CHAPTER 28

....

PRAGMATICS AND PROSODY

....

JULIA HIRSCHBERG

28.1 INTRODUCTION

VARIATION in prosody, or intonation, can influence the interpretation of linguistic phenomena in many languages (Ladd 1980, 1996; Bolinger 1986, 1989). Type and location of prosodic prominence and prosodic phrase boundaries, different overall intonational contours, and changes in intensity, duration, and speaking rate can, in appropriate contexts, help to indicate syntactic attachment, disambiguate scope of modifiers and negation, signal information status, type of speech act, and propositional attitude, and contribute to the licensing of implicatures and to reference resolution, inter alia.

In this chapter, we discuss aspects of prosodic variation and pragmatic meaning which have been explored by researchers in linguistics, computational linguistics, and psycholinguistics. We begin by describing some commonly used frameworks for representing prosody, including the ToBI framework, which we will use to identify prosodic variation throughout. We then examine the pragmatic influence of prosody on the interpretation of syntactic, semantic, and discourse phenomena. We conclude by suggesting some new avenues for research in the relationship between prosody and pragmatic interpretation.

28.2 CONVENTIONS FOR DESCRIBING PROSODY

Many conventions have been developed for describing prosodic variation so that this variation can be documented and compared among different researchers. These conventions are also employed to annotate corpora for statistical analysis, automatic classification experiments, and speech technologies such as text-to-speech synthesis. They vary in terms of the decisions they make about what prosodic features to annotate and which labels to employ. They can be continuous or categorical in nature, either

by describing prosodic variation such as Fo (fundamental frequency) as a sequence of continuous variations or by identifying prosodic events as tokens from a given inventory of prosodic phenomena—e.g. one may describe a prosodic prominence in terms of the relative height of its Fo or as a particular type of pitch accent.

Continuous descriptions typically focus on describing the Fo contour and often have been developed for Fo modelling in text-to-speech synthesis. The Fujisaki superpositional model (1982), developed for Japanese synthesis, is based primarily on acoustico-physiological observation and describes prosody as a superposition of accent contours on a phrasal contour via phrase and accent commands that control Fo; parameterizations of phrase and accent signals are continuous; it has rarely been used for annotation. Hirst's INTSINT (Hirst and Di Cristo 1998) describes intonation in terms of sequences of labels representing target points for Fo, which are defined in terms of the speaker's current and previous pitch range. Taylor's (1998) TILT model, also developed for synthesis, describes the pitch contour of prominent and phrase-ending syllables using a compact parameterization of the slope, height, and peak position of the rise and fall of the pitch contour within a syllable.

Categorical systems include the British (Halliday 1967a; O'Connor and Arnold 1961) and American (Pierrehumbert 1980) schools, both designed for their respective variants of English. The former defines prosody in terms of pitch movements (e.g. rises and falls) while the latter uses pitch targets or tones (e.g. high vs low tones) to describe prosodic phenomena. The currently most popular instance of this school is the ToBI (Tones and Break Indices) system, developed for Standard American English (SAE) (Silverman et al. 1992; Pitrelli et al. 1994).[1] This convention includes a set of labelling tiers in which orthography time-align with the wave form, *break indices* (five degrees of juncture), tones (*pitch accents, phrase accents*, and *boundary tones*, largely following Pierrehumbert 1980), and miscellaneous phenomena such as disfluencies are labelled, together with wave files and a representation of the Fo contour. Differences among ToBI break indices denote differences in word or phrase boundaries, with higher numbers marking greater degree of juncture. An SAE standard declarative contour, which typically falls, is modelled with one or more simple high pitch accents, a low phrase accent, and a low boundary tone (Figure 28.1); a yes-no question contour is modelled by sequences of low pitch accents and a high phrase accent and high boundary tone (Figure 28.2).[2]

Other SAE variants are shown in Figure 28.3.

In SAE ToBI, words may be prominent (*accented*) or not (*deaccented*), and, if accented, may bear one of five different accents, including the two simple accents (H* and L*) and three complex accents (L+H*, L*+H, and H+!H*), where the '*' indicates

---

[1] A fuller description of the ToBI systems may be found in the ToBI conventions document and the training materials available at <http://www.ling.ohio-state.edu/~tobi/ame_tobi/>. Newer materials are available from <http://anita.simmons.edu/~tobi/index.html>.

