#### Intonation

Intonation is typically defined as the systematic and linguistically meaningful use of pitch movement at the phrasal or **suprasegmental** level. In this way, intonation is contrasted with tone which refers to the linguistically meaningful use of pitch movement at the lexical level in languages such as Chinese or Vietnamese. This narrow definition of intonation is usually expanded, particularly in pedagogical treatments, to encompass stress and intonation group analysis, i.e. the alignment of word groups and pitch contours. This broader definition recognises that meaning-bearing elements of the intonation contour select sites of lexical stress and that intonation contours defined by pitch movement often coincide with phrasal or clausal groups separated by pauses. Thus, **intonology** is concerned with subjective perceptions of pitch, stress and pause and their equivalent acoustic parameters of fundamental frequency (F0), intensity (volume) and duration (both vowel and pause lengths). 

A consensus has yet to be reached as to the precise description and unique functions of the intonation systems of languages. In perhaps the most comprehensive survey of intonation systems comprising more than twenty languages, Hirst and DiCristo (1998) outline some of the issues involved in creating a 'prosodic typology'. These include the difficulty of integrating findings from research traditions employing different theoretical frameworks and transcription systems and embracing the very different pitch and stress characteristics of languages as typologically different as English and Chinese. As an illustration of the kinds of concerns that are typically addressed in models of intonation, the following discussion summarises the history of English intonation study and the current state of the field.

The analysis of the intonation system in English is commonly divided into two broad traditions: British and American. Perhaps the most influential early twentieth-century phonetician was Henry Sweet (1878, 1892) whose **tonal analysis** became the basis for much of the later work in the British tradition. Sweet identified five possible tones, three **single tones** (level, rise and fall) and two **compound tones** (rise– fall and fall-rise). Each tone projected a largely attitudinal meaning, and labels varied quite widely. A rising tone, for instance, could indicate 'an expectant or suspensive attitude' or communicate 'a character of cheerfulness or geniality' (1898: 39). Palmer (1922b) added the tone group as the unit within which the five tones functioned. This was a group of words usually separated by pauses which comprised three segments: the nucleus (the stressed or prominent syllable), the **head** which consisted of anything before the nucleus and the **tail** which included anything after it. Over the next few decades, intonologists added the prehead to Palmer's original categories. The resulting structure is shown in Table 1 (adapted from Tench 1996: 12).

A second, pedagogically oriented system was developed several years after Palmer's work by Armstrong and Ward (1926). They posited two basic **tunes** with a limited number of variations. Tune I was a falling tone used in declarative statements and commands, and Tune II was a rising tone signalling uncertainty or incompleteness. Although this kind of contour analysis continued in the work of O'Connor and Arnold (1961/1973), it is fair to say that it has been eclipsed in more recent approaches by **componential** systems.

One of the most significant contributions to intonation in the British tradition was made by Halliday (1967) as part of his framework of systemic grammar. Systemic grammar unites form and function, and begins with the general principle that intonational contrasts are grammatical in nature and can be shown to be as independently formalised as syntactic choices. Intonation structure comprises three separate systems: tonality (tone unit division), tonicity (internal structure of tone units) and **tone** (pitch movement on the final tonic). Together, the systems unite syntactic, prosodic and information structure. Halliday proposes a marked/ unmarked distinction in which unmarked tone **units** comprising prosodic feet are coextensive

Table .	1
---------	---

Pre-head	Head	Nucleus	Tail
a	DOG is a person's best	FRIEND	I reckon

with information units and syntactic clauses. The internal structure of the tone unit comprises 'given' information followed by a 'new' or focal element coinciding with the tonic syllable on the last lexical item of the tone unit. The tonic syllable carries one of five possible tones, and a number of secondary tones may appear on both the tonic and pretonic to indicate affective meaning.

In the following example, the tone group (indicated by a double slash) is divided into rhythmic feet (indicated by a single slash); each foot contains one stressed syllable and one or more unstressed syllables. The pretonic segment, comprising everything before the tonic, 'mind', carries the secondary tone marked by the full stop and specifying a 'neutral' pretonic. The tonic also carries a neutral, falling tone [1] and is glossed as 'unemotional'. Thus, the symbols describe 'tone 1 with neutral pretonic and neutral tonic' (Halliday 1970: 14).

