

# Intonation

Introductory article

*Sun-Ah Jun*, UCLA, Los Angeles, California, USA

## CONTENTS

*What is intonation?*

*Prosodic constituency*

*Relation to syntactic structure*

*Semantic and pragmatic functions*

*Intonation is the melody of a sentence or a phrase. It marks a grouping of words or the prominence relations among words, and reflects the syntactic structure and the semantic and pragmatic meaning of a sentence.*

## WHAT IS INTONATION?

Traditionally, intonation has been defined as the melody of a sentence or a phrase. A melody is, in a general sense, a pattern of pitch changes. Some parts of a sentence are produced with higher or lower pitches than other parts of a sentence, and the overall pitch pattern of a sentence delivers the speaker's intention as well as certain aspects of the meaning of the sentence or sentence type. However, intonation involves more than just the global changes in pitch over the course of a sentence. The overall pitch pattern has an internal structure. Intonation marks groupings of words and defines a hierarchy for these groupings. This means that, while the acoustic realization of pitch is continuous, intonation is not in fact an undividable whole – instead it is decomposable into subcomponent pitch events. Some pitch events mark the boundaries between groupings of words, either small boundaries or large, and others mark the prominence relations within a group of words.

The boundary of a group of words, marked by pitch patterns called boundary tones, is sometimes also marked by duration. For example, syllables at the end of a group of words are often substantially lengthened, and this is known as preboundary lengthening. Further, prominent words that are marked by pitch are also often characterized by higher intensity and longer duration. In this way, pitch patterns carry the tune or melody of a sentence or a phrase and, together with intensity and duration, they also convey the prominence relations among the words in a phrase. Thus, intonation, whether it marks a phrase boundary or prominence relations, is comprised not only of

pitch changes but also of other prosodic features such as intensity and duration. For this reason, the terms 'intonation' and 'prosody' are often used interchangeably. In sum, intonation has an internal structure, and the structure is cued by pitch, intensity, and duration. (See **Prosody**)

Intonation can change the meaning of a sentence but not the meaning of a word. In this sense, it is different from (lexical) pitch accent or tone, both of which can change the meaning of a word. In a pitch accent language, such as Japanese or Swedish, each word has a distinctive pitch pattern. That is, words with different pitch patterns do not have the same meaning, even when the words have the same sequence of consonants and vowels. For example, in Japanese, 'kami' means 'God' when high pitch occurs on the syllable 'ka', but the word means 'paper' when high pitch occurs on the syllable 'mi'. Similarly, in a tone language such as Mandarin, each syllable of a word has a distinctive pitch pattern. For example, one syllable 'ma' can mean 'mother' when it is produced with a high pitch, 'horse' when it is produced with a low rising pitch, 'hemp' when it is produced with a high rising pitch, and 'scold' when it is produced with a falling pitch. In languages such as these, where pitch is distinctive at the word level, pitch variations at the phrasal level, i.e. intonation, are not as easily observable as they are in nontonal and/or nonpitch accent languages such as English, French, or Korean. Still, all languages have intonation, and intonation conveys the structure and meaning of a sentence as well as the speaker's intention.

Though there are certain features of intonation that are similar across languages, many attributes of the intonation patterns found within a given language are determined on a language-specific basis. For example, in most languages pitch rises at the end of a yes–no question and falls at the end of a statement. However, in Chickasaw, an American Indian language of the Western Muskogean

language family spoken in south-central Oklahoma, the reverse pattern is found. That is, pitch falls at the end of a yes–no question and rises at the end of a statement. Languages also differ in terms of sentence-medial pitch patterns. This depends on the intonational groupings within a sentence and the types of pitch movement permitted within a word in the language. For instance, in English, the stressed syllable of a word can be realized with several different pitch patterns (e.g. high, mid, low, rising), and the edge of an intonational grouping can be marked by either a high or a low pitch. On the other hand, in Seoul Korean, pitch patterns are not linked to specific syllables within a word. Instead, they are linked to certain locations within a phrase. The phrase initial syllable can be either high or low, but the phrase final syllable is almost always high. This suggests that the pitch patterns of sentences within a given language reflect the prosodic system of that language, i.e. the intonational structure of that language and whether the language has stress, tone, pitch accent, or none of these.