[2] The examples in Figures 28.1–28.3 are taken from the ToBI training materials, prepared by Mary Beckman and Gail Ayers.
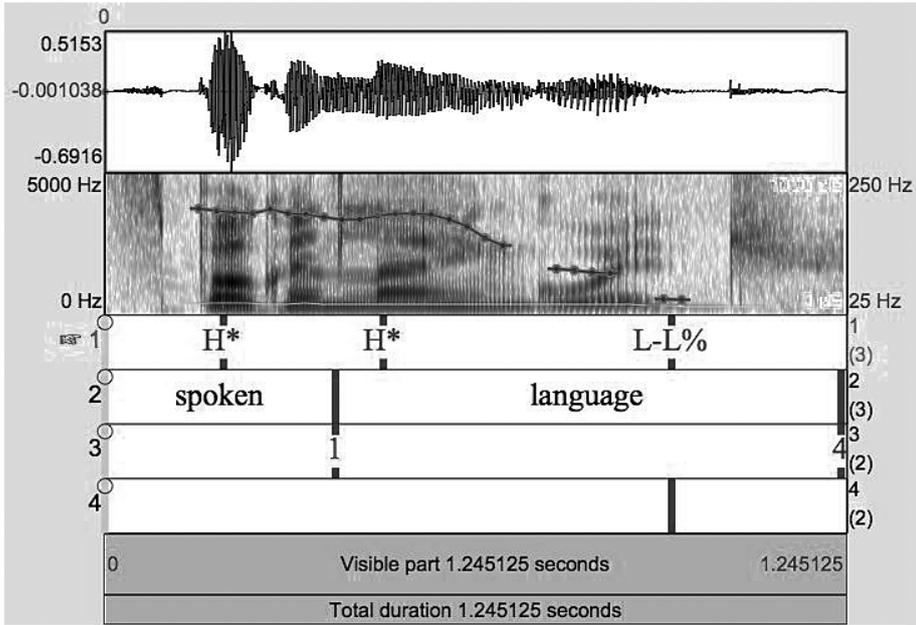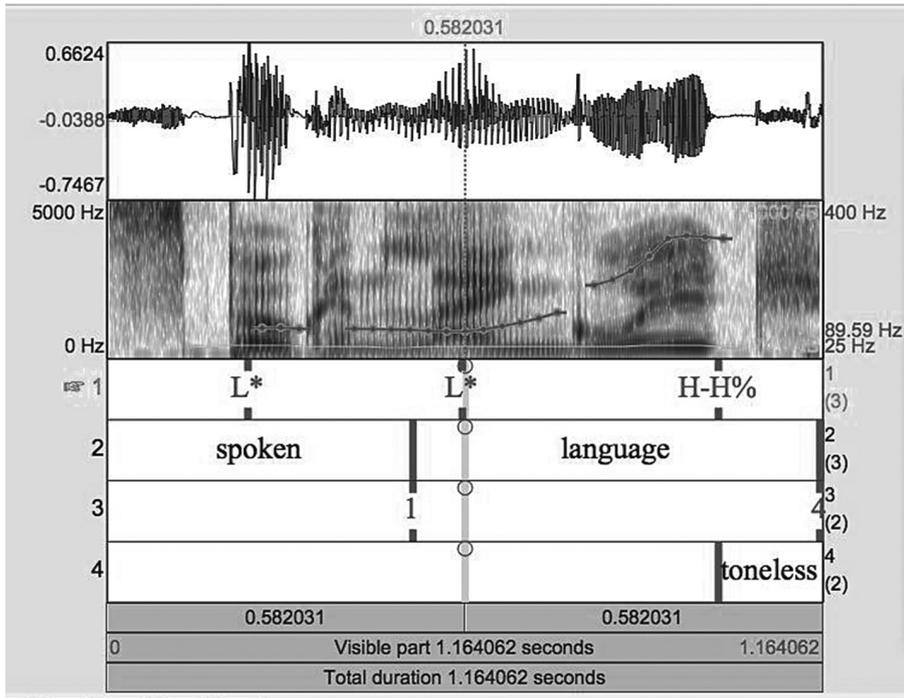
FIGURE 28.1 A declarative contour H* L-L%

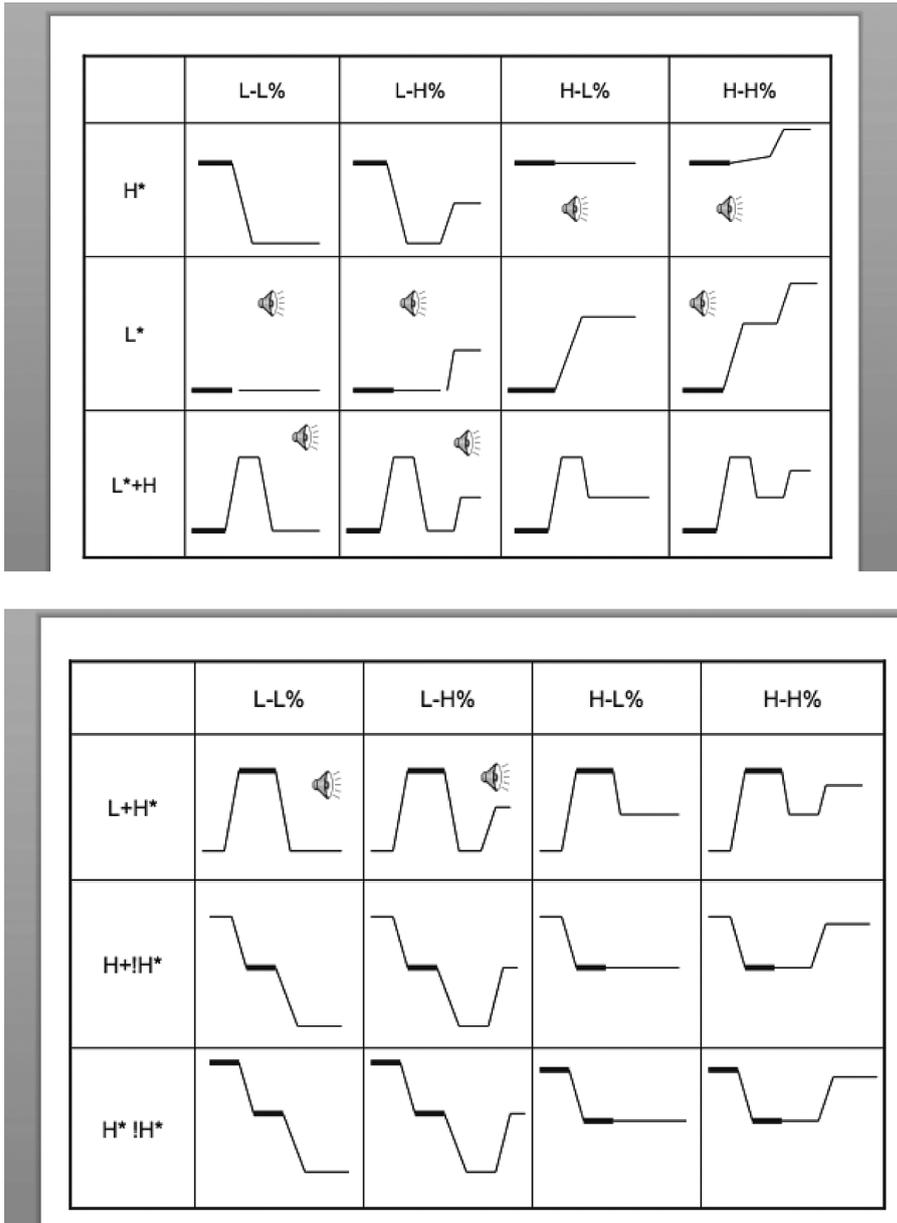**FIGURE 28.2** A yes–no question contour L* H-H%

FIGURE 28.3 Schematic ToBI contours

alignment of the accent with the primary stress of the word. *HIF0* is annotated on the highest accent peak in a minor (*intermediate*) phrase and serves the function of indicating variation in pitch range from phrase to phrase.

