituent heads (usually stressed syllables) and constituent edges. The tones are called pitch accents and are labelled with a star *, e.g., L*, H* are monotonal pitch accents. The final pitch accent in a phrase is the most important one and is called a nuclear accent.

When two tones combine, they form a bitonal pitch accent. The stronger of the two is followed by a star in the notation and is the one which is associated with the constituent head, e.g., L*+H, where

the weaker tone H is a trailing tone, or L+H*, where the weaker tone L is a leading tone. In Pierrehumbert's (1980) theory, the weaker tone precedes or follows the starred tone by a fixed amount of time. Research has shown that there may exist different relations between the two tones of a bitonal pitch accent. Some bitonal accents are of the type originally suggested by Pierrehumbert, that is, a metrically strong tone which is preceded or followed by a weaker one: an example of such a pitch accent is L*+H in English. However, there is at least one more type of bitonal pitch accent in which the two constituent tones form a looser unit, and each tone is aligned with a particular segmental landmark: examples are L*+H in Greek and L+H* in English. Tritonal accents have also been occasionally proposed, but those have not been widely used.

Edge tones are of two types: phrase accents, e.g., H-, L-, and boundary tones, e.g., H%, L%. While both types of edge tones are associated with the right boundary of the phrase, boundary tones can also associate with its left boundary as well, e.g., %H. Bitonal and multitonal combinations of edge tones have also been proposed. Some Autosegmental-Metrical systems do not recognise phrase accents at all, but the majority recognise the association of phrase accents with the boundaries of the intermediate phrase (ip). Boundary tones typically associate with the right edge of an intonational phrase (IP). Pierrehumbert and Beckman (1988) suggest that phrase accents may also have secondary association either to a given tone bearing unit or to another boundary: they claim that this is the case in English, where phrase accents associate to the right edge of the intermediate phrase, but also to the left edge of the word which carries the nuclear pitch accent, and thus spread over the whole stretch of syllables between it and the end of the intermediate phrase. Secondary association of post-nuclear tones with metrically strong syllables or with syllables in specific positions in the tone bearing unit have also been reported in a number of languages.

The phonetic realisation of pitch accents is in terms of specific points in the contour called tonal targets. The transition from one tonal target to the next is derived by interpolation between the successive targets. The tonal targets constitute turning points in the intonation contour, and phonetically are described in terms of their scaling (the F0 value of the targets) and their alignment (their position relative to specific tone-bearing units (TBUs). Pitch accents are typically realised on stressed syllable nuclei, and edge tones are realised on peripheral tone-bearing units, e.g., phrase-final syllable nuclei.

The Autosegmental-Metrical model tries to avoid some of the problems encountered by the levels approach to do with the necessary and sufficient number of levels needed for an adequate description of (American English) intonation by using scaling and downstep. Scaling refers to the pitch level of a tonal target: it can vary, as evidenced by the distinction drawn between high and low rises and falls in many of the English intonation models discussed previously. Examples of pitch scaling are downstep and upstep – respectively, the lower or higher scaling of pitch accents. There is no unanimity amongst researchers on the issue of whether downstep should be treated as a phonological phenomenon or a matter of phonetic scaling, that is, whether the normal H* and the downstepped !H* in English should be considered different pitch accents or phonetic realisations of one and the same pitch accent.

In Autosegmental-Metrical phonology, Ladd (1996) was the first who made a distinction between pitch span and pitch range. Pitch span refers to the extent of range of frequencies used by a speaker, while pitch range refers to whether these frequencies are high or low. For example, two speakers may use the same pitch span of 160 Hz, but the first may use a low range (e.g., 120 – 280 Hz) and the second one may use a higher range (e.g., 160 – 320 Hz).

The grammar for the composition of tunes in English proposed by Pierrehumbert (1980) can be seen in Figure 3.3. All elements can

combine freely. Left-edge boundary tones are optional, whereas all other elements are compulsory, so a well-formed tune includes at least one pitch accent (which in that case will be nuclear), a phrase accent and a boundary tone. Nuclear pitch accents are not given any special status. Unlike a British School-type of analysis, the Autosegmental-Metrical model treats pre-nuclear and nuclear pitch accents as equal. Although all elements can combine freely, it has been shown that some combinations in a given language occur far more frequently than others. For example, Dainora (2001, 2006) found that in the Boston University Radio Speech Corpus H* L-L% occurred 33% of the time, while L*+H L-H% occurred less than 0.5% of the time.

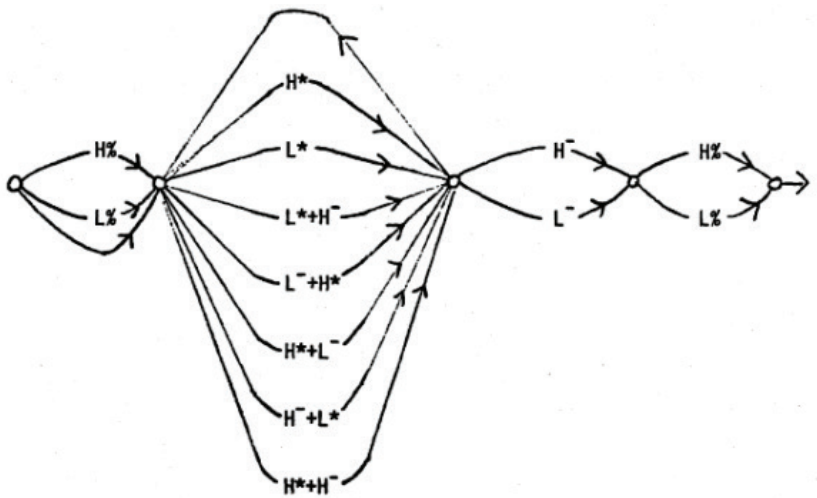


Figure 3.3. Pierrehumbert's finite state grammar of tone sequences (from Pierrehumbert 1980, p. 29)

As far as metrical structure is concerned, the Autosegmental-Metrical model maintains that prosodic phrasing should be based first and foremost on properties of the tune (though the importance of duration, sandhi phenomena, etc. has been shown as

well). This distinguishes it from phonological models (Selkirk 1984, Nespor and Vogel 1986) in which prosodic phrasing is based on syntactic structure.

The ToBI (Tones and Break Indices) system is the best-known application of research carried out within the Autosegmental-Metrical framework. It is a collection of conventions specially designed for transcribing and annotating speech prosody. It was originally intended for the annotation of American English speech corpora (Silverman et al. 1992, Beckman et al. 2005) and has provided the basis for development of similar systems for a number of languages (Jun 2005, 2014). Towards the end of the millennium, the system started being used for labelling other English varieties, as well as other languages, and thus soon established itself as a general framework for prosodic analysis and intonational annotation.

A ToBI representation includes a sound file with its associated spectrogram and pitch track, a tonal tier on which the pitch accents, phrase accents and boundary tones are marked, as well as a break index tier on which the strength of prosodic boundaries is indicated with the help of numbers. Other tiers may be added on which syllables, words, the orthographic text of the utterance, etc. are shown. All of these will be illustrated in the discussion of the original system for the transcription of Mainstream American English MAE_ToBI in Chapter 4.

3.6 ToBI as a basis for an International Prosodic Alphabet (IPrA)

Because ToBI was originally conceived as a transcription system and a tool for the annotation of the phonological prosodic features of a given language or language variety, there have been suggestions for the creation of an International Prosodic Alphabet (IPrA) based on the ToBI conventions (e.g., Hualde and Prieto 2016).

Every phonetician and phonologist is of necessity well acquainted with the International Phonetic Alphabet (IPA). The Alphabet was first published in the late 19th century and was intended as a tool for the accurate representation of the pronunciation of languages, especially those like English, in which the standard orthography is an "unreliable guide to pronunciation. One of the original goals was to provide unique characters for the symbolic representation of the distinctive sounds – the phonemes of a given language, and thus to standardise the representation of spoken language, avoiding the confusion created by the variety of transcription systems which existed at the time. The first version of the IPA was developed by A. J. Ellis, H. Sweet, D. Jones, and P. Passy, and was published in 1888. The IPA was revised several times throughout the 20th and the beginning of the 21st century by the International Phonetic Association – the body responsible for it, which also publishes the charts summarising the alphabet. The last set of charts was published in 2020. One could thus argue that today the IPA offers a synthetic representation of our scientific knowledge about the composition of human language at the segmental level. The symbols (100 in total) and the additional diacritical marks (more than 30) allow scholars to adequately present in writing the most important characteristics of all existing vowels and consonants that have been found in human languages.

However, things are different as far as the prosodic characteristics of speech are concerned. The IPA offers only a limited number of symbols for the written representation of a small selection of prosodic features, namely, ['] primary and [ˌ] secondary lexical stress, [ː] duration, [|] minor (foot) group and [‖] major (intonation) group boundary, [.] syllable break, [‿] linking. In addition, there are a small number of “tones and accents” symbols which can be used for the representation of pitch movement in intonation languages like English and Bulgarian – [↓] downstep, [↑] upstep, [↗]

global rise, and [↘] global fall. (IPA tonal symbols and diacritics for the representation of lexical tone in tone languages are not listed here.) It is obvious that this limited set of symbols is insufficient for the adequate representation of language prosody details as revealed by contemporary research.

This lack of an adequate, universally agreed notational system is amongst the main reasons why different systems for the representation of suprasegmentals abounded in research on the topic prior to the time the Autosegmental-Metrical model of intonational phonology and the ToBI prosodic annotation framework became popular.

Bulgarian intonation has likewise been transcribed in different ways by different researchers. For example, Stoykov (1966, p. 156) illustrates in writing the intonation of the sentence "Той чете." ("He is reading.") pronounced with a final fall of the pitch as follows:

Тòй
че-
mè

Penchev (1980) describes a total of 10 types of melodic contours in Bulgarian, in which F0 differences are represented on five pitch levels, ranging from 1 (the lowest one) to 5 (the highest one). In addition, he also uses a numerical representation of the pitch changes in each contour. These alternative transcriptions are shown in Figure 3.4 (adapted from Penchev's chapter on intonation in Tilkov 1982, p. 263):

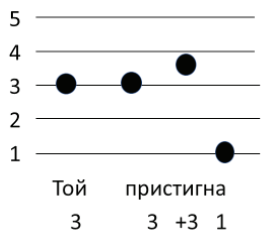


Figure 3.4. Representation of the pitch change in the utterance "Той пристигна." pronounced with a final fall (adapted from Tilkov 1982, p. 263).

If we connect the four points in Figure 3.4. which represent the pitch height of the respective syllables with respect to the five pitch levels, we will get Penchev's melodic contour Type 1 which typically occurs with Bulgarian declaratives. In it, the main highlighted part is realised above the medium high level (+3), any unstressed parts which occur before the center are at mid level (3), and the end of the post-central part reaches a low level (1).

It is obvious that such diverse notation systems do not allow for any systematic comparisons between the findings of different researchers for one and the same language (in our example – Bulgarian). Neither do they enable cross-linguistic analyses: although Penchev's notation borrows heavily from the American "levels" approach to intonational analysis, the different number of levels which he postulates for Bulgarian, as well as the nature and type of his essentially syntactically-based contours do not make it possible for any systematic comparisons between English and Bulgarian intonation to be drawn.

Pierrehumbert's dissertation (Pierrehumbert 1980) as well as subsequent research within the new Autosegmental-Metrical framework for analysis and annotation of intonation in (American) English made particularly popular the ToBI system. "ToBI" stands for "Tones and Break Indices". Over the last 20 years, this approach has been used as a methodological framework for the analysis of intonation in a number of languages other than English, resulting in modifications for German (GToBI), Greek (GRToBI), Dutch (ToDI), Japanese (JToBI), Korean (K-ToBI) and many more. Figure 3.5 shows a typical (American) pronunciation of the English phrase "Tones and break indices", pronounced as two intonation phrases. The first one – "Tones" is pronounced with a low nuclear pitch accent L* followed by two rising edge tones H-H%. The second intonation phrase "and Break Indices" is pronounced with the nuclear pitch accent H+!H* on "Break" and a low boundary L-L%

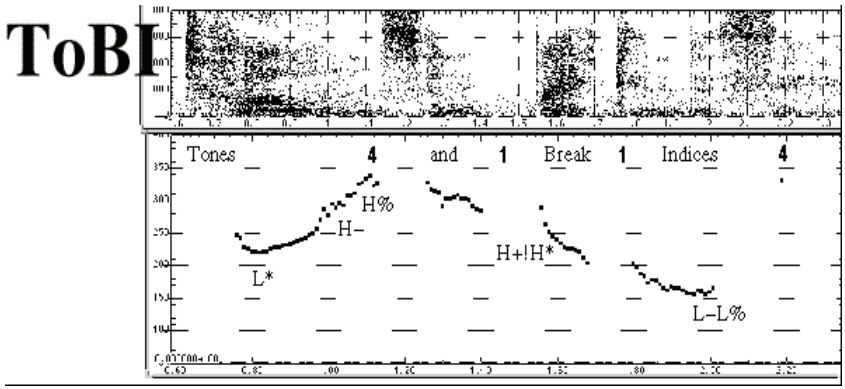


Figure 3.5. Basic pitch movement and intonation transcription of the phrase "Tones and Break Indices" (ToBI website 1999).

Figure 3.6 (Dimitrova and Jun 2015) is an illustration of the use of ToBI conventions for the annotation of the intonation of the simple Bulgarian declarative sentence "Милена намери лимони", pronounced as a single intonational phrase, in which the two pre-nuclear pitch accents are labelled as L*+H, the nuclear accent – as !H*, the phrase accent is L- and the boundary tone is L%.

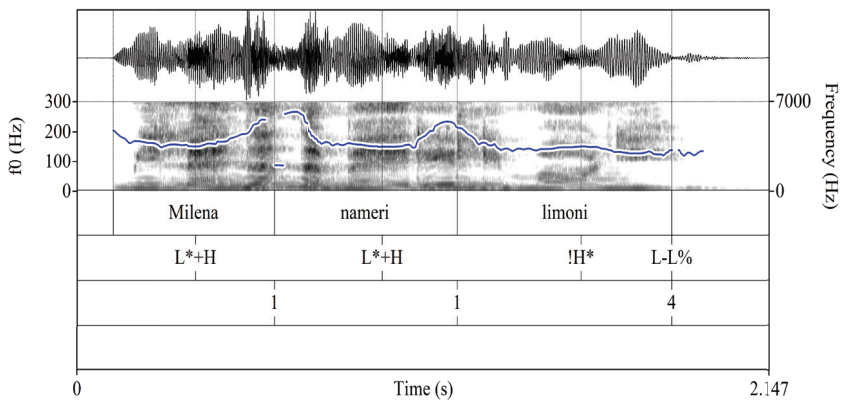


Figure 3.6. Sound wave, spectrogram with pitch track and ToBI labelling of the sentence "Милена намери лимони." (Milena found lemons.)

It can be seen that the pitch contour in Figure 3.6. is different from Penchev's contour discussed earlier. While some of the variability can be explained by the lexico-syntactic differences between Penchev's and Dimitrova and Jun's example utterances which result in different distribution of the pich accents in each of them, the explanation for other differences has to be sought elsewhere. Therefore, in the course of intonation analysis it is necessary to properly record the details of the pitch movement in order not to lose important information in the early stages of analysis which may turn out to be distinctive at a later stage. This is where a reliable system of intonation transcription would come to the aid of the intonologist.

However, the ToBI website (ToBI 1999) explicitly emphasises that ToBI should not be seen as a prosodic version of the International Phonetic Alphabet, as intonation and prosodic structure can differ even across dialects of the same language, which in turn necessitates the development of different, language- and accent-specific versions of ToBI.

Nevertheless, it can be rightly argued that ToBI has already proved its suitability and effectiveness as an intonation transcription system, as well as its great adaptability to the needs of typologically diverse languages. It is therefore not surprising that the first official proposal for the creation of an international prosodic annotation system – an International Prosodic Alphabet, or IPrA, was based on the principles of the Autosegmental-Metrical model of intonational phonology and the ToBI conventions. The working proposal was first put forward at a seminar held during the 18th International Congress of Phonetic Sciences which was held in Glasgow, U.K., from 10 August to 14 August 2015, and subsequently presented in written form by Hualde and Prieto (2016).

It is only natural that, in spite of the general consensus amongst researchers regarding the advantages of the broad Autosegmental-Metrical approach and the ToBI annotation system, a number of un-

resolved issues still remain. Hualde and Prieto (2016) point out two characteristics of the application of the Autosegmental-Metrical approach to different languages which lie at the base of some of the current misunderstandings:

(i) the adoption of language-specific phonological labels and phonetic realisation rules which are non-transparent cross-linguistically, and

(ii) the ambiguity of some ToBI annotations which appear to be a compromise between broad phonetic and phonological levels of transcription.

Hualde and Prieto's proposal for overcoming these problems is the addition of two distinct levels of prosodic representation – a phonological and a broad phonetic one, and the use at the phonetic (surface) level of labels which are consistent and transparent cross-linguistically.

To conclude, Hualde and Prieto's proposal for better adaptation of the existing ToBI conventions to the needs and purposes of a newly created International Prosodic Alphabet (IPrA) is in agreement with the main principles of the Autosegmental-Metrical analysis of intonation, but whether and when such an alphabet will be officially adopted by the wider researcher community for the time being remains unclear.

The development and adoption of an International Prosodic Alphabet would have many advantages. First, such a universal tool for prosodic transcription would make it possible to ensure a much greater degree of comparability of analyses within a given language or dialect. It would be possible to exchange and compare data, including available corpus data currently annotated in many different ways, which makes it impossible to use them for the purposes of contrastive analysis. The advantages of a single transcription system are also obvious when it comes to further work in the field of into-

nation typology. The same applies to contrastive studies of prosody in two or more languages or dialects. Last but not least, such a transcription system would be extremely useful in the study of prosody for the purposes of foreign language learning.

It should be remembered that since its creation at the end of the 19th century, the International Phonetic Alphabet (IPA) itself has been repeatedly revised and improved. We have every reason to expect that the same will happen with a future International Prosodic Alphabet (IPrA) – namely, that it will develop and improve in parallel with the development of our knowledge about the prosodic organisation of human speech.

4. The Autosegmental-Metrical analysis of English and Bulgarian intonation

Having introduced the central tenets of the Autosegmental-Metrical model of intonational phonology and the general principles of the ToBI framework and tool for intonational labelling and annotation, we next present in some more detail the current Autosegmental-Metrical descriptions of the intonation of General American (GA), Received Pronunciation (RP) / Southern Standard English (SSE) and Standard Bulgarian. This is needed as a prerequisite for the comparison between them which we draw in Chapter 5 in an attempt to predict some of the difficulties that Bulgarian learners are likely to experience in the acquisition of the prosodic system of English.

We discuss RP and GA prosodic descriptions within the ToBI framework separately for two main reasons. The first reason has to do with teaching the intonation of English as L2 to advanced learners at tertiary level. Polls of English Philology students' preferences at a Bulgarian university about their choice of reference accent of English have shown that about half of the students select RP as their pronunciation model when speaking English, and the other half opt for General American (Dimitrova 2018). Given the relative scarcity of advanced English intonation courses based on the prosody of RP, and the unavailability of such courses featuring GA suprasegmental features, an informed decision on the part of the teacher regarding potential problematic areas for both groups of students becomes essential. However, and this brings us to the second reason for considering the intonation systems of RP and GA separately, few systematic comparisons of the prosodic systems of the two English accents exist to date. The currently proposed ToBI systems for the two spoken varieties are outlined below, but it can easily be seen

that they diverge in several important respects and therefore cannot provide a systematic basis for making generalisations about a common model of English intonation to be used in teaching English prosody to L2 learners.

4.1 American English

We begin by looking at General American (GA), or "Mainstream American English" (MAE), to use the terminology of Beckman et al. (2005) whose extensive outline of the prosody of this particular standard English accent we present below. Beckman and colleagues start by reminding that the term ToBI can be used in two different senses: to refer to the original annotation system developed in the early 1990s for the labelling of spoken corpora of Mainstream American English, and for the general framework for developing prosodic annotation systems for other varieties of spoken English as well as for other languages. To distinguish the general developmental ToBI framework from the system for labelling Mainstream American English prosody, we follow Beckman et al. in adopting the term "MAE_ ToBI" for the latter.

The main tenets of the ToBI model provide the basis for MAE_ ToBI as well. In brief, they include the following assumptions:

1. The intonation contour of an utterance can be represented linearly as an autosegmental string of tones, while prosodic units (e.g., intonation phrases and lower-level units which are part of the metrical hierarchy) are represented on a separate metrical tier with the help of break indices.

2. The intonation contour is broken down into a sequence of static relatively high and relatively low pitch levels: H and L tones.

3. Local pitch range depends on a number of effects such as prominence relations and upstep or downstep which are specified independently of the tone level.

4. There are two kinds of tones in a phrase: pitch accents and edge tones; pitch accents align with stressed syllable nuclei, while edge tones align with segments at the respective phrase boundary.

5. In English, two levels of intonational phrasing are distinguished – the intermediate phrase and the intonational phrase. The edge tones which correspond to the two levels are the phrase accent, which is aligned to the stretch immediately after the last tone target in the syllable with nuclear stress, and the boundary tone, which is aligned to the final syllable of the intonational phrase.

In effect, the phrase accent marks the beginning of the post-nuclear tail in the phrase. The contour over the pre-nuclear head can be described with the help of the same set of pitch accent types which are used for the nuclear accent of the phrase. However, it must be remembered that although some of the terms which are used, such as nucleus, head and tail, are reminiscent of the British School of intonation analysis, MAE_ToBI and the underlying Autosegmental-Metrical theory do not subscribe to the holistic dynamic pitch contour view typical of the British approach. Rather, they adhere to Pierrehumbert's (1980) tone target model which decomposes the intonation contour into tones instead.

The conventions which MAE_ToBI uses are of several kinds. First, a full ToBI record must include at least six obligatory parts:

- an audio recording of the utterance,
- the F0 contour (pitch track),
- a Tones tier with the transcription of the intonation contour,
- a Words tier which gives an orthographic transcription of all words in the utterance,
- a Break Index tier which shows the numerical indices for the different degrees of juncture between words, and
- a Miscellaneous tier for marking disfluencies and comments.

Figure 4.1. shows a full ToBI record of the phrase “wrapped in a warm cloak” from Aesop’s fable “The North Wind and the Sun”, pronounced by a native English speaker. The top panel shows the waveform, the next one – the spectrogram, followed by the pitch track. The top transcription tier shows the tones – in this example, they are the pre-nuclear pitch accents L* and L+H* and the nuclear pitch accent H*, as well as the phrase accent and the boundary tone L-L%. (The full tonal inventory of MAE_ToBI is given in Table 4.1.) The second tier in Figure 4.1. is the Words tier which contains an orthographic transcription of all words in the phrase. The third tier is the Break Index tier on which the degree of boundary strength is indicated by a number from 0 to 4. (A full list of the break indices for the annotation of juncture in American English is given in Table 4.2.) The bottom tier is the Miscellaneous tier, which in this particular example is empty.

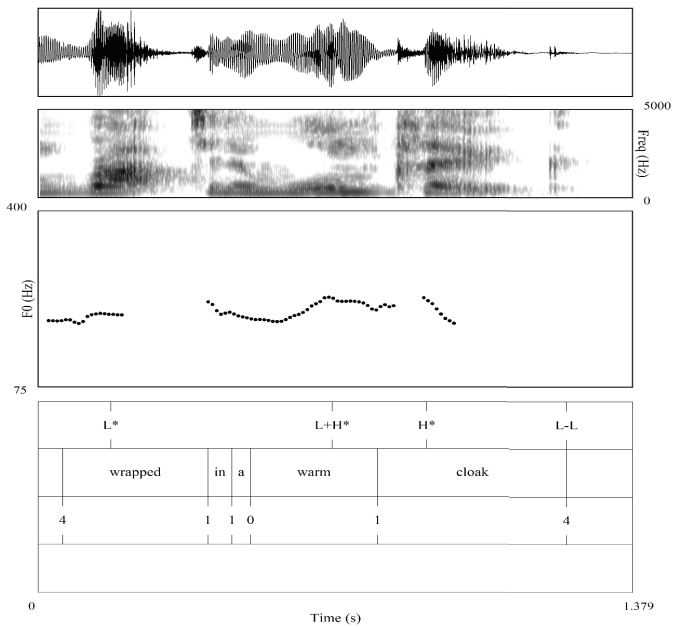


Figure 4.1. An example of MAE_ToBI labelling of the intonation phrase “wrapped in a warm cloak”

Table 4.1. presents the complete inventory of tones used by MAE_ToBI for transcribing intonation contours in American English.