## //.1 why don't you/ make up your/ mind// (Halliday 1970: 32)

Halliday's model remains in use, and although researchers continue to disagree as to its internal structure and its pragmatic meaning, the basic concept of the tone group or tone unit continues to be the unit of analysis in much British English work to the present day.

A somewhat different tradition developed in American linguistics. The influential structural theorist Bloomfield (1933) regarded intonation and stress as secondary phonemes both because they could not be attached to a particular segment and because he regarded intonation as a 'modification' of speech. Thus, much of the work that directly followed Bloomfield was concerned with assuring the status of stress and intonation as distinctive linguistic features. The most thorough description of the system at that time was given by Pike (1945) whose comprehensive phonemic treatment of intonation, stress and pause and accompanying transcription methods assured prosodic features a place in mainstream linguistics. Pike posited four relative but significant levels of pitch. These pitch **phonemes** were described as the basic building blocks for intonation contours and shown as a

series of connected numbers representing the particular levels, e.g., 2-4; 1-3. In addition, he stipulated two pause phonemes - a tentative and a final pause. In terms of function, Pike viewed intonation as attitudinal. He listed approximately thirty primary contours, and a number of modifications variously labelled in attitudinal terms such as 'endearment', 'detachedness', and 'incomplete deliberation'. A strong critic of this kind of analysis was another American, Bolinger, who argued that pitch levels themselves were not meaningful and that configuration was the key: 'the basic entity of intonation is the pattern ... the fundamental, down to earth sense of a continuous line that can be traced on a piece of paper' (1951: 206).

With the publication of The Sound Patterns of English (Chomsky and Halle 1968) intonation was again consigned to the edge of linguistics and purposefully omitted by the authors. Some generative scholars attempted to generate intonational contours via transformational rules; however, this was problematic as there was no way to incorporate the acknowledged attitudinal function of intonation. As a way to manage this difficulty, researchers attempted to separate out 'linguistic' and 'non-linguistic' aspects of intonation and ignored the latter.

As these traditions demonstrate, two distinct approaches emerged in English intonation analysis. While the British tradition was criticised for a lack of a theoretical basis and an oversimplification of description, the American system, which had a strong theoretical basis, tended to characterise as 'extralinguistic' those features that did not fit neatly into the proposed framework. From a pedagogical perspective, Levis (2005) suggests that current materials and approaches continue to reflect these different orientations. He notes a bias toward **descrip**tion in British English-based materials and prescription in equivalent American English texts.

Despite these differences, significant agreement has been reached on both sides of the Atlantic regarding the **multifunctional role** of intonation in discourse (Chun 2002; Tench 1996). The grammatical function of intonation encompasses a number of structures including the use of a final rising or falling pitch to distinguish utterances as statements or questions,

and the employment of tone unit and pause structure to disambiguate relative clauses such as:

//My sister who lives in Connecticut is the
oldest// the youngest lives in California//
 //My sister// who lives in Connecticut//
is coming for Thanksgiving//

As previously noted, the **attitudinal function** of intonation is widely recognised. However, more recent treatments of intonation have followed Crystal (1969) in emphasising that care needs to be taken in separating intonational effects from the effects of the lexical items themselves. While it is clear that intonation has an affective function, there is a danger in applying too many precise labels and unnecessarily complicating the tonal inventory. Affective meaning is communicated by a cluster of **prosodic** and **paralinguistic** variables that include loudness, stress, rate, kinesics and contextual expectation among others.