Pierrehumbert and Hirschberg (1990) proposed a set of possible meanings for different pitch accent types in SAE which have formed the basis for a number of perception

experiments. In their proposal, H* accents, which are typically found in standard declarative utterances, convey that the accented item should be treated as *new* information in the discourse—as part of what is being asserted by the speaker. L* accents, on the other hand, convey that an accented item should be treated as *salient* in the discourse but not part of what is being asserted. These are typically seen in prominent items in *yes–no* question contours and are often employed to make prepositions or adverbs or discourse markers prominent in utterance-initial position. L+H* accents can be used to produce a sense of *contrast*, often termed *contrastive stress*. In this complex accent, the H tone is aligned with the stressed syllable of the prominent word and the Fo rise from the initial L tone is rapid. In the L*+H accent, the L tone is associated with the word's stressed syllable and the rise to the H is gentler, resulting in more of a 'scooped' accent. This difference in alignment is critical to interpretation, conveying an impression either of uncertainty of or incredulity depending upon other prosodic and contextual features. H+!H* accents are realized as a fall onto the stressed syllable and have been proposed by Pierrehumbert and Hirschberg (1990) as conveying some sense that speaker and hearer are already familiar with the mentioned item, as in a 'reminding' context. Contours which include these pitch accents are shown schematically in Figure 28.3.

While ToBI was originally developed to describe the prosody of American English, the standard has been extended and adapted for many other languages, including Portuguese, Chinese, Korean, Japanese, German, Greek, Italian, Spanish, Bengali, Chickasaw, Korean, French, Bininj Gun-wok, and Standard Serbo-Croatian (cf. Jun 2005 for descriptions of a number of these). Similar but distinct categorical systems have been developed for Dutch (Gussenhoven 2005's ToDI system), British English (the IViE system, Grabe 2004), and Bruce's model of Swedish prosody (Bruce et al. 1994). Köhler's KIM annotation scheme (Köhler 1991), developed for German, describes intonation in terms of peaks and valleys but also distinguishes a variety of pitch prominence categories. The IPO standard, developed for Dutch, describes prosody as an inventory of pitch movements assigned to syllables and classified as rises or falls by the range of movement and whether these are accent-lending or not ('t Hart et al. 1990). Dilley and Brown (2005) have developed RaP, a labelling system for the prosody of American English that includes the annotation of rhythm as well as prominence and phrasing.

28.3 PROSODY, SYNTAX, AND AMBIGUITY

Research on the prosody–syntax interface has focused on two goals. One is to identify systematic relationships between prosodic variation—particularly phrase boundaries but also prominence patterns—and syntactic structure, while the other focuses on ways in which prosody can influence the perception of ambiguous syntactic phenomena.

Most researchers with the first goal in mind look to see how prosodic phrases, whether intermediate or intonational, divide an utterance into meaningful 'chunks' of

information (Bolinger 1989): the greater the perceived phrasing juncture, the greater the discontinuity between segments or constituents. While some have sought to identify simple syntactic constraints on phrase location (Crystal 1969; Downing 1970; Cooper and Paccia-Cooper 1980; Selkirk 1984; Croft 1995), others have focused on more practical matters, such as the possibility that prosodic phrase boundaries can aid syntactic parsing (Marcus and Hindle 1990; Steedman 1991; Oehrle 1991; Abney 1995) or that syntactic information can identify appropriate locations for phrase boundaries in text-to-speech synthesis (Dirksen and Quené 1993; Prevost and Steedman 1994; Hirschberg and Prieto 1996; Boula de Mareüil et al. 1998). Still others have investigated whether there are regularities in the placement of *nuclear* (main sentence) *stress* based on syntactic structure (Bresnan 1971) or whether instead, in Bolinger's (1972) famous phrase, 'Accent is predictable (if you're a mindreader)'. There is also work on the role of prominence in the different pragmatic interpretations of English *cleft* sentences (Herment and Leonarduzzi 2012), who find several prosodic patterns which they see as serving different functions in a corpus study. They found that prosody can serve to distinguish what is informative vs what is not via a falling tone over informative elements in *it-* and *wh-*clefts, where the syntactically focal element may be given and the presuppostional element may be new (e.g. *So you mean that it's the nerves that FEED them*—with a falling tone on *feed*). Prosody may also mark the syntactically focused element as intonationally focused or not, thus negating the effect of syntactic position (as in the example above in which *nerves* is given in the context and deaccented) or reinforcing it. Alternatively, in another context where *feed* is given, a speaker might mark the focal element more strongly by saying *So you mean that it's the NERVES that feed them*—and not something else.

Currently there is considerable interest in and controversy around the Implicit Prosody Hypothesis (Fodor 2002), which proposes a default prosodic phrasing in phenomena such as attachment preferences which is language-specific. The claim researchers are investigating is that differences in individual languages' attachment preferences for syntactically ambiguous sentences are due to the languages' different prosodic systems. So far, empirical evidence for this hypothesis appears mixed (Bergmann et al. 2008; Gryllia and Kügler 2010; Jun 2010).