Finally, the pitch categories that appear in descriptions of intonation, such as high tone or low tone, are not absolute but are in fact relative. They are relative to adjacent pitch values within the sentence or phrase and are relative to the speaker's pitch range. In a given speaker's production, a syllable can be said to have a high tone intonationally if its pitch is higher than that of the immediately preceding syllable within the same prosodic phrase or if the pitch is realized at the top of the speaker's current pitch range. On the other hand, the interpretation of pitch categories in intonation is not gradient but categorical. The categorical nature of intonation is determined by the categorical nature of its component pitch events. If a statement ending in a low pitch is interpreted as a declarative, the absolute lowness of the pitch at the end of the sentence does not influence the degree of 'declarativeness' of that statement. It simply is or is not declarative. This contrasts with the paralinguistic features of speech which convey a speaker's emotional state (e.g. angry, fearful, joyous) and his or her attitude towards the listener (e.g. friendly, appeasing, aggressive). The interpretation of paralinguistic features is gradient. So, for instance, the higher the absolute pitch or intensity in a word or a phrase, the angrier or more aggressive the speaker is likely to be. In this article, we will discuss the categorical properties of intonation, and in the following sections we will show the role of intonation in grammar, especially with respect to its

role in prosodic grouping, its relation to syntactic structure, and its semantic and pragmatic functions. (See **Emotion**)

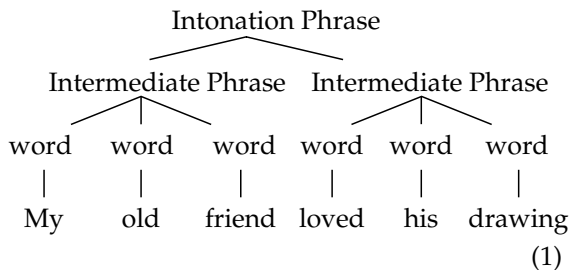
## PROSODIC CONSTITUENCY

When we produce a sentence, we tend to form subgroups of words within the sentence. This grouping reflects how close adjacent words are in terms of their meaning and function relative to one another. The degree of juncture or distance between each word is not always equal. Words belonging to the same group are closer together, i.e. have smaller juncture between them, than words belonging to different groups which have a larger juncture between them. The degree of juncture between words and groups of words is abstract and subjective, but studies have found that native speakers of the same language generally agree on the degree of juncture between the words in a given utterance. This is because the abstract entity termed 'juncture' reflects the organizational structure of the spoken utterance and is realized by consistent acoustic features such as pauses, pre-boundary lengthening, and other intonational attributes. (See **Prosody**)

The grouping of words within an utterance is called prosodic grouping or phrasing. There is typically more than one level of grouping within an utterance, reflecting the different degrees of juncture. That is, there are subgroupings within various groups within an utterance. For example, a sentence such as 'The lady with flowers loves John's brother' could be divided into two large groups, (The lady with flowers) and (loves John's brother), but the degree of juncture among the four words in the first group is not equal. The juncture between the first two words, 'the lady', or the next two words, 'with flowers', is smaller than the juncture between the second and the third word, 'lady with'. This shows that a whole utterance can be divided into groups which in turn can be divided into smaller groups, and so on. The hierarchical structure of groupings within an utterance is called the prosodic hierarchy, and each group in the prosodic hierarchy is called a prosodic constituent.

In English, investigators disagree on exactly how many levels of prosodic constituents there are above the word. We will not discuss the different proposals or the differences among the proposed levels here except to say that there are in general two prosodic levels above the word. The higher of the two is often called an intonation phrase and the lower one is called an intermediate phrase (or a

phonological phrase), and these are marked by intonation. The prosodic grouping for an example sentence is shown in (1). The whole sentence forms one intonation phrase, and is further divided into two intermediate phrases. Each intermediate phrase includes three words. This is one of many possible phrasings for this sentence.



Words can also be subdivided into prosodic constituents such as the foot (a sequence of strong and weak syllables), the syllable, and the mora. However, in this article we will focus on the prosodic constituents above the word level, since prosodic constituents below the word level are not directly related to intonation.