Table 4.1. The tonal inventory of MAE_ToBI

pitch accents	L*, H* (!H*), L+H* (L+!H*), L*+H (L*+!H), H+!H*
phrase accents	H- (!H-), L-
boundary tones	H%, L%, %H

As can be seen from Table 4.1., MAE_ToBI posits a system of five phonologically distinct pitch accents; three of them have allophonic realisations (shown in brackets) in which the H tone is downstepped (shown by adding the downstep diacritic [!] in front of the H symbol). In two of the pitch accents F0 is low or falls from a higher to a lower level on the accented syllable – these are the simple low L* and the early-peak fall H+!H*, respectively. In two of the pitch accents F0 is high or rises from a lower to a higher level, reaching its peak on the accented syllable – the simple high H* and the rising L+H*. Finally, L*+H is often described as a "scooped rise" accent (Beckman et al. 2005, p. 25, Grice et al. 2020, p. 287) whose low target is within the accented syllable and the F0 peak is after it.

A pitch accent label on the Tones tier is optimally placed within the accented syllable, thus marking it unambiguously for pitch accent, and should coincide with the F0 minimum or maximum, if it occurs within the accented syllable.

The phonemic phrase accents are two – high (H-) and low (L-), and the first one of them also has a downstepped high-mid allophone (!H-). The phonological domain for the distribution of the phrase accent is the intermediate phrase (ip), which means that a phrase accent is placed at every Break Index which has a value of 3 or 4.

The boundary tones in the MAE_ToBI system are also two – high (H%) and low (L%). The domain for their distribution is the intonational phrase (IP), which means that a boundary tone is marked at every break index marked "4". The IP-initial high boundary tone %H can mark the left edge of the intonational phrase, but is relatively rare in American English.

In addition, MAE_ToBI uses some diacritics in order to show peculiarities of tonal realisation such as downstep, delayed tone < and early tone >. Any uncertainty on the part of the transcriber regarding the occurrence of a tone or about tone type is marked [?], e.g., *? means that the transcriber is not sure about the occurrence of the pitch accent.

The basic break index values are the following (Table 4.2.):

Table 4.2. The break index inventory of MAE_ToBI

0	very close connection between words
1	ordinary phrase-internal juncture
2	tone-breaks mismatch (e.g., 1 with unexpected tonal marker, or 3 or 4 without phrase accent / boundary tone)
3	intermediate phrase end
4	intonational phrase end

Quoting Pierrehumbert (2000), Beckman et al. (2005) strongly emphasise the fact that the original MAE_ToBI system "is at the level of abstraction of a broad phonemic transcription, or rationalized spelling system, such as those of Korean and Finnish. Just as a broad phonemic transcription for any language must be guided by the phoneme inventory of that language (as revealed by the lexical

contrasts), a ToBI-style transcription of the prosody and intonation of any language must be guided by an inventory of its prosodic and intonation patterns" (Pierrehumbert 2000, p. 26). The authors point out that when labellers use ToBI-style symbols which represent symbolic categories, this should imply that they have access to research showing that those categories are indeed part of the inventory of categories in the respective language which is being transcribed.

Researchers have occasionally added a Phonetic tier to their ToBI frameworks (see, e.g., Jun 2000) in cases when possible phonological analyses of a tune need to be recorded but their status needs further analysis and justification. Other tiers have also been added to the four compulsory ones originally proposed by MAE_ToBI, such as a syllable tier, a tier for marking sandhi phenomena, etc. But as pointed out by Jun (2022, p. 162), "The strengths of the ToBI system come from the transcription of tones being phonological, especially based on the AM model of intonational phonology, and from the architecture and mechanism of the transcription system." The symbols for the tones do not just mark turning points in the F0 curve, they transcribe distinctive tonal categories which have linguistic functions in the given language. The tones mark prominence and information structure, or they mark prosodic constituency, thus capturing prosodic and metrical structure.

4.2 Southern Standard British English (RP)

The other major variety of spoken English – Southern Standard British English (SSBE), or Received Pronunciation (RP) was described within the Autosegmental-Metrical framework by Grabe (1997, 1998, 2000, 2001). Her system was originally applied to the comparison of the prosody of English and German, and is therefore of particular interest to anyone planning to conduct cross-language contrastive research. It was also used as a tool for the comparison of the prosody of several different varieties of spoken English as part

of the Intonational Variation in English (IViE) project, and is sometimes referred to as "the IViE system". Grabe's system for transcribing Standard British English intonation is outlined below.

First, unlike the original ToBI system, Grabe's analysis treats nuclear pitch accents separately from pre-nuclear ones. Second, Grabe (1998) adopts a modified version of Gussenhoven's (1984) autosegmental-metrical account of British English in which rises are transcribed as L*+H and falls – as H*+L phrase-medially, and (respectively) as L*+H H% and H*+L 0% phrase-finally. In the original ToBI system, on the other hand, a nuclear fall will be represented as H* L-L%, and a nuclear rise – as L* H-H%. Besides, for marking the absence of a rise at an intonational phrase boundary, Grabe uses 0%. She motivates the differences between her system and the original ToBI one by pointing out that "The arguments for the different analyses of English are based on slightly different accounts of the semantics of English intonational contours" (Grabe 1998, p. 130).

Grabe's system for transcribing the intonation of Southern Standard British English also differs considerably from most other systems based on the Autosegmental-Metrical model in that it only assumes two levels of phonological representation and uses a very restricted inventory of tonal categories. At the underlying level of phonological representation, the following tonal categories are posited:

(i) only two basic pitch accents: H*+L, L*+H

The nuclear H*+L comprises a high target which is starred, followed by a low target. In non-final position the peak is aligned within the stressed syllable. In final position a rise-fall in F0 may be observed. On shorter syllables F0 undergoes compression in English (but truncation in other languages like German). The pre-nuclear H*+L does not differ from the nuclear one in terms of its realisation.

As far as the nuclear L*+H is concerned, Grabe actually found a very small number of nuclear realisations of this tone in her corpus,

and no examples of pre-nuclear realisations. However, she adds a note that this does not imply the absence of IP-final rises in the English corpus: "fall-rises" represented as H*+L H% actually occurred quite frequently in the corpus.

(ii) IP boundary tones: H%, 0%, no phrase accents

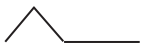
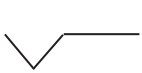
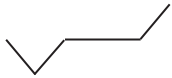
In the system proposed by Grabe for the transcription of Southern Standard British English / RP, it is suggested that the basic intonational inventory of English involves pitch patterns which can be accounted for as combinations of H*+L, L*+H and a boundary tone H%, without the need of any phrase accents. Additionally, it is posited that tonal specifications of IP boundaries are not obligatory.

The model involves a single level of intonational phrasing – the intonational phrase, whose edges can be marked by an initial or a final high boundary tone %H or H%, respectively, but can also remain tonally unspecified and therefore marked as %0 or 0%. It is important to note, however, that "0" is not a phonological category, but just an intonational phrase edge marker.

At the "phonological surface level", the underlying inventory can be changed by a set of modifications, all of which are optional and generate further patterns. Grabe (1997) compares the concept of a tonal modification to that of a connected speech process such as assimilation, like the one affecting the final /s/ of "packs" in the phrase "he packs shorts", causing it to assimilate to the following consonant – the initial /ʃ/ of "shorts". An example of tonal modification is Grabe's underlying phonological H*+L pitch accent which covers a range of phonetic realisations, including H*+L itself, but also !H*+L (with downstep), !H*+^L (half-completion), H*+>L (displacement), H*+L (delay) and H*> (deletion).

Grabe suggests the following correspondences between her system and the original ToBI system (Table 4.3.), claiming that "the alternative transcriptions reflect the auditory patterns more closely than the ToBI transcriptions" (Grabe 1997):

Table 4.3. ToBI and alternative transcriptions of contours of pitch accent + boundary tone in Standard Southern British English (adapted from Grabe 1997)

	Boundary type	
Pitch accent	**„level“**	**„rising“**
„falling“		
ToBI alternative	H* L-L% H*+L 0%	H* L-H% H*+L H%
„rising“		
ToBI alternative	L*+H H-L% L*+H 0%	L*+H H-H% L*+H H%

Although the system proposed by Grabe for Southern Standard British English offers a considerable simplification in terms of the basic notation compared with the original MAE_ToBI, it must be borne in mind that for the description of the actual realisations of the possible contours, the addition of modifications is essential, and so the notational conventions for the written representation of any such modifications must also be considered part of the basic notational system.

Unfortunately, to our knowledge, Grabe's model has never served as the basis for materials for teaching and learning Standard British English / RP intonation, so one can only speculate about its strengths and shortcomings in the context of teaching and learning L2 prosody.

4.3 "Standard British" vs. ToBI

The need to annotate the prosody of large-scale (English) speech databases has prompted Roach (1994) to examine the possibility of converting automatically between the Standard British system and ToBI. In his discussion of the ToBI pitch accents, Roach presents them as follows:

H* – non-low accented syllable;

L* – accented, but using the lowest pitch of a speaker's range;

L* + H – "scooped accent", a low tone which is followed immediately by a sharp rise to a high peak;

L + H* – "rising peak accent", a high peak target on the accented syllable which is immediately preceded by a sharp rise from a valley in the lowest part of the pitch range;

H + !H* – a clear step down onto the accented syllable from a high pitch which cannot itself be accounted for by any preceding accent.

Combinations of a phrase accent and a boundary tone at the end of an intonational phrase include:

L-L% – low ending;

L-H% – "continuation rise";

H-H% – high-rising ending, as in yes-no questions;

H-L% – falling; the H tone raises the final L to mid.

Combining the pitch accents with ToBI's phrase accents H- and L-, and with the boundary tones H% and L%, Roach suggests the following ToBI equivalents for the traditional (nuclear) tonetic stress marks used by proponents of the British school (see Table 4.4.).

The high level tone of the British school ends with L% because, according to the traditional ToBI conventions, if the sequence ends with a H% boundary tone, it would be a rise.

Table 4.4. Nuclear tones (left-hand column) and their suggested ToBI equivalent representations (after Roach 1994).

Tone name	Pitch accent	Phrase accent	Boundary tone
Low level	L*	L-	L%
High level	H*	H-	L%
(High) rise-fall	L*+H	L-	L%
High fall-rise	H*	!H-	H%
High fall	H*	L-	L%
Low fall	!H*	L-	L%
High rise	H*	H-	H%
Low rise	L*	L-	H%
Low fall-rise	!H*	L-	H%

Pre-nuclear tones are treated by Roach as combinations of a pitch accent and a phrase accent (Table 4.5.)

Table 4.5. Pre-nuclear tones (left-hand column) and their suggested ToBI equivalent representations (after Roach 1994).

Tone name	**Pitch accent**	**Phrase accent**
(High) rise-fall	L*+H	L-
High fall-rise	(?)	(?)
High fall	H*	L-
Low fall	!H*	L-
High rise	!H*	H-
Low rise	L*	H-
Low fall-rise	(?)	(?)

Roach argues that the pre-nuclear fall-rise tones present a problem (hence the (?) notation in Table 4.5.) because the obvious choice for their representation is the sequence H*+L H-, but the ToBI system has dispensed with Pierrehumbert's (1980) H*+L pitch accent. Level pre-nuclear tones have been excluded from the conversion, but if necessary, they could be represented as L* L- for low level and H* H- for high level.

The experiment which Roach conducted for limited automatic conversion of the prosodic symbols of the Spoken English Corpus showed that such conversion is indeed feasible. Unfortunately, the applicability of the equivalences between the British School system for intonation analysis and annotation and the ToBI system proposed by Roach remains to be tested outside the realms of speech corpora annotation.

To summarise, as a result of the extensive work on analysing and describing the prosody of English speech, traditional approaches to teaching it in the L2 classroom, and in particular the British School approach, still predominate today. As admitted by some language educators, although work within the framework of the Autosegmental-Metrical model of intonational phonology has offered new perspectives on research into language acquisition, it is also true that "language instructors who attempt to apply this model to teaching intonation may be daunted by the terminology, notation and abstractness of the theory" (Chun 2002, pp. 42-43). They find it difficult to learn the rules that need to be applied to the string of Hs and Ls in order for the surface-level pitch contour to be obtained – which is what is ultimately taught to the learner. Therefore, with some minor exceptions, ToBI and Autosegmental-Metrical phonology remain largely for the researcher to use. Nevertheless, they can be extremely useful for describing and comparing the similarities and differences between the prosodic systems of the mother tongue and the foreign language, and for predicting the occurrence of deviations from the prosody of the L2 in learners'productions.

In an attempt to make some predictions regarding the difficulties of Bulgarian learners in the field of English prosody, we next use the Autosegmental-Metrical theoretical approach and the ToBI framework in order to describe the intonational system of Contemporary Standard Bulgarian, and to compare it with that of English.

4.4 An Autosegmental-Metrical analysis of the prosody of Contemporary Standard Bulgarian

The ToBI system for Bulgarian (BG_ToBI) is based on the autosegmental-metrical framework of intonational phonology and follows the conventions described in Pierrehumbert (1980), Beckman and Pierrehumbert (1986), Beckman et al. (2005), Ladd (1996, 2008), among others. It is hoped that, among other things, it will equip language educators and researchers who work in applied linguistics with a useful tool for carrying out contrastive analyses of Bulgarian and other languages with a view to facilitating the acquisition of prosody of the foreign language, or of that of Bulgarian as L2.

The linguistic variety which is described is contemporary standard Bulgarian, as spoken by educated Bulgarians mainly in the capital Sofia, and also by the majority of newsreaders and announcers on major radio and TV networks in the country, such as Bulgarian National Radio and Bulgarian National Television.

While largely taking into account previous research on the prosody of Bulgarian, the system is predominantly based on more recent empirical research, and thus attempts to reflect the latest prosodic characteristics and developments which have taken place in the language in the last couple of decades. Such recent research has been based on corpora specially collected for the purpose which include read as well as semi-spontaneous speech.

The outline of BG_ToBI below follows closely the description in Andreeva and Dimitrova (2022b).

It is a prerequisite and a general requirement that a ToBI system for a given language be based on a sound body of previous research. Previous work on Bulgarian intonation has included a number of configuration-based accounts. Among the first scholars who discuss Bulgarian intonation is Stoykov (1942, 1966) who adopts a syntactically based approach to the description of the main melodic contours of declaratives, imperatives and interrogatives. Us-

ing instrumental and auditory analysis, Stoykov shows that the contour of simple declarative utterances is a gradually falling one, and that of non-final clauses in a complex or compound declarative sentence is a rising one signaling incompleteness. Questions formed with a question word or the interrogative particle „li" have the same contour as simple declaratives, while those without a question word or particle have a rising intonation contour. Imperatives are distinguished from declaratives on the basis of their wider pitch range and more abrupt changes of pitch direction (Stoykov 1966, p. 157).

Several later studies carried out in the 1960s and 1970s follow the same descriptive approach based on sentence types and syntactic categories, and rely heavily on auditory analysis, often supported by acoustic measurements (Popov 1963, Georgieva 1967, 1970, Georgieva 1974, Mahrova 1978). In her typological account of the intonation of Slavonic languages, Nikolaeva (1977) measures the fundamental frequency, intensity and duration of accented and unaccented syllables, and on the basis of the experimental results describes the melody of different sentence types.

Tilkov also views intonation as a complex linguistic phenomenon which involves fundamental frequency features (melody, range and register), intensity characteristics (word stress, 'logical stress' and 'phrasal stress') and temporal characteristics (pauses and tempo) (Tilkov 1981, p. 23). He uses the term 'logical stress' for the emphatic realisation of the nuclear syllables in narrowly focused words, and 'phrasal stress' – for non-emphatic nuclear accent. Although Tilkov describes the general direction of the pitch movement in the pre-nuclear part of the utterance, he doesn't explicitly mention pre-nuclear accent(s) and/or their type(s).

Following the tradition of syntactically based investigation of intonation, Misheva (1991) uses short utterances comprising one to five syllables, and varies the position of stress in order to study the

role of the intonation contour in signaling the communicative type of utterance: statement vs. command vs. question. Misheva analyses both their global and local prosodic characteristics. She finds the same global rising-falling contour in all three utterance types. However, she reports statistically significant differences in local characteristics such as the alignment of the F0 maxima with the accented syllable: in statements the peak is reached early in the accented syllable or in the pre-tonic one, whereas in questions it is reached late in the accented syllable or in the post-tonic one. Misheva also manipulates the word order (SVO and OVS) in short declarative sentences to investigate what she terms the 'focusing' function of intonation and interprets the results using the concepts of theme and rheme. Like Tilkov (1981), she describes the general direction of the F0 movement in the thematic part of the utterance with respect to the focused syllable, and concludes that it is the absence of accentual prominence which is the linguistically relevant feature of themes (Misheva 1991, p. 137). In the rheme, on the other hand, she notes that there is always accentual highlighting. Misheva and Nikov (1998) observe that rhemes are phonetically characterized by the same accentual pattern independent of the focus type – neutral (in broad focus), contrastive, emphatic (in narrow focus), the only difference being the tonal contrast between the accented and the unaccented syllables of the rheme as well as differences in phrasing. This has been disproved by later autosegmental-metrical analyses of Bulgarian intonation.

Misheva and Nikov (1998) distinguish semantic accents (corresponding to pitch accents in Autosegmental-Metrical analysis) and phrase accents (corresponding to boundary tones). The main unit of phrasing in their system is the syntagm (corresponding to the intonation phrase). The syntagm is defined in semantic and syntactic terms as consisting of one or more prosodic words organized by intonation and delimited by two types of pauses: objective (silent) pauses and

subjective pauses, caused by an interruption of the continuity of the prosodic features in the speech signal.

A major problem of these configuration-based approaches to the analysis of Bulgarian intonation is that they do not attempt to establish an inventory of intonation units typical of the language. Also, the fact that they are all essentially phonetic approaches leads to rather general conclusions, most of which only confirm already well-known generalisations.

The most significant early level-based account of Bulgarian intonation is that of Penchev (1980). Much of Penchev's work is dedicated to information structure and the role of intonation in indicating theme – rheme relations in Bulgarian sentences, starting from the assumption that both intonation and word order are very important for signaling information structure in the language. His investigation of Bulgarian intonation follows the American structuralist tradition (Pike 1945, Trager and Smith 1951). Penchev posits five pitch levels (level 1 is the lowest and level 5 – the highest) and describes the main intonation contours in Bulgarian in terms of the movement of the pitch from one level to another in several parts of the phrase, namely, the beginning, the pre-central part, the centre, and the post-central part.

Penchev describes six neutral (depending on the focus position) and four emphatic (regardless of the focus position) intonation contours. Contours 1 and 2 are both falling, but while in contour 1 the pre-central part is lower than the centre, contour 2 typically has a high beginning and the pitch gradually falls to the bottom of the speaker's range. Contour 3 begins at a mid level and rises gradually. Contours 4 and 6 are similar to contours 1 and 2, respectively. The difference is that contour 4 and 6 both have final rises. Contour 5, like contour 3, begins at a mid level but instead of gradual rise it ends with a fall-rise. Contours 4, 5, and 6 typically signal non-finality and occur in utterances which are divided into two or more intonation phrases.

The emphatic contours in Penchev's system are modifications of the neutral ones and signal contrast or speaker attitude. Penchev views contrast as the negation of the information provided in the previous context. It can be signaled by strengthening, either through an increase of F0 on the accented syllable, or through a decrease of F0 (deaccentuation or, in Penchev's terminology, 'de-rhematisation') on the preceding content word, both of which result in a greater pitch difference between the focused and given parts of the utterance.

A disadvantage of Penchev's approach is the fact that his classification criteria are so interrelated that, given the appropriate linguistic context, one can ultimately assign several meanings to almost any contour without knowing which elements contributed to their interaction.

The autosegmental-metrical approach and the ToBI transcription system have provided a basis for research on Bulgarian intonation carried out by Andreeva et al. (2001, 2016), Andreeva (2007, 2009), Dimitrova and Jun (2015), Dimitrova et al. (2018), among others, establishing relations between phonological tonal categories, their phonetic realisation and their information-structural functions. The corpora on which the analyses are based contain both semi-spontaneous speech acquired in map tasks (Anderson et al. 1991) and strictly-controlled read speech data (sentences in different focus conditions, passages) produced by speakers of contemporary standard Bulgarian born and raised in Sofia.

In line with the original MAE_ToBI system (Beckman et al. 2005), the model builds upon an on-ramp analysis and employs both leading and trailing tones. An inventory of five pitch accents (L*, L*+H, L+H*, H*, and H+!H*), two phrase accents (L-, H-), and one initial and two final boundary tones (%H, L%, H%) is derived from the combined analysis of the data. The inventory is defined with respect to the information structure and the various communicative sentence types which are investigated.

In addition to the audio recording and the pitch track, a full BG_ ToBI record of an utterance includes the following tiers:

(1) A tone tier on which the pitch accents and the edge tones (phrase accents and boundary tones) are transcribed using the conventional ToBI symbols and the modifiers for alignment (early/late) and scaling (upstep/downstep);

(2) A syllable tier with the phonetic transcription of each syllable. The division into syllables follows the conventions of Bulgarian phonetics;

(3) A word tier: Because the Bulgarian alphabet is based on the Cyrillic script, our examples are transliterated in italics following the Bulgarian Transliteration Law conventions (2009). A translation is also given in the figure captions;

(4) A break index tier: it is used for recording phrase boundary strength information;

(5) A miscellaneous tier: it is used if the transcriber needs to include notes, mark disfluencies, etc.

BG_ToBI includes the following phonemic pitch accents:

1. H*

This pitch accent is predominantly used in declarative sentences in Bulgarian and signals new information. The accented syllable is perceived as high. Phonetically, the H* is manifested as a peak on the accented syllable and is preceded by a slightly rising onglide with no clear low target before the peak (Figure 4.2.). The phonetic realisation of this accent is variable. This variability is especially noticeable in contexts of tonal crowding as reported by Andreeva et al. (2016). In their data, when the accented syllable is early in the sentence, the H target is reached close to the end of the accented syllable in 57 % of the cases, in the middle – in 23% of the cases, and close to the beginning – in 20 % of the cases. When the focus

is realised late in the sentence, the H target is reached close to the beginning of the accented syllable in 75% of the cases, in the middle – in 21% of the cases, and close to its end – in only 4 % of the cases. Results from auditory tests reported by Misheva and Nikov (1998) show that statements in which the focused word occurs early and the peak is aligned at the beginning of the syllable are perceived as more confident answers compared with statements of the same structure in which the peak is aligned later.

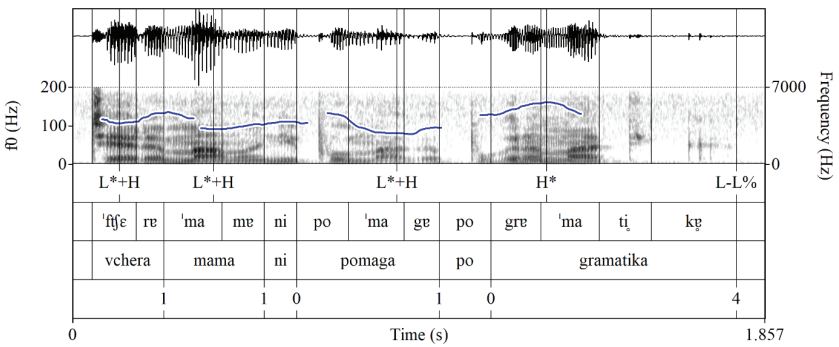


Figure 4.2. Waveform, spectrogram, and F0 contour of the broad-focus statement "Вчера мама ни помага по граматика." ('Yesterday mum helped us in grammar.')