Whatever the source of the relationship between prosody and syntactic disambiguation, there have been many studies substantiating such a relationship between prosody and syntax in many languages. Corpus-based studies (Altenberg 1987; Bachenko and Fitzpatrick 1990; Ostendorf and Veilleux 1994; Fujio et al. 1997) and laboratory production and perception experiments (Grosjean et al. 1979; Wales and Toner 1979; Gee and Grosjean 1983; Price et al. 1991; Beach 1991; Hirschberg and Avesani 2000; Warren et al. 2000; Edmonds et al. 2008; Wagner and Crivellaro 2010; Kaland and Van Heuven 2010; Gollrad et al. 2010) have variously found that the discontinuity indicated by a phrase boundary may serve to favour various differences in the interpretation of syntactic attachment ambiguity for phenomena such as prepositional phrases, relative clauses, parentheticals, appositives, and adverbial modifiers. The presence or absence of a phrase boundary can also distinguish prepositions from particles, can indicate the

scope of modifiers in conjoined phrases, and can disambiguate case. Some minimal pairs illustrate the role that prosodic boundary location can play (where boundaries are indicated by |):

VP-attachment: *Anna frightened the woman | with the gun* (Anna held the gun)

NP-attachment: *Anna frightened | the woman with the gun* (the woman held the gun)

Complementizer: *Mary knows many languages you know* (you both know many)

Parenthetical: *Mary knows many languages | you know* (as you are aware …)

The animal that usually fights the lion is missing (the lion's normal opponent is missing)

The animal that usually fights | the lion | is missing (Appositive: the lion is missing)

Restrictive relative: *My brother who is a writer needs a new job* (I have multiple brothers but the writer brother is the one who needs a job

Non-restrictive: *My brother | who is a writer | needs a new job* (I may or may not have other brothers)

Preposition: *John laughed | at the party* (John laughed while at the party)

Particle: *John laughed at | the party* (John ridiculed the party)

Modifier scope: *This collar is dangerous to younger | dogs and cats* (the collar is dangerous to both younger dogs and younger cats)

This collar is dangerous to younger dogs | and cats (it's dangerous to all cats)

Simple complement: *We only suspected | they all knew that a burglary had been committed* (we only suspected that … they all knew that a burglary had been committed)

Parenthetical: *We only suspected | they all knew | that a burglary had been committed* (they all knew that we only suspected that a burglary had been committed)

Prosodic information is also useful in disambiguating *garden-path sentences*, such as *The horse | raced past the barn | fell*, where insertion of prosodic boundaries can block a garden-path interpretation. Ning and Shih (2012) showed that garden-path interpretations could be enhanced or suppressed by manipulating phrasing features using insertion of pause and pitch re-set in Chinese perception experiments. Pauker et al. (2011) found evidence from ERP data that conflicting prosodic boundaries were less likely to be judged acceptable by subjects and produced a stronger effect in N400/P600 components than cooperating prosodic boundaries. Even mathematical syntax has been shown to be disambiguable via phrasing variation (O'Malley et al. 1973), as in examples such as 'ten minus eight | divided by two is one vs ten | minus eight divided by two | is six', where phrase boundaries can serve to realize the scope of parentheses.

Prosodic variation other than phrasing can also influence disambiguation of syntactic ambiguity. For example, range and rate can also distinguish phenomena such as parenthetical phrases from others (Kutik et al. 1983; Grosz and Hirschberg 1992): parentheticals are generally uttered in a compressed pitch range and with a faster speaking rate than other phrases. The location of pitch accents can cue the right node raising reading (Marcus and Hindle 1990) of utterances like *WE only SUSPECTED | THEY all*

KNEW | that a BURGLARY had been committed (we suspected but they in fact knew that a burglary had been committed) vs other readings of the sentence (e.g. *We only SUSPECTED they all knew that a BURGLARY had been committed* (we suspected that they all knew that a burglary had been committed).

Pitch accent location may also convey the structure of complex nominals (Liberman and Sproat 1992; Sproat 1994) and distinguish among part-of-speech ambiguities, as in the noun/adjective ambiguity of *GERMAN teachers* (teachers of German) vs *German TEACHERS* (teachers who are German) and the preposition/particle interpretations of *LEAVE in the LIMO* (the limo is the vehicle you should leave in) vs *LEAVE IN the LIMO* (the limo should not be removed from perhaps a wedding budget). But note that prepositions may also be accented to convey focus of attention or contrast, as in *I didn't shoot AT him, I shot PAST him*. So, the relationship between accent and part-of-speech is dependent upon context.

While prosodic variation **can** disambiguate syntactically ambiguous utterances, evidence that it does so reliably is mixed (Wales and Toner 1979; Cooper and Paccia-Cooper 1980; Nespor and Vogel 1983; Schafer et al. 2000; Peters 2006). Speakers often manage to convey the distinctions illustrated above without employing particular prosodic means. Often speakers may not recognize the potential ambiguity of such utterances or context may disambiguate. Even when explicitly asked to disambiguate, they may choose other, sometimes lexical, methods of disambiguation.

28.4 Prosody and Semantic Interpretation

Prosodic prominence and phrasing can also influence the semantic interpretation of utterances. Prosody can signal focus, define the scope of negation, quantification, and modals, and influence the interpretation of presupposition (Lakoff 1971; Schmerling 1971; Jackendoff 1972; Ball and Prince 1977; Wilson and Sperber 1979; Enkvist 1979; Gussenhoven 1983; Rooth 1985, 1991, 1992; Baart 1987; Rochemont and Culicover 1990; Dirksen 1992; Zacharski 1992; Birch and Clifton 1995). Changing the location of nuclear stress in an utterance can change the semantic interpretation of utterances containing *focus-sensitive operators* (Halliday 1967a; Jackendoff 1972; Rooth 1985, 1992; Sgall et al. 1986; Partee 1991; Vallduví 1992; Selkirk 1995; Schwarzschild 1999; Büring 1999; Wagner et al. 2010). Dimitrova et al. (2010) in fact have found from ERP studies that focus-sensitive particles appear to create an expectation of a focal accent in studies of Dutch.

