

1

A survey of intonation systems

DANIEL HIRST and ALBERT DI CRISTO

1. Background

The description of the intonation system of a particular language or dialect is a particularly difficult task since intonation is paradoxically at the same time one of the most universal and one of the most language specific features of human language.

Intonation is universal first of all because every language possesses intonation. Hockett (1963) made this one of his list of ten significant empirical generalisations about languages: generalisations which we should not necessarily want to include in the definition of what constitutes a language but which just happen to be true. Intonation is universal also because many of the linguistic and paralinguistic functions of intonation systems seem to be shared by languages of widely different origins. It has often been noted, for example, that in a vast majority of languages some sort of raised pitch (final or non-final) can be used in contrast with lower pitch to indicate that an utterance is intended as a question rather than as a statement. In this sense the universal status of intonation is rather different from that observed for other phonological systems such as vowels or consonants for example. While it is true that all languages have vowel and consonant systems, and even that similar patterns of vowels and consonants can be found in languages which are only very distantly related, these systems do not convey meanings directly in the way that intonation seems to. There is, for example, no systematic universal meaning which can be ascribed to

the difference between front vowels and back vowels or between stops and fricatives.

Despite this universal character, the specific features of a particular speaker's intonation system are also highly dependent on the language, the dialect, and even the style, the mood and the attitude of the speaker. Experimental research has shown (Ohala and Gilbert 1981, Maidment 1983) that speakers are capable of distinguishing languages in which utterances are spoken on the basis of their prosody alone. Recent results obtained using low-pass filtered recordings (Mehler *et al.* 1988) suggest the striking fact that as early as four days after birth infants have already acquired (presumably during the last months of pregnancy) the ability to distinguish the prosody of their native language from that of other languages. The prosodic characteristics of a language are not only probably the first phonetic features acquired by a child (Kaplan 1970, Crystal 1973, Lieberman 1986, Levitt 1993, Konopczynski forthcoming), but also the last to be lost either through aphasia (Caplan 1987) or during the acquisition of another language or dialect (Cruz-Ferreira 1984, Touati 1990).

In recent years there has been an increasing awareness of the importance of intonation not only for phoneticians and linguists, but also for psycholinguists (cf. papers in Cutler and Ladd 1983) and for speech engineers working on technological applications such as automatic speech synthesis and speech recognition by machines (cf. Lea 1980, Holmes 1988, Waibel 1988, Vaissière 1989 and papers in Bailly and Benoît 1992). It has become obvious, for example, that to synthesise a continuous text in such a way that a listener can understand it without making a strenuous effort, needs a fairly sophisticated approach to the intonation of the text. In the same way, the fact that listeners obviously pay a great deal of attention to prosodic cues in the process of perceiving and understanding spoken language (cf. Darwin 1975, Cutler 1984, Bannert 1987, House 1990) seems to imply that automatic speech understanding systems should be drawing more information from prosodic features than is currently the case.

Paradoxically, it is still difficult to find in the literature a succinct and precise statement of the specific characteristics which make one language sound prosodically different from another. This is true not only of the vast majority of the world's languages whose intonation has never been described at all, but even for those languages which have been the object of considerable research.

There are probably a number of reasons which can explain this state of affairs. First of all it often seems to be felt that it is difficult, if not impossible, to describe the intonation of a language without being a native or near-native speaker. A consequence of this is that comparatively few linguists have undertaken comparative studies of intonation systems dealing with more than two languages (although cf. Delattre 1965, Gårding 1981, 1984, 't Hart *et al.* 1990) or typological studies of prosodic systems (but cf. Bolinger 1978b, 1989,

Cruttenden 1981, Fox 1985, Bruce *et al.* 1988, Bertinetto 1989). Linked to this, as both cause and consequence, is the fact that there have been strikingly few attempts to provide a language independent prosodic transcription system comparable to that of the International Phonetic Alphabet for segmental transcription (cf. Bruce 1989). The fact that intonation is not written down means that it is difficult for a non-native speaker to decide if two utterances are tokens of the same intonation pattern or not. A preliminary proposal for an international transcription system for intonation is given below (§1.2) and the system described here has been used by several of the authors in their contributions to this book.

The aim of this volume is to assemble, for the first time, a sample of descriptions of the intonation systems of a number of different languages, written by specialists of intonation, most of whom are also native speakers of the language described and/or working in the country where the language is spoken. We thus hope to have made a first step in the direction of establishing a prosodic typology by bringing together some of the material necessary for describing the variability of intonation systems across languages.

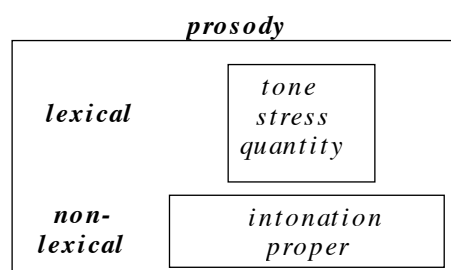
Although we tried to include as wide a sample of languages as possible, we are perfectly conscious that the descriptions presented here are more a reflection of the present state of the art in the field of intonation studies than a statistically significant sample of the variety of intonational forms in human language. A recent evaluation (Ruhlen 1987) estimates that there are about 5000 distinct extant languages in the world which can be grouped into 17 major groupings or phyla. Thirteen of the twenty languages in our sample are from the Indo-European phylum.

We try to emphasise in this survey the ways in which the intonation systems of the different languages described in this volume differ, rather than ways in which they are similar to each other, in the hope that by describing some of the ways in which the individual languages vary we can make some progress towards identifying the different dimensions along which they can be contrasted. At the same time we attempt to provide a thematic guided tour of the material contained in the individual chapters.

Before we look at the different descriptions, a few words concerning terminology may prove useful. We warn the reader, however, that here as in the rest of this chapter, the distinctions we make and the conclusions we draw are in no way intended to imply that all the contributors to this book would agree with us.