2. L+H*

This pitch accent signals new information and is realised as a sharp rise (or a jump) from a low target in the preceding syllable or at the very beginning of the accented syllable up to a high target reached late in or just after the accented syllable (Figure 4.3.).

Researchers have different views on whether H* and L+H* are categorically different pitch accents or whether they are just two extremes of one and the same accent type. Pierrehumbert (1980) and Pierrehumbert and Hirschberg (1990) claim that only L+H* can

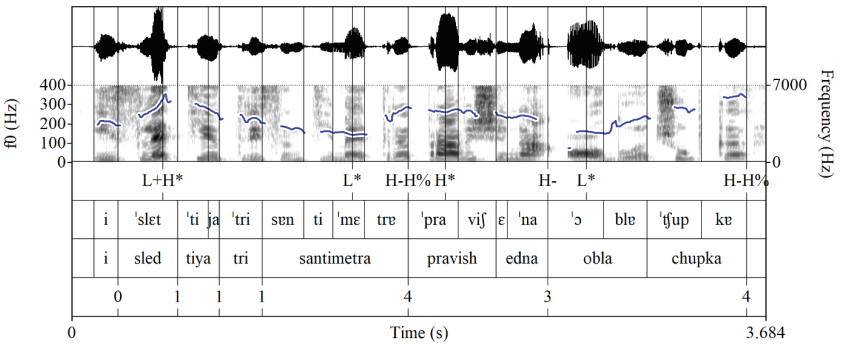


Figure 4.3. Waveform, spectrogram, and F0 contour of the utterance "И след тия три сантиметра правиш една обла чупка …" ('And after these three centimeters, you make a round turn …')

be preceded by a low target, while the statistical evidence provided by Ladd and Schepman (2003) shows that this is also true for H*. The question whether these two accents are associated with different meanings is also debatable. For Bulgarian, Andreeva and Oliver (2005) and Oliver and Andreeva (2008) have shown that the domains of interpretation of H* and L+H* overlap. Both accent types can signal either new information or a presence of contrast. Some speakers show a clear preference for L+H* pitch accents in narrow contrastive focus conditions, and for H* in narrow non-contrastive focus conditions, whereas other speakers show a strong tendency towards realising L+H* in both narrow non-contrastive and contrastive focus.

Dimitrova and Jun (2015) report a variant of the bitonal L+H*, namely LH*, where both the L and the H target are aligned with the syllable edges, and there is at least 10 Hz F0 rise within the accented syllable. However, their perception experiments demonstrate the functional similarity of L+H* and LH* in narrow non-contrastive and narrow contrastive focus, and of H* and LH* in broad as well as in narrow non-contrastive and contrastive focus. Thus, Dimitrova

and Jun's results confirm previous findings regarding the functional equivalence of H* and L+H*.

3. L*

This pitch accent is characterized by a shallow fall and is realised as a local low pitch in the lower third of the speaker's range. This accent appears mostly in nuclear position in some open questions with small degree or absence of speakers' confidence, and before a continuation rise (Figure 4.3.).

4. L*+H

The starred L tone of this bitonal pitch accent is aligned within or slightly before the accented syllable and the trailing H tone is aligned in the first post-tonic syllable or after it. In contrast to L+H*, the perceived pitch of the accented syllable in L*+H is low. This accent type is one of the frequently used pitch accents in the pre-nuclear position, but L*, H* and L+H* are also observed (Andreeva 2007, Andreeva et al. 2016, Dimitrova and Andreeva 2017, Dimitrova et al. 2018). These findings refute earlier claims by Misheva (1991),

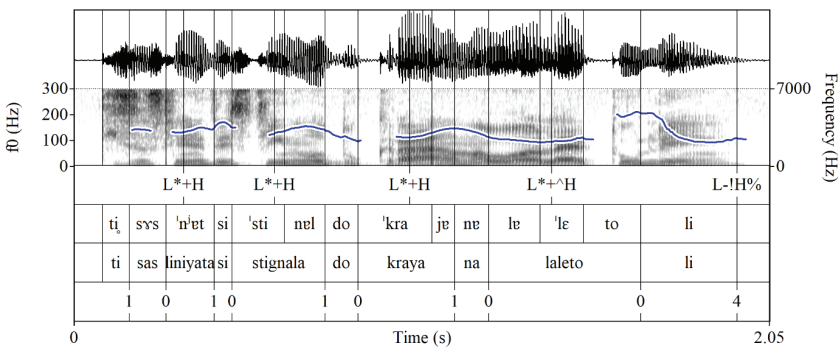


Figure 4.4. Waveform, spectrogram, and F0 contour of the narrow focus yes-no question "Ти с линията си стигнала до края на лалето ли?" ('Have you got with the line to the end of the tulip?')

Misheva and Nikov (1998) and Penchev (1980) that the pre-nuclear part of the Bulgarian declarative utterance is unaccented. L*+H also occurs in nuclear position in yes-no questions but is usually realised with a wider pitch range (Figure 4.4.).

Dimitrova and Jun (2015) discuss the variable alignment of the high trailing tone in the L*+H pitch accent, which in their data was sometimes realised as far to the right as the second post-tonic syllable. They suggest that the H tone may be a phrasal accent. Such variability of tonal alignment in free stress languages is not unknown (see, for example, Themistocleous 2016 on Cypriot Greek). Dimitrova and Andreeva (2017) find a tendency for the trailing tone of the L*+H pitch accent to align later with slower speaking rate for most but not all of their speakers. This shows that the H target of the bitonal pitch accent is not separated by a fixed distance and/or a fixed time interval from the starred tone, as postulated by Pierrehumbert's invariance hypothesis (Pierrehumbert 1980, p. 80). Therefore, Andreeva and Dimitrova (2022b) hypothesise an anchorage domain (Welby and Loevenbruck 2005, 2006, Themistocleous 2016) of the L*+H, namely, the prosodic word, where the L* is aligned with the onset of the lexically stressed syllable (or just before it), while the H trailing tone aligns with a following unstressed syllable within the domain of the prosodic word. However, they also report counter-examples in which the H spreads to the first or second syllable of the next prosodic word.

Data in the research literature designed to test the stability of tonal alignment often involves the repetition of long sentence lists. Andreeva and Dimitrova's (2022b) data, on the other hand, come from a variety of sources, including semi-spontaneous speech, which is why they suggest that the repetition of long sentence lists may lead to training effects and more stable productions by speakers than is usual in everyday speech, and that this may provide an explanation for the variability in their data.

Figure 4.5. illustrates the variability in the alignment of the H target in three consecutive pre-nuclear L*+H pitch accents. It may be a clearly high target, which in the first pitch accent is aligned at the beginning of the second post-tonic syllable, in the second pitch accent – with the vowel of the first post-tonic syllable and in the third pitch accent – with the vowel onset of the third post-tonic syllable. Sometimes the high target may be even hard to detect, as in the third pitch accent in Figure 4.2., where it is part of the gradual rise to the next upstepped high target. It is important to note that upstep in Bulgarian is a gradual phonetic modification of the pitch which depends on the communicative intention of the speaker.

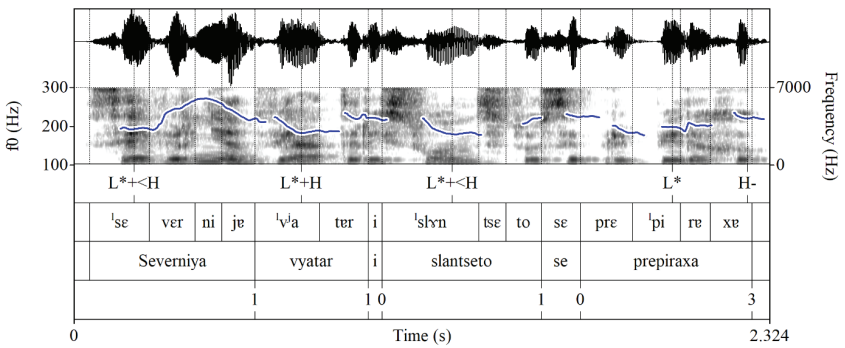


Figure 4.5. Waveform, spectrogram, and F0 contour of the utterance "Северният вятър и слънцето" ('The North Wind and the Sun')

5. H+!H*

This pitch accent is realised as a fall from a high pitch target on the pre-tonic syllable which continues throughout the accented syllable and usually ends in the post-accented one. The height of the pre-tonic syllable may be a high target or the end of a plateau (Figure 4.6.). Some speakers may complete the fall earlier, around the offset of the accented syllable. This is the reason why in earlier

autosegmental-metrical analyses this realisation of the tone has been labeled as H+L* (Andreeva et al. 2016). Indeed, the accented syllable in H+L* is perceived as much lower than in H+!H*, which may be due to the gradient difference in the prominence of the pitch accented syllable. The two realisational variants are predominantly used in nuclear position in broad focus. H+!H* is used to confirm a fact and is more neutral. Its variant realisation as [H+L*] is more typical of concluding statements, expressing definiteness. The difference between the two is thus stylistic. However, a dedicated perception experiment is needed to confirm that speakers reliably hear the difference between the two variants. These realisations of early peaks are labelled as H+!H* because this variant is the unmarked one.

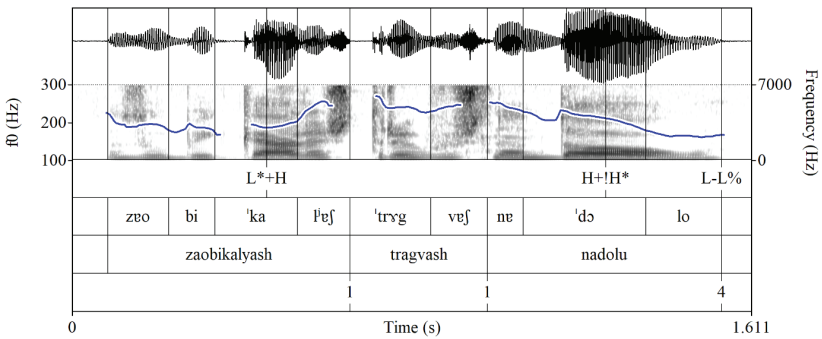


Figure 4.6. Waveform, spectrogram, and F0 contour of the broad focus statement "Заобикаляш, тръгваш надолу." ('You go round, go down.')

BG_ToBI posits two phrase accents associated with the right edge of the intermediate phrase:

L- is realised as a F0 minimum low in the speaker's range.

H- is realised as a F0 maximum which is roughly equal in height to the peak of the preceding H tone in the phrase. When the syllable

on which the nuclear H tone is realised is followed by several unstressed syllables, there is a high plateau which spreads to the right edge of the intermediate phrase.

In the BG_ToBI system there are also two boundary tones, low L% and high H%, associated with the right edge of the intonation phrase, and one initial boundary tone, high %H, associated with its left edge. A mid or low initial boundary tone is not labelled.

Combinations of a phrase accent and a boundary tone at the right edge of an intonation phrase yield more complex tonal configurations. An example of the L-H% final low rise is given in Figure 4.7. It is realised on a final unstressed syllable: in the first intonation phrase, this is the final syllable of a three-syllable word “обаче“ (‘however’) and is immediately preceded by the H* nuclear accent, whereas in the second intonation phrase, the low rise is preceded by as many as five unstressed syllables. For an example of the initial %H boundary tone see Figure 4.8.

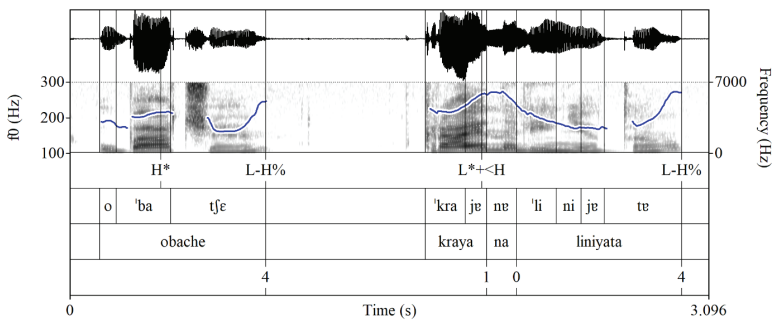


Figure 4.7. Waveform, spectrogram, and F0 contour of the utterance “Обаче … в края на линията …” (‘However … at the end of the line …’)

The H targets of the pitch accents, phrase accents and boundary tones described above may be modified through scaling of the

pitch range (downstep or upstep). In the case of downstep (!H), the top line of the pitch range is lowered, so that the H target is shifted downwards with respect to the preceding H tone (Figure 4.8.). In the case of upstep (^H), the H target is shifted upwards.

Additionally, the pitch accents can be modified through different timing (early or late) of the tonal targets. If the tonal target of a pitch accent occurs outside the syllable, the MAE_ToBI conventions (Beckman et al. 2005) are followed, and the label '>' is used when the tone occurs before the relevant syllable, or '<' when it occurs after it.

The symbols for the modifiers (!, ^, <, >) are placed before the affected tone.

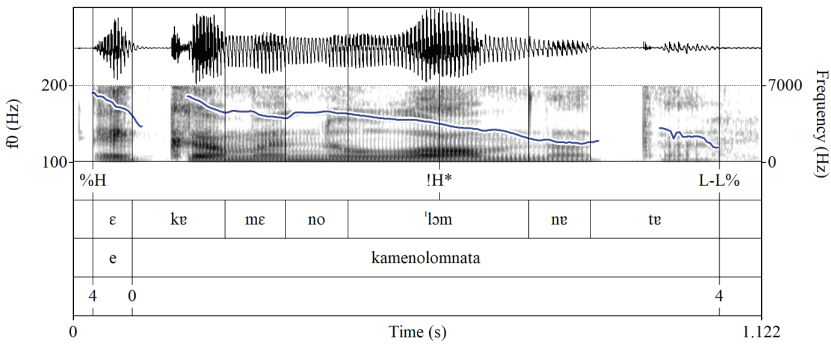


Figure 4.8. Waveform, spectrogram, and F0 contour of the narrow focus yes-no question with 'closed' meaning "… е каменоломната?" ('…. is the quarry?').

Tables 4.6. and 4.7. provide a summary of the phonological pitch accent and edge tone categories proposed for Bulgarian.

Table 4.6. Inventory of pitch accent categories in Bulgarian (the accented syllable is shaded)

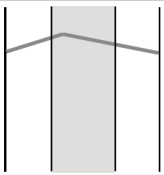
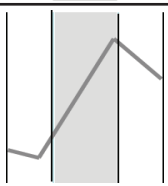
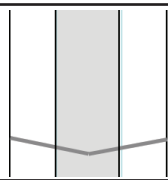
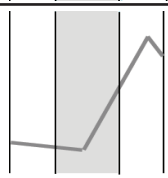
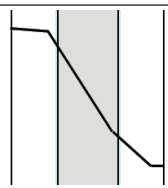
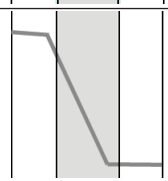
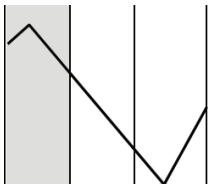
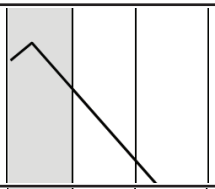
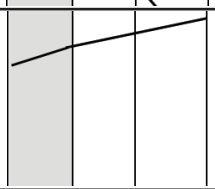
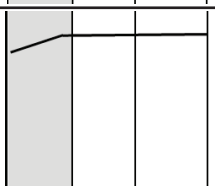
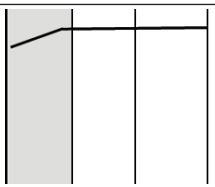
schematic representation σ ˈσ σ	**pitch accent type**	**description**
	H*	peak on the accented syllable, preceded by a slightly rising onglide with no clear low target before the peak
	L+H*	sharp rise (or jump) from a low target in the preceding syllable or at the very beginning of the accented syllable up to a high target late in or just after the accented syllable
	L*	local low pitch in the lower third of the speaker's range
	L*+H	valley within or slightly before the accented syllable, the trailing H tone is aligned in the first post-tonic syllable or after it
	H+!H*	a) fall from a high pitch target on the pre-tonic syllable to a downstepped target in the accented syllable
	[H+L*]	b) fall from a high pitch target on the pre-tonic syllable to a low target in the accented syllable

Table 4.7. Inventory of edge tone categories in Bulgarian, shown in the vicinity of H* (the accented syllable is shaded)

schematic representation σ σ σ	edge tones	description
	%H	phrase initial high boundary, realised on the first syllable of the intonation phrase
	L-L%	low stretch at the end of the intonation phrase
	L-H%	fall-rise ending in the upper part of the speaker's range
	H-H%	high rise at the end of the intonation phrase which reaches the upper part of the speaker's range
	H-L%	plateau

In accordance with general ToBI conventions, the degree of juncture perceived between words and phrases is encoded by means of the following break indices:

'0' is used for any juncture smaller than a lexical word boundary (i.e., for a juncture between clitics and their hosts, or between two clitics);

'1' marks the juncture corresponding to a prosodic word boundary;

'2' marks perceived juncture with no intonation effect, or apparent intonational boundary without a pause or any other clues;

'3' marks the juncture corresponding to an intermediate phrase boundary;

'4' marks the juncture corresponding to an intonational phrase boundary.

To sum up, BG_ToBI is a system proposed for transcribing Bulgarian intonation based on the general principles of the framework of Autosegmental-Metrical theory. The Bulgarian intonational system comprises an inventory of five pitch accents (L*, L*+H, L+H*, H*, H+!H*), two phrase accents (L- and H-) and three boundary tones (L%, H% and %H). The prosodic units are the intonation phrase and the intermediate phrase, with the prosodic word a hypothetical third constituent (see below). The default pre-nuclear pitch accent is L*+H, and the default nuclear one is H*. The default nuclear pattern for declaratives is H*/H+!H* L-L%, and for yes/no questions L*+H L-L%. Table 4.8. shows the nuclear configurations and stylised tunes along with their meanings.

Table 4.8. Inventory of nuclear configurations and stylised tunes and their meanings

BG_ToBI	Context/Meaning
H* L-L%	declaratives with broad, narrow non-contrastive and narrow contrastive focus, closed yes/no questions

H* L-H%	polite yes/no questions with 'closed' meaning, non-finality
H* H-H%	non-finality
^H* L-L%	emphasis
%H ^H* L-L%	lively introduction of new information
!H* L-L%	matter-of-fact statements
H+!H* L-L%	declaratives with broad and narrow non-contrastive focus, confirmation of a fact, concluding statement, definiteness
L+H* L-L%	declaratives with narrow non-contrastive and narrow contrastive focus, 'open' yes/no questions, wh-questions
L*+H L-L%	'extra-open' yes/no questions with and without Q-li
L*+H L-H%	yes/no questions with and without Q-li signalling incredulity, non-finality
L* H-H%	'extra-open' yes/no questions with and without Q-li, non-finality
(L*+H) !H* L-L%	wh-questions with post-focal accent (The focus-associated pitch accent on the wh-word is given in brackets.)
L+H* L-	neutral vocatives
L+<H* L-L% [+long]	insistent vocatives
L* H* L-L% [+long]	challenging chant
L+H* !H* H-L% [+long]	vocative chant

Bulgarian pitch accents are often characterized by variable alignment of the tonal target with the tone-bearing unit, which constitutes a challenge for the 'classical' autosegmental-metrical theory. This variability of the phonetic realisation is triggered by speaker-specific production strategies as well as by the position of the accented syllable within the phrase. For example, when the default H* pitch accent is early in the phrase, the H target is usually reached close to the end of the accented syllable. When it is late in the phrase, the H target is reached close to the beginning of the accented syllable. The trailing tone (H) of the L*+H pitch accent can be shifted to the right when it is followed by several unaccented syllables. An anchorage domain is hypothesised, namely, the prosodic word, where the L* is aligned with the onset of the lexically stressed syllable (or just before it), while the H trailing tone aligns with a following unstressed syllable within the domain of the prosodic word. However, counter-examples were also found in which the H spreads to the first or second syllable of the next prosodic word. What is needed are experiments with strictly controlled material in order to shed more light on the nature and size of this anchorage domain.

The combined analyses of data from read and semi-spontaneous speech corpora have made possible the study of the interaction between information structure and intonation, the importance of which is broadly recognised in the literature. The analyses show that there are different strategies for signaling the information structure of an utterance by implementing different combinations of word order variation and intonational category choices. Results obtained so far have shown that, on the one hand, speakers use different pitch accent categories to signal a specific focus type, and on the other hand, they use the same pitch accent category to signal different focus types but manipulate the strength of one or more of the acoustic properties, or suppress or enhance the prominence of surrounding words. Previous analyses of Bulgarian speech reveal

that nominal material replicated by a pronominal clitic in statements and yes-no questions, as well as repeated constituents in narrow-focused read sentences with canonical word order are usually accented in pre-nuclear and deaccented in post-nuclear position. Andreeva (2017) reports that speakers produce pre-nuclear accents on contextually given material in 89% of the cases in non-contrastive and 86% of the cases in contrastive narrow focus. These findings on lack of deaccentuation in Bulgarian are confirmed by the AM analysis of Bulgarian intonation presented above. In addition, the analysis of material after the focused interrogative word even in short wh-questions was found to have post-focal prominence. This phenomenon has also been reported for other Balkan and Slavonic languages (Ladd 2008).

The analysis of Bulgarian vocatives found additional prominence on the last unstressed syllable which is attributable to the lack of qualitative vowel reduction of Bulgarian unstressed vowels, accompanied by lengthening of the syllable necessary for the realization of the vocative chant tune. The combined effect of the two creates the perceptual impression of prominence, and eventually results in the readjustment of the metrical structure through the addition of a metrically strong final syllable which has been analyzed as also carrying a pitch accent.

In conclusion, the BG_ToBI system outlined above reflects the current state of knowledge about the prosody of contemporary Standard Bulgarian. However, more systematic research is needed on both the phonetics and phonology of intonation and the interplay of intonation, syntax and information structure in the language.

5. The prosody of English and Bulgarian: a comparison and some predictions

This chapter draws a comparison between the prosodic characteristics of English and Bulgarian with the aim of reaching some conclusions and making predictions about some of the possible difficulties which (advanced) Bulgarian learners of English may experience with the acquisition of the prosodic features of the L2. The accents which are compared are (Mainstream) American English and Contemporary Standard Bulgarian. For the time being, these are the relevant accents whose prosodic systems have systematically been described within the same analytical framework, as discussed in the previous chapter – that of Autosegmental-Metrical phonology, using the ToBI annotation tool. The comparison will be drawn following the methodology outlined by Mennen (2015) in the L2 Intonation Learning theory, as summarised briefly below.

Languages (and accents of a given language) can differ both in terms of the number of structural elements (pitch accents, phrasal accents, or boundary tones) in them or in terms of the phonetic implementation of those elements. Prosodic phonological differences between languages or language varieties are categorical and can involve differences in the inventory of distinct phonological elements, their form, or their meaning. Prosodic phonetic differences, on the other hand, are considered to be gradient in nature (Ladd 1996). This means that prosodic interference on the phonological level would involve transfer as a consequence of differing inventories of tunes in the L1 and the L2, differing forms of these tunes, or different meanings of the tunes, whereas interference on the phonetic level would involve transfer due to a different phonetic implementation of the same tune in the L1 and the L2.

Modifying Ladd's (1996) parameters of cross-language variation, the L2 Intonation Learning Theory recognises the following four dimensions in terms of which the prosodic similarities and differences between a learner's L1 and L2 can be described:

(i) the systemic dimension, which involves the inventory of phonological categories and their distribution in the two languages;
(ii) the realisational dimension, which characterizes the phonetic implementation of the languages' categorical phonological elements;
(iii) the semantic dimension, which is concerned with the functional importance of the phonological categories;
(iv) the frequency dimension, which considers the frequency of use of the categorical elements.