For example, *John only introduced MARY to Sue* conveys that Mary is the only person John introduced to Sue, but *John only introduced Mary to SUE* suggests that Sue is the only person John introduced Mary to, and *JOHN only introduced Mary to Sue* indicates that John is the only one who introduced Mary to Sue. Other such operators include other quantifiers (*all, most, some, frequently, rarely, sometimes*), adverbs of quantification (*sometimes, most often*), modals (*must*), emotive factives/attitude verbs (*It's odd*

that), and counterfactuals. With prominence on *night*, for example, the following question/answer sequence is felicitous:

> *When do ships go through the lock?*
> *Most ships pass through the lock at NIGHT.*

However, with prominence on *lock*, the second sentence becomes a felicitous answer to the question, *Where do most ships pass through at night?* Temporal quantification behaves similarly, as illustrated in:

> *Who goes to Brighton? LONDONERS most often go to Brighton.*
> *Where do Londoners go on vacation? Londoners most often go to BRIGHTON.*

Other temporal quantifiers include *occasionally*, and so on.

The well-known example *DOGS must be carried* was observed on a sign on a British train by Halliday (1967a), who was confused to learn that every train rider should carry a dog. He noted that the intended meaning is more likely realized as *Dogs must be CARRIED*. Other modals which associate with focus in a similar way include *can, should*, and *may* and phrases such as *it's odd that*.

Although most research on the role of prosody in semantic interpretation has concentrated on prosodic prominence variation, location of phrase boundaries can also change the semantic interpretation of an utterance (cf. Schafer 1998). For example, the interpretation of negation in a sentence like *Bill doesn't drink because he's unhappy* is likely to vary, depending upon whether it is uttered as a single intonational phrase or two phrases. As a single phrase the negation has wide scope: Bill does indeed drink, but the cause of his drinking is not his unhappiness. When there is a phrase boundary after *drink*, the negation has narrow scope: Bill's unhappiness has led him *not* to drink. However, like other interpretations that may be favoured by prosodic variation, if context itself can disambiguate a potentially ambiguous sentence, speakers may produce more neutral phrasings. For example, an utterance of *Bill doesn't drink because he's unhappy* when produced as a single phrase may be interpreted with narrow scope of negation as well as wide if context disambiguates (Hirschberg and Avesani 2000). Such cases where a particular prosodic realization may be interpreted in several ways—but its contrast is less likely to—give rise to the notion of 'neutral' intonation, a notion whose evidence is probably more persuasive for phrasing variations than for accent variation. In Brazilian Portuguese, for example, negation may be indicated pre-verbally, post-verbally, or pre- and post-verbally, where the negation structure may be signalled prosodically (Armstrong et al. 2008).

28.5 PROSODY AND DISCOURSE PHENOMENA

The pragmatic effect of prosodic variation is perhaps most studied in the interpretation of discourse phenomena. The interpretation of pronominal referents can

vary depending upon whether the pronoun is accented or not in different contexts. Information status such as *focus of attention, topic/comment, theme/rheme, given/new* is often correlated with variation in accent or phrasing. The relation between the relative accessibility or salience of information in a discourse and prosody has been discussed in theories of *communicative dynamism, focus of attention*, and *centring* in discourse and in models of sentence production.

Variation in overall discourse structure has also been found to be conveyed prosodically, by variation in pitch range, pausal duration between phrases, speaking rate, and other prosodic phenomena. *Discourse markers* or *cue phrases*, which can also signal discourse structure, are also prosodically disambiguated from their adverbial interpretations. Variation in tune or contour has been widely associated in the literature with different speech acts. Other correlations between prosodic features and phenomena such as turn-taking behaviour have also been noted.

28.5.1 Reference resolution

Pronouns may be interpreted differently depending upon whether they are prominent or not, in varying contexts (Asher and Wada 1988; Gundel et al. 1993; Venditti et al. 2002). Empirical studies have verified Lakoff's (1971) proposal that pronouns are more likely to be deaccented than accented; they are accented to convey so-called *marked* or less likely interpretations. In *John called Bill a Republican and then he insulted him*, the pronouns will be resolved differently if they are intonationally prominent. With both deaccented, the likely referent of *he* is *John* and *him* is *Bill*, but if both are accented, the preferred referents of each are switched. The interpretation of *bound anaphora* are also affected by the accenting of the anaphor, although in such cases the most likely reading of an utterance such as *John likes his colleagues and so does Sue*—strict: Sue likes John's colleagues vs. sloppy: Sue likes her own colleagues—when the pronoun *his* is prosodically prominent will be the opposite of the most likely interpretation of the sentence in the absence of prosodic information (i.e. read) (Hirschberg and Ward 1991). More recently, Braun and Chen (2008) have found evidence that listeners can make use of prosodic information on material occurring **before** the referential expression to identify an upcoming referent: for example, adverbial *nu* (*now*) can focus listener attention on upcoming contrastive information in Dutch eye-tracking experiments.

28.5.2 Focus of attention and contrast

Many authors have examined the role of prosodic prominence in signalling the relative importance of an item in relation to other entities in the discourse (Grosz 1977; Ladd 1979; Bing 1979; Bardovi-Harlig 1983b; Sidner 1983; Eady and Cooper 1986; Grosz and Sidner 1986; Hajicova et al. 1990; Gordon et al. 1993). Dimitrova et al. (2010) have found not only that prosodic prominence can signal focus but that inappropriate deaccenting

of items in focal contexts or accenting of items in non-focal contexts show differences in brain activity in ERP experiments with Dutch speakers.