Experiments have shown that prosodic constituency and the hierarchical structure are psychologically real. This means that these abstract entities are part of native speakers' tacit knowledge about their language. Phonetic studies on speech production have shown that prosodic constituents are phonetically cued so as to mark the prosodic hierarchy. For example, the duration of a single sound segment is longer in the initial position of a higher prosodic phrase than it is in the initial position of a lower prosodic phrase, which in turn is longer than that in the initial position of a word occurring somewhere in the middle of a lower prosodic phrase. The same pattern is found for the duration of phrase final sounds. That is, the higher the prosodic constituent is in the prosodic hierarchy, the greater the degree of lengthening is at the edges of this prosodic constituent. (See **Phonetics**)

Psycholinguistic experiments have also shown that native speakers of a language are sensitive to the boundaries of prosodic constituents when they are processing sentences of their language. For example, when English speakers listen to sentences such as the one in (2) below, the processing time, i.e. the time taken to understand the meaning of the sentence, is shorter if the prosodic phrasing (marked by acoustic features such as preboundary lengthening, boundary tones, and pause) matches the syntactic or semantic grouping of the sentence (see next section for more detail). Thus, English

speakers process (2a) more quickly than (2b) because there is a match between the prosodic phrasing and the syntactic/semantic grouping in (2a), while there is a mismatch between the prosodic phrasing and syntactic/semantic grouping in (2b). (See **Sentence Processing: Mechanisms; Sentence Processing**)

Sentence: When George left the house, it was dark.

- a. expected prosodic phrasing: (When George left the house) (it was dark)
- b. unexpected prosodic phrasing: (When George left) (the house it was dark) (2)

In sum, utterances are prosodically divided into large and small phrases, and these prosodic constituents are hierarchically structured. Furthermore, native speakers phonetically cue the prosodic hierarchy in speech production by manipulating pitch, duration, and intensity, and they are also able to perceive these cues when processing a sentence.

## RELATION TO SYNTACTIC STRUCTURE

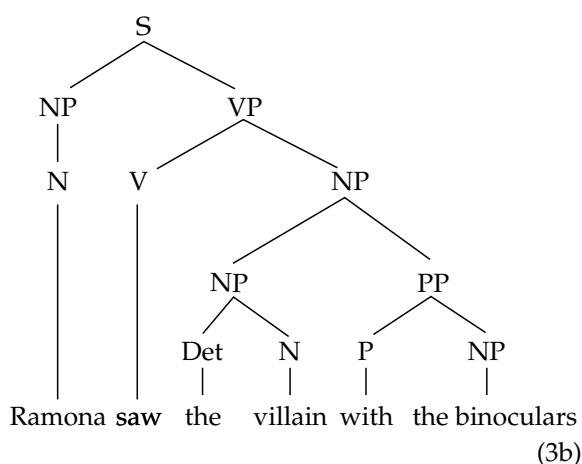
We have seen that sentences, when produced, are divided into intonational phrasings, or prosodic groupings, marked by intonation. (Thus, the term 'prosodic phrase' is used interchangeably with 'intonational phrase' below.) However, a sentence can be phrased in various ways depending on multiple factors such as syntax, the location of a focused word within the sentence, speech rate, and/or the length of the words and phrases. Therefore, prosodic phrasing does not always match the syntactic phrase of a sentence but is heavily influenced by syntax. In this section, we will show how the intonational phrasing is influenced by the syntactic structure of the sentence. (See **Syntax**)

Though there are many ways to phrase a sentence prosodically, prosodic phrase boundaries cannot be placed at random but are constrained by syntactic constituents. Specifically, there are certain syntactic constituents that tend to trigger phrase boundaries and others that do not. For example, speakers generally put a prosodic phrase boundary at a sentence-internal clause boundary (e.g. the sentence 'When you are tired, it's better to rest' is phrased as '(When you are tired) (it's better to rest)'), at the boundary of a parenthetical phrase (e.g. 'Disneyland is, Jane said, the most popular place to visit in LA' is phrased as '(Disneyland is) (Jane said) (the most popular place to visit

in LA)'), and before a tag question (e.g. 'He will win, won't he?' is phrased as '(He will win)(won't he)'). Similarly, speakers generally put a prosodic boundary between two words if the two words belong to two different syntactic phrases unless each syntactic phrase has only one word (e.g. between subject noun phrase and verb phrase).