The term **intonation** has often been used interchangeably in the literature with that of **prosody**. When a distinction is made between the two words, this is often not explicit. The difference of usage varies considerably from one author to another and can, in our opinion, be traced to a double ambiguity in the use of the term **intonation** itself.

The first ambiguity depends on whether or not intonation is defined in a **broad** sense, that is as including factors such as word-stress, tone and quantity which can be an essential part of the lexical identity of words, or in a **narrow** sense, as excluding such factors. The term *prosody*, like that of *suprasegmentals* can be reserved for the broad sense as opposed to *intonation proper* which is then restricted to what are sometimes called **supralexical**, **postlexical** or simply **non-lexical** characteristics, consisting of such phenomena as the overall form of pitch patterns, declination, boundary phenomena *etc.*, which we describe in more detail in §2 below. This usage can be summarised by the following diagram:



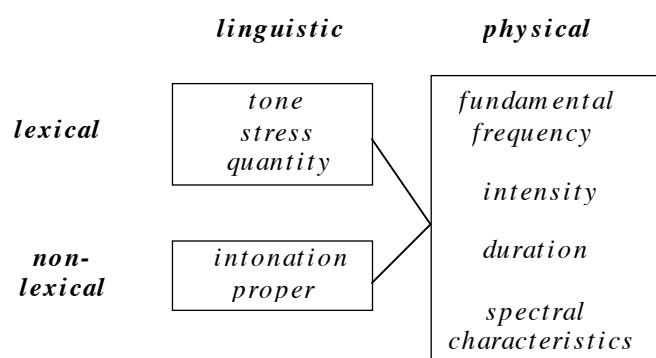
The second ambiguity depends on a distinction between levels of analysis and description. In phonetics, as in all sciences, a distinction may be made between the **physical** level, that of observable and measurable physical parameters, and the **formal** level, which is a rather more abstract level of representation set up as a model in an attempt to describe and explain the observed data. In the case of language, the abstract **linguistic** level attempts to account for a speaker's linguistic **competence**, the implicit knowledge about the language which he is assumed to possess.

On the physical level, intonation is used to refer to variations of one or more acoustic parameters. Of these, fundamental frequency (F_0) is universally acknowledged to be the primary parameter. Many authors, however, have drawn attention to the pluriparametric nature of intonation which besides fundamental frequency involves variations of intensity and segmental duration (Rossi *et al.* 1981, Beckman 1986). Some authors in particular include under the term intonation aspects of temporal organisation or **rhythm** which besides intensity and duration may be reflected in variations of spectral characteristics such as for example distinctions between full and reduced vowels (Crystal 1969).

A distinction between physical and linguistic levels of representation is perhaps more obvious on the level of segmental phonology, corresponding to the distinction between the acoustic data obtained for example from spectrographic analyses (physical) and phonological transcriptions (linguistic). It is worth noting that linguists and phoneticians have always approached

segmental phonology by setting up formal linguistic categories (phonemes, distinctive features *etc.*) and then describing the physical characteristics of these categories. Trying to establish the inventory of abstract categories by direct observation of speech signals would amount to assuming that all problems of automatic speech recognition have been solved. Ladd and Cutler (1983) proposed a distinction between what they described as **concrete** and **abstract** approaches to prosody. In our view, for the reasons we have just exposed (although we are aware that this in no way reflects a universal consensus), any attempt to define intonation on a physical basis (Ladd and Cutler's "concrete approach") necessarily implies a formal (abstract) definition, even if this is never made explicit (for an interesting discussion of the abstract nature of intonation cf. Collier 1972).

The two distinctions we made above: **lexical** vs. **non-lexical**, and **linguistic** vs. **physical** are not in fact entirely independent of each other, since something like the following pattern is commonly assumed:



Among non-specialists, a particularly widespread hypothesis concerning intonation was that of a one to one correspondence between formal prosodic characteristics in general and their physical manifestation, corresponding to a physical equivalent of the formal lexical/non-lexical distinction. It was thus often believed that in English, for example, the formal exponents of lexical prosodic characteristics (word stress) and non-lexical prosodic characteristics (intonation) are mapped onto the physical parameters of intensity and fundamental frequency respectively. Something like this is implicit in Trager and Smith's (1951) description of English by means of four stress phonemes and four entirely independent pitch phonemes. This assumption, which could not have been made if English had been a tone language, has been extremely hard-wearing. In recent years, however, it has been demonstrated that the correspondence between abstract prosodic characteristics and acoustic features is far from simple. On the one hand it has been known for a long time that fundamental frequency (F_0) is a far more efficient cue for stress than either

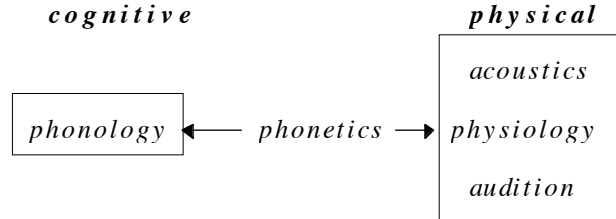
duration or intensity alone (Jassem 1952, Bolinger 1958, Fry 1958, Lehiste 1970, Faure *et al.* 1980). On the other hand, many writers have observed that intensity and duration are more systematically correlated with stress in a language such as English than is F_0 (Beckman 1986). A possible explanation for this was proposed by Hirst (1983a) who suggested that there is an asymmetry between production and perception, so that while duration and intensity differences are the most systematic correlates of stress in speech production, the dominant perceptual cue is fundamental frequency. For a slightly different interpretation, however, cf. Botinis (this volume chapter 16).

One other way to attempt to establish a physical definition of intonation has been to maintain that there is a difference of **scale** between **global** prosodic properties of an utterance put down to intonation proper, and **local** properties which are lexically determined, together with a third lower order level of **segmental** or microprosodic properties (Di Cristo and Hirst 1986, Grønnum this volume). It seems clear, however, that a distinction between microprosody, lexical prosody and intonation cannot be maintained on purely physical grounds since it depends on a prior identification of the relevant linguistic constituents: phoneme, morpheme, word, phrase, sentence *etc.* which are clearly of a formal linguistic nature.