There is also considerable research on the realization of broad and narrow focus by prosodic means. For example, *I talked to the girl in the RED dress* represents a case of narrow focus, felicitous when there is only one girl in the context whose dress is red, while I talked to the girl in the red DRESS conveys no such contrast (broad focus). In English, there has been considerable discussion of whether F0 height or accent type signals focus, with evidence that the L+H* accent signals focus in some cases (Ito and Speer 2008a), while F0 peak as well as duration and intensity appear to perform a similar function in others. Gili Fivela (2008a) also found that two pitch accents (H* and H+L*) can convey narrow and contrastive focus in Pisan Italian, while Wang et al. (2012) found results similar to English in perception experiments in the Anduo dialect of Tibetan. Broad vs narrow focus realization in Dutch has been studies by Hanssen et al. (2008), who found differences in accent shape as well. Dimitrova et al. (2008) have examined the interactions between semantically unexpected information and other discourse factors including syntactic status in the production of contrast in Dutch as well. Baumann et al. (2006) have studied the phenomenon in German, discovering that speakers used both categorical as well as gradient prosodic means to indicate broad/narrow/contrastive focus differences. Kügler (2008) found evidence of durational cues to focus also in German. Greif (2010) studied the prosodic realization of different types of contrastive (corrective) focus relative to *wh*-focus in Chinese, learning that semantic alternative-based contrast vs pragmatic contrast based upon assumptions about speaker–hearer expectations in the common ground were distinguished prosodically. Hellmuth (2006) finds that focus is signalled more by pitch range (i.e. the relative F0 height of a pitch accent) as well as alignment in Egyptian Arabic, while Genzel and Kügler (2010) observe that pitch height and duration appear to convey focus in Hindi. Chen and Destruel (2010) found that for Toulousean French, like Standard French, post-focus deaccentuation and phrasing mark focus; however, unlike Standard French, the Toulouse variety also uses variation in phrasing type to indicate whether the subject or object of the verb is focused. Finally, Zerbian et al. (2010) examine several African languages to identify prosodic marking of focus but uncover little evidence in Afro-Asiatic or Niger-Congo, concluding that the relationship between prosody and focus varies considerably across languages. In related research, Ambrazaitis (2006) found that unexpected information was signalled by a more prominent pitch peak in South Swedish.

28.5.3   Topic/comment and theme/rheme

There is much evidence that prosodic prominence plays an important role in conveying information about discourse topic (Schmerling 1975; Chafe 1980; Prevost 1995). Terken (1985) found that in task-oriented monologues, speakers referred to local topics with deaccented pronominal expressions, and accented, full NPs otherwise, even though

many of these full NPs evoked entities which had already been mentioned in the previous discourse (i.e. represented given information). Pitch accent on pronouns has also been found to be correlated with changes in attentional state in studies by Cahn (1995) and Nakatani (1997). In centring theory (Grosz et al. 1983, 1995; Kameyama 1986; Brennan et al. 1987), what the current conversation is 'about' (i.e. its *backward-looking centre*) can be altered by the way pronominalized referents are produced prosodically (Grosz et al. 1995). In addition, theme/rheme and focus/background distinctions have been found to be distinguished by duration and F0 range in Chinese by Chen and Braun (2006).

28.5.4 Given/new information

There has been considerable laboratory and corpus-based research on the role of prosody in distinguishing given (old) from new information, where givenness may be defined as simple repeated mention or as something more complex (Lehman 1977; Gundela 1978; Allerton and Cruttenden 1979; Fuchs 1980; Chafe 1980; Nooteboom and Terken 1982; Bardovi-Harlig 1983a; Brown 1983; Terken 1985; Kruyt 1985; Fowler and Housum 1987; Horne 1991; Terken and Hirschberg 1994). Halliday (1967b) noted that an expression may be prosodically marked as given by deaccenting if the information conveyed by the expression is situationally or anaphorically *recoverable* on the basis of the prior discourse or by being salient in the situation. Chafe (1974, 1976) proposed similarly that an expression may be deaccented if the information is in the listener's consciousness. However, it has become clear from numerous studies that there is not a simple one-to-one mapping between givenness and prosodic realization, however givenness is defined. Factors that appear to influence the prosodic marking of given information include: (1) whether or not a given item is part of a complex nominal; (2) the location of such an item in its prosodic phrase; (3) whether preceding items in the phrase are 'accentable' due to their own information status; and (4) the grammatical function of an item when first and subsequently mentioned. Other higher-level discourse factors may also influence whether given items are accented or deaccented, including distance from previous evocation of the item, the frequency with which the item has been evoked, whether the evocation is a return to a previous topic or not, and whether the item is used contrastively. There also appear to be differences in production of given vs new information in terms of genre (read vs conversational speech) and context.

Röhr and Baumann (2010) found a correlation in proportion of accented items as well as higher and later accentual peaks from textually given items through textually 'accessible' items to new items in read speech. In studies of spontaneous laboratory speech, Brown (1983) found that expressions used to evoke items which had been explicitly mentioned in the previous discourse were deaccented, but that expressions used to refer to *inferable* items were usually accented by British English subjects. And Terken and Hirschberg (1994) in their spontaneous speech experiments with American

English subjects found that differences between the grammatical function of a previously mentioned item and their subsequent mention significantly influenced how they were produced; for example, items introduced in subject, object, or PP-object position and then mentioned in a different position were much more likely to be accented than items subsequently mentioned in the same grammatical role as their first mention. Empirical results from the Edinburgh Map Task dialogues (Bard and Aylett 1999) found little evidence of *within* speaker deaccentuation of given information (i.e. when a speaker him/herself repeats a word or phrase) although they did find some evidence of reduction and rare instances of deaccentuation of given information across speakers. Other studies of repeated information in Japanese subjects when a speaker was confirming or questioning previously mentioned information find differences in prosodic features as well (Shimojima et al. 2002).

28.5.5 Discourse/topic structure

Variation in overall discourse structure has also been found to be conveyed prosodically in many languages, by variation in features such as pitch range, pausal duration between phrases, and speaking rate (Woodbury 1987; Grosz and Hirschberg 1992; Savino et al. 2006; Gili Fivela 2008b; Fezza and Auran 2012; Silverman 1987; Avesani and Vayra 1988; Ayers 1992; Swerts et al. 1994; Swerts 1997; Brown et al. 1980; Lehiste 1979; Passonneau and Litman 1993; Hirschberg and Nakatani 1996; Koiso et al. 1998b; van Donzel 1999; Venditti 2000). For SAE read speech, it has been found that phrases beginning new topics are begun in a wider pitch range, are preceded by a longer pause, and are louder and slower than other phrases; narrower range, longer subsequent pause, and faster rate characterize topic-final phrases (Grosz and Hirschberg 1992). In a study of British English read speech, Brown et al. (1980) also found that subjects typically started new topics relatively high in their pitch range and finished topics by compressing their range; they hypothesized that internal structure within a topic was similarly marked. Lehiste (1979) found similar results for single paragraphs, also in read speech. Silverman (1987) found in production studies that manipulation of pitch range alone, or range in conjunction with pausal duration between utterances, allowed subjects to reliably disambiguate utterances that were intuitively potentially structurally ambiguous; he used a small pitch range to signal either continuation or ending of a topic or quotation, and expanded range to indicate topic shift or quotation continuation in studies focused on producing synthetic speech with natural prosody. Avesani and Vayra (1988) also found variation in range which appeared to correlate with topic or discourse structure in productions by a professional speaker of Tuscan Italian. Swerts et al. (1992) also found that F0 scaling was a reliable indicator of discourse structure in spoken instructions, although the structures tested were quite simple. However Ayers (1992) found that pitch range appeared to correlate more closely with hierarchical discourse structure in read speech than in spontaneous speech.