5.1 The systemic dimension

We start by comparing the inventories of prosodic phonological elements of English and Bulgarian along the systemic/phonological dimension. These include the pitch accents, phrasal accents and boundary tones, as well as the units of the prosodic hierarchy. A side-by-side comparison of the tone inventories of the two languages (see Table 5.1) reveals a lot of similarities.

As can be seen from Table 5.1. (Mainstream American) English and (Contemporary Standard) Bulgarian share virtually the same inventory of pitch accent types. All five pitch accents can be found in both nuclear and pre-nuclear position, though their frequency of occurrence in the two languages may be different (see below). Although the table for American English lists all variants of the pitch accents which involve a downstepped H pitch target, while the Bulgarian one mentions only the downstepped allophone [!H*] of the monotonal H*, virtually all of the other downstepped realisations of the American pitch accents have been found in Bulgarian as well.

Table 5.1. The phonological tones in English and Bulgarian (allophones are shown in brackets)

	(Manistream American) English	**Contemporary Standard Bulgarian**
pitch accents	L* H* (!H*) L+H* (L+!H*) L*+H (L*+!H) H+!H*	L* H* (!H*) L+H* L*+H H+!H* (H+L*)
phrase accents	L- H- (!H-)	L- H-
boundary tones	L% H% %H	L% H% %H

Based on the above comparison, our prediction is that on the systemic phonological level Bulgarian learners of English are unlikely to face any major problems with the acquisition of the English pre-nuclear and nuclear pitch accent types. The same prediction can be made with regard to the inventories of phrase accents and of boundary tones, since they are also phonologically the same in the two languages which are being compared.

A comparison of the inventory of boundary tones for marking juncture in English and Bulgarian reflects the recent proposal made by Andreeva and Dimitrova (2022b) that in Bulgarian, the prosodic

word may be part of the prosodic hierarchy. Therefore, the definitions of the numerical values for lower-level juncture demarcation are different in the two languages. Any juncture which is smaller than a regular boundary between two lexical words, such as the boundary between a host (lexical word) and a clitic, or between two clitics, is marked as '0' in both, but in MAE_ToBI (Beckman et al. 2005) it is defined simply as "very close connection between words". Note, however, that earlier accounts of juncture in MAE, such as Hirschberg and Beckman (1994) do reserve this annotation for "cases of clear phonetic marks of clitic groups", e.g., the medial affricate in contractions of "did you", or a flap as in "got it".

In Bulgarian, "1" is reserved for juncture corresponding to prosodic word boundary, whereas in English it is used to indicate "ordinary phrase-internal juncture". The other three numerical values for the annotation of constituent boundaries are defined in the same way:

"2" is reserved for marking cases of tone – breaks mismatch,

"3" is used for marking an intermediate phrase boundary,

"4" is used for the annotation of an intonational phrase boundary.

The prosodic hierarchy for English posits the following constituents:

σ = syllable;

W = (lexical) word;

ip = intermediate phrase;

IP = Intonational Phrase;

T = tones: T* = pitch accent, T- = phrase accent, T% = boundary tone (Figure 5.1).

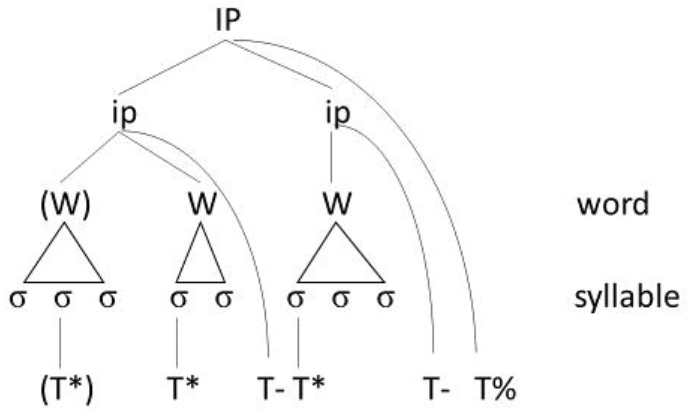


Figure 5.1. Hierarchical structure of English intonation

By comparison, the prosodic hierarchy for Bulgarian will include an additional level between that of the word and the intermediate phrase – the level of the prosodic word (ω / PW) – a unit composed of a host and its clitic(s) (see Figure 5.2.).

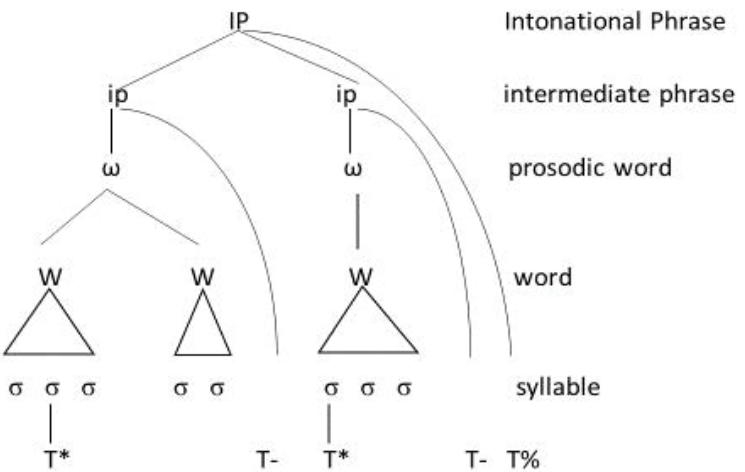


Figure 5.2. Proposed hierarchical structure of Bulgarian intonation, including the prosodic word.

We predict that the different hierarchies of prosodic units in the two languages are likely to cause problems for Bulgarian learners of English, and deviances from the native English norms are to be expected in the speech of Bulgarian learners of the language due to L1 interference at the level of the prosodic word.

A discussion of the similarities and differences in the systemic dimension must also include the ways in which the structural elements combine with each other, that is, it must specify the "tonal phonotactics" which tells us what tunes are permissible in the language. It must also specify the "tune-text association" (Ladd 1996) of the tunes – how the tones are realised with respect to the segmental string of the utterance. A detailed and systematic comparison between English and Bulgarian along these lines is rather difficult to draw at the moment because we lack comprehensive lists of all tonal combinations allowed in the two languages. However, some parallels can be made on the basis of existing analyses.

All four possible combinations of a phrase accent and a boundary tone are permissible in both languages:

L-L% – a low ending typically occurring at the end of declaratives;

L-H% – the typical "continuation rise";

H-H% – high-rising ending, as in yes-no questions; it can also signal non-finality;

H-L% – a mid plateau.

Some frequently occurring combinations of nuclear pitch accents, phrase accents and boundary tones attested in both languages are given below, along with their interpretation (the semantic dimension):

H* L-L% – the neutral declarative pattern;

H* H-H% – a combination used to signal non-finality, or for questioning;

L* H-H% – the canonical yes-no question tune;

L*+H L-H% – a combination which can be used in both languages to convey uncertainty or incredulity.

These examples do not by any means exhaust the list of similarities between English and Bulgarian in terms of the language-specific combinations of tones. The compilation of exhaustive lists of the permissible combinations in English and Bulgarian is a topic for further research, and can reveal important differences between the two languages.

The main tune-text association patterns in MAE_ToBI and BG_ToBI also share a number of similarities:

- In both languages, a starred tone is associated with a lexically stressed syllable, whereas the leading or the trailing tone of a bitonal pitch accent is associated with an unstressed syllable which precedes or, respectively, follows the stressed syllable;
- L*+H shows F0 minimum within the stressed syllable;
- L+H* shows F0 minimum before the stressed syllable and F0 peak (maximum) within the stressed syllable;
- The starred tone of the H+!H* pitch accent is realised as mid and is preceded by a high target;
- Syllables not associated with a tone (tonally underspecified syllables) receive their surface F0 by interpolation between the pitch accents which precede and follow them (however, F0 may sag between two consecutive H* pitch accents);
- Phrase accents are realised over the (stretch of) syllable(s) immediately after the nuclear pitch-accented word up to the end of the phrase;
- A boundary tone is an obligatory property of the intonational phrase and is realised on the final syllable of the intonational phrase;

- The IP-initial boundary tone %H is optional in both languages;
- Upstep and downstep occur in both languages.

A comparison of the phonological models of (American) English and Bulgarian intonation reveals a number of similarities between the two languages in terms of their tonal inventories, the tonal phonotactic possibilities and tune-text association. The major difference which emerged concerns the hypothesised existence in Bulgarian of the Prosodic Word as part of its prosodic hierarchy. Although further research is needed in order to confirm the hypothesis, this difference can cause deviations in Bulgarian-accented English from the native prosodic patterns of the target language.

5.2 The realisational dimension

The realisational, or phonetic, dimension of the LILt model enables the researcher to reveal cross-language similarities or dissimilarities in the way in which the elements of the prosodic system are phonetically implemented. Examples of phonetic implementation are the actual tonal alignment of the pitch accents – that is, how they are lined up with the segmental string of an utterance; their scaling – that is, their height relative to neighbouring pitch events in an utterance; their shape or slope – that is, whether they rise or fall steeply or gradually. Mennen (2015) reviews research on cross-language differences in the realisational dimension of intonation, reporting numerous instances of dissimilarities, especially in the way in which the tune is coordinated with the segments of an utterance or with syllable boundaries.

An example of dissimilarity between English and Bulgarian in the realisational dimension is the alignment of the H* peak: in English, the high target of the H* pitch accent is reported to occur towards the end of a syllable (Pierrehumbert 1980, Beckman and

Pierrehumbert 1986). In Bulgarian, Andreeva and Dimitrova (2022b) report that when H* is early in the phrase, the H target is usually reached close to the end of the accented syllable, but when it is late in the phrase, the H target is reached close to the beginning of the accented syllable.

Another difference between English and Bulgarian concerns the alignment of the H tone in the bitonal L*+H pitch accent when the accent is pre-nuclear. In Bulgarian, the trailing tone (H) can be shifted to the right when it is followed by several unaccented syllables. In English, the high target is usually reached within the first post-accented syllable. This difference can also cause deviations in the English speech of Bulgarian learners due to L1 interference.

In general, Bulgarian pitch accents are described by Andreeva and Dimitrova (2002b) as often being characterized by variable alignment of the tonal target with the tone-bearing unit. This variability of the phonetic realisation is triggered by speaker-specific production strategies as well as by the position of the accented syllable within the phrase, and due to transfer from the L1 is likely to cause deviations in Bulgarian-accented English speech.

5.3 The semantic dimension

Before the advent of Autosegmental-Metrical phonology, researchers tended to describe the functions of the categorical elements – usually whole tunes – in terms of speaker attitudes and emotions (e.g., showing surprise or being polite), or in terms of speech acts (e.g., making a statement or asking a question). However, this approach does not provide us with a sound basis for unambiguously characterizing the semantics of tunes. For example, the "falling contour" H* L-L% is used both in English and in Bulgarian as the default tune for declaratives as well as wh-questions.

Pierrehumber and Hirschberg (1990) take a different approach to the description of contours within the Autosegmental-Metrical

approach. They treat the meaning of the contour as compositional, resulting from the combined contribution of each of its parts – pitch accents, phrase accents and boundary tones. The choice of tune on the part of the speaker is determined by his/her desire to convey a particular relationship between an utterance, the current mutual beliefs of the participants in the discourse, and the anticipated contributions of subsequent utterances. The speaker accents an item in order to signal its salience in the ongoing discourse. The type of accent conveys information status – whether the accented item should be included amongst the participants' mutual beliefs, or whether it should be excluded, or made inferable. For example, items marked with H* should be treated as adding "new" information to the discourse which should be added to the hearer's mutual belief space. The speaker would mark items with L* in order to make them salient, but also to signal that they convey "old" information which he/she believes is already part of the hearer's mutual beliefs. Combining a pitch accent with different edge tones, however, can bring about different interpretations, as illustrated by the two tunes (i) and (ii) with which B's reply in the following short dialogue can be uttered:

A: Did you like the film?

B: I thought it was good.

(i) H* H* L-L%

(ii) H* H* H-H%

If pronounced with the tones in (i), B's reply will be a straightforward answer to A's question which gives the required new information. However, if said with the tones in (ii), it can be glossed as "I thought it was good, but do you agree with me" – that is, it will still provide the required information, but in addition will ask for comment (example adapted from Pierrehumbert and Hirschberg 1990, p. 290).

One problem with Pierrehumbert and Hirschberg's approach is that it was never developed further after the publication of their programmatic paper in 1990. Another problem is posed by its relative complexity from the point of view of teaching intonational meanings to L2 learners. That is why in practice a combination with glosses for intonational functions from older descriptions of English intonation is often to be found both in teaching materials and in research work on the topic, including the present one.

The current state of our knowledge about the functionality of the structural elements or tunes hardly allows a detailed comparison between English and Bulgarian, and hence – any predictions about possible deviations in the L2.

Some examples of similarities between the two languages under investigation in the semantic dimension were already mentioned above.

An example of a dissimilarity is the "contradiction contour" in English, transcribed as L* L-H%, which is interpreted as conveying information which should have been mutually shared, but in fact is not. To quote another example from Pierrehumbert and Hirschberg (1990, p. 293):

A: Let's order the Chatteaubriand for two.

B: I don't eat beef.

L* L* L-H%

The use of the L+H* accent in combination with the L- phrase accent and the H% boundary tone to make a correction or contrast in English, as in B's reply in the next short conversation, is another potential case of dissimilarity between the two languages compared here.

A: Jane has passed all her exams.

B: That's what she claims.

L+H* L-H%

Yet another example of a dissimilarity likely to lead to deviation from the L1 tune in the L2 production of Bulgarian learners of English is the tune L*+H L-H% which can be used in English to express uncertainty, as in B's reply in the next dialogue (Pierrehumbert and Hirschberg 1990, p. 295):

A: Did you take out the garbage?

B: Sort of.

L*+H L-H%

The use of the tune to express incredulity has in fact been reported in both English and Bulgarian. Pierrehumbert and Hirschberg (1990) refer to work by Ward and Hirschberg who unify the "incredulous" and "uncertain" meaning of the tune and call it "lack of speaker commitment", with the difference between the two meanings being signaled by differences in pitch range and speech tempo. From the perspective of the LILt model, this is a further example of the interrelations between the model's dimensions. Such interrelations should always be kept in mind when making predictions about prosody in L2. In any case, what is needed in order to make informed predictions is a body of research into the intonational functions in both languages which has been conducted within the same analytic framework in order to enable systematic comparisons.

5.4 The frequency dimension

This dimension of the LILt is an addition to the dimensions initially proposed by Ladd (1996) and looks into cross-language similarities and differences in how often the elements of the prosodic system are used in a given language. It is well known that variability in the frequency of use exists even amongst language varieties which share the same tonal inventory. Thus, Grabe and Post (2002) found that, in their data from the IViE corpus, speakers from Cambridge pronounced declaratives with a fall (H*L % in the IViE nota-

tion) over 90% of the time, and with a fall-rise (H*L H%) the rest of the time. In comparison, English speakers from Belfast produced declaratives with rise-plateaux nuclear accents (L*H %) in 80% of the cases, and with a fall (H*L %) the rest of the time.

Im, Cole and Baumann's (2018) analysis of public speech from a TEDTalk by a male speaker of American English focuses on the relationship between pitch accent assignment and information status, and makes the following predictions:

H* – the pragmatically neutral pitch accent – was expected to mark all types of information status categories except givenness;

L+H* – which is typically used to mark contrastive or corrective focus – was expected to be associated with discourse new elements and alternative concepts;

!H* was expected to be used to mark accessible information;

L* was expected to be found with given categories – salient items which, however, are not to be added to a predication made by the speaker.

What Im et al. (2018) found was that (i) the assignment of a pitch accent to a word significantly distinguished it from words which do not carry information status; (ii) the assignment of a pitch accent significantly distinguished words that are given from words that are not given in the referential but not in the lexical condition; (iii) the type of pitch accent does not predict the meaning of the word in relation to the discourse context, so there is no strict one-to-one mapping between accent type and information status condition.

A comparison with data from the Buckeye corpus of conversational speech (Pitt et al. 2005) reveals that accent patterns vary across speech styles: about 50% of all words in the public speech sample were unaccented, while in the conversational speech sample they were about 75%. H* and L+H* were the most frequent pitch accents in the TEDTalk, they were followed by !H*, and the least

frequently used tone was L*. In conversational speech, the most frequently used pitch accent was H*, while L+H* was used more than twice less frequently than in public speech.

Dainora (2006), whose results strongly suggest that the nuclear pitch accent is a significant determinant of the following boundary tone, gives the following frequencies of distribution of the combination of nuclear pitch accent + phrase accent before a high boundary tone H% in a sample of 1207 phrases:

Table 5.2. Frequency of distribution of nuclear pitch accents and phrasal tones (adapted from Dainora 2006)

	Frequency of occurrence in sample of 1207 phrases	**Frequency of occurrence before H%**
H* L-	56%	41%
L+H* L-	23%	58%
L* L-	7%	84%
H+!H* L-	6%	26%
H* H-	5%	16%
L+H* H-	2%	12%

Some combinations were excluded from the above table, because their frequency of occurrence was less than 0.5% in Dainora's data, namely, L*+H L-, L* H-, L*+H H-, H+!H* H-.

In Bulgarian the most frequently occurring pre-nuclear pitch accent is L*+H, and the default nuclear one is H*. The default nuclear

pattern for declaratives is H*/H+!H* L-L%, and for yes/no questions L*+H L-L%. We can make some tentative predictions about possible deviations from the native English norms in the speech of Bulgarian learners, namely, that due to L1 interference they will use more frequently L*+H as a pre-nuclear pitch accent and H+!H* as a nuclear one in declaratives than native speakers of English. However, it should be borne in mind that the above results are based on analyses of speech samples from a relatively small number of speakers. Until results based on larger, comparable as well as stylistically diverse datasets become available, it will be difficult to draw reliable comparisons of the frequency of use of prosodic constituents in the frequency dimension.

5.5 Focus marking

Although focus and information structure were already mentioned briefly, it is worth discussing in some more detail the marking of focus and information status in English and Bulgrian, again with the purpose of predicting potential problems for Bulgarian learners. To cite from Gussenhoven, “Speakers conduct conversations so as to establish a common understanding with their hearers about some aspect of the world. In a discourse model, the speaker keeps track of the development of this common understanding, and labels his linguistic expressions for the way the information they convey relates to the information in the discourse model as developed at that point. In English, pitch accents are used for this purpose. Broadly, their location indicates the size of the ‘focus constituent’, which contains the constituent(s) whose information status is being signalled, while their distribution within the focus constituent expresses the type (or meaning) of the focus.” (Gussenhoven 2008, p. 83).

Depending on size, the focus can be “broad” or “narrow” (Ladd 1980, 1996), also sometimes referred to as “normal” or “contrastive”. In English, in broad focus the nuclear pitch accent is usually

placed on the last lexical item in the intonation phrase. Pre-nuclear accents, though not obligatory, often perform information structural functions such as topic marking. They are often placed on content words such as nouns, rather than on function words such as prepositions and conjunctions. Discourse-given items are regularly deaccented if they occur after the focused constituent. The default neutral declarative intonation pattern is H* L-L%. The most frequent pitch accent type used to mark a correction or contrast is L+H* (Pierrehumbert and Hirschberg 1990).

Early work on focus marking and information structure in Bulgarian has found that the underlying intonational pattern in broad focus declaratives is H* L-L%. In narrow focus the underlying nuclear pitch accent H* is realised with an emphasis [+raised peak]. In the case of contrastive narrow focus, the phonetic realisation of the underlying H* is <H* [+raised peak; +delayed peak]. In later studies, this pitch pattern is re-analysed as L+H*, and the nuclear pitch accent in broad focus as H+!H* or H+L*. Moreover, all studies report interspeaker variation in the phonological choice of the pitch accent type: occasional use of H* and !H* in broad focus, as well as variability between H* and L+H* in narrow contrastive and noncontrastive focus (Andreeva et al. 2001, Andreeva and Oliver 2005, Andreeva 2007, 2009, Oliver and Andreeva 2008, Dimitrova and Jun 2015). Apart from the secondarily associated phrase accent (L-), the tail remains phonologically unspecified.

To summarise, in order to signal information structure in Bulgarian, speakers use different pitch accent types in the same focus condition and the same pitch accent types in different focus conditions. When using one and the same pitch accent, speakers employ both peak alignment and peak height to discriminate between the different focus conditions. They use later or higher peak in contrastive compared to noncontrastive focus conditions, and higher peak in narrow compared to broad focus.

Andreeva et al. (2016) who investigated the use of global and local cues to signal information structure in Bulgarian, found that narrow-focused syllables were consistently realised with longer duration, later peak alignment (but still early in the syllable), greater mean F0 and greater pitch change (including the previous and next syllable), higher intensity and spectral balance than syllables in broad focus. This reflects the use of different pitch accent types to signal narrow vs. broad focus: predominantly H* (with peak alignment close to the beginning of the accented syllable) vs. H+!H*/H+L* (with early peak alignment on the pre-tonic syllable), respectively. The 'phonetic strength' of the default pre-nuclear accent (L*+H) in the narrow focus condition is reduced, and in this way the difference between the nuclear and the pre-nuclear accented syllables is smaller in the broad focus condition, and greater in the narrow focus condition. Contrastive and noncontrastive narrow focus accents are differentiated by local cues (longer vowel, syllable and word durations) when the focus is early in the sentence. Since the speakers use similar accent types in both focus conditions (mostly H* with peak alignment close to the end of the accented syllable, but also L*+(!)H, L+(!)H* and H+!H*), this indicates phonetic variation of phonological categories in order to reinforce the contrastive condition. When focus is late in the sentence, contrastive and noncontrastive focus are distinguished by global cues, i.e. by enhancing the tonal contrast between the nuclear and the pre-nuclear prominence.

Dimitrova and Jun (2015), who found no significant perceptual distinctness between L+H* and H* in broad, narrow contrastive and narrow non-contrastive contexts, conclude that in the absence of additional cues, e.g. longer durations, extended pitch range, and greater intensity differences in narrow compared to broad focus in their data, as reported by Andreeva et al. (2016), the different F0 shapes of the pitch accents seem to be a weak cue to focus type in Bulgarian.

To conclude, the differences in focus marking strategies which exist between English and Bulgarian, emerge as another potential source of difficulty for Bulgarian learners acquiring the prosody of English as L2.

5.6 Stress and rhythm

Although (lexical) stress and speech rhythm are not part of Mennen's LILt model, they constitute an integral part of the prosody of a language, and should therefore constitute part of the comparison of the prosody of English and Bulgarian.

Word stress, or lexical stress – the increased prominence of a syllable in a word in comparison with the rest of the syllables in it, occurs in both English and Bulgarian. Both languages have free lexical stress, that is, the position of the stressed syllable is not fixed with respect to the word boundaries. Stress position has a distinctive function in both English and Bulgarian, e.g., the noun "'import" and the verb "im'port" in English, or the nouns "'пара" (steam) and "па'ра" (a coin) in Bulgarian are distinguished from one another in speech by the position of stress.

Stressed syllables are potential locations for pitch accent realisation. A pitch accent usually associates with the primary-stressed syllable, but it is also possible for a secondary-stressed or even an unstressed syllable to be accented if contrast or emphasis is intended. Figure 5.3. illustrates this in the sentence "I said DEport, not EXport", which is pronounced with lexical stresses and pitch accents on the initial syllables of the two verbs in order to convey the contrast. This shift of stress is accomplished in spite of the fact that the second verb thus acquires the stress pattern of the respective noun. (A change of primary stress position for rhythmic reasons in English, known as "stress shift", is discussed in connection with rhythm later in this section.)