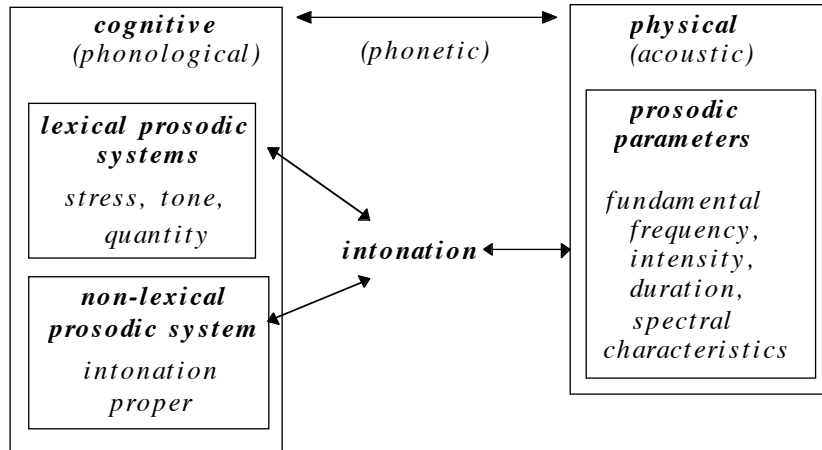
The dichotomy between linguistic and physical levels of analysis, like most dichotomies, is not as water-tight as it might look at first sight. Many, if not most, definitions of intonation fall somewhere in between the formal and the physical extremes, and refer to the speaker's or listener's **impression** of physical characteristics. The terms **pitch**, **loudness**, **length** and **timbre** are often used in this sense as auditory correlates of **fundamental frequency**, **intensity**, **duration** and **spectral characteristics** respectively. Such impressions are evidently determined not only by the physical characteristics of the speech signal but also by the speaker's linguistic knowledge and they somehow straddle the boundary between the physical world and a speaker's abstract (cognitive) representation of that world.

This no-man's-land between the formal and the physical has been the object of much discussion in recent years. A number of writers (Fromkin 1975, Keating 1988) have been concerned with exploring the "phonology/phonetics interface". By contrast, Ohala (1990) has recently suggested that "there is no interface between phonology and phonetics" and has rather pleaded for the integration of phonetics and phonology into a single field of research. Our own view is that while Ohala is right in claiming that phonetics and phonology do not constitute autonomous domains, the concept of an interface is a useful metaphor to describe the link between, on the one hand, an abstract, cognitive level of phonological description and, on the other hand, the physical levels of description provided by acoustics and physiology *etc.* It should be clear, however, that in this sense it is the whole field of phonetics which should be

seen as constituting this interface between the cognitive and the physical levels as in the following:



We propose, then, to continue to use the term **prosody** in its most general sense to cover both the abstract cognitive systems and the physical parameters onto which these systems are mapped. On the abstract, phonological level, prosody consists of a number of lexical systems (tone, stress and quantity) and one non-lexical system: **intonation**. We also propose to use the term **intonation** with a second meaning, to refer to a specifically **phonetic** characteristic of utterances, a construction by which the prosodic primitives on the lexical level and the non-lexical level, however we choose to represent these formally, are related to acoustic prosodic parameters. This phonetic interpretation of intonation as the interface between prosodic systems and prosodic parameters is illustrated in the following figure:



1.1 General prosodic characteristics of the languages

It follows from the definitions given in the preceding section that it is impossible to describe the intonation system of a language without at the same

time giving an account of the other relevant prosodic characteristics, since at a physical level, which after all is the only level which is directly observable, the two sets are inextricably mingled.

Classical phonological descriptions provide a typological framework (for a thorough overview cf. Fox 1985) for discussing these lexical properties, depending on whether the language in question makes lexical distinctions based on quantity, stress or tone. In the case of stress, a distinction is often made (Trubetzkoy 1939; Garde 1968) between languages with fixed stress and languages with free stress. It seems probable however that such distinctions can only be made on formal grounds. There is not, in other words, necessarily any *acoustic* cue for the fact that word stress is lexically distinctive in certain languages such as German, Greek, Russian, Spanish, Arabic, Chinese but not in others such as French, Hungarian, and Vietnamese. In the same way there is no logical necessity for there to exist an *acoustic* distinction between tone languages and stress languages. As Wang (1967) remarked:

It is extremely difficult to distinguish utterances in certain types of tone languages (*e.g.* Mandarin) from those in a non-tone language (*e.g.* English) by just examining the pitch measurements of these utterances. (pp. 99–100).

One possibility (developed in Hirst 1987) is that in fact all languages make use of tone and stress (and presumably quantity) at some point in the representation of utterances. Prosodic differences between languages, under this interpretation, would arise from the fact that the different prosodic primitives can be introduced into the phonological representation at different levels. When the prosodic characteristics of a word are not lexically specified, they will need to be introduced by rules which are assumed to convert an underlying representation into a surface representation. It would follow from this that the fact that surface forms are similar in different languages is no guarantee that the lexical representations will be the same. It is not obvious, in other words, that lexical characteristics will translate in any simple one to one fashion into acoustic characteristics, although of course there must be some way in which the language learner makes a choice between different possible underlying lexical representations. It seems more reasonable, however, given the present state of knowledge, to assume that the distinguishing criteria are formal rather than physical ones (cf. discussion by Van der Hulst and Smith 1988).

Most recent work in the framework of non-linear (generative) phonology has assumed that tone is formally represented in the lexicon of a tone language as a sequence of tonal segments (H and L for example), together with language specific rules specifying how the tones are to be associated to the segments or syllables of the word. Word stress, on the other hand, has been represented in a number of different ways: as a distinctive feature of segments (Chomsky and Halle 1968), as an abstract diacritic symbol (*) associated with one syllable of a

lexical item (Goldsmith 1976), or, in more recent work, as a **hierarchical prosodic structure** in which a sequence of syllables is grouped into a higher order unit, one syllable of which is specified as the strongest element or **head** of the sequence (cf. work in the framework of **metrical phonology**: Liberman 1975, Selkirk 1981, Giegerich 1985, Nespor and Vogel 1986, Halle and Vergnaud 1987, Goldsmith 1990).