On the other hand, speakers do not in general put a prosodic phrase boundary between words if they belong to the same syntactic phrase, unless the phrase is too long or each word is emphasized or produced very slowly. For example, it is very unlikely that a determiner and the following noun (e.g. 'a book', 'the teacher'), or an adjective and a noun (e.g. 'pretty girl', 'last year'), or a possessive pronoun and a noun (e.g. 'my book') would belong to two different prosodic phrases. For the same reason, it is unlikely for speakers not to put a prosodic boundary between two words if each belongs to a different syntactic phrase and the second syntactic phrase has other component words. For example, for the sentence 'The child with asthma outgrew the condition last year', combining 'asthma' and 'outgrew' together in the same prosodic phrase while putting a phrase boundary after 'outgrew' is not likely (i.e. \*(The child) (with asthma outgrew) (the condition) (last year)). In sum, this suggests that prosodic phrasing reflects the syntactic grouping of the words.

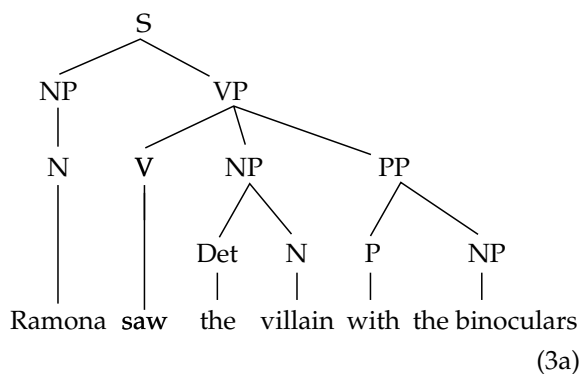
Psycholinguistic and phonetic studies have shown that speakers disambiguate a syntactically ambiguous sentence by producing different prosodic phrasings for each syntactic structure. For example, the sentence 'Ramona saw the villain with the binoculars' can mean two things. Namely, Ramona could have used the binoculars to see the man, or Ramona could have seen the man who possessed the binoculars. For the first meaning the prepositional phrase 'with the binoculars' modifies the verb 'saw', as shown in (3a), and for the second meaning the prepositional phrase modifies the preceding noun phrase, 'the villain', as in (3b).



It has been found that speakers may produce the sentence differently depending upon which of the two syntactic structures is intended. For the structure in (3a), a prosodic phrase boundary is placed between 'the villain' and 'with the binoculars', i.e. (Ramona saw the villain) (with the binoculars), while for the structure in (3b), no such boundary is produced. Figure 1a shows a pitch track for the utterance with the first meaning, i.e. (3a), and Figure 1b shows a pitch track for the utterance with the second meaning, i.e. (3b). The horizontal axes in the figures show the time in milliseconds and the vertical axes show the pitch in hertz (Hz). A higher value in Hz indicates higher pitch. Above the pitch tracks, waveforms are presented which indicate the amplitude (vertical axes) over time (horizontal axes). Each word of the sentence is written over the waveforms, and a vertical line marks the end of each word or phrase. (The same format is used in Figure 2.) The word 'villain' in Figure 1a is longer than the same word in Figure 1b. This is because the word in Figure 1a is the final word in the prosodic phrase, but it is not in Figure 1b. The word in Figure 1a also shows a falling and slightly rising pitch, indicating a phrase boundary after the word.

Psycholinguistic studies have also found that native speakers perceive these disambiguating intonational and timing cues in ambiguous sentences and that these acoustic cues help native speakers process the sentences to arrive at the intended meanings. This shows that syntax imposes some constraints on the production of prosodic structure and that the prosodic structure is parsed or accessed by native speakers together with the syntactic structure when processing a sentence.