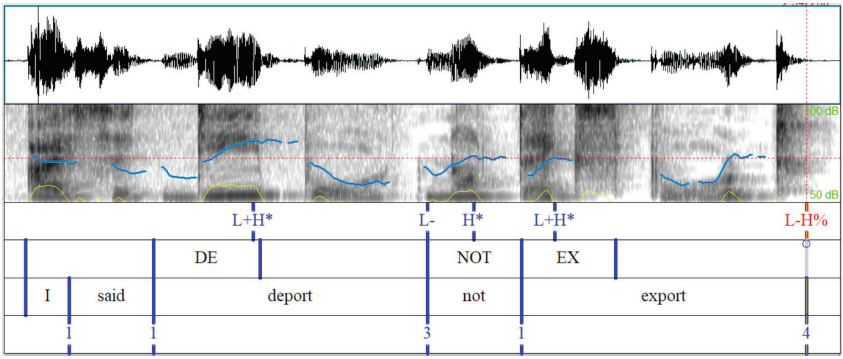


Figure 5.3. The sentence "I said deport, not export" pronounced with lexical stresses and pitch accents on the initial syllables of the two verbs for contrast

English word stress is weight-sensitive, that is, it falls on heavy syllables with branching rhymes, while light syllables are always unstressed. Stress position in English is said to be predictable if information about syllable weight, word class and word morphology is taken into account: for example, a morphologically simple two-syllable noun such as "castle" will have stress on the initial syllable, if that syllable is heavy.

If we exclude pitch change, the main acoustic correlates of stress in English are duration (perceived as length), intensity (perceived as loudness) and vowel quality (lexically stressed syllables have full, unreduced vowels). Of these, duration and vowel quality have been claimed to play a major role.

Lexical stress in Bulgarian, on the other hand, is weight-insensitive: it is not dependent on syllable weight and can occur on both heavy and light syllables. However, according to Andreeva (2017), the kind of stress shift illustrated in Figure 5.3 above is impossible in Bulgarian: the language will employ a strategy whereby the word(s) will be pronounced syllable-by-syllable. Intensity has been claimed to be the main acoustic correlate of lexical stress in Bulgarian (Tilkov and Misheva 1978).

These differences between English and Bulgarian in terms of lexical stress characteristics and position imply that Bulgarian learners are likely to have difficulties with the acquisition of English word stress. The weight-insensitive system of their mother tongue, along with the rather complex information needed for the identification of stress position in an unfamiliar English word, make it difficult to teach any rules for stress placement in the English language classroom. The different cues to stress reported in the research literature to be of importance in the two languages are also a potential source of production as well as perception problems.

The rhythm of English connected speech has long served as the prototypical example of "stress-timing" – a kind of rhythmic organization in which the stressed syllables in the flow of speech tend to recur at roughly equal (isochronous) intervals of time. The view that the languages of the world can be classified as being of two types – stress-timed or syllable-timed according to their rhythmic organization, has been popular amongst researchers and foreign language educators since the middle of the twentieth century.

In spite of the impressive body of experimental evidence which has demonstrated that neither interstress intervals in allegedly stress-timed languages, nor syllable durations in so-called syllable-timed languages are in fact isochronous, the theory of stress- / syllable-timing persists, not least because a "weak" version of it does receive some support from work on both rhythm production and rhythm perception. In stress-timed languages like English, there is indeed a tendency for unstressed syllables to be shortened, for their vowels to be reduced and for their segmental composition to be simplified through elision and assimilation, as a result of which interstress intervals tend to be perceived as more equal than they actually are. In order to reconcile the popularity of the theory with the impossibility to assign a language unambiguously to one of the two rhythm types on the basis of objective measurements of inter-

stress intervals or syllable durations, the view of rhythm as a scalar rather than a dichotomous feature of connected speech has become very popular as well.

Dimitrova (1998) compared the rhythm of Contemporary Standard Bulgarian with that of English (claimed to have stress-timed rhythm) and French (a popular example of a syllable-timed language), using the characteristics of speech rhythm proposed by Dauer (1987). Table 5.3. summarises the results from the comparison.

Table 5.3. Bulgarian speech rhythm compared with that of English and French (adapted from Dimitrova 1998)

Components of language rhythm	**French**	**Bulgarian**	**English**
Duration	N	0	+
Syllable structure	-	-	+
Intonation	-	+	+
Vowels	N	0	+
Consonants	-	-	+
Function of accent	-	+	+

When Dimitrova's assessment of Bulgarian for the relevant rhythm components is compared with Dauer's assessment for English and French, it becomes evident that on a scale of rhythm Bulgarian will occupy an intermediate position between these two prototypical examples of rhythm types. English undoubtedly receives a higher rhythm „score" than Bulgarian: it has six out of six "plus" marks, while Bulgarian has only two. Consequently, English is the more stress-timed of the two.

According to Dauer, it is to be expected that naïve native speakers and trained non-native speakers will be able to identify accented syllables in connected English speech fairly easily and consistently – more easily and consistently than they identify accented syllables in connected Bulgarian speech. At the same time, when listening to English and Bulgarian, one can perceive certain rhythmic differences between the two languages which cannot be explained solely in terms of stress-dependent consonant and vowel allophones and stressed syllable durations.

Although the comparison of the prosodic features of English and Bulgarian predicts deviations from stress-timing in the L2 speech of Bulgarian learners of English due to L1 interference, the exact nature of the deviations, and their respective classification in terms of the four dimensions of LILt remain a topic for further investigation.

6. The prosody of Bulgarian-accented English: experimental evidence

The aim of this chapter is to present relevant findings from recent research on the prosody of L1 Bulgarian and L2 Bulgarian-accented English which can (i) support or refute some of the predictions made in the previous chapter, and (ii) help scholars and educators investigate further the similarities and differences between the prosody of English and Bulgarian in order to make informed predictions about the difficulties which Bulgarian learners are likely to experience with the acquisition of the prosody of English. The first part of the chapter reviews relevant experimental results pertaining to stress, rhythm and accentuation, while the second part of the chapter presents in some detail the most recent research carried out on the intonation of Bulgarian-accented English.

6.1 Metrical structure, rhythm, accentuation

In chapter 5, it was predicted that in the English speech of Bulgarian learners of the language there will be deviations from the stress-timed rhythm typical of the L2. Teachers of English as a foreign language will agree that, on the one hand, it is much harder for the listener to process arhythmic than rhythmic speech and, on the other hand, speech production errors involving stress and rhythm cause incomprehensibility more often than errors involving individual sounds. Brown, among others, claims that rhythm is "the guide to the structure of information in the spoken message" (1977, p. 42).

The somewhat impressionistic observation made above that stress and rhythm errors cause incomprehensibility more often than segmental errors is supported by the findings reported in speech rhythm studies of more general nature. Thus, Lehiste (1977) claims

that listeners seem to impose a rhythmic structure on sequential stimuli and thus "hear sequences of only approximately equal time intervals as more equal than they really are" (1977, p. 258). One way of accounting for this perceptual illusion is by accepting that regularity exists at an underlying level, but is then distorted in performance because of speech production constraints. If, however, this underlying (quasi-)periodic pattern is absent, as is often the case when one produces non-native utterances with inappropriate rhythmic patterns, then the listener will find it difficult, and sometimes even impossible, to restore an underlying regularity which is actually not there.

As far as Bulgarian learners of English are concerned, the validity of the observation that, for them, the production and perception of rhythm is a problem even at a fairly advanced level of proficiency is proved by the results from a Diagnostic Test specially designed to identify problematic areas of pronunciation (Rogerson and Gilbert 1992) reported by Dimitrova (2022). Twenty-five first-year university students of English at a Bulgarian university took the test which included, among other tasks, the identification of

A – the number of syllables in words said in isolation,

B – the stressed syllables of isolated words,

C – the unstressed syllable(s) of isolated words,

D – whether the rhythm of a pair of phrases is the same or different,

E – sentence stress in each sentence in a short dialogue.

The results from these five test tasks are shown in Table 6.1. and Figure 6.1.

Table 6.1. % correct responses for 5 tasks of the Diagnostic Test

Type of task	A	B	C	D	E
% correct responses	90,8	88,3	88,3	53,3	80,0

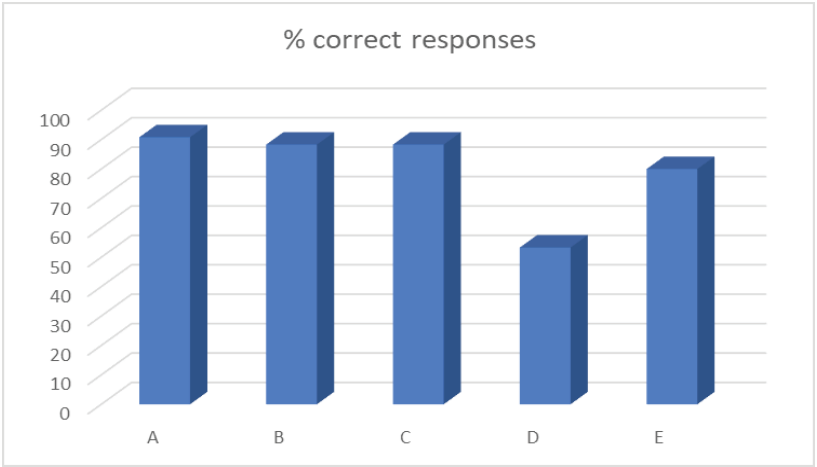


Figure 6.1. % correct responses for 5 tasks of the Diagnostic Test

It is obvious that, whereas the students had few problems identifying stress in isolated words (tasks A, B and C), identification of stress and rhythm in connected speech was much more problematic for them. It could be speculated that the better performance on the first three tasks might be due to the students' knowledge of the correct stress patterns of the test words. Nevertheless, the fact remains that the students' performance on the rhythm task (D) was almost twice as bad as that on the first three tasks. This result supports the theoretical prediction made in the previous chapter on the basis of the comparison between English and Bulgarian speech rhythm. Moreover, Dimitrova's result shows that rhythm is likely to pose a problem not only in terms of production, but also in terms of perception.

Dimitrova (1998b) studied the tolerance towards stress clashes in Bulgarian-accented English in comparison with the speakers' mother tongue (L1 Bulgarian) and target language (L2 English). In Metrical Phonology, a stress clash occurs when two syllables bearing primary stress are adjacent in the same phonological domain, for example, in a phonological phrase, e.g., [thir'teen 'men]$_{PhP}$.

In English, clashes are resolved by the operation of the Rhythm Rule which moves the first of the two adjacent primary-stressed syllables to the left to another heavy syllable, e.g., ['thirteen 'men]$_{PhP}$.

Using somewhat simpler terms, it is often said that "stress shift" has taken place in order to eliminate the stress clash.

Other languages may use different strategies for remedying a clash, such as Beat Deletion – weakening of one of the stresses, that is, destressing, or Beat Insertion – inserting some "phonological distance" between the clashing stresses (Liberman and Prince 1977, Nespor and Vogel 1989, Selkirk 1984).

For English, Dimitrova's study used some of Grabe and Warren's (1995) sentence materials, e.g., "As John had not practised his Chinese cooking, they had a pizza.", in which the phrase "Chinese cooking" contains a potential stress clash.

The Bulgarian material consisted of sentences which contained words with variable stress, e.g., "Този втренчен орлов поглед я смущаваше.", where the phrase "орлов поглед" contains a potential stress clash and the word "орлов" is a variable stress word which can be stressed either on the first or on the second syllable. The hypothesis was that if Bulgarian is similar in terms of its rhythmic organisation to English and if a rule similar to the Rhythm Rule operates in the language, then a big majority of the speakers will prefer to avoid the stress clash by producing the variable stress word with a stress pattern which prevents the occurrence of a clash. A group of university students of English read and recorded both the English and the Bulgarian list of sentences, and three phonetically trained listeners participated in the perceptual analysis of the recorded data.

Grabe and Warren (1995) report that 84% of the words in their data were transcribed by their listeners (three trained phoneticians) as shifted, while genuine stress clashes were rare, constituting only 4% of the data. (In the remaining 12% of the cases, there was nei-

ther a stress shift nor a stress clash, because of the position of the potential stress shift word in the sentence.) In the English sentences read by the Bulgarian university students of English, however, only 58% of the potential stress shift words actually had, according to the listeners, their stress moved to an earlier syllable. The rest of the words (42%) were produced with late stress, as a result of which stress clash occurred and the rhythm of the respective sentence was disrupted.

For Bulgarian, the percentage of sentences which contained a stress clash was even higher: 56.3%. In only 43.7% of all cases did the Bulgarian students choose to produce a variant of the variable stress word which would ensure that no stress clash would occur in the respective sentence.

Finally, the production results obtained for Bulgarian-accented English were analysed and compared with the answers of another group of advanced university students of English who were asked to indicate, on a sheet of paper, the stressing of either the potential stress shift English word in isolation, or of the word when it appeared in a phrase containing a potential stress clash. This was done in order to control for the possible influence of wrongly learned stress patterns of some of the stress shift words used in the experiment.

The results show that the students who took part in the experiment were aware of the phenomenon of stress shift in English. They even produced stress shift more often than they thought they did (in 58% of the read sentences vs. 48.9% of the written phrases), but nevertheless far less often than native English speakers (who, in Grabe and Warren's experiment, produced 84% of the words with shifted stress).

As far as the influence of the students' mother tongue is concerned, it is evident that at least one kind of rhythmic disruption, namely, the occurrence of two strong stresses next to each other (that is, the occurrence of stress clash), is tolerated to a much

greater extent in contemporary standard Bulgarian (56.3%) than in (Southern British) English (4%, Grabe and Warren 1995). A rule somewhat similar to the Rhythm Rule in English may sometimes operate to remedy the clash in Bulgarian, but speakers use it far less frequently than English speakers do. For example, the word "молив" in the phrase "син молив" was produced with stress on the first syllable, creating a clash, by 41% of the speakers, and with stress on the second syllable in order to avoid the clash – by 59% of the Bulgarian students.

Dimitrova's (1998b) results support the prediction that rhythm poses a problem even for Bulgarian learners of English who have reached, in many other respects, a fairly advanced level of proficiency in the L2. Rhythm in Bulgarian-accented English shows deviations from the target language norms due to interference from the learners' L1 (Bulgarian). Another reason for the deviations may be sought for in perception: as demonstrated, identifying the correct rhythmic patterns constitutes a perception problem for Bulgarian learners of English almost half of the time.

Stoykova (2018) investigated the use of weak forms by Bulgarian learners of English. She analysed the readings of "The North Wind and the Sun" by 20 first-year university students of English, all of whom were at B2/C1 level of proficiency according to the European Framework of Reference for Languages. The text contains 113 words in all, 19 of which are function (grammatical) words which occur 47 times throughout the text in different contexts, but are always unstressed and should therefore be pronounced in their weak forms.

Stoykova's analysis shows that 85% of the function words in the text were actually pronounced in their strong forms by the students, including grammatical words such as "should" and "and". There were only 3 speakers who produced less than 30% of the function words in their strong forms. The big majority – 15 students in all – used strong forms between 32-47% of the time. In addition, some

of the strong-form productions involved segmental errors of vowel quality as well.

Amongst the explanations which Stoykova offers in order to account for her findings are lack of knowledge / no prior instruction on the topic of weak form usage in English, fear on the part of the students not to be misunderstood, and – last but not least – the different rhythmic organization of English than the students' L1. These deviations in the production of appropriate weak forms are also part of the general tendency for Bulgarian learners to use destressing to a much smaller extent than native speakers. This tendency has also been observed in research on the intonation of Bulgarian-accented English.

6.2 Intonation

Dimitrova (2019) presents some preliminary results from a pilot study into the prosodic characteristics of "Bulgarian English" – the spoken language of Bulgarian learners of English. The analysis is based on speech data obtained from six female speakers aged 19-23, who were all born and were living in Sofia, and at the time of recording were undergraduates at Sofia University. The speakers read and recorded Aesop's fable "The North Wind and the Sun" (a standard text routinely used in phonetic research) in Bulgarian and in English. All recordings were made during a single session in the Language Lab at Sofia University using the Audacity software.

The recordings were analysed, segmented and labelled in Praat using the ToBI labelling conventions for English as outlined in, e.g., Beckman et al. (2005) and for Bulgarian used in Andreeva (2007) and Dimitrova and Jun (2015). Figure 6.2 shows the waveform, spectrogram and pitch track (F0) of the phrase „северният вятър беше принуден да признае", along with several labelling tiers. The top labelling tier contains the ToBI pitch and phrase accent labels, the second tier from top shows the segmentation of the phrase into syl-

lables, with the pitch accented syllables transcribed, and the bottom two tiers contain the orthographic Bulgarian text and its English translation.

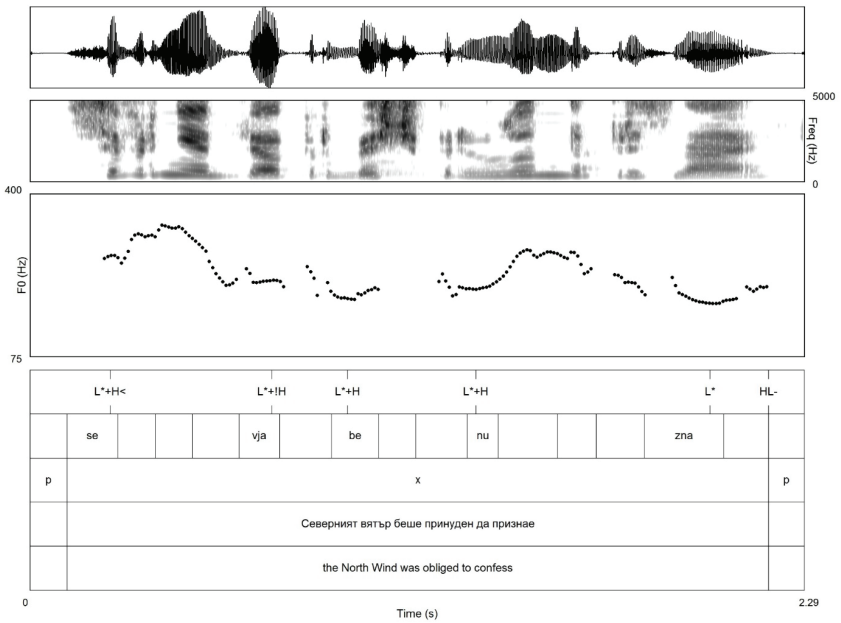


Figure 6.2. Illustration of the segmentation and the ToBI labelling of the speech data

Fundamental frequency and durational measures were obtained using Praat scripts. Pitch level and pitch span were analysed separately: according to Ladd (1996), they both are characteristics which are partially related, but nevertheless should be viewed as distinct. Pitch level was defined as the overall height of a speaker's voice, whereas pitch span was taken to refer to the range of frequencies typically covered by a speaker.

The Long-Term Distributional (LTD) measures in which the analysis was interested were the following:

Fundamental frequency (F0) characteristics

- for pitch level – mean and median F0, measured in Hertz (Hz),
- for pitch span – F0 excursion, measured in Hz and converted to semitones (ST) using the formula given by Reetz (1999); Fundamental frequency excursion was calculated as the difference between the maximum and minimum F0 values in a given intonation phrase.

Temporal characteristics

- mean syllable duration, measured in milliseconds (ms),
- speech tempo, measured in number of syllables per second,
- intonation phrase (IP) and pause duration, measured in ms.

The results obtained for the Bulgarian and the English readings of the undergraduate students were compared, in search for differences between the prosody of their L1 and L2. At the same time, the English readings of the participants in the experiment were also compared with the publicly available IPA recordings of "The North Wind and the Sun" for native British and American English. However, it should be borne in mind that whereas the comparison between L1 Bulgarian and L2 English can yield statistically significant results (because what is being compared are the L1 and the L2 performance of the same group of speakers), the results from the second kind of comparison (between L1 English and L2 English) will be tentative and should be interpreted with caution: this is because the native speaker data come from a single British Received Pronunciation (RP) and General American (GA) speaker, and also because there is little background information available about the two speakers (age, background, education, etc.). Finally, the tentative nature of the latter group of results is also due to the fact that few cross-varietal comparisons of even isolated aspects of (standard) British and (standard) American prosody are currently available.

The preliminary results refer primarily to the Long-Term Distributional measures obtained for the Bulgarian and the English readings of the speakers. The results obtained by measuring fundamental frequency (F0) are shown in Table 6.2.

Table 6.2. Fundamental frequency (F0) results (values in Hz rounded to the nearest whole)

	Bulgarian (B)	**Bulgarian English (BE)**	**American English (AmE)**	**Received Pronunciation (RP)**
Mean F0 (Hz)	226	220	198	188
Median F0 (Hz)	223	215	199	179
Min F0 (Hz)	177	185	138	135
Max F0 (Hz)	297	275	240	269
Pitch span (ST)	9.1	6.7	10.7	11.7

For Mean F0, the results obtained for the Bulgarian (B) and the Bulgarian English (BE) readings of the female undergraduates were very similar: about 226 Hz for the Bulgarian reading of "Северният вятър и слънцето", and about 220 Hz for Bulgarian English (the same six female Bulgarian speakers reading the fable "The North Wind and the Sun" in English). This similarity is altogether an unsurprising result given that the Bulgarian and the Bulgarian English readings were produced by the same speakers. The respective result for standard British English (RP) is approximately 188 Hz, while for standard American English (AmE) it is approximately 198 Hz. How-

ever, although the differences between L1 and L2 Mean F0 look substantial, they may also be idiosyncratic for the reasons explained above.

The results for Median F0 are similar to those obtained for Mean F0, namely, 223 Hz for Bulgarian, 215 Hz for Bulgarian English, 199 Hz for American English, and 179 Hz for RP. Whether these mean and median F0 differences between the native and the non-native readings are statistically significant, however, is a question that can only be answered by an analysis which involves comparisons with readings from groups of native English speakers whose backgrounds are comparable to those of the Bulgarian participants.

It is interesting to note that both the RP and the American English speaker used considerably lower F0 (135 Hz and 138 Hz, respectively) than the Bulgarian English undergraduates whose minimum F0 value was found to be much higher – 185 Hz. The Bulgarian speakers also used much higher F0, especially in their L1 Bulgarian readings (297 Hz on average). In the reading of "The North Wind and the Sun", the Bulgarians used a maximum F0 of 275 Hz, which is also slightly higher than the 269 Hz maximum used by the native RP speaker, and considerably higher than the 240 Hz maximum F0 used by the native American English speaker. If further analyses show that these dissimilarities are not age-related for example, then this may suggest an interesting difference between native and non-native (Bulgarian) English on the one hand, and possibly between native English and native Bulgarian prosody, on the other hand.

Such a possibility is also suggested by the pitch span results which were obtained: the span (F0 excursion) average for the six Bulgarian English speakers was only 6.7 semitones, compared to 10.7 semitones for the American English speaker, and 11.7 semitones for the British RP speaker, on the one hand, and 9.1 semi-

tones for Bulgarian, on the other hand. This suggests that the six Bulgarians who took part in the experiment used a narrower (6.7 semitones) pitch range when they read the fable in English than when they read the Bulgarian version (9.1 semitones average). The observed Bulgarian English span of 6.7 semitones is also considerably narrower than the spans used by both the British (11.7 semitones) and the American (10.7 semitones) native speakers. The Bulgarian English results for F0 span in fact confirm a well-known impressionistic observation regarding the intonation of Bulgarians speaking English, namely, that their pitch range is noticeably "flatter", that is, narrower than that typical of native English speakers. The data also seem in line with another casual observation, namely, that British English RP has a wider pitch span than standard American English. However, all caveats regarding the native English data used in this study remain valid for the above findings for pitch span.

The F0 results were also analysed statistically: Linear Mixed Models (LMMs) with the respective measure as dependent variable, "Speaker" as random factor, and "Language" (Bulgarian – B, Bulgarian English – BE, British RP, and standard American English – AmE, that is, GA) as fixed factors were calculated, and Post-hoc tests were carried out. The above findings and observations regarding pitch span differences between Bulgarian English and the other three "Language" factors were statistically significant ($F(3, 10.64) = 11.33$, $p = 0.0012$).

Thus, the difference in pitch span between Bulgarian and Bulgarian English appears as a valid, statistically confirmed one for the group of six female participants. The "Language" pitch range differences (in semitones) between American English, Bulgarian, Bulgarian English and RP, along with the respective standard error values (bigger in native-speaker English than in both Bulgarian English and Bulgarian) are shown in Figure 6.3.