Of the languages described in this book, three are clear-cut cases of tone languages. These are: Vietnamese (Îô, Trân and Boulakia), Thai (Luksaneeyanawin) and Chinese (Kratochvil). Chinese is traditionally described as possessing four lexical tones: High, Rising, Low and Falling, although as Kratochvil demonstrates, an adequate characterisation of tonal phenomena in Chinese needs to account for both pitch and intensity variations. Thai is described as possessing five distinctive tones: High, Mid, Low, Rising and Falling; while (North) Vietnamese has six distinctive tones: Rising, Static and Glottalised with High and Low variants of each category. Besides their tonal characteristics, all three tone languages are described as possessing a distinction between stressed and unstressed syllables which is lexically distinctive in Chinese but not in Vietnamese or Thai.

Two other languages presented in this volume: Japanese (Abe) and Swedish (Gårding) are notorious for the fact that they are often described as being somehow intermediate between stress and tone systems. It has been suggested (cf. McCawley 1978, Van der Hulst and Smith 1988) that the classical typological distinction between **stress** languages and **tone** languages should be extended to a three-way distinction between **stress** languages like English, Dutch, Russian *etc.*, sometimes called “dynamic stress” languages or “stress-accent” languages,¹ **tone** languages (like Chinese, Vietnamese and Thai) and **pitch accent** or **tonal accent** languages (like Japanese and perhaps Swedish). In support of such a distinction, Beckman (1986) has presented experimental evidence that accentual contrasts in Japanese make less use of differences in what she calls “total amplitude” (a function of intensity and duration) than they do in English. It remains, however, to be seen whether comparable experimental data from other languages will provide direct evidence for a binary distinction of this sort, or whether it would be preferable to think of languages as forming a continuous scale defined by the average ratio between the duration of stressed and unstressed syllables in the language. Botinis (this volume) suggests such a scale for Swedish > Danish > Italian > Greek > Spanish.

Abe (this volume), points out that contrary to what is observed in stress systems, where only the **position** of stress is significant, words in Japanese need also to be characterised by a contrast between **presence** and **absence** of stress (his T1 and T2 words). This difference has been accounted for in the literature in a number of different ways. In a phonological analysis of the lexical prosodic systems of various dialects of Japanese, Haraguchi (1977) adopted

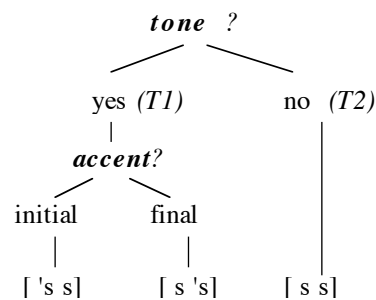
Goldsmith's (1976) use of a diacritic symbol (*) to indicate the place of the accent. Other writers, (cf. Pulleyblank 1986), have suggested that rather than use an abstract diacritic symbol, a pitch accent system can be better accounted for by assuming that for some words a single high tone (H) is "pre-linked" to one vowel in the lexical representation. A possible way to account for the distinction between dynamic stress systems and pitch accent systems would be to suppose, as suggested above, that in dynamic stress languages one syllable is lexically marked as the "head" of the word, whereas in pitch accent systems the relevant lexical characteristic is the presence or absence (and if present the position) of a single lexically specified tone. This allows a simple explanation for the fact that in a stress system the maximum number of potential contrasts is equal to the number of syllables (in practice in many stress systems the position of the head is restricted to a smaller number of possible positions). Thus for disyllabic words in a stress language like Greek (Botinis this volume) we find a two-way distinction such as /nómos/ (law) and /nomós/ (county). In a pitch accent system, by contrast, the potential number of contrasts is one more than the number of syllables since it is possible for accentless words to occur. Thus in Japanese, disyllabic words show a potential three-way lexical distinction with examples like *káki* (oyster) *kakí* (fence) and *kaki* (persimmon). In the dialect of Tokyo, described in this book, the distinction between final accent and no accent is only manifested when there is a following syllable such as the subject particle *-ga*.

Japanese thus appears in some sense halfway between a tone language and a stress language or, as Abe puts it, as "a tonal language but not strictly a tone language". Contrasts in Japanese are **syntagmatic** as in a stress language, rather than **paradigmatic** as in tone languages (Garde 1968) but, as in a tone language, the lexical specification directly encodes relative pitch height rather than simply an abstract position. This also accounts for the fact that in a stress language the actual pitch accent associated with accented syllables may vary according to the intonation (see section 2.1 below) whereas in a tonal accent language this does not appear to be the case.

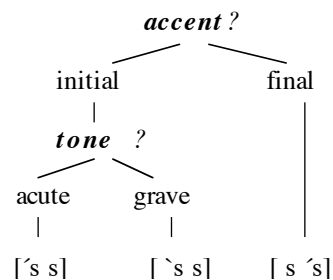
Swedish (Bruce and Gårding 1978, Gårding this volume) possesses two distinct word accents called Accent 1 (acute) and Accent 2 (grave) which can be contrastive except that only Accent 1 can occur on word-final syllables. Gårding discusses several different analyses which have been made of these accents: as an underlying distinction between High and Low, as a sequence High + Low with different association lines linking the tones to the syllables, or as peaks with delayed onset *etc.* The tonal nature of these word accents is apparent from the fact that, just as in a tone language, their overall shape is not modified by the overall intonation pattern. One possibility which is suggested by the present typology is that as in German, Dutch, Danish and English *etc.*, syllables in Swedish can be marked syntagmatically as prosodic heads, but that as in

Japanese, Swedish also allows a paradigmatic contrast between presence and absence of lexically distinctive (High) tone on non-final stressed syllables. For a similar analysis of East Norwegian pitch accents cf. Withgott and Halvorsen (1988).