In sum, intonation can help cue the syntactic structures of a sentence, although not all syntactic structures are cued by intonation. Additionally, intonation plays a further role in conveying the



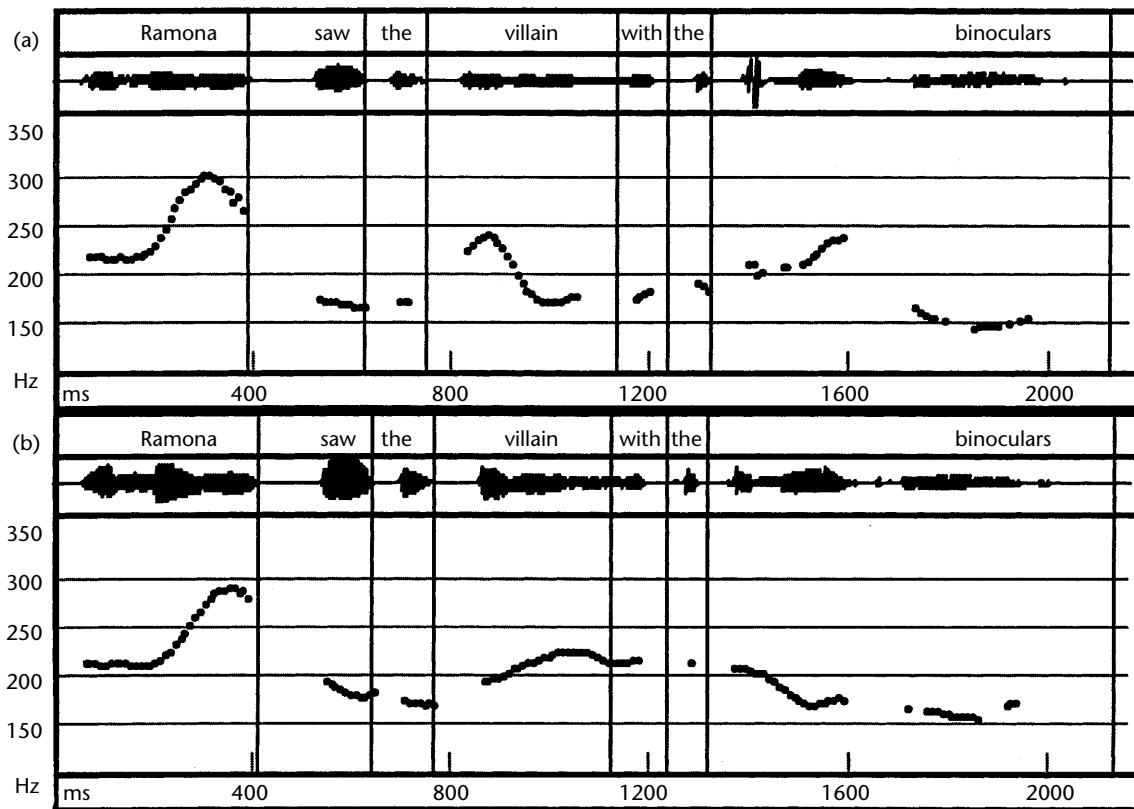


Figure 1. (a) A pitch track of a sentence with (3a) structure, and (b) the same sentence with (3b) structure.