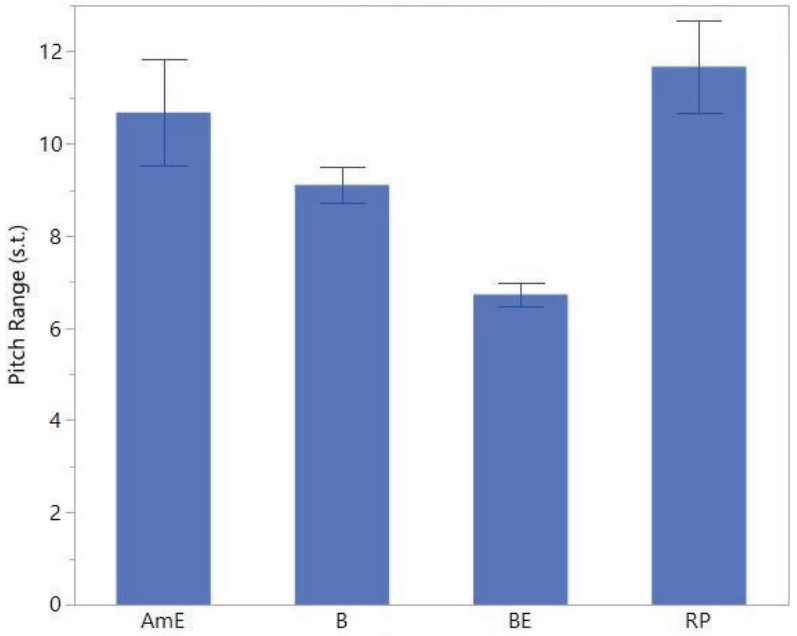


Figure 6.3. Pitch span differences (and standard error values) between American English (AmE – GA), Bulgarian (B), Bulgarian English (BE) and British English (RP) (in semitones)

The other F0 characteristics which were measured, namely, mean F0, median F0, F0 standard deviation, minimum and maximum F0, were also found to differ significantly. However, due to the nature of the native English speaker data, no statistically valid generalisations based on these findings can be made. Some interesting observations, however, concern

(i) the larger F0 standard deviation which was found in RP (33.8 Hz) than in Bulgarian English (24.7 Hz) or American English (22.7 Hz), and

(ii) the F0 standard deviations which the Bulgarian speakers used in their mother tongue (28.9 Hz) and in their L2 English speech (24.7 Hz).

This means that the Bulgarian learners of English who took part in the investigation tended to use more pitch variation in their mother tongue than when they spoke English, and the latter variation was smaller than the one used by the native RP speaker when reading "The North Wind and the Sun".

As regards minimum F0, RP (134.7 Hz) and American English (138 Hz) were significantly different from L1 Bulgarian (177 Hz) and L2 English (184.7 Hz) ($F (3, 15.92) = 4.75$, $p = 0.015$). The higher minima used by the Bulgarian speakers when speaking both their mother tongue and L2 English, however, may be due to sociolinguistic variables such as age.

The temporal characteristics which were investigated were (i) mean syllable duration, which was measured in milliseconds (ms), speech tempo, measured in number of syllables per second, (iii) intonation phrase (IP) duration, and (iv) pause duration, the latter two also measured in ms. As expected, the results reflected the noticeably faster speech tempo of the American English speaker. Nevertheless, some interesting preliminary observations can be made regarding the mean Intonation Phrase duration values which were obtained for the four "Language" conditions (see Table 6.3.).

Table 6.3. Mean Intonation Phrase (IP) durations (ms)

	Bulgarian (B)	**Bulgarian English (BE)**	**American English (AmE)**	**Received Pronunciation (RP)**
Mean IP duration (ms)	1181	989	877	1271

The Intonation Phrases produced by the fast-talking native speaker of AmE are on average only 877 ms long, while those in the speech of the slowest speaker – the one speaking RP – are 1271 ms long. What is most interesting from the point of view of the cur-

rent research, however, is the fact that the Bulgarian group of speakers produced considerably shorter IPs when speaking English (989 ms) than when speaking their L1 (1181 ms). The results are statistically significant ($F (3, 10.19) = 7.55$, $p = 0.006$). This finding seems to confirm another well-known empirical observation, namely, that foreign learners of a language tend to produce shorter chunks of speech, and also tend to pause more often when speaking in the foreign language. And although this is commonly explained with the need for more planning time, the data from the pilot investigation reported here demonstrate that foreign learners use shorter speech chunks in a reading task as well.

On a more general note, as far as the analysis of the "temporal dimension(s)" of foreign learner prosody is concerned, the study suggests a number of further characteristic features which should form an integral part of L2 prosody research, namely, rhythm and timing characteristics such as, for example, stressed and unstressed syllable duration.

The major findings which emerge from Dimitrova's (2019) pilot study are the difference in pitch span between Bulgarian and Bulgarian English, the small pitch span standard deviations in Bulgarian English, and the higher F0 minima used by the Bulgarian as compared with the native English speakers. Taken together, these results can go a long way towards explaining the popular impressionistic observation that when speaking English, Bulgarian learners of the language tend to sound "flat", and even "dull" and "uninterested", according to some native speakers. Conversely, Bulgarians often tend to perceive native English intonation as "exaggerated". A similar observation regarding German listeners who think that English intonation is "over the top" has been noted by Eckert and Laver (1994, reported in Mennen 2007, p. 64).

Mennen offers an insightful discussion of current evidence on the influence of speakers' language background on their pitch range, or span. Drawing on Ohara's (1992) study of gender-dependent use of

pitch levels in English vs. Japanese, she observes that "It is thought that cultures or languages have their particular 'vocal image', which reflects socio-culturally desired personal attributes and social roles, and that speakers choose a pitch (within their anatomical/physiological range) that approximates the vocal image they want to project" (Mennen 2007, p. 64). Ladd (1996, 2008) likewise discusses these pitch changes, considering them variations in terms of pitch level (overall pitch height) and pitch span (frequency range). The findings presented above add further evidence to support the importance of these parameters as already emerging in studies such as Mennen (2007, 2015), Zimmerer et al. (2014, 2015) and Andreeva (2016), among others.

The prosodic characteristics analysed in the above pilot study are primarily related to the phonetic, or realisational, dimension of Mennen's L2 Intonation Learning theory (LILt). For drawing comparisons between the tonal inventories of Bulgarian and Bulgarian English, and of L1 and L2 English, what is needed is native speech data from speakers comparable to the Bulgarian group of participants. Further analyses of the phonetic implementation of the categorical phonological elements comprising the "systemic dimension" of LILt (e.g., the timing of pitch accents) are also necessary. Finally, comparative investigations of the functioning of pitch accents, phrasal accents and boundary tones, as well as of their frequency of use, that is, of the "semantic" and "frequency" dimensions of LILt also need to be carried out.

In a follow-up to her 2019 pilot study, Dimitrova (2021) investigated the prosodic characteristics of speech data obtained from two comparable groups of six Bulgarian and six English female speakers. The Bulgarian females were aged 19-23 at the time when they took part in the research, they were speakers of Contemporary Standard Bulgarian, and were university undergraduates born and living in Sofia. The six English speakers were all born and living in England at the time of recording, they were of comparable age to that of the Bulgarian participants, and like them were also university

undergraduates. Their accent was judged by two lecturers in English Phonetics to be representative of the kind of pronunciation that has been dubbed "Modern Received Pronunciation", "Southern Standard Pronunciation", or "Southern Standard British English" – an educated accent virtually devoid of any salient regional pronunciation features. Their accent will be referred to in the rest of the discussion as just "Received Pronunciation", or RP.

All speakers read and recorded Aesop's fable "The North Wind and the Sun", and the Bulgarian speakers recorded the Bulgarian text – "Северният вятър и слънцето" as well. The recordings were analysed, segmented and labelled in Praat using the same ToBI (Tone and Break Indices) labelling conventions used in the pilot study.

Figure 6.4. shows the waveform, and the spectrogram, intensity and pitch tracks of the phrase „Северния вятър", as well as five of the labelling tiers: the top one (marked "PAs") contains the ToBI pitch accent and edge tone labels, the second one from the top (marked "syl") shows the segmentation into syllables, with the pitch accented syllables transcribed in block capitals, the third tier from the top shows the boundaries of the intonation phrase ("IP"), and the bottom two tiers contain the orthographic Bulgarian text ("text") and its English translation ("transl").

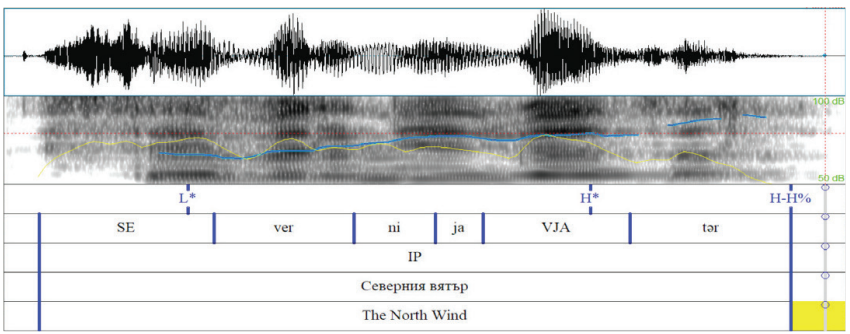


Figure 6.4. Illustration of the segmentation and the ToBI labelling of the speech data.

The manually segmented and labelled files were analysed using dedicated Praat scripts, with the help of which a number of fundamental frequency (F0) and durational measures were obtained. The Long-Term Distributional (LTD) measures in which this study was interested were the following:

- Fundamental frequency characteristics: for pitch level – mean and median F0, measured in Hertz (Hz), for pitch span – F0 excursion, measured in Hz and converted to semitones (ST).

Pitch level and pitch span were again analysed separately, following Ladd (1996). Fundamental frequency excursion (pitch span) was calculated as the difference between the maximum and minimum F0 values in a given intonation phrase; it was measured in Hz, and then converted to ST using the formula given by Reetz (1999).

- Temporal characteristics: mean syllable duration, measured in milliseconds (ms); intonation phrase (IP) and pause duration, measured in ms.

In addition, counts of the number of IPs, pauses, stressed and unstressed syllables in the reading of each speaker were also performed.

The results which were obtained for the Bulgarian (B) and the Bulgarian English (BE) readings of the undergraduate students will be compared first. In addition, a comparison will be made between the English readings of the Bulgarian participants and the readings of the native British English (RP) speakers in an attempt to shed light upon the way(s) in which the prosodic characteristics of Bulgarian English differ from those of native English modern RP pronunciation.

Both comparisons are expected to yield statistically significant results: in the first case, the comparison is between the L1 and the

L2 performance of the same group of speakers, whereas in the second case, it is between groups of L1 and L2 speakers who are similar in terms of group size, speaker age, gender, and education level. A comparison will also be made with the data reported in the pilot study (Dimitrova 2019).

The results obtained by measuring fundamental frequency (F0) are shown in Table 6.4., where "RP 1" stands for the group of 6 undergraduate native RP speakers, and "RP 2" indicates the results for the single native RP female reader of the official IPA version of "The North Wind and the Sun" fable reported in Dimitrova (2019).

Table 6.4. Fundamental frequency (F0) results (values rounded to the nearest whole in Hz; pitch span results shown in semitones ST). The features for which significant differences were found are marked with *

	Bulgarian (B)	**Bulgarian English (BE)**	**RP 1**	**RP 2**
*Mean F0 (Hz)	226	220	199	188
*Median F0 (Hz)	223	215	197	179
*StDev F0 (Hz)	28.8	24.6	17.8	n.a.
*Min F0 (Hz)	177	185	169	135
*Max F0 (Hz)	297	275	240	269
*Pitch span (ST)	9.1	6.7	6.1	11.7

The F0 results were also analysed statistically: Linear Mixed Models (LMMs) with the respective measure as dependent variable, "Speaker" as random factor, and "Language" (Bulgarian – B, Bulgarian English – BE and British Received Pronunciation – RP 1) as fixed

factors were calculated, and Post-hoc tests were carried out. (The "RP 2" data come from a single speaker, and were therefore not included in the statistical analysis.)

For Mean F0, the results obtained for Bulgarian (B) and Bulgarian English (BE) were similar, and virtually the same as those obtained in the pilot study, since the Bulgarian and the Bulgarian English readings were produced by the same speakers. The respective result for standard British English (RP) is 199 Hz – for the group of undergraduates (see Figure 6.5.), and 188 Hz – for the single speaker. The difference between Bulgarian and Bulgarian English, on the one hand, and RP 1, on the other hand, is statistically significant [$F(2, 25.43) = 45.5$; $p<0.0001$].

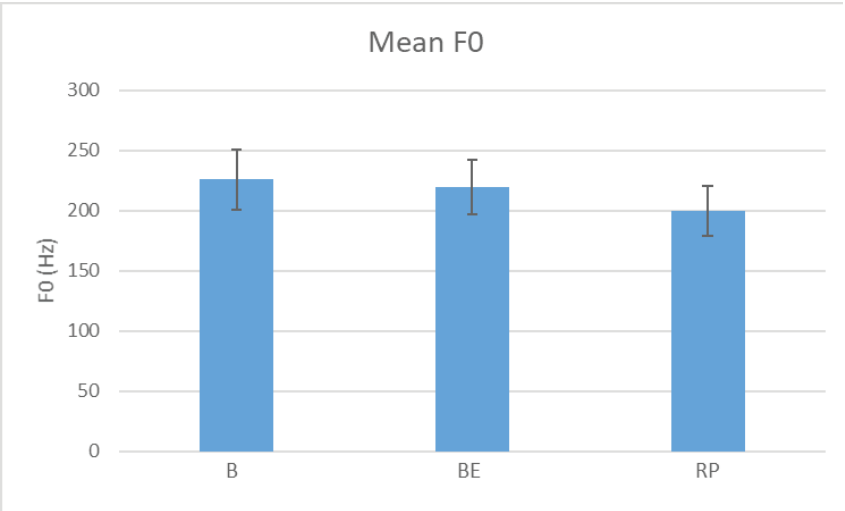


Figure 6.5. Mean F0 values and standard deviations (in Hz) for Bulgarian (B), Bulgarian English (BE) and British English Received Pronunciation 1 (RP)

The results for Median F0 are similar to those obtained for Mean F0, namely, 223 Hz for Bulgarian, 215 Hz for Bulgarian English, 197 Hz for the RP 1 group of speakers, and 179 Hz for the single RP 2 speaker. Again, the difference between Bulgarian and Bulgarian

English, on the one hand, and RP 1, on the other hand, is statistically significant [F (2, 25.42) = 40.1; p<0.0001].

It is worth noting at this point that the RP 2 speaker (the speaker who read the representative IPA version of the fable) used lower Mean and Median F0 which, together with the considerably lower Minimum F0 measured for this speaker, gives us grounds to suppose that this was an older speaker than the undergraduates who comprised our current RP reference speaker group. The F0 minimum values for Bulgarian, Bulgarian English and RP 1 were all significantly different from each other as well [F (2, 25.78) = 4.6837; p = 0.0184].

The Standard Deviation of F0 (StDev) also differed significantly [F (2, 23.82) = 27.0339; p<0.0001]. It was biggest in Bulgarian spoken as a mother tongue – 28.8 Hz, and smallest in native RP speech – 17.8 Hz. In terms of this feature, the value obtained for Bulgarian English (24.6 Hz) was again closer to Bulgarian, rather than to English. This result suggests that there is more F0 variation in both Bulgarian and Bulgarian English speech, than in native RP pronunciation, which does not support the frequent observation that English intonation "goes up and down all the time", unlike Bulgarian intonation (see more on this below).

The Maximum F0 values which were obtained were the highest for the Bulgarian speakers in their L1 Bulgarian readings (297 Hz). In the reading of the English text "The North Wind and the Sun", the Bulgarians used a maximum F0 of 275 Hz, which is considerably higher than the 240 Hz maximum used by the group of native RP 1 speakers. This may suggest the existence of an interesting suprasegmental difference between native Bulgarian and non-native Bulgarian English speech on the one hand, and native English RP speech, on the other hand. Tests show the difference to be statistically significant [F (2, 25.3) = 43.5734; p<0.0001].

The pitch span results which were obtained are as follows: the average span (F0 excursion) for the six Bulgarians speaking their

mother tongue is 9.1 semitones, and only 6.7 semitones in Bulgarian English. For the British RP 1 group, the span is even narrower – 6.1 semitones. Statistical tests show that with respect to this suprasegmental feature, Bulgarian spoken as L1 is significantly different from Bulgarian English on the one hand, and from British English on the other hand [$F(2, 22.34) = 22.7072$; $p<0.0001$]. The highest measure for this characteristic was 11.7 ST, and it was obtained for the individual RP 2 speaker: this seems an idiosyncratic feature of the speaker. The pitch span results (in semitones) for Bulgarian, Bulgarian English and RP, along with the respective standard deviation values are shown in Figure 6.6.

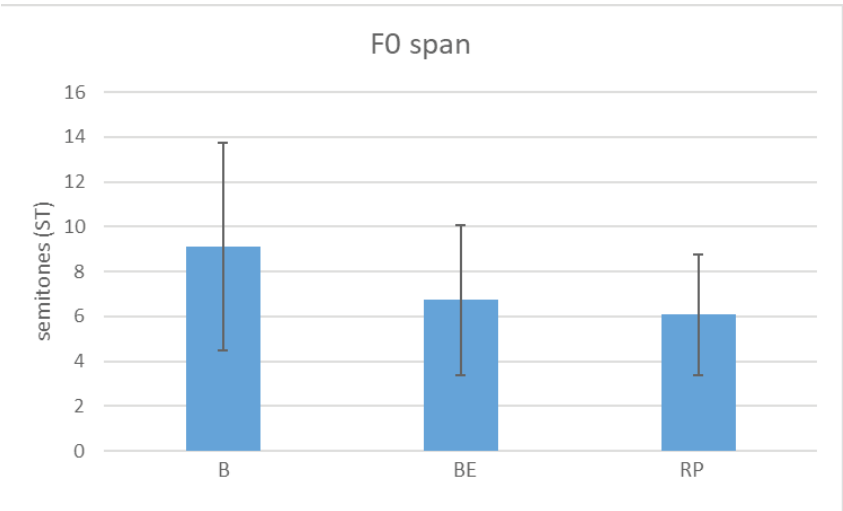


Figure 6.6. Pitch span measures (and standard deviations) for Bulgarian (B), Bulgarian English (BE) and British English (RP) (in semitones ST)

The results for the L1 (Bulgarian vs. British English RP) readings of the two groups of speakers are in agreement with earlier results reported by Andreeva et al. (2014), who found significant differences between two language groups of speakers: in their study, the speakers of Germanic languages (German and English) used narrow-

er pitch span (and lower pitch maxima) than the speakers of Slavic languages (Bulgarian and Polish), supporting the hypothesis that "linguistic communities tend to be characterized by particular pitch profiles" (Andreeva et al., 2014, p. 776).

However, the use of narrower F0 span in English than in Bulgarian found in the study reported here does not corroborate the observation sometimes made by native English speakers that Bulgarian-accented English sounds "flat" and "monotonous". It may well be the case that such impressions are due not to long-term F0 characteristics but to the use of certain pitch accents or "tones" which are absent from the pitch accent inventory of the learner's L1. Such an account will be in line with the systemic dimension of Mennen's model and the inventory and distribution of the respective pitch accents. It may also be due to differences in the frequency dimension, in that a pitch accent may be phonemic in both the L1 and the L2, but may be less frequently used in the learner's mother tongue. For English, we can tentatively hypothesise that one such pitch accent could be the "fall-rise" tone.

It may also be worth applying a somewhat different approach to pitch range modelling, as suggested by Patterson and Ladd (1999), who measured F0 values not just at the highest and the lowest point in an intonation phrase, but used other well-established landmarks such as initial peaks, as well as other accent peaks, valleys, and final lows in a sentence. It is clear that, with respect to F0 span, further research is needed to account for the seeming discrepancy between acoustic measurements and auditory impressions.

The temporal characteristics investigated by Dimitrova (2021) were mean intonation phrase duration (measured in milliseconds – ms), overall number of intonation phrases and pauses, speech tempo (measured in number of syllables per second, but not reported in the paper), mean syllable duration, and mean pause duration (the latter two also measured in milliseconds). These are shown in Table 6.5.

Table 6.5. Temporal measures for Bulgarian, Bulgarian English and RP. Duration measures are given in milliseconds (ms). The features for which significant differences were found are marked with *.

	Bulgarian (B)	**Bulgarian English (BE)**	**Received Pronunciation (RP)**
*Mean intonation phrase (IP) duration (ms)	1181	989	1621
*Number of IPs	140	178	98
*Number of pauses	68	88	55
*Mean syllable duration (ms)	137.7	203.6	183.8

The results for the temporal features were also analysed statistically: Linear Mixed Models (LMMs) with the respective measure as dependent variable, "Speaker" as random factor, and "Language" (Bulgarian – B, Bulgarian English – BE and British Received Pronunciation – RP) as fixed factors were calculated, and Post-hoc tests were carried out.

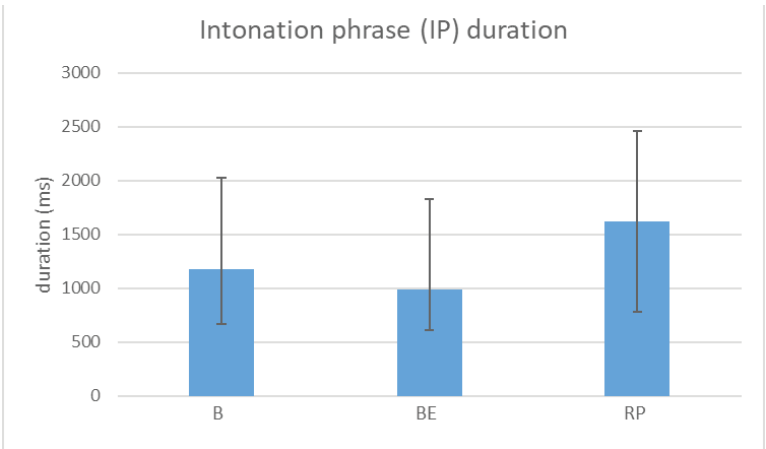


Figure 6.7. Intonation phrase (IP) duration differences (and standard deviations) between Bulgarian (B), Bulgarian English (BE) and British English (RP) (in milliseconds).

Intonation phrase duration differences between Bulgarian, Bulgarian English and RP (Figure 6.7.) were all statistically significant [$F(2, 26.33) = 37.3680$; $p<0.0001$]. However, the results for L1 Bulgarian are not directly comparable with those for L1 and/or L2 English: they are mostly attributable to lexico-syntactic differences between the two languages. The comparison which is of interest here is the one between intonation phrase durations in L1 and L2 English. The intonation phrases produced by the Bulgarian speakers of English were much shorter – they were only about 60% of the duration of the intonation phrases produced by the native RP speakers. Also, the overall number of intonation phrases in the reading of the English text by the Bulgarians was much higher (178 vs. only 98 IPs produced by the RP speakers). The number of pauses was also higher (88 in Bulgarian English vs. 55 in RP). These findings are in accord with the results for IP duration in the pilot investigation and confirm previous reports in the research literature that L2 learners tend to produce shorter chunks of speech, and also tend to pause more often than L1 speakers of the same language.

Mean syllable durations likewise differed significantly in Bulgarian, Bulgarian English and RP [$F(2, 10) = 106.5554$; $p<0.0001$], whereas mean pause duration was not significant and has therefore not been included in Table 6.5.

Finally, the difference between Bulgarian English and RP speakers in terms of the number of accented syllables in the English text was also significantly different: the native speakers accented on average 45 syllables, while the Bulgarian speakers of English put a pitch accent on 58 syllables, which is about 23% more in comparison with the native RP speakers [$F(1, 5) = 78.6691$; $p=0.0003$]. This result confirms yet another well-known empirical observation about the prosodic characteristics of Bulgarian English, and of L2 speech in general – namely, that non-native speakers highlight more words. An alternative explanation for this and other similar

results which have been reported in the research literature with regard to accentuation in L2 is that non-native learners fail to de-stress words which constitute old information in connected speech. Indeed, in the Bulgarian English reading of the text, there were on average 86 unstressed syllables per reading/speaker, compared with 99 unstressed syllables found on average in the reading of the RP speakers. (The number of syllables counted in the English text amounted to a total of 144, with the word "traveler" considered 3-syllabic irrespective of whether a given speaker pronounced it with 3 or with only 2 syllables. By comparison, the Bulgarian text of the fable comprised 200 syllables.)