Swedish and Japanese thus both appear to possess characteristics of both paradigmatic (tonal) and syntagmatic (accentual) prosodic systems. A difference between the two systems seems to be that Japanese is predominantly tonal and only secondarily accentual whereas Swedish is the opposite. The possible patterns for disyllabic words could be summarised as follows for Japanese:



whereas the patterns for Swedish disyllabic words would be:



To distinguish the two systems, we could call the prosodic system of Swedish a **tonal**↗**accent** system, since the tonal contrast is restricted to a subset of the accentual contrasts, whereas we could describe Japanese as an **accentual**↗**tone** system since the accentual contrast is restricted to a subset of the tonal contrasts.

The different types of lexical prosodic systems described above can be summarised as in table 1 below.

It should be noted that it is only the **lexically**↗**distinctive**↗characteristics of a given prosodic system that are recorded in this table so that it is claimed that in a language like Finnish, for example, words have neither lexically distinctive tone, nor lexically distinctive stress, although as suggested above the

Table 1. A classification of languages on the basis of their lexical prosody.

type	example	number of lexical tones	lexical stress
fixed stress	<i>Finnish</i>	0	no
free stress	<i>Greek</i>	0	yes
accentual tone	<i>Japanese</i>	1	no
tonal accent	<i>Swedish</i>	1	yes
tone	<i>Thai</i>	>1	no
tone and stress	<i>Chinese</i>	>1	yes

phonological system of Finnish will specify that stress will be assigned to the first syllable of each word and that this stress will be manifested by a particular tonal configuration. Similarly, both Japanese and Swedish are described as possessing a single lexically distinctive tone which is present only on a subset of lexical items (T1 words in Japanese, A2 words in Swedish).

Once again it is assumed that the phonological systems of these languages will determine those accentual and tonal properties of words which are not lexically specified.

A number of linguistic models have been proposed in recent years to account for the way in which phonologically distinctive features such as accent and tone are converted into the relevant acoustic characteristics of an utterance. Ladd (1983a) proposed a distinction between two basic types of intonation models which he called the **Contour Interaction** type and the **Tone Sequence** type. In the Contour Interaction type, the intonation contour is seen as the result of superposing on each other pitch configurations of different sizes. In the Tone Sequence type:

The pitch movements associated with accented syllables are themselves what make up sentence intonation. (p. 40)

Contour Interaction models have their origin in attempts to apply techniques of analysis by synthesis to fundamental frequency curves, factoring out the observed values into two interacting components: word intonation and sentence intonation. Such a model, building on earlier work by Öhman (1967), has been developed in a number of publications by Fujisaki and his colleagues (Fujisaki and Nagashima 1969; Fujisaki *et al.* 1979; Fujisaki 1988) who argue in particular that the model adequately reflects the dynamic characteristics of the physiological mechanisms underlying pitch control in speech. The same basic idea, that an intonation contour is the result of the superposition of contours defined on different hierarchical levels, has been applied on a slightly more

abstract level to the analysis of intonation contours of Swedish and several other languages by Gårding and her colleagues (Gårding *et al.* 1982, Gårding 1983, Gårding this volume) as well as to Danish (Thorsen 1983b, Grønnum [Thorsen] this volume).

Unlike Contour Interaction models (as well as most descriptive accounts of intonation in particular languages), Tone Sequence models have been particularly concerned with integrating the description of intonation into an overall view of phonological representation. This is particularly evident in the version presented by Pierrehumbert (1980) which builds on earlier work by Goldsmith (1974, 1976) and Leben (1976), themselves following in the tradition of Newman (1946) and Trager and Smith (1951). Pierrehumbert's work explores the possibility that the linguistic primitives involved in models of intonation are not formally distinct from those involved in lexical tone systems. In addition to an inventory of tones, restricted for English to H(igh) and L(ow), Pierrehumbert makes use of diacritic symbols, distinguishing for example between "H%" representing a **boundary tone**, "H*" representing the "strong" tone of a pitch accent and "H-" representing a phrase accent. She points out that:

both H* and H% are equally H tones but they differ in how they are associated with the text. (p. 29)

A synthesis of aspects of both Contour Interaction and Tone Sequence models seems the most promising direction for future research. In particular, as we suggested above, rather than make use of ad-hoc diacritic symbols, the representation of hierarchical prosodic structures makes it possible to distinguish the different types of tones directly by associating them with different levels of prosodic structure (Hirst 1983a, 1987, Bruce 1988, Pierrehumbert and Beckman 1988), assuming that the same formal apparatus used to describe lexically distinctive prosodic characteristics (tone, quantity and stress) is available for the phonological representation of utterances in all languages. Just as it is obviously impossible to speak without variations of fundamental frequency, intensity and segmental duration, so it is impossible, following this idea, to speak without tonal segments and hierarchical prosodic structures. Typological distinctions would arise then not from what phonological primitives are used in a representation, but at what level of the representation (lexical or nonlexical) these primitives are introduced.

1.2 Theoretical background

Despite our attempts to facilitate comparisons between the different chapters, and despite the fact that the authors were specifically asked to emphasise description rather than theory, it nonetheless remains inescapable that the description of a language is a complex interaction between the language itself and the linguist

who describes the language from the viewpoint of his own theoretical commitments and convictions. A wide variety of different approaches are represented in this volume. We have made no attempt to harmonise the theoretical standpoints of the individual authors, feeling that such an attempt would be premature, given the limited character of our present day knowledge of the nature of intonation in general. To improve the state of this knowledge, considerable research remains to be done into the ways in which intonation systems vary before we may even begin to formulate an overall theory of the different formal parameters involved in prosodic typology.