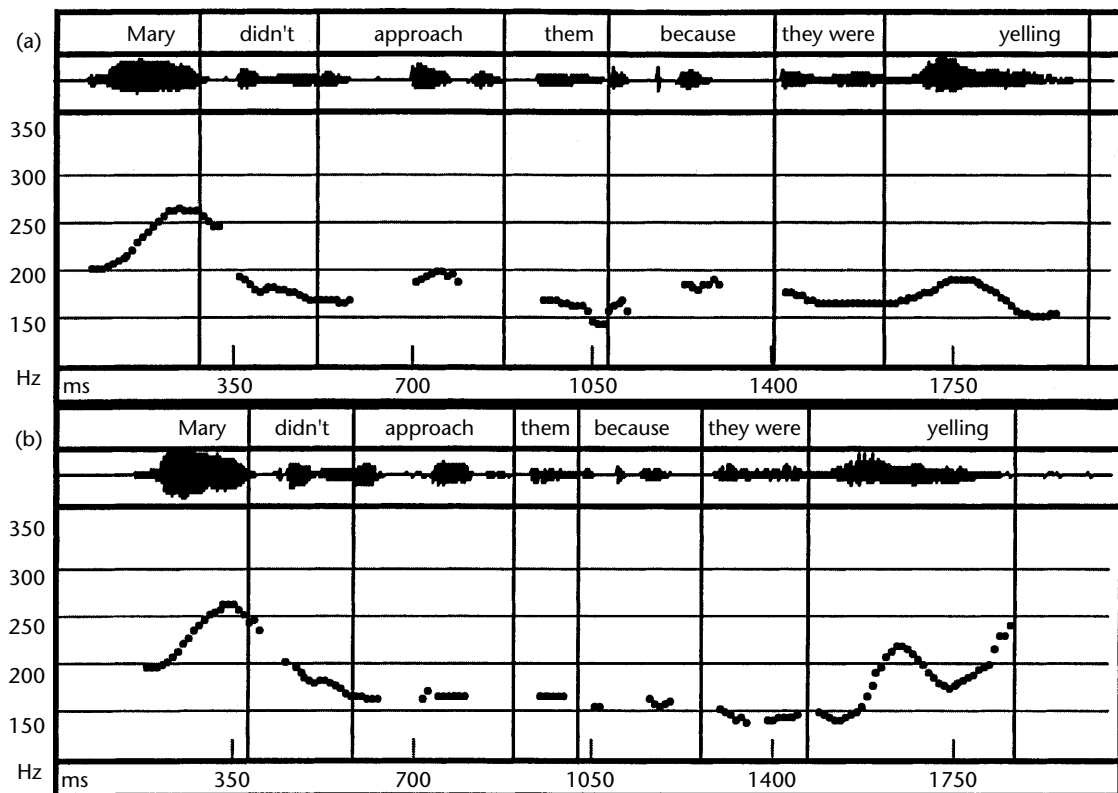
semantic and pragmatic meaning of a sentence, and this will be discussed in the next section.

## SEMANTIC AND PRAGMATIC FUNCTIONS

It is well known that native speakers emphasize or focus an important word within a sentence by means of pitch, intensity, and duration. For example, in English a focused word is realized with a sharply rising pitch and has longer duration and higher intensity than a nonfocused word. Listeners interpret the focused word by comparing it with other items belonging to the same category as the focused word. For instance, the sentence 'John only introduced Bill to Mary' can mean different things depending upon which word is focused. If 'Bill' is focused ('John only introduced BILL to Mary'), the sentence is true only when John introduced exactly one person to Mary and that person was Bill. On the other hand, if 'Mary' is focused ('John only introduced Bill to MARY'), the sentence is true only when John introduced Bill to exactly one person and that person was Mary. Thus, it can be seen that focus intonation changes the truth conditions of the sentence, suggesting that the

semantic interpretation of a sentence is influenced by intonation. (See **Semantics and Cognition; Pragmatics, Formal**)

Intonation can also influence the relationship between words within a sentence. For example, the sentence 'Mary didn't approach them because they were yelling' can have two meanings depending upon how the sentence is produced. It can mean either that Mary did not approach them and the reason was because they were yelling, or it can mean that Mary approached them not because they were yelling but for some other reason. For the former meaning, 'not' modifies the verb 'approach', while for the latter meaning, 'not' modifies the word 'because'. Example pitch tracks for this sentence conveying these two different meanings are presented in Figure 2. (These are not the only possible intonation patterns that would convey these two meanings.) The utterance in Figure 2a conveys the former meaning and that in Figure 2b conveys the latter meaning. There is a phrase boundary after 'them' in Figure 2a, but not in Figure 2b. (Note the lengthening on 'them' in Figure 2a.) Further, the sentence ends with a falling boundary tone in Figure 2a, but with a rising boundary tone in Figure 2b. These



**Figure 2.** A pitch track of the sentence meaning (a) 'Mary did not approach them and the reason was because they were yelling' and (b) 'Mary approached them not because they were yelling, but for some other reason'.

examples illustrate that intonation can change the modification relations between words.