The prosodic features analysed by Dimitrova (2021) are primarily related to the phonetic, or realisational, dimension of Mennen's L2 Intonation Learning theory (LILt). Further investigation into the type and number of pitch accents (systemic dimension), their phonetic implementation in terms of alignment and timing (realisational dimension), their functions and frequency of occurrence (semantic and frequency dimension) are needed in order to shed more light on the nature and origin of the prosodic deviations observed in L2 Bulgarian-accented English.

Andreeva and Dimitrova (2022c) investigated the amount of interference of L1 F0- and duration-related characteristics in the speech of advanced Bulgarian learners of two related Germanic languages – German and English. Previous research has found that Bulgarian speakers use wider pitch range and are more variable compared to German and English speakers (Andreeva et al. 2015) Assuming that there is transfer of F0-related characteristics from the L1, the expectation was to find expansion of the L2 target norms for pitch range. Alternatively, there may be adaptation of the native language pitch range to that of the target language.

The research methods and the results for Bulgarian (abbreviated in this study to "BG_L1"), Bulgarian-accented English ("GB_L2")

and Standard Southern British English RP pronunciation ("GB_L1") are presented and discussed below.

The "English data set" for this study consisted of ten Bulgarian speakers of English and six English native speakers as controls. All speakers were female university students of comparable age (average 20.7 years) and spoke the respective standard language varieties. The Bulgarian participants had some knowledge of the phonetics and phonology of English. The material recorded was Aesop's fable "The North Wind and the Sun", with the Bulgarians reading the text in Bulgarian as well as in the L2 (English). The results for L2 English are thus directly comparable with the findings of Dimitrova (2019, 2021).

Measurements

Syllable and Intonation Phrase (IP) boundaries as well as pauses were segmented, and lexically stressed syllables were labelled manually in Praat (Boersma and Weenink 1992–2022). All accented syllables were marked and counted, including those in lexical words with double prominence and in prominent function words. In addition, the pauses and IPs per reading were also counted.

Pitch analysis

Pitch analysis was performed as follows. First, F0 was extracted automatically from all recordings by means of the ESPS algorithm "get_f0" (Talkin 1995) with time steps of 5 ms. Secondly, a manual inspection and correction (i.e., removal of data points) of the extracted pitch contours was performed in Praat. The corrections included the removal of octave jumps as well as other artefacts (e.g., due to creaky voice). From the cleaned data the following F0 long-term distributional (LTD) measures per IP were calculated using Praat scripts: mean, median, minimum, maximum, standard deviation (all in Hz), and span in semitones (ST). The conversion from Hz was performed with the following formula (Reetz 1999):

Pitch Span [ST] = 39.863 * log10(MaximumF0/MinimumF0)

Temporal features

The durations of the IPs, pauses and prominent syllables were extracted per reading, speaker and native/target language using Praat scripts. Mean syllable duration, mean duration of accented syllables as well as accented/unaccented syllable duration ratios were computed. In addition, two measurements of speaking rate were calculated: (a) speech rate (SR, the number of canonical syllables divided by the duration of the respective recording) and (b) articulation rate (AR, the number of canonical syllables divided by the sum of IP durations per recording).

Statistical analyses

For statistical validation, the software JMP 16 (SAS 2021) was used. For both the F0- and the duration-related parameters a model for the English data set was calculated, comparing native language and target language. Linear mixed-effect models (LMM) were fitted for the duration-related parameters, with the respective log-transformed measure as dependent variable, speaker as random factor, and language (native language/target language) as fixed factor. Separate Tukey post-hoc tests were carried out per variable, if appropriate. The confidence level was set at $\alpha = .05$.

For the analyses of the F0-related parameters a nonparametric Kruskal-Wallis test was used because the data were not equally distributed. To determine differences between speaker groups, post-hoc Dunn's pairwise tests with Bonferroni adjustment were performed.

Results

F0-related parameters

Means and standard deviations for the English data set for each of the F0-related parameters are presented in Table 6.6. Because nonparametric tests were used, the median values are also reported. Following Ladd (1996), the measures for mean and median

(related to pitch level) and span are considered to be attributes of pitch range, and the standard deviation (SD) – an attribute of pitch variation.

For the English data set, a main effect of language was found (see Table 6.6. and Figure 6.8.) on all measurements (for mean: ($\chi^2(2) = 39.86$, $p < .0001$); for median: ($\chi^2(2) = 39.29$, $p < .0001$); for SD ($\chi^2(2) = 36.09$, $p < .0001$); for span ($\chi^2(2) = 30.99$, $p < .001$). Post-hoc tests revealed that: (a) for pitch level, BG_L1 and GB_L2 have significantly higher values than GB_L1; (b) for pitch span, BG_L1 has significantly higher values than GB_L2 and GB_L1, and (c) for pitch variation, the three groups have significantly different values with BG_L1 having higher values than GB_L2, which in turn has higher values than GB_L1.

Table 6.6. F0-related parameters for Bulgarian (BG_L1), Bulgarian-accented English (GB_L2) and RP English (GB_L1)

parameter	**BG_L1**	**GB_L2**	**GB_L1**
mean values			
mean	223 (32.0)	220 (33.4)	200 (21.9)
median	219 (32.1)	217 (35.5)	198 (21.7)
SD	27.79 (27.8)	23.7 (14.2)	18.1 (8.1)
span (ST)	7.78 (3.2)	6.6 (3.0)	6.1 (2.8)
median values			
mean	217 (32.0)	214 (33.4)	198 (21.9)
median	216 (32.1)	213 (35.5)	197 (21.7)
SD	24.1 (15.3)	20.1 (14.2)	17.7 (8.1)
span (ST)	7.4 (3.2)	6.1 (3.0)	5.5 (2.8)

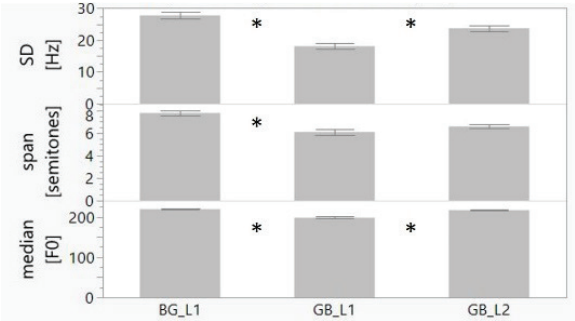


Figure 6.8. Median, span and SD values for Bulgarian (BG_L1), Bulgarian-accented English (GB_L2) and RP English (GB_L1)

The comparison between BG_L1 and GB_L1 corroborates the results in Andreeva et al. (2014) who report the use of wider pitch range and higher variability in two Slavic languages (Bulgarian and Polish) compared to two Germanic languages (German and English). The latter findings can also account for the higher F0-related parameter values which were found in GB_L2 , compared to GB_L1, respectively.

Duration-related parameters

The duration-related parameters in the native RP English readings were compared with those in the respective target language readings by the Bulgarian speakers. At this point BG_L1 is not included in the analyses because of the differences between the respective texts in terms of number of words and syllables, syllable complexity, etc. (Table 6.7.)

The Bulgarian speakers of English produced a higher number of IPs and pauses than the native English speakers (26.0 vs. 16.8 IPs, and 13.3 vs. 9.3 pauses, respectively). They also produced more accented syllables than the natives (58.5 vs. 49.2). The analyses show a main effect of language with respect to the accented/unaccented syllable duration ratio ($F [1, 7] = 8.61$, $p<0.0109$), which is signifi-

Table 6.7. Duration-related parameters for Bulgarian and English

Parameter	**BG_L1**	**GB_L2**	**GB_L1**
mean accented syllable duration	209.7 (16.3)	338.1 (34.8)	305.0 (25.7)
accented/unaccented ratio	1.5 (0.3)	2.2 (0.2)	2.5 (0.9)
articulation rate	6.2 (0.8)	4.4 (0.4)	5.5 (0.3)
speech rate	5.3 (0.4)	3.9 (0.6)	4.5 (0.4)

cantly lower in GB_L2 than in GB_L1. On the other hand, the mean accented syllable durations in GB_L2 and GB_L1 do not differ significantly. This suggests that the Bulgarian speakers of English do not reduce sufficiently the duration of unaccented syllables in the target language. As for articulation rate, the Bulgarian speakers of English were significantly slower than the English native speakers ($F [1, 7] = 27.12$, $p<0.0001$), whereas the speech rates of the two groups turned out to be non-significant. This can be attributed to the longer pause durations between utterances produced by the native English speakers.

Analyses were also carried out by collapsing the speaking rate measurements in the L2 realisations of all the Bulgarian speakers in one group and comparing them to the speaking rate mesurements in their L1 realisations. Pearson correlations were used to determine whether L2 speaking rate can be predicted on the basis of L1 speaking rate. A strong positive correlation between L1 and L2 speaking rates was found (speech rate: $r(24) = .783$, $p < .0001$; articulation rate: $r(24) = .791$, $p < .0001$).

In conclusion, Andreeva and Dimitrova's (2022c) study shows that all F0-related parameters in the L2 speech of the Bulgarian learners of English were lower than in their L1 but higher than those of the native English speakers. However, the latter group of learners had higher values for level compared to the English native speakers, and higher values for span compared to their L1. Thus, the authors' first assumption that the L2 target norms for pitch range will be expanded due to L1 influence was confirmed for the pitch level and variability of the Bulgarian learners of English. With regard to the pitch span used by the Bulgarian students, Andreeva and Dimitrova observed adaptation of the native language norms towards those of the target language.

With regard to the duration-related parameters, the study found that the Bulgarian speakers used slower articulation rate, more IPs and pauses in their L2 than the native speakers. They also failed to deaccentuate: more accented syllables were found in L2. In the absence of statistically significant difference between accented syllable duration in L1 and L2 English, the lower ratio between accented and unaccented syllable durations in L2 English is interpreted as an indication of the smaller amount of reduction of unaccented syllables. The strong correlation found between L1 and L2 speaking rates of the Bulgarian speakers constitutes evidence that L1 speaking rate can indeed predict the speaking rate in L2. The results suggest that the so-called L2 speaking style is influenced by L1 prosody with respect to F0-related features. As for duration-related characteristics of L1, they can explain some of the variability found in L2 speech.

7. Summary and outlook

A quick survey of materials for teaching prosody as part of foreign language instruction reveals that, if a broad definition of the term "prosody" is adopted, then some suprasegmental features do find their place in the curriculum. These usually include lexical stress as well as some basic intonation patterns, such as question intonation, the intonation of lists, etc. However, teaching materials based on in-depth comparisons between learners' L1 and L2 are relatively sparse.

Although the prosodic features of L2 have frequently been neglected in the past, the advancement of the Autosegmental-Metrical theory of intonational phonology and the ToBI analytical framework and annotation system have offered new perspectives on the research and teaching of L2 prosody.

In the past, models of second language acquisition tended to focus on the pronunciation of segmental sounds – the vowels and consonants of the target language. However, the last few decades have witnessed a renewal of interest in the study of L2 prosody. Descriptions based on Ladd's (1996/2008) dimensions in which cross-linguistic intonation differences can occur and Mennen's (2015) L2 Intonation Learning Theory have additionally boosted this interest.

The present investigation was focused on the prosodic characteristics of Bulgarian-accented English. After reviewing some of the most popular second language acquisition theories, and describing the best-known models of English intonation, the methodology for comparison of the prosodic systems of L1 and L2 which was chosen was the one put forward by the L2 Intonation Learning Theory, and the preferred theoretical framework was that of Autosegmental-Metrical phonology. Using ToBI as an analytical framework and an

annotation tool, the intonation systems of English and Bulgarian were compared with the aim of revealing similarities and differences between the two languages which can provide a basis for predictions where L2 intonation deviation is likely to occur.

In the systemic dimension, a comparison of the phonological models of English and Bulgarian intonation revealed a number of similarities between the two languages in terms of their tonal inventories, the tonal phonotactic possibilities and tune-text association. One difference which emerged concerns the hypothesised existence of the prosodic word as a unit in the prosodic hierarchy in Bulgarian. Although further research is needed in order to confirm the hypothesis, we predict that the different hierarchies of prosodic units in the two languages are likely to cause problems for Bulgarian learners of English, and deviances from the native English norms are to be expected in the speech of Bulgarian learners of the language due to L1 interference at the level of the prosodic word.

Besides, a more detailed comparison of the tonal combinations allowed in the two languages is also necessary. The analysis of the tonal phonotactics in English and Bulgarian appears as another topic for future investigation in the systemic dimension which can reveal important differences between the two languages.

The differences in focus marking strategies which exist between English and Bulgarian, such as the rare de-accentuation of given information in Bulgarian, emerge as a further potential source of difficulty for Bulgarian learners acquiring the prosody of English as L2.

In the realisational dimension, some examples of differences in the phonetic implementation of pitch accents in the two languages are:

(i) the alignment of the H* peak which is reported to occur towards the end of the syllable in English, whereas in Bulgarian its alignment depends on the position of the word / syllable in the phrase;

(ii) the alignment of the H tone of the pre-nuclear bitonal L*+H pitch accent: in English, the high target is usually reached within the first post-accented syllable, whereas in Bulgarian it can be shifted further to the right when it is followed by several unaccented syllables.

In general, Bulgarian pitch accents are described by Andreeva and Dimitrova (2022b) as often having variable alignment of the tonal target with the tone-bearing unit. This variability is triggered by speaker-specific production strategies as well as by the position of the accented syllable within the phrase. Due to transfer from the L1, the variable alignment of the tonal targets is likely to cause deviations in Bulgarian-accented English speech.

Another realisational difference which should be noted concerns the wider pitch span and higher pitch level used in Bulgarian and Bulgarian-accented English compared with native L1 English. This finding does not corroborate the observation sometimes made by native English speakers that Bulgarian-accented English sounds "flat" and "monotonous". Such impressions may be due not to long-term F0 characteristics but to differences in the frequency dimension: a pitch accent may be phonemic in both the L1 and the L2, but may be used less frequently in the learner's mother tongue. For English, we can tentatively hypothesise that one such pitch accent could be the "fall-rise" tone.

The current state of our knowledge about the functionality of the structural elements or tunes allows only a preliminary comparison between English and Bulgarian in the semantic dimension. Still, some similarities between the two languages (also noted by Andreeva 2017 in her comparison of Bulgarian and German intonation) include:

(i) A given focus type can be expressed with different pitch accents in both languages;

(ii) Broad focus is marked by early peak accents with a falling onglide;

(iii) Givenness lowers pre-nuclear pitch accents and cancels post-nuclear ones.

Several examples of dissimilarity can also be singled out: they are likely to cause deviations in the L2.

(i) the English "contradiction contour" L* L-H%;

(ii) the pattern L+H* L-H% which is used in English to make a correction or to express contrast;

(iii) the tune L*+H L-H% which can be used in English to express uncertainty.

What is needed in order to make further informed predictions about deviations in L2 prosody in the semantic dimension is a more substantial body of research into the intonational functions in the two languages conducted within the same analytic framework.

In the frequency dimension, we can make a few tentative predictions about possible deviations from the native English norms in the speech of Bulgarian learners, namely, that due to L1 interference they will use more frequently L*+H as a pre-nuclear pitch accent and H+!H* as a nuclear one in declaratives than native speakers of English. However, these predictions are currently based on analyses of speech samples from a relatively small number of speakers. Until results based on larger, comparable as well as stylistically diverse datasets become available, it will be difficult to draw reliable comparisons of the frequency of use of prosodic constituents in this dimension.

On the basis of the differences between English and Bulgarian in terms of lexical stress cues and position we can predict that Bulgarian learners are likely to have difficulties with the acquisition of English word stress. The weight-insensitive system of Bulgarian and the rather complex information needed for the identification of

stress position in an unfamiliar word in English will make it difficult to teach any rules for stress placement in the English language classroom. The different cues to stress reported to be of importance in the two languages are also a potential source of production as well as perception problems. Lexical stress is thus likely to cause problems in both the systemic and the realisational dimension.

The comparison of the rhythmic characteristics of English and Bulgarian predicts deviations from stress-timing in the L2 speech of Bulgarian learners of English due to L1 interference. However, the exact nature of the deviations, and their classification in terms of the four dimensions of LILt also remain topics for further investigation.

In conclusion, the L2 Intonation Learning theory (LILt) proposed by Mennen (2015) has gained popularity and is beginning to be widely used as a tool for comparison of the similarities and differences between L1 and L2 intonation, and for the formulation and testing of research hypotheses regarding the difficulties experienced by foreign learners when they acquire the prosody of L2. A number of questions remain to be addressed, such as the extent to which L2 intonation acquisition depends on the acquisition of the segmental system and of other prosodic properties of the L2 like prosodic length and prosodic structure; the role of universal constraints on L2 intonation learning; similarities and differences between learners from different L1 backgrounds, etc. In addition, stress, rhythm and other temporal aspect of language prosody need to become a part of any model of prosodic learning. In spite of the many questions which still seek an answer, however, LILt remains the most thorough and well-developed theory of L2 intonation learning to date. It provides a sound starting point for L2 prosody research but, as noted by Mennen (2015) herself, it should be treated as "an evolving or 'working' model, which is subject to change when more data are published".

References

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Anderson, A., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. and Weinert, R. (1991). 'The HCRC Map Task Corpus'. *Language and Speech* 34(4), pp. 351-366.

Andreeva, B. (2007). 'Zur Phonetik und Phonologie der Intonation der Sofioter-Varietät des Bulgarischen', PhD dissertation. *PHONUS 12*, Institute of Phonetics, University of the Saarland, Saarbrücken.

Andreeva, B. (2009). 'Towards the intonational phonology of the Sophia Variety of Bulgarian'. In: Zybatow, G., Junghanns, U., Lenertová, D. and Biskup, P. (Eds.) *Studies in Formal Slavic Phonology, Morphology, Syntax, Semantics and Information Structure*. *Proc. of FDSL 7*. Frankfurt/Main: Peter Lang, pp. 357-371.

Andreeva, B. (2017). 'Prosodic Encoding of Phrasal Prominence and Information Structure in a Second Language: When Bulgarian and German Prosody meet'. *Contrastive Linguistics* XLII, 2017/1, pp. 59-83.

Andreeva, B., Avgustinova, T. and Barry, W. J. (2001). 'Link-associated and focus-associated accent patterns in Bulgarian'. In: Zybatow, G., Junghanns, U., Mehlhorn, G. and Szucsich, L. (Eds.) *Current Issues in Formal Slavic Linguistics* (Linguistik International; 5). Frankfurt/Main: Peter Lang, pp. 353-364.

Andreeva, B. and Oliver, D. (2005). 'Information Structure in Polish and Bulgarian: Accent Types and Peak Alignment in Broad and Narrow Focus'. In: Franks, S., Gladney, F. Y. and Tasseva-Kurktchieva, M. (Eds.) *Formal Approaches to Slavic Linguistics 13: The South Carolina Meeting 2004*, Ann Arbor: Michigan Slavic Publications, pp. 1-12.

Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Oleskowicz-Popiel, M. and Trouvain, J. (2014). 'Comparison of pitch range and pitch variation in Slavic and Germanic languages'. In: Campbell, N., Gibbon, D. and Hirst. D. (Eds.) *Prodeedings of the 7th International Conference on Speech Prosody*. TCD, Dublin, Ireland, pp. 776-780.

Andreeva, B., Möbius, B., Demenko, G., Zimmerer, F. and Jügler, J. (2015). 'Linguistic measures of pitch range in Slavic and Germanic languages'. *Proceedings of Interspeech 2015. Speech beyond Speech: Towards a Better Understanding of the Most Important Biosignal.* ISCA, Dresden, pp. 968-972.

Andreeva, B., Barry, W. J. and Koreman, J. (2016). 'Local and global cues in the prosodic realisation of broad and narrow focus in Bulgarian'. In: Żygis, M. and Malisz, Z. (Eds.) *Special Issue, Slavic Perspectives on Prosody*. *Phonetica* 73, pp. 260-282.

Andreeva, B. and Dimitrova, S. (in press, 2022). 'Intonation and information structure'. In: Bermel, N. and Fellerer, J. (Eds.) *The Oxford Guide to the Slavonic Languages*. Oxford: Oxford University Press.

Andreeva, B. and Dimitrova, S. (submitted, 2022b). 'Bulgarian Tones and Break Indices (BG_ToBI): a system for intonational annotation'. *Philologia* 41.

Andreeva, B. and Dimitrova, S. (2022c). 'The influence of L1 prosody on Bulgarian-accented German and English.' *Speech Prosody,* 23-26 May 2022, Lisbon, Portugal.

Aoyama, K. and S. G. Guion, S. G. (2007). 'Prosody in second language acquisition: Acoutic analyses of duration and F0 range'. In: Bohn, O.-S. and Munro, M. J. (Eds.) *The role of language experience in second-language speech learning – In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 281-297.

Armstrong, L. E. and Ward, I. C. (1926). *Handbook of English Intonation*. Cambridge/Leipzig: M. Heffer & Sons Ltd. / B. G. Teubner.

Beckman, M. and Pierrehumbert, J. (1986). 'Intonational structure in Japanese and English'. *Phonology Yearbook* 3, pp. 15-70.

Beckman, M. E., Hirschberg, J. and Shattuck-Hufnagel, S. (2005). 'The original ToBI system and the evolution of the ToBI framework'. In: Jun, S.-A. (Ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, pp. 9-54.

Best, C. (1995). 'A direct realist view of cross-language speech perception'. In: Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, Baltimore, pp. 171-204.

Best, C. and Tyler, M. D. (2007). 'Non-native and second-language speech perception'. In: Bohn, O.-S. and Munro, M. J. (Eds.) *Language experience in second language speech learning: In honor of James Emil Flege.* Amsterdam: J. Benjamins, pp. 13-34.

van Bezooijen, R. (1995). 'Sociocultural aspects of pitch differences between Japanese and Dutch women'. *Language and Speech*, 38(3), pp. 253-265.

Boersma, P. and Weenink, D. (1992-2022). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.06 Available at: http://www.praat.org/ (Accessed: 23 February 2022)

Bolinger, D. L. (1958). 'A Theory of Pitch Accent in English', *Word*, 14:2-3, pp. 109-149, DOI: 10.1080/00437956.1958.11659660

Bolinger, D. L. (Ed.) (1972). *Intonation*. Harmondsworth: Penguin Books.

Bolinger, D. L. (1986). *Intonation and its parts: Melody in Spoken English*. London: Edward Arnold.

Bolinger, D. L. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse.* Stanford, California: Stanford University Press.

Bradford, B. (1988). *Intonation in Context*. Cambridge: Cambridge University Press.

Brazil, D. (1994). *Pronunciation for Advanced Learners of English.* Cambridge: Cambridge University Press.

Brazil, D., Coulthard, M. and Johns, C. (1980). *Discourse Intonation and Language Teaching.* London: Longman.

Brown, G. (1983). 'Prosodic structure and the given/new distinction'. In: Cutler, A. and Ladd, D. R. (Eds.). *Prosody: Models and measurements*. Berlin: Springer Verlag, pp. 67-77.

Brown, G. (1977). *Listening to Spoken English*. (2nd edn. 1990), Longman.

Bulgarian Transliteration Law (2009). *Darzhaven vestnik*19, 13.03.2009. Available at: https://www.lex.bg/bg/laws/ldoc/2135623667 (Accessed: 11 January 2021). Закон за транслитерацията (2009), Държавен вестник 19, 13.03.2009.

Cauldwell, R. (2002). *Streaming Speech: Listening and Pronunciation for Advanced Learners of English*. British/Irish version. Birmingham: speechinaction.

Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. (Language learning and language teaching 1.) Amsterdam: John Benjamins.

Coupé, C., Oh, Y. M., Dediu, D. and Pellegrino, F. (2019). 'Different languages, similar encoding efficiency: Comparable informationrates across the human communicative niche'. *Science Advances*, vol. 5, no. 9, eaaw2594, DOI: 10.1126/sciadv.aaw2594.