We mentioned above the quite remarkable absence of any consensus concerning the transcription of intonation. The absence of any other transcription system led us to develop our own system which, following a suggestion by Hans 't Hart, we call: **INTSINT** (an **I**nternational **T**ranscription **S**ystem for **I**NTonation). INTSINT obviously owes a great deal to a number of other transcription systems which have been proposed in the past although most of these have generally been designed for the transcription of a single language. An international transcription system needs to embody a certain number of hypotheses about what possible variations of prosodic features are significant across languages. Since the original development of the INTSINT system a new system called ToBI (for **T**one and **B**reak **I**ndices) has been proposed for transcribing the intonation of American English (Silverman *et al.* 1992) based on research by Pierrehumbert (1980), Pierrehumbert and Beckman (1988) Wightman *et al.* (1991) and others. There has been much interest in the last few years in the possibility of adapting this system to other languages (see Gibbon (this volume) for its application to German), although the authors of ToBI have pointed out on several occasions that they do not believe it can be used directly for describing other languages or dialects, since, like a broad phonemic transcription, it presupposes that the inventory of tonal patterns of the language is already established. By contrast, INTSINT can be considered the equivalent of a narrow phonetic transcription and can consequently be used for gathering data on languages which have not already been described.

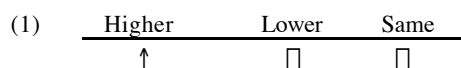
One specific original motivation for INTSINT was an attempt to develop a system which could be used for transcribing both English and French. Transcription systems devised for English intonation are not generally suitable for transcribing French intonation, since in French (Di Cristo, this volume) rhythmic groups **culminate** with a prominent syllable rather than beginning with one as in English (Wenk and Wioland 1982), for further discussion see below §2.1). This characteristic of French intonation makes it impossible to use the same symbols to mark the boundaries of the groups, the position of the prominent syllables, and the pitch movement involved, as is the case with many transcription systems devised for English. To avoid this problem, in INTSINT the transcription symbols are written on a separate line below the orthographic

or phonetic text so that pitch patterns can be transcribed independently of the division into stress groups.

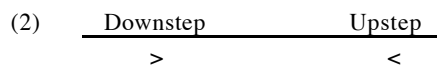
In contrast with many transcription systems in which pitch **movements** are taken to be the primitive elements, it is assumed in INTSINT that the phonetic representation of an utterance is most adequately defined as a sequence of static points (cf. the “turning points” of Gårding 1977a) each of which is linked to the neighbouring points by an appropriate transition function. We have used a phonetic representation of this type for a number of years now for modelling F_0 curves of utterances in several languages (Hirst 1983a, 1987, 1991, Hirst *et al.* 1993). Recent work has concentrated on developing algorithms for automatic analysis of prosodic features. For an overview and a discussion of the relationship between different levels of representation cf. Hirst *et al.* (in press).

All the pitch symbols in INTSINT are used to define a pitch **point** or **target**, the height of which is determined in one of two ways.

The first possibility is for pitch points to be defined as relatively **higher**, **lower** or the **same** as the immediately preceding pitch point:



Two further symbols make it possible to represent a slight **Downstepping** (lowering) or **Upstepping** (raising) of pitch relative to the preceding pitch point:



In most cases, Higher and Lower correspond to peaks and valleys, respectively, whereas Downstep and Upstep correspond to a levelling off in a falling or rising stretch of pitch.² The possibility is also, however, left open to make a quantitative distinction, in that Downstep and Upstep are assumed to imply a smaller pitch change than that transcribed as Lower or Higher.

A second possibility is for the symbol to refer more globally to an extreme value with respect to the speaker’s range of voice, in which case it may take the value **Top** or **Bottom**:



Beyond a certain length, utterances tend to be divided into units which can be, but are not necessarily, separated by pauses. Square brackets are used in INTSINT to mark the **boundaries** of these units which we shall refer to as **intonation units** (for discussion of this and other terminology see below §2.1 and §2.4).

An extreme **initial pitch** in an intonation unit can be marked inside the initial bracket as **Top** or **Bottom**. An unmarked initial bracket is taken as meaning a **Mid** initial pitch.

- (4)

Top	Bottom	Mid
[	[	[

Final pitch in an intonation unit can be marked inside the final boundary as **Top**, **Bottom**, **Higher**, **Lower**, **Upstep**, **Downstep**. An unmarked final boundary is interpreted as **Same**.

- (5)

Top	Bottom	Higher	Lower	Downstep	Upstep	Same
[	[	↑	[	>	<	]

A typical pitch pattern such as that of the Finnish sentence *Laina lainaa Lainalle lainan*. (Laina lends Laina a loan) (Iivonen this volume):

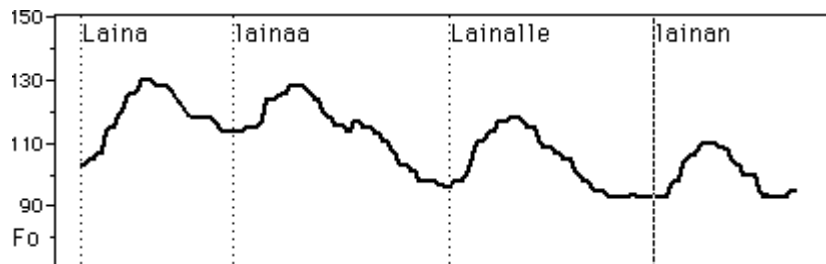


Figure 1. the fundamental frequency curve for a non-emphatic declarative sentence in Finnish (from Iivonen this volume).

can consequently be transcribed simply as:

- (6)

LAIIna	LAIInaa	LAIInalle	LAIInan
[	[	↑	[

The relative scaling of the points within an Intonation Unit need not be specifically marked since it is assumed that the most important factor is the height of each successive point relative to the previous point.

It is probable that no language will need to make use of all of the potential contrasts provided by INTSINT, just as no language uses all the segmental symbols provided by IPA. The description of the intonation system of a given language will consequently need to specify which sequences of symbols constitute well formed intonation patterns in that language and how the symbols relate to the prosodic structure of the utterance.

Note, finally, that the names of the different symbols have been chosen so that the initial letter of the name: **T**op, **B**ottom, **H**igher, **L**ower, **S**ame, **U**pstep,

Downstep can be used for transcription when an appropriate set of graphic symbols is not available.