The discourse or textual function of intonation encompasses both informational and **interactional** aspects of pitch and pause structure. Production and perception studies investigating the role of discourse prosodics in information-structuring suggest that systematic pitch and pause characteristics are linked to topic structure at both the local (utterance) level and global (discourse) level. Speakers use pitch range and pause length to mark boundary strength, and listeners use prosodic cues to parse incoming information and predict up-coming discourse structure (Cutler et al. 1997). Nonreferential or interactional functions of intonation include the use of pitch variation to regulate turn-taking in conversation, to communicate sociolinguistic information such as status differences, solidarity or social distance between interlocutors and in general terms to contribute to relationship-building between discourse participants. Research additionally points to an indexical function of intonation associated with the use of specific intonational patterns to mark a speaker's affiliation with a regional or socio-cultural group. Among the most notorious patterns discussed in both the linguistic and nonlinguistic press is the high-rising terminal tone (HRT) also variously known as Valley girl, Mallspeak, Uptalk or Upspeak.

Much of the most recent research regarding intonation has also taken advantage of increased access to technology. Historically, assessments of pitch movements relied on the impressionistic judgment of the intonologist, while developments in the field of acoustic phonetics had little impact on theories of intonation. However, rapid improvements in digital speech processing and synthesised speech have encouraged researchers to bridge the gap between model building and the physical correlates of intonation. This is not without its difficulties. As is true of any model where a fit is attempted between theoretical categories and actual data, particularly phonetic realisations of gradient phenomena such as pitch change, decisions must be warranted and reliable enough to be replicated. Thus, although it is becoming progressively more common to see instrumental support in the form of phonetic diagrams for claims of significant theoretical primitives, the use of instrumentation and the importance it is given in terms of support for any given claim varies considerably among researchers.

The most recent models of intonation structure and function are exemplified below in a discussion of two different yet comprehensive frameworks developed by David Brazil (1985/ 1997) and Janet Pierrehumbert (1980/1987). The models evolved with very different purposes in mind. Brazil's model closely follows the British functionalist and pedagogical traditions and prioritises the description of naturally occurring discourse. His concerns are to both elucidate the role of intonation in communication and develop a model that can be used as a basis for teaching English intonation to language learners (as evidenced by his 1994 publication, Pronunciation for Advanced Learners of English). Pierrehumbert, on the other hand, follows the American, generative tradition. She builds a theoretical model of intonational phonology using language examples largely created and tested in the laboratory. Theoretical primitives and phonetic implementation rules allow the complete phonetic contour to be reconstructed, and applications of this model have included work in synthesised speech. Despite these very different orientations, where both models address the pragmatic function of intonation in discourse, they reach similar conclusions. It

should be noted, however, that this has been a comparatively limited concern of the American model in contrast to its importance for Brazil's **discourse intonation** model.

Brazil proposes that intonation structure directly contributes to the **pragmatic** message of the discourse by linking the information to a world or context that the hearer can make sense of. The speaker chooses from a series of formal options which operate at the same level of abstraction as syntactic and lexical choices and have independent implications for discourse structure. The speaker's choices project a context of interaction based on the ongoing situated context of the discourse and the speaker's assessment of the hearer's knowledge state. As this context is constantly changing, intonation choices are relevant only at the moment of speaking, and the speaker is involved in a continuous assessment of the relationship between the message and the hearer. Therefore, within the context of any given interaction, the participants are in the process of negotiating a common ground or background to which new or unknown information is added, contributing to the structure both within and between intonation units. It is this negotiation toward a state of convergence, a roughly mutual understanding of what is being said in the discourse, that allows for successful communication between participants.

In the British tradition of tonal analysis, Brazil adopts pitch-defined tone units. Unit boundaries are identified solely on the basis of pitch level and movement on stressed or **prominent** syllables. One or two points of prominence, representing the speaker's assessment of the relative information load carried by the elements in the utterance, are identified from the surrounding content. For example, given a potential tone unit such as 'a parcel of books lay on the table', at least two possible prominence selections could be made:

- a. a parcel of BOOKS lay on the TAble
- b. a PARcel of books lay on the TAble

In (a) the speaker presents a prominent choice of 'BOOKS' as opposed to, for example, flowers or cups, and makes a similar prominence choice regarding the location, i.e. on the table as opposed to on the floor or the chair. The choice of prominence on both syllables projects a context in which both these pieces of information are unrecoverable for the hearer. Equally, by choosing not to make prominent certain other words in the unit, the speaker assumes that this information is recoverable for the hearer, either because of non-linguistic factors, e.g., books can be assumed to lie on the table as opposed to stand up, or for linguistic factors, e.g., constraints on the language system limit the choice of function words such as 'of' and 'on'. In (b), a context is projected in which 'books' has already been negotiated between participants, but the two other prominence choices are new:

- A. Was there a book on the doorstep when you came in?
- B. There was a PARcel of books on the TABle.