Couper-Kuhlen, E. (1986). *An introduction to English prosody.* Tübingen: Max Niemeyer Verlag. / London: Edward Arnold.

Cruttenden, A. (1986). *Intonation. (Cambridge Textbooks in Linguistics)*. Cambridge: Cambridge University Press.

Cruttenden, A. (2008). *Gimson's Pronunciation of English.* 7th ed. Revised by Alan Cruttenden. London: Hodder Education.

Crystal, D. (1969*)*. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Crystal, D. and Quirk, R. (1964). *Systems of Prosodic and Paralinguistic Features in English*. The Hague: Mouton.

Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. PhD dissertation, University of Chicago.

Dainora A. (2006). 'Modeling intonation in English: a probabilistic approach to phonological competence'. In: Goldstein, L., Whalen, D.

and Best, C. (Eds.) *Laboratory Phonology 8*. New York: Mouton de Gruyter, pp. 107-132.

Dauer, R. (1983). 'Stress-Timing and Syllable-Timing Reanalyzed.' *Journal of Phonetics*, vol. 11, pp. 51-62.

Dauer, R. (1987). 'Phonetic and Phonological Components of Language Rhythm'. In: *Proceedings of the 11th International Congress of Phonetic Sciences*. Tallinn: Academy of Sciences of the Estonian SSR, pp. 447-450.

Dimitrov, G. (2017): Димитров, Г. (2017). 'Английската интонация в български контекст'. В: Данова, М., Димитрова, С. и Ковачева, М. (Ред.) *Нови направления в синхронната и диахронната фонетика и фонология*. София: УИ „Св. Климент Охридски", с. 167-182.

Dimitrova, S. (1998). 'Bulgarian speech rhythm: stress-timed or syllable-timed?' *Journal of the International Phonetic Association*, vol. 27, no. 1-2, pp. 27-34.

Dimitrova, S. (1998b). 'Stress shift in English, Bulgarian English and Bulgarian'. *Proceedings of the International Conference "70 Years of English and American Studies in Bulgaria"*, Sofia, 1-3 October 1998, pp. 30-34.

Dimitrova, S. (2016): Димитрова, С. (2016). 'Международна система за прозодична транскрипция IPrA: предимства, проблеми, предизвикателства'. В: Стаменов, М. и Панчев, И. (Ред.) *Актуални проблеми на съвременната лингвистика. Юбилеен сборник в чест на проф. д.ф.н. д-р хон. кауза Стефана Димитрова.* София: Издателство на БАН „Проф. Марин Дринов", с. 90-97.

Dimitrova, S. (2018). 'The choice of model in teaching English pronunciation at tertiary level'. In: Kovacheva, M. (Ed.) *Language Close Up. Papers in Honour of Christo Stamenov*. Sofia: St Kliment Ohridski University Press, pp. 249-255.

Dimitrova, S. (2019). 'Some prosodic characteristics of Bulgarian English'. In: Slavova, E., Bagasheva, A., Slavova, K., Tincheva, N., Muharska, R., Ishpekova, R. and Catalan, Z. (Eds.) *Traditions and Transitions*, *vol. 2.* Sofia: St Kliment Ohridski University Press, pp. 125-138.

Dimitrova, S. (2021). 'Suprasegmental features of Bulgarian English Speech'. *Philologia*, vol. 40, pp. 73-92.

Dimitrova, S. (2022). *In Search of Speech Rhythm in Bulgarian (in comparison with English).* Sofia: Polis.

Dimitrova, S. and Jun, S.-A. (2015). 'Pitch accent variability in focus production and perception in Bulgarian declaratives'. In: Wolters, M., Livingstone, J., Beattie, B., Smith, R., MacMahon, M., Stuart-Smith, J. and Scobbie, J. (Eds.) *Proceedings of the 18th ICPhS*. Glasgow: University of Glasgow, paper no. 832, 5 pp. Available at: https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2015/Papers/ICPHS0832.pdf (Accessed: 10 October 2021.)

Dimitrova, S. and Andreeva, B. (2017): Димитрова, С. и Андреева, Б. (2017). 'Пренуклеарни акценти в българския език'. В: Данова, М., Димитрова, С. и Ковачева, М. (Ред.) *Нови направления в синхронната и диахронната фонетика и фонология*. София: УИ „Св. Климент Охридски", с. 152-166.

Dimitrova, S., Andreeva, B., Gabriel, C. and Grünke, J. (2018). 'Speaker age effects on prosodic patterns in Bulgarian'. *Proc. Speech Prosody 2018* (Poznań, Poland), pp. 709-713.

Eckert, H. and Laver, J. (1994). *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation*. Weinheim: Psychologie Verlags Union.

Eckman, F. R. (1977). 'Markedness and the Contrastive Analysis Hypothesis'. *Language Learning,* 27 , pp. 315-330.

Estebas – Vilaplana, E. (2015). 'The learning of English intonation by Spanish speakers in a distance education environment'. *PTLC 2015*, UCL, London, pp. 39-43.

Estebas – Vilaplana, E. (2018). 'The teaching and learning of L2 English intonation in a distance education environment: TL_ToBI vs. the traditional models'. *Linguistica* 57 (1), pp. 73-91.

Flege, J. E. (1995). 'Second-language speech learning: Theory, findings, and problems'. In: Strange, W. (Ed.) *Speech perception and linguistic experience: Issue in cross-language research*. Timonium, MD: York Press, Baltimore, pp. 229-273.

Flege, J. E. (1997). 'English vowel productions by Dutch talkers: More evidence for the "similar" vs "new" distinction'. In: James, A. and Leather, J. (Eds.) *Second-Language Speech: Structure and Process*. Berlin, Germany: Mouton de Gruyter, pp. 11-52.

Flege, J. E. (2007). 'Language contact in bilingualism: Phonetic system interactions'. In: Cole, J. and Hualde, J. I. (Eds.) *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, pp. 353-380.

Flege, J. E. and Bohn, O.-S. (2021). 'The revised speech learning model (SLM-r)'. In: Wayland, R. E. (Ed.) *Second Language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, pp. 3-83.

Georgieva, M. (1967): Георгиева, М. (1967). 'По въпроса за мелодиката на българския език'. *Български език* 1, с. 13-22.

Georgieva, M. (1970): Георгиева, М. (1970). 'Мелодия на простото изявително разширено изречение в българската литературна реч'. *Известия на Института за български език* 19, с. 83-89.

Georgieva, M. (1974): Георгиева, М. (1974). 'Интонацията на един вид елиптични въпросителни изречения в български и руски език'. *Сборник в памет на проф. Ст. Стойков. Езиковедски изследвания*. София, с. 77-82.

Gimson, A. C. (1970). *An Introduction to the Pronunciation of English*. 2nd ed. London: Edward Arnold.

Grabe, E. (1997). 'Comparative intonational phonology: English and German'. In: *Intonation: Theory, Models, and Applications*. Athens, Greece, 18-20 September, 1997, pp. 157-160. Available at: ISCA Archive https://www.isca-speech.org/archive_open/archive_papers/int_97/inta_157.pdf (Accessed: 5 May 2022.)

Grabe, E, (1998). 'Pitch accent realisation in English and German'. *Journal of Phonetics* 26, pp. 129-143.

Grabe, E. and Warren, P. (1995). 'Stress shift: Do speakers do it or do listeners hear it?' In: Connell, B. and Arvaniti, A. (Eds.) *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, pp. 95-110.

Grabe, E., Post, B., Nolan, F. and Farrar, K. (2000). 'Pitch accent realisation in four varieties of British English'. *Journal of Phonetics* 28 (2), pp. 161-185.

Grabe, E., Post, B. and Nolan, F. (2001). 'Modelling intonational variation in English. The IViE system'. In: Puppel, S. and Demenko, G. (Eds.) *Proceedings of Prosody 2000*. Adam Mickiewitz University, Poznan, Poland, pp. 51-57.

Grabe, E. and Post, B. (2002). 'Intonational Variation in the British Isles'. *Speech Prosody 2002, Aix-en-Provence, France, April 11-13, 2002*. Available at: https://www.isca-speech.org/archive_open/sp2002/sp02_343.pdf (Accessed: 17 April 2022).

Grice, M. and Baumann, S. (2007). 'An introduction to intonation – functions and models'. In: Trouvain, J. and Gut, U. (Eds.) *Non-Native Prosody. Phonetic Description and Teaching Practice*. Vol. 186, series 'Trends in Linguistics. Studies and Monographs [TiLSM]. Mouton De Gruyter, pp. 25-52.

Grice, M., German, J. S. and Warren, P. (2020). 'Intonation systems across varieties of English.' In: Gussenhoven, C. and Chen, A. (Eds.) *The Oxford Handbook of Language Prosody*. Oxford: Oxford University Press, pp. 285-302.

Gussenhoven, C. (1984) *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, C. (1991). 'The English rhythm rule as an accent deletion rule.' *Phonology*, vol. 8, no. 1, pp. 1-35.

Gussenhoven, C. (2008). 'Types of Focus in English'. In: Lee, C., Gordon, M. and Büring, D. (Eds.) *Topic and Focus. Studies in Linguistics and Philosophy*, vol 82. Dordrecht: Springer, pp. 83-100. DOI: 10.1007/978-1-4020-4796-1_5

Gut, U. (2009). *Non-native Speech. A corpus-based analysis of the phonetic and phonological properties of L2 English and L2 German*. Frankfurt: Peter Lang.

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. The Hague: Mouton.

Hirschberg, J. and Beckman, M. (1994). *The ToBI Annotation Conventions*, online MS. Available at: http://www.cs.columbia.edu/~julia/files/conv.pdf (Accessed: 26 April 2022)

Hualde, J. I. and Prieto, P. (2016). 'Towards an International Prosodic Alphabet (IPrA)'. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 7(1), pp. 1-25, DOI: 10.5334/labphon.11

Im, S., Cole, J. and Baumann, S. (2018). 'The probabilistic relationship between pitch accents and information status in public speech'. *9th International Conference on Speech Prosody 2018, 13-16 June 2018, Poznań, Poland*, pp. 508-511. DOI: 10.21437/SpeechProsody.2018-103.

Ivanchev, Sv. (1957/1978). 'Nablyudenia varhu upotrebata na chlena v balgarskia ezik'. In: Pashov, P. (Ed.). *Pomagalo po balgarska morfologia. Imena*. Sofia: Nauka i izkustvo, pp. 186-211. Иванчев, Св. (1957/1978). 'Наблюдения върху употребата на члена в българския език'. В: Пашов, П. (Ред.) *Помагало по българска морфология. Имена.* София: Наука и изкуство, с. 186-211.

Jacewicz, E., Fox, R. A., O'Neill, C. and Salmons, J. (2009). 'Articulation rate across dialect, age, and gender'. *Language Variation and Change*, vol. 21(2), pp. 233-256. DOI: 10.1017/S0954394509990093.

Johannsson, F. (1973). *Immigrant Swedish Phonology: A study in multiple contact analysis*. Lund, Sweden: CWK Gleerup.

Jones, D. (1976). *An Outline of English Phonetics*. Cambridge: Cambridge University Press.

Jun, S.-A. (2000). 'K-ToBI (Korean ToBI) labelling conventions: Version 3'. *Speech Sciences* 7, pp. 143-169.

Jun, S.-A. (Ed.). (2005), *Prosodic Typology: the Phonology of Intonation and Phrasing*, Oxford: Oxford University Press.

Jun, S.-A. (Ed.). (2014), *Prosodic Typology II: the Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, S.-A. (2022). 'The ToBI Transcription System: Conventions, Strengths, and Challenges'. In: Barnes, J. and Shattuck-Hufnagel,

S. (Eds.) *Prosodic Theory and Practice*. Cambridge, MA: MIT Press, pp. 151-181.

Kingdon, R. (1958). *The Groundwork of English Intonation*. London: Longmans.

Kuhl, P. K. (1991). 'Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories. Monkeys do not.' *Perception and Psychophysics* 50 (2), pp. 93-107.

Kuhl, P. K. (1992). 'Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the "centers" of speech categories'. In: Tohkura, Y., Vatikiotis-Bateson, E. and Sagisaka, Y. (Eds.) *Speech, Perception, Production and Linguistic Structure*. Tokyo: IOS Press, pp. 240-264.

Kuhl, P. K. (2000). 'A new view of language acquisition'. *PNAS – Proceedings of the National Academy of Sciences of the United States of America, October 24, 2000*, vol. 97 (22), pp. 11850-11857. Available online at: https://doi.org/10.1073/pnas.97.22.11850 (Accessed: 28 September 2021).

Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. (1996/2008). *Intonational Phonology*. (2nd edition 2008). Cambridge: Cambridge University Press.

Ladd, D. R. and Schepman, A. (2003). '"Sagging transitions" between high pitch accents in English: Experimental evidence'. *Journal of Phonetics* 31, pp. 81-112.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press: Ann Arbor.

Lecumberri, M. L. G. (1997). 'Traditional British analyses of intonation: a review'. *Atlantis* XIX (2), pp. 103-111.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1977). 'Isochrony reconsidered'. *Journal of Phonetics* 5, pp. 253-263.

Lenneberg, E. (1969). *Biological Foundations of Language*. New York: Wiley.

Liberman, M. and Prince, A. (1977). 'On stress and linguistic rhythm'. *Linguistic Inquiry* 8, 2, pp. 249-336.

Ladd, D. R. and Schepman, A. (2003). '"Sagging transitions" between high pitch accents in English: Experimental evidence'. *Journal of Phonetics* 31, pp. 81-112.

Mahrova, T. (1978): Махрова, Т. (1978). 'За интонационните конструкции'. Български език 4, с. 303-312.

Misheva, A. (1991): Мишева, А. (1991). Интонационна система на българския език. София: БАН.

Misheva, A. and Nikov, M. (1998). 'Intonation in Bulgarian'. In: Hirst, D. and Di Cristo, A. (Eds.) *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, pp. 275-287.

Nikolaeva, T. (1977): Николаева, Т. (1977). Фразовая интонация славянских языков. Москва: Наука.

Mennen, I. (2007). 'Phonological and phonetic influences in non-native intonation'. In: Trouvain, J. and Gut, U. (Eds.) *Non-Native Prosody: Phonetic Descriptions and Teaching Practice*, Berlin: Mouton De Gruyter, pp. 53-76.

Mennen, I. (2015). 'Beyond segments: towards an L2 intonation learning theory (LILt).' In: Delais-Roussarie, E., Avanzi, M. and Herment, S. (Eds.) *Prosody and Languages in Contact: L2 acquisition, attrition, languages in multilingual situations.* Springer Verlag, pp. 171-188. DOI: 10.1007/978-3-662-45168-7_9

Mennen, I., Schaeffler, F. and Docherty, G. (2007). 'Pitching it Differently: A Comparison of the Pitch Ranges of German and English Speakers.' *Proceedings of ICPhS XVI*, Saarbrücken, 6-10 August 2007, pp. 1769-1772.

Mennen, I., Schaeffler, F. and Docherty, G. (2008). 'A methodological study into the linguistic dimensions of pitch range differences between German and English.' *Proc. IV Conference on Speech Prosody*, University of Campinas, 2008, pp. 527-530.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Foris: Dordrecht.

Nespor, M. and Vogel, I. (1989). 'On clashes and lapses'. *Phonology* 6, pp. 69-116.

Nolan, F. (2006). ‚Intonation'. In: Aarts, B. and McMahon, A. (Eds.). *The Handbook of English Linguistics*. Hoboken: Blackwell, pp. 433-458.

Nolan, F. (2022). 'The Rise and Fall of the British School of Intonation Analysis'. In: Barnes, J. and Shattuck-Hufnagel, S. (Eds.) *Prosodic Theory and Practice*. MIT Press, pp. 319-350.

O'Connor, J. D. and Arnold, G. F. (1973). *Intonation of Colloquial English*. London: Longman.

Ohara, Y. (1992). 'Gender dependent pitch levels: A comparative study in Japanese and English'. In: Hall, K., Bucholtz, M. and Moonwomon, B. (Eds.) *Locating Power: Proceedings of the Second Berkeley Women and Language Conference* 2, Berkeley, pp. 478-488.

Oliver, D. and Andreeva, B. (2008). 'Peak Alignment in Broad and Narrow Focus. A Cross-language Study'. In: Zybatow, G., Szucsich, L., Junghanns, U. and Meyer, R. (Eds.) *Formal Description of Slavic Languages: The Fifth Conference (FDSL5), Leipzig 2003*. Frankfurt am Main: Peter Lang, pp. 134-148.

Palmer, H. E. (1922). *English intonation: with systematic exercises*. Cambridge: Heffer.

Patterson, D. and Ladd, D. R. (1999). 'Pitch range modelling: linguistic dimensions of variation'. *ICPhS 99: Proceedings of the 14th International Congress of Phonetic Sciences,* University of California, Berkeley, San Francisco, pp. 1169-1172.

Penchev, Y. (1980): Пенчев, Й. (1980). *Основни интонационни контури в българското изречение*. София: БАН.

Pešková, A. (2020). 'Towards the Developmental L2 Intonation hypothesis: Evidence from L2 Italian and L2 Spanish'. *Going Romance* 34, Paris (France), 26-28 November 2020.

Pešková, A. (2021). 'L2 varieties in contrast and Developmental L2 Intonation Hypothesis.' *1st International Conference on Tone and Intonation (TAI) 2021 "Tone and Intonation in a globalized, digital world"*, Sonderborg (Denmark), 6-9 December 2021.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*, PhD dissertation. Cambridge, MA: MIT Press.

Pierrehumbert, J. (2000). 'Tonal elements and their alignment'. In: Horne, M. (Ed.) *Prosody: Theory and Experiment. Studies Presented to Gosta Bruce*. Dordrecht: Kluwer, pp. 11-26.

Pierrehumbert, J. and Beckman, M. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press. Cambridge, MA Cambridge, MA.

Pierrehumbert, J. and Hirschberg, J. (1990). 'The meaning of intonational contours in the interpretation of discourse'. In: Cohen, P. R., Morgan, J. and Pollack, M. E. (Eds.) *Intentions in communication*, Cambridge, MA: MIT Press, pp. 271-311.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Piske, T. (2007). 'Implications of James E. Flege's research for the foreign language classroom'. In: Munro, M. J. and Bohn, O.-S. (Eds.) *Language experience in second language speech learning. In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 301-314.

Piske, T., MacKay, I. R. A. and Flege, J. E. (2001). 'Factors Affecting Degree of Foreign Accent in an L2: A Review'. *Journal of Phonetics* 29, pp. 191-215.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S. and Raymond, W. (2005). 'The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability'. *Speech Communication*, vol. 45, no. 1, pp. 89-95.

Popov, K. (1963): Попов, К. (1963). *Съвременен български език. Синтаксис.* София: Наука и изкуство.

Reetz, H. (1999). *Artikulatorische und akustische Phonetik*. Trier: Wissenschaftlicher Verlag.

Roach, P. (1994). 'Conversion between prosodic transcription systems: "Standard British" and ToBI'. *Speech Communication* 15, pp. 91-99.

Roach, P. (1998). 'Some languages are spoken more quickly than others'. In: Bauer, L. and Trudgill, P. (Eds.) *Language Myths.* Penguin, pp. 150-158.

Roach, P. (2009). *English Phonetics and Phonology: a practical course*. 4th edn. Cambridge: Cambridge University Press.

Rogerson, P. and Gilbert, J. (1992). *Speaking Clearly*. Cambridge: Cambridge University Press.

SAS (2021), "JMP" 16 ed. Cary (NC): SAS Institute.

Selinker, L. (1972). 'Interlanguage.' *International Review of Applied Linguistics in Language Teaching* 10 , pp. 209-231.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992). 'ToBI: A standard for labeling English prosody'. *Proceedings of ICSLP 92*, Banff, Alberta, Canada, pp. 867-870.

Stockwell, R. P. and Bowen, J. D. (1965). *The Sounds of English and Spanish*. Chicago: University of Chicago Press.

Stoykov, St. (1942): Стойков, Ст. (1942). 'Български книжовен изговор. Опитно изследване'. *Сборник на Българската академия на науките и изкуствата* 37 (3), с. 281-415.

Stoykov, St. (1966): Стойков, Ст. (1966). *Увод във фонетиката на българския език*. София: Наука и изкуство.

Stoykova, N. (2018). 'English weak forms – a challenge for Bulgarian learners of English?: a pedagogical perspective'. *Philologia* 16(16), pp. 99-106. DOI:10.18485/philologia.2018.16.16.7

Talkin, D. (1995). 'A robust algorithm for pitch tracking (RAPT)'. In: Kleijn, W. B. and Paliwal, K. K. (Eds.) *Speech Coding and Synthesis.* New York: Elsevier.

Themistocleous, Ch. (2016). 'Seeking an Anchorage. Stability and Variability in Tonal Alignment of Rising Pre-nuclear Pitch Accents in Cypriot Greek'. *Language and Speech* 59(4), pp. 433-461.

Tilkov, D. (1981): Тилков. Д. (1981). *Интонацията в българския език.* София: Народна просвета.

Tilkov, D. (Ed.) (1982): Тилков, Д. (Ред.) (1982). *Граматика на съвременния български книжовен език. Том 1: Фонетика*. София: БАН.

Tilkov, D. and Misheva, A. (1978): Тилков, Д. и Мишева, А. 'Физически корелати и перцепция на словното ударение в българския език'. *Slavica Pragensia XXI, Acta Universitatis Carolinae – Philologica* 3-5, pp. 83-87.

ToBI website (1999). Ohio: The Ohio State University Department of Linguistics. Available at: http://wwwhomes.uni-bielefeld.de/gibbon/Docs/ToBI.htm (Accessed: 10 May 2022).

Trager, G. L. and Smith, H. L. (1951/1957). *An Outline of English Structure*. Republished with corrections 1957. Washington: American Council of Learned Societies.

Trofimovich, P. and Baker, W. (2006). 'Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech'. *Studies in Second Language Acquisistion* 28, pp. 1-30.

Tyler, M. D. (2019). 'PAM-L2 and phonological category acquisition in the foreign language classroom'. In: Nyvad, A. M., Hejná, M., Højen, A., Jespersen, A. B. and Sørensen, M. H. (Eds.) *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*. Aarhus: Aarhus University, pp. 607-630.

Welby, P. and Lœvenbruck, H. (2005). 'Segmental "anchorage" and the French late rise'. *Proc. Interspeech* 2005, pp. 2369-2372.

Welby, P. and Lœvenbruck, H. (2006). 'Anchored down in Anchorage: Syllable structure and segmental anchoring in French'. *Italian Journal of Linguistics / Rivista di linguistica*, Pacini Editore S.p.A, 2006, 18, pp. 74-124.

Wells, J. C. (2006) *English Intonation: An Introduction.* Cambridge: Cambridge University Press.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B. and Trouvain, J. (2014). 'Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers'. In: Campbell, N., Gibbon, D. and Hirst, D. (Eds.) *Proceedings of the 7th Speech Prosody Conference, Trinity College Dublin, Ireland*, Dublin: International Speech Communication Association, pp. 1037-1041.

Zimmerer, F., Andreeva, B., Jügler, J. and Möbius, B. (2015). 'Comparison of pitch profiles of German and French speakers speaking French and German'. In: Wolters, M., Livingstone, J., Beattie, B., Smith, R., MacMahon, M., Stuart-Smith, J. and Scobbie, J. (Eds.) *Proceedings of the 18th ICPhS.* Glasgow: University of Glasgow, paper no. 615, 5 p. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0615.pdf (Accessed: 10 October 2021).

Snezhina Dimitrova

Prosody in L2.
Bulgarian-Accented English

Първо издание
Българска

Художник *Константин Жеков*

Формат 60×90/16. Печ. коли 11
Печатница „Симолини–94“ – София

„ИК Полис“ ООД
1000 София, ул. „Парчевич“ 37, вх. Б, ет. 3
тел.: (02) 981-76-27

www.polis-publishers.com

Цена 16,00 лв.

ISBN 978-954-796-096-1