One voluntary limitation of INTSINT is that it is (at least in its present form) restricted to the transcription of pitch. There is of course no logical reason why it should not be extended to include other prosodic features such as duration and loudness although these can perhaps be more easily integrated into the segmental transcription of an utterance.

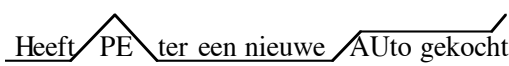
In its present form INTSINT provides no way of scaling relative pitch heights **between** Intonation Units although this obviously will be necessary when dealing with the intonation of continuous texts or dialogues. It might however seem reasonable to assume that in the absence of any other specification the **Top** level in successive Intonation Units is gradually lowered. We should then simply require a symbol to indicate resetting: one possibility would be to use double initial brackets (`[[`) to indicate this. In the same way a double final bracket (`]]`) could be used to indicate prosodic structuring on the paragraph level when this is prosodically signalled *e.g.* by extra-high or extra-low pitch (cf. Hirst this volume). We need, consequently to add the following possibilities:

- (7)

Resetting	Extreme Top	Extreme Bottom
<code>[[</code>	<code>]]</code>	<code>]]</code>

The last of these symbols can also be used in such cases as that described in European Portuguese by Cruz-Ferreira (this volume) for example, where there is a contrast between a low final pitch and an extra-low final pitch, the latter often accompanied by creaky voice.

We mentioned above that INTSINT has been used in ten of the chapters of this book (Hirst, Alcoba and Murillo, Cruz-Ferreira, Moraes, Di Cristo, Dascălu-Jinga, Svetozarova, Misheva and Nikov, Benkirane, Abe). Most of the other chapters use raw or stylised acoustic data for illustration (Gårding, Grønnum, Botinis, Fónagy, Luksaneeyanawin, Îo•, Traşn and Boulakia, Kratochvil). Two other authors, ('t Hart, Iivonen) make use of a simple two-level curve through the text similar to that which had been used by Pike (1945).

- (8) a 

This could of course be transcribed directly into INTSINT as follows:

- (8) b

Heeft	PEter	een nieuwe	AUto	gekocht?
<code>[</code>	<code>]</code>	<code>]</code>	<code>]</code>	<code>↑</code>

Note that when pitch is described as remaining static, as in the sequence *-ter een nieuwe* above, we need to specify this explicitly by means of the symbol Same (`]`) since without this the pitch would be assumed to rise continuously from the low pitch on *-ter* to the high pitch on *AU-*.

Rossi's description of Italian intonation makes use of a rather more abstract system of representation which he has in previous studies applied to the analysis of French intonation. According to this system, an intonation pattern consists of a linear sequence of stress and intonation morphemes functioning at different syntactic levels.

Gibbon's chapter on German, as we mentioned above, uses an adaptation for German of the ToBI system which has recently been proposed as a standard for transcribing the intonation of American English (Silverman *et al.* 1992).

Bolinger, finally, transcribes pitch using the squiggly lines of text which he introduced in Bolinger (1955).³ This system of representation is in fact very similar to the analogical systems described above and the same remarks apply concerning the fact that behind such a transcription there is an implicit system of discrete contrasts. Notice for example in the following illustration (from Bolinger this volume):

(9) I've lost ^{more} ^{pa} _{tients} that way!

the accented syllable **pa-** is represented with a static pitch slightly lower than the level of the preceding accented syllable *more*, whereas the following unstressed syllable *-tients* is represented as a continuously descending sequence interpolating between *pa-* and *that*.

With INTSINT the sentence could be transcribed:

(10) I've lost more patients that way!
[□ □ > □ ↑]

2. Description of intonation patterns

2.1 Description of a basic non-emphatic pattern

Some authors have cast doubt on the existence of such a thing as a "basic neutral unmarked intonation pattern", cf. in particular Schmerling (1976) and Selkirk (1984). For a critical discussion of Schmerling's arguments see Ladd (1980, pp. 73–76). The concept is, however, one which has proved a useful starting point for much research and probably will continue to do so in the future.

A "simple basic pattern" can be defined in several different ways. It can be identified with the pattern observed on "simple syntactic structures" such as simple sentences, clauses ('t Hart), noun phrases (Di Cristo), or even single words (Luksaneeyanawin). Another approach is to define it as the pattern observed on simple semantic or pragmatic units such as "sense groups", "information units" *etc.* (Luksaneeyanawin). It can be defined finally by phonetic

criteria, as a sequence not containing pauses, a “breath group” (Kratochvil) or as a sequence containing only a single major pitch movement or nuclear tone (Cruz-Ferreira). In most cases all these approaches converge on the same **prosodic** unit, variously called by such names as “prosodic phrase”, “intonation unit”, “tone group” *etc.* In this section we shall be concerned with the way in which the general prosodic characteristics of languages described in §1.1 contribute to the overall intonation pattern of simple declarative non-emphatic utterances constituting a single Intonation Unit. The question of the complex pragmatic, semantic, syntactic and phonological constraints on the constitution of intonation units is beyond the scope of this chapter, but for some discussion see §2.4 below.

To describe the variable intonation patterns observed in the different languages we make a distinction between **global characteristics** – affecting the whole intonation unit, **local characteristics** – affecting a single point in the sequence and **recurrent patterns** which occur several times on a smaller sequence usually containing just one stressed syllable. This classification leads us to examine the way in which various authors have made use of the notion of **rhythmic grouping** and this in turn suggests an interesting typological classification.

Many of the intonation patterns we describe below as typical for a given language are also to be found as stylistic or dialectal variants for other languages. Unless otherwise mentioned, we shall be concerned with the pattern which is described as being stylistically the most neutral for each language.

Practically all the languages in this sample are described as having a globally rising-falling pitch movement in simple unemphatic declarative utterances which form a single intonation unit. This overall pattern generally finishes on an extreme low pitch. In Finnish (Iivonen) this low pitch is often accompanied by aperiodic voicing or “creaky voice”. Exceptions to the general rule are mentioned for dialectal variants as in some Midland and Northern dialects of British English as well as in the Extremadura dialect of Spanish and in the Corfu dialect of Greek where declaratives are said to end with a raised final pitch.