Prominent syllables are divided into two categories based on where they appear in the tone unit and comprise the first prominent onset syllable (key choice) and the final tonic syllable (termination choice). Both key and termination choices are analysed using a three-term system (high, mid and low) that is based on relative pitch height for any given speaker. High pitch indicates that the material is contrastive or highlighted in relation to the surrounding information. Mid choices are glossed as additive and denote an expansion or enlargement of significant information. A low-pitch choice signifies a reformulation or 'equative' function indicating that no new information is added. In addition, low termination is used as a cue to the end of an interaction. In the following example of a typical teacher-student interaction, the student responds to the teacher's mid-key invitation with a mid key rather than a low key as this would imply the end of the exchange and no necessity for teacher feedback. The teacher confirms the correct answer with a mid key repetition and closes the interaction with a positive evaluation with a low termination:

T: H M //what's the final ANSwer?// L S: H M //sixTEEN//

The third and final system, tone choice, is realised on the tonic syllable, the prominent syllable on which the maximum, sustained pitch movement is identified. There are five possible tone choices. Tones that end in falling movements (fall or rise-fall) are designated as proclaiming and contain new or asserted information. Tones with a rising movement (rise or fall-rise) are described as referring and mark information as already 'conversationally in play', i.e. assumed to be known or recoverable. Thus, tone choice summarises the common ground between speakers, i.e. what is assumed to be known and unknown in the context of any given interaction. A specific choice of tone can also reflect sociolinguistic variables between discourse participants such as differences in social status or social distance. Brazil proposes that rise and rise-fall tones carry an additional value of dominance, and choice between these four tones is the prerogative of the controller of the discourse; for example, the teacher in teacherstudent interaction. The final level, or neutral tone indicates a withdrawal from the unique context of any given interaction. In agreement with some previous treatments of the level tone, Brazil proposes that it is used in semi-ritualised or routinised language behaviour such as repeating formulas or equations or giving directives in the classroom (Brazil 1997: 138):

```
//→stop WRITing →PUT your pens
down//
```

Tone, key and termination are interlocking systems which combine to produce the **communicative value** of the utterance, and discourse genres can be characterised by particular kinds of prosodic composition. In teaching discourse, for example, the model predicts that a teacher will use a low termination and falling tone to end an exchange. Students, on the other hand, are likely to use a mid (agreement) key and rising tone in response to teacher elicitation to show that they are expecting teacher feedback. In addition to the tone unit, Brazil identifies the **pitch sequence**. This is a second, larger unit of measurement which comprises a stretch of consecutive tone units that falls between two low termination choices and delineates longer sections of speech. Points of maximal disjunction (paragraph beginnings and endings) are marked with a high initial key and closed with a low final termination. Essentially equivalent to the paragraph in written discourse, it is consistent with other proposals describing larger units variously labelled as **speech paragraphs**, **intonational paragraphs**, or **major** and **minor paratones**.

Pierrehumbert's (1980/1987) approach to intonation in discourse is usually referred to as the **autosegmental-metrical** (AM) approach. In 1990, Pierrehumbert and Hirschberg published a paper that directly addressed the meaning of intonation in discourse as it was constituted using this model. In agreement with Brazil, they present an independent system which assigns a primarily pragmatic function to intonation choices: we propose that a speaker chooses a particular tone to convey a particular relationship between an utterance, currently perceived beliefs of a hearer and the anticipated contribution of subsequent utterances' (Pierrehumbert and Hirschberg 1990: 271).