Contour type has been mentioned as a potential correlate of discourse or topic structure (Brown et al. 1980; Hirschberg and Pierrehumbert 1986; Swerts et al. 1992). In particular, Hirschberg and Pierrehumbert (1986) suggest that H*!H* L-L% (*downstepped*) contours commonly appear either at the beginning or the ending of topics. Swerts et al.'s empirical studies showed that 'low' vs 'not-low' boundary tones were good predictors of topic endings vs continuations, but there is little other empirical evidence for these proposals to date.

Duration of pause between utterances or phrases has also been identified as an indicator of topic structure (Lehiste 1979; Chafe 1980; Brown et al. 1980; Silverman 1987; Avesani and Vayra 1988; Swerts et al. 1992; Passonneau and Litman 1993), although Woodbury (1987) found no similar correlation. Brown et al. (1980) found that longer, topic pauses (.6–.8 sec.) marked major topic shifts. Passonneau and Litman (1993) also found that the presence of a pause was a good predictor of perception of segment boundaries in Chafe's pear stories. Another aspect of timing, speaking rate, was found by Lehiste (1979) and by Butterworth (1975) to be associated with perception of discourse structure: both found that utterances beginning segments exhibited slower rates and those completing segments were uttered more rapidly. Amplitude was also noted by Brown et al. (1980) as a signal of topic shift; they found that amplitude appeared to increase at the start of a new topic and decrease at the end.

Another role played by prosody in conveying discourse structure is its use in disambiguating potentially ambiguous *discourse markers*, or *cue phrases*—phrases such as *now, well, in the first place*, which can either function as explicit indicators of discourse structure (a *discourse* use) or can be interpreted in a *sentential* reading, often as adverbials. Variation in phrasing and pitch accent are significantly correlated with this disambiguation (Hirschberg and Litman 1993). Tokens interpreted as discourse uses are commonly produced either as separate phrases or as part of larger phrases; in the latter case they tend to be deaccented or uttered with an L* accent. However, when cue phrases are produced with H* prominence, they tend to be interpreted as temporal adverbs. So phrases starting with cue phrases are likely to be interpreted as starting a new topic or subtopic in a discourse.

28.5.6 Prosody and turn-taking

There is evidence that prosodic information is importantly involved in a variety of conversational phenomena observed by researchers in a variety of areas (Sacks et al. 1974; Auer 1996; Selting 1996; Koiso et al. 1998a; Gravano and Hirschberg 2011). Work by Curl (2004), for example, has shown that repetitions after next-turn repair initiators such as *huh?* and *what?* differ in their prosody from the utterance which occasioned the query in pitch range, loudness, articulatory characteristics, and duration. Differences in this features indicated whether a response was *fitted* (the initial turn was produced in the clear or overlapped by the listener) or *disjunct* (the initial turn itself overlapped the conversational partner's talk). In German, emphatic reduplication can signal that

important information is coming up (Niebuhr et al. 2012). Golato and Fagyal (2006) found that the intonation of 'ja ja' sequences in German telephone conversations carried different meanings based upon their intonation and their position in a conversational sequence: when uttered with H* L-% contours (GToBI), they conveyed that the previous speaker had spoken at too great length; uttered with L+H* L-% intonation they appeared to indicate some conversational misalignment. Finally, a number of studies have found that *final lowering*, a compression of the pitch range during the last half-second or so of an utterance, can also convey structural information to hearers, by signalling whether or not a speaker has completed a topic or a turn (Silverman 1987). Pitch contour and range as well as timing have also been shown to correlate with turn-final vs turn-keeping utterances (Sacks et al. 1974; Geluykens and Swerts 1994; Auer 1996; Selting 1996; Koiso et al. 1998a; Caspers 1998; Hjalmarsson 2011; Gravano and Hirschberg 2011), as well as marking backchannels in dialogue (Benus et al. 2007).

28.5.7   Contour interpretation and speech acts

There is considerable evidence that full intonational contours can, in the appropriate context, signal syntactic *mood, speech act*, or speaker attitude, belief or emotion in many languages (O'Connor and Arnold 1961; Bolinger 1986, 1989; Ladd 1980, 1996; Syrdal and Kim 2008; Rilliard et al. 2012). For example, the L*H L-H% contour may convey uncertainty or incredulity, depending upon acoustic realization and context (Ward and Hirschberg 1985; Hirschberg and Ward 1992). In the following example it appears to convey uncertainty:

> A: *Did you feed the animals?*
> B: *I fed the L*+H GOLDFISH* L-H% (is that what you meant?)

In the next example, it appears to convey incredulity:

> A: *I'd like you here at eight tomorrow.*
> B: *L*+H EIGHT in the L*+H MORNING L-H%* (you've got to be kidding)

The high-rise question contour (H* H-H%) has been claimed by Hirschberg and Ward (1995) to signal that the propositional content of the utterance is in fact to be added to speaker and hearer's mutual beliefs while questioning only whether that content is already part of the hearer's beliefs already. Exchanges such as the following illustrate this phenomenon:

> A (showing up for an appointment) to receptionist: *My name is Gregory Ward?*
> Waitress (delivering meals) to customer: *This is the chicken chermula?*

But while there has been considerable additional effort to associate particular contours with particular interpretations, most empirical studies conclude that interpretation is

significantly affected by context and that a given contour may convey different meanings in different contexts (Liberman and Sag 1974; Sag and Liberman 1975; Ladd 1977, 1978, 1980; Bing 1979; Bouton 1982; Ward and Hirschberg 1985; Grabe et al. 1997; Gussenhoven and Rietveld 1997; Sorianello 2012).

More general attempts have been made to identify compositional meanings for contours within various systems of intonational analysis (Gussenhoven 1983; Pierrehumbert and Hirschberg 1990). In particular, researchers have examined how different languages produce declarative statements vs *yes–no* and *wh*-questions—with rising or falling contours—as a step in characterizing the prosody of the language (Barrett and Hata 2006; Ma et al. 2006; Falé and Faria 2006; Keane 2006; Applebaum 2010; Chen 2010; Kung et al. 2010). More recently, variations have been observed in the contours and accentuation patterns of question types with different pragmatic functions in languages such as American and Australian English, Russian, Japanese, Swedish, Korean, Dutch, Italian, Puerto Rican Spanish, Portuguese, and Catalan (Chen 2006; Crocco 2006; Giordano 2006; Rathcke 2006; Hara and Kawahara 2008; Lee 2008; Armstrong 2010; Crespo-Sendra et al. 2010; Hedberg et al. 2010; Fletcher and Loakes 2010; Mata and Santos 2010; Vanrell et al. 2010; Edlund et al. 2012).

In addition to investigations of contour meaning, studies have also been made on the disambiguating role various contours may play in distinguishing between *direct* and *indirect* speech acts—between what might be taken as the 'literal meaning' of a sentence and some other illocutionary use of that sentence by a speaker (Searle 1969). For example, a sentence with the form of a *yes–no* question, such *Can you lift that box?*, might be interpreted in its literal sense (*Are you able to lift that box?*) or as a request to perform the action. Of the considerable literature on how hearers distinguish between the two readings, there is some evidence that prosody can play a role. Perception studies done by Sag and Liberman (1975) examined whether *yes–no* questions and *wh*-questions which were lexically ambiguous between direct and indirect speech act interpretations could be disambiguated by varying intonational contour. Their preliminary findings were inconclusive but did indicate tendencies for such prosodic-based disambiguation. In a later corpus-based study of *yes–no* questions involving modals in a radio call-in show, Steele and Hirschberg (1987) found that tokens uttered with L* H-H% (class rising question contour) of H* L-L% (classic declarative contour) tended to be interpreted as requests for a simple *yes* or *no*, while indirect speech act readings tended to be uttered with H* L-H% (continuation rise) or H* H-L% (plateau contour). Additionally, the modal in tokens interpreted as direct speech acts was more likely to be reduced than in tokens interpreted as indirect speech acts (*Cn yuh carry that suitcase?* (a request) vs *Can you carry that suitcase?* (questioning ability). Following up on this study, Nickerson and Chu-Carroll (1999) found somewhat different results in a series of production experiments. These showed that utterances realized with a low boundary tone (L%) were more likely (73 per cent) to be used to convey an indirect reading and that those with a high boundary tone (H%) were slightly more likely (54 per cent) to be used to convey a direct meaning. Lai (2008) also found that direct and indirect agreement/contradictions differed intonationally. So, while prosodic differences between

productions of direct vs indirect speech acts have been found, these are not consistent across studies. Computational studies have also found differences in the acoustic/prosodic characteristics of different types of dialogue acts (statements, different question types, interjections, and so on) to be useful in automatic classification of such phenomena, such as Shriberg et al. (1998), Taylor et al. (1998), and Lai (2012).

28.5.8 Prosody and implicature

There has been considerable speculation and some empirical studies in the linguistic and psycholinguistic literature on the relationship between conversational implicature and prosody. Horn (1972) noted that contrastive accent helped to license a lower-bound quantity implicature in utterances like *I don't have THREE friends* (I have more than three). Hirschberg (1991) noted that such an interpretation is more likely when the sentence is uttered with *fall–rise* intonation (L*+H L-H%) but also claimed that an upper-bound as well as a lower-bound interpretation was possible, depending upon context. Ward and Hirschberg (1985) explored the role of this contour in implicating uncertainty. Hara and Kawahara (2008) examined the relationship between contrastive accent, scope of negation, and implicature in Japanese. Grodner et al. (2008) and Huang and Snedeker (2009b) have examined the role of prosody in the interpretation of scalar implicatures, finding that such implicatures are calculated by listeners more rapidly when the quantifier is phonologically reduced (e.g. *summa* vs *some of*) in some contexts but not in others. Dennison and Schafer (2010) compared the affect of L+H* L-H% contours vs L+H* L-L% in implicature interpretation and compare this to explicit negation.

28.6 Additional Areas of Interest

There have been a number of other studies of the role of prosody in signalling phenomena such as politeness, charisma, speaker style, and emotional state. For example, Grawunder and Bodo (2010) have found prosodic and voice quality correlates of perceived politeness in the speech of Korean and German speakers. Biadsy et al. (2008) discovered correlations between acoustic and prosodic variation and perceptions of charisma in SAE, Swedish, and Palestinian Arabic-speaking listeners. Kim et al. (2010) found evidence of a rhetorical prosody style in French courtroom speech. There is evidence of prosodic differences between truth and lies (Hirschberg et al. 2005; Fadden 2006). There is a vast literature in the speech community investigating prosodic correlates of emotional speech, with studies of classic emotions and 'derived' emotions such as uncertainty. Most of these studies have found evidence that low-level prosodic information such as mean and maximum pitch and intensity can distinguish different types of emotional state reasonably well (Pereira and Watson 1998; Koike et al. 1998).

A few studies have also proposed that higher-level prosodic information (pitch contour) may also distinguish different emotions, especially those focused on synthesizing emotion (Cahn 1989; Murray and Arnott 1993). However, a few studies have in fact found correlations between contour and emotion (Mozziconacci 1998; Liscombe et al. 2003; Garrido et al. 2012).