Following the British tradition of analysis, we shall refer to the pitch movement associated with the final stressed syllable as the **nuclear** pitch, or **nucleus**. Standard Danish is claimed to be unlike most Germanic languages in that in most dialects (with the exception of that of Bornholm) there is no specific nuclear pitch movement at all. In most theoretical frameworks the nuclear pitch movement is given a special phonological status (referred to as “sentence accent” or “primary stress”). An alternative analysis is to treat the nucleus as the combination of a normal pitch accent and the boundary of an Intonation Unit. Within such a framework Danish would need to be analysed as exceptional in not having any specific pitch manifestation for the end of Intonation Units.

Tone and intonation in tone languages interact to a certain extent so that when rising tones are associated with final falling intonation the result can override the expected tonal configuration as in Chinese (Kratochvil), although without destroying the tonal identity of the morpheme. Abe states that:

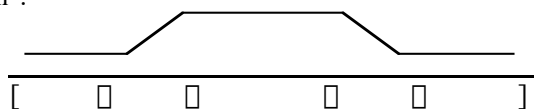
Tones by their nature resist being perturbed by intonation.

Similarly, Luksaneeyanawin notes that in Thai:

The **Tune System** of intonation does not contaminate the phonological system of tones in the language. Each phonological tone still keeps its phonetic features distinct from the other phonological tones.

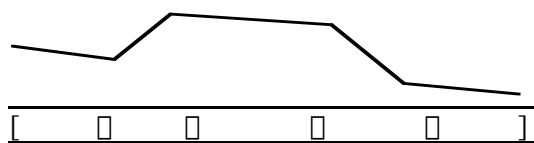
In most languages the falling nucleus is generally prepared by a rising pitch occurring on the first stressed syllable of the unit. Following Crystal (1969) we shall refer to this early rise as the pitch **onset**. One exception to the general tendency for a rising onset is Western Arabic where the pre-nuclear pattern is described as usually more or less flat followed by a nuclear pitch itself consisting of a rising-falling pitch movement. The combination of rising onset and falling nucleus however is an extremely common feature of most other languages in the sample.

When no other pitch movement occurs in the Intonation Unit between the rising onset and the falling nucleus, the resulting pitch pattern corresponds to what has been described for Dutch intonation ('t Hart this volume) as the “hat-pattern”:



Besides Dutch, this pattern is described as common in European Portuguese, German, British and American English.

Between the onset and nucleus an overall globally declining pattern seems to be the unmarked case, so that the “hat” is in fact slightly cocked:



We leave it an open question as to whether the difference between these two patterns should be explicitly marked in the transcription (for example by replacing the symbol Same (□) by the symbol Downstep (>), or whether such general declination should be assumed by default.

In Danish the **degree** of declination is said to vary in function of the modality of the sentence (see §2.2 below), whereas in German it is described as a

stylistic variable. Greek is described as having declination only at final junctures but Botinis suggests that the declination observed in many languages is perhaps an artefact of laboratory speech where individual sentences constitute a complete “turn-unit”, and that work on spontaneous rather than laboratory speech may find little or no declination at all.

Although there is quite a considerable literature describing F_0 declination in various languages, recent work has shown that the nature and generalisation of the phenomenon are far from uncontroversial. First of all, declination has been interpreted either as an actively controlled process or as a by-product of a peripheral mechanism (see, for a review and discussion: Cohen *et al.* (1982), Gelfer *et al.* (1983), Ladd (1984), Bruce (1984a), 't Hart (1986)). Secondly, while declination has been regularly observed in laboratory speech, its manifestation in other circumstances is less evident. Spontaneous speech samples that typically lack declination have been reported by Lieberman and Tseng (1981) and by Umeda (1982). To what extent declination is situation-dependent remains an open question, as does its relevance for phonological and phonetic descriptions of intonation. It seems probable in fact that the notion of declination covers quite a range of different phenomena which interact at different levels. Ladd (1984) gives a useful discussion of the different theoretical assumptions that this notion has been made to cover. Connell and Ladd (1990) suggest that instead of declination, a more general term “downtrends” should be used to cover a number of phonological and phonetic characteristics including in particular the following:

- (i) **declination**: a strictly phonetic characteristic of utterances consisting of a continuous lowering from the beginning to the end of the intonation unit.
- (ii) **downdrift**: an iterative lowering of successive high pitches within an intonation unit separated by intermediate low points.
- (iii) **downstep**:⁴ an iterative lowering of successive high points within an intonation unit without intervening low tones.
- (iv) **final lowering**: a rapid lowering occurring before the final boundary of an intonation unit.

For many of the languages in this volume, besides a globally falling pattern over the Intonation Unit, a series of pitch movements is described, typically associated in some way with the stressed syllables. In the case of tone languages, the form of these pitch movements is determined by the lexically specified tone of each lexical item. For languages with no lexical tone, the same pattern usually re-occurs on each stressed syllable of an Intonation Unit (often with the exception of the final nuclear syllable which may be given special treatment). Many languages have a choice of various recurrent pitch movements for different expressive and stylistic functions but by far the most common pattern described is a rising pitch on each stressed syllable except for the last. A

pattern of this type seems to be basic for Dutch, German, French, Italian, Russian, Greek, Brazilian Portuguese, and Finnish.

In some languages, a falling rather than a rising pattern occurs on the stressed syllables. This seems to be the most common pattern in American English, Spanish and Hungarian as well as a fairly common variant for German.

Another common pattern is a step-down in pitch for each stressed syllable between the onset and the nuclear movement. This seems to be the most common pattern in British English and perhaps Romanian. A similar pattern is described for Danish although unlike the basic pattern for British English this downstep to the stressed syllables is followed by a step-up and a falling pattern on any subsequent unstressed syllables.

A distinction between stressed and unstressed syllables does not logically require reference to any higher-level structuring. Stress is, however, often seen as fulfilling some sort