Furthermore, intonation reflects the structure of a discourse and the relative informativeness of words or constituents in a sentence (e.g. new versus old information). Words or constituents of a new topic or new information become prominent by pitch range manipulation or by getting pitch accent. In order to cue a discourse structure, a pitch range is manipulated in such a way that sentences with a new topic are produced with an expanded pitch range while those with an old topic (or a nonmajor topic as in parenthetical phrase) is produced in a reduced pitch range. That is, the pitch peak is higher at the beginning of a paragraph introducing a new topic than in the middle of the paragraph, and the pitch valley is lower at the end of the paragraph finishing a major topic than in the middle of the same paragraph.

Similarly, in English, a word with new information is produced with pitch accent, and a word that is repeated or 'given' in the discourse does not get pitch accent. When the sentence 'I don't read

German' is produced with broad focus, the sentence final noun, i.e. 'German', would receive pitch accent. However, when this word is repeated and becomes old information, it does not get any accent, as in (4B). Here, the verb 'read' is accented and the object noun 'German' is deaccented. Deaccenting is also found in word(s) after contrastive focus. For example, when the word 'John' is contrastively focused (for example, against 'I') as in (5), it will get pitch accent and the rest of the sentence will get deaccented. Thus, accentedness reflects the relative semantic weight and the informativeness within a constituent.

A: I found an article for you in a German journal.

B: I don't READ German. (4)

JOHN reads German. (5)

This relationship of semantic weight versus accent, however, is not universal. In Italian and Romanian, for example, a word with given information still receives accent, and in French and Korean, which do not have lexical stress, new

versus given information is delivered by phrasing, not by accenting. A word with new information begins a new prosodic phrase, while a word with old information is dephrased, i.e. does not begin a new phrase but belongs to the preceding phrase. Thus, beginning a new phrase in French and Korean is equal to receiving pitch accent in English.

To conclude, we have seen how intonation is related to prosodic structure, syntactic structure, and the semantic and pragmatic meaning of a sentence. Intonation marks a structure as well as prominence relations among words. It also conveys semantic and pragmatic meaning. It is related to all subareas of grammar and is psychologically real. Thus, intonation is a crucial part of a native speaker's knowledge of his or her grammar, and it is also essential for effective communication between the members of a speech community.

### Further Reading

- Beckman M (1996) The parsing of prosody. *Language and Cognitive Processes* **11**(1/2): 17–67.
- Beckman ME and Pierrehumbert J (1986) Intonational structure in Japanese and English. *Phonology Yearbook* **3**: 255–309.
- Cutler A, Dahan D and Donselaar W. van (1997) Prosody in the comprehension of spoken language: a literature review. *Language and Speech* **40**: 141–201.
- Ferreira F (1993) The creation of prosody during sentence production. *Psychological Review* **100**: 233–253.
- Gee TP and Grosjean F (1983) Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology* **15**: 411–458.
- Jun S-A (1996) *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*. New York: Garland Publishing.
- Jun S-A and Fougeron C (2000) A phonological model of French intonation. In: Botinis A (ed.) *Intonation: Analysis, Modeling and Technology*, pp. 209–242. Dordrecht: Kluwer Academic Publishers.
- Kjelgaard MM and Speer SR (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* **40**: 153–194.
- Ladd DR (1996) *Intonational Phonology*. Cambridge, UK: Cambridge University Press.
- Nespor M and Vogel I (1986) *Prosodic Phonology*. Dordrecht: Foris.
- Price P, Ostendorf M, Shattuck-Hufnagel S and Fong C (1991) The use of prosody in syntactic disambiguation. *Journal of Acoustical Society of America* **90**(6): 2956–2970.
- Rooth M (1996) Focus. In: Lappin S (ed.) *Handbook of Contemporary Semantic Theory*. London: Blackwell.
- Selkirk EO (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Shattuck-Hufnagel S and Turk A (1996) A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* **25**(2): 193–247.
- Ward G and Hirschberg J (1985) Implicating uncertainty: the pragmatics of fall-rise intonation. *Language* **61**(4): 747–776.