Also similarly to Brazil's framework this is a componential model; however, individual components of the pitch contour are constituted within the tradition of **pitch phonemes** or intonational morphemes. Unlike the tonal contour analyses of the British tradition, the model comprises a series of static tones or tonal targets that together with a series of phonetic implementation rules, determine the shape of the pitch contour. There are two groups of tones: pitch accents and boundary tones. Pitch accents occur on stressed or 'salient' syllables and mark the information status of the lexical item on which they appear. High pitch accents (H\*) mark the new information in the following (Pierrehumbert and Hirschberg 1990: 286):

The train leaves at seven.  $H^* H^* H^*$ 

The second group of low and high tones associate with the right edge, or closing boundary of either intermediate phrases or intonational phrases (L per cent, H per cent). Phrases are identified by phonetic criteria and pausing, and as the end of an intonational phrase is also the end of an intermediate phrase, this creates four possible complex tones at the end of an utterance. The following exemplifies a typical declarative contour (Pierrehumbert and Hirschberg 1990: 286)

The train leaves at seven.  $H^*$   $H^*$   $H^*$  L L%

Final boundary tones also indicate whether a section of the discourse is complete. If completed, a low boundary tone marks off the semantically related sections of the discourse (LL per cent); if further discourse is required for its interpretation, a high boundary tone is used (HH per cent). Each component – pitch accent, phrase accent and boundary tone - contribute to a distinct type of information to the overall interpretation of the pattern. Pitch accents convey information on the status of individual referents, and phrase accents convey information as to the degree of relatedness of one intermediate phrase to the surrounding ones. Boundary tones convey information about relationships among intonational phrases whether a phrase is to be interpreted either with respect to a succeeding phrase or not. In addition, a number of automatic phonetic implementation rules also apply that allow the complete phonetic shape of the contour to be recreated. Two of the most significant are an **upstep rule** which raises a L boundary tone after a H phrase accent and a **catethesis rule** which causes a gradual declination of pitch across a phrase. However, unlike Brazil's conception of a high, initial key to mark larger, pitch sequence units, there is no discussion of a phrase initial, left edge boundary tone.

The meaning of the intonation contour is derived from the particular sequence of pitch accents, phrase accents and boundary tones that occur, and many of the tonal combinations that are identified by Pierrehumbert and Hirschberg and the values attached to them bear comparison to Brazil's interpretations. For example, the following contour – an H\* pitch accent followed by an L phrase accent and an L per cent boundary tone – is said to 'convey new information' in much the same way that Brazil's proclaiming, falling tone adds a new variable to the background (Pierrehumbert and Hirschberg 1990: 272)

Legumes are a good source of vitamins  $H^* \perp L\%$ 

If the L phrase accent were followed by a high boundary tone (H per cent), the contour would be essentially equivalent to a mid termination, referring tone in Brazil's model which carries a pragmatic meaning synonymous with Pierrehumbert and Hirschberg's gloss of 'when S believes that H is already aware of the information, if S wishes to convey that it is mutually believed (1990: 290). Thus, both models, while varying substantially in the theoretical constructs that they employ, share a similar conception of the function of intonation in discourse, i.e. that the speaker is focused on fitting their message into their understanding of the current beliefs of the hearer and the weight of subsequent utterances.

Both models of discourse intonation are in use in varying degrees in research and pedagogy and have sophisticated transcription systems. The system associated with the AM model is known as **ToBI** (Tone and Break Indices.) Both frameworks have been applied cross-linguistically and to the analysis of non-native speaker discourse. They have also been used to transcribe corpora of read and spontaneous speech. It remains to be seen if one will ultimately prove to be more explanatory in these diverse applications than the other.

L. P.

#### Suggestions for further reading

- Brazil, D. (1997) The Communicative Value of Intonation in English, Cambridge: Cambridge University Press.
- Ladd, D.R. (1996) Intonational Phonology, Cambridge: Cambridge University Press.
- Pierrehumbert, J. and Hirschberg, J. (1990) 'The Meaning of Intonational Contours in the

Interpretation of Discourse', in P.R. Cohen, J. Morgan and M.E. Pollack (eds), *Intentions in Communication*, Cambridge, Mass: MIT Press.

- Tench, P. (1996) The Intonation Systems of English, London: Cassell.
- Wennerstrom, A. (2001) *The Music of Everyday Speech*, Oxford and London: Oxford University Press.

# Taylor & Francis Not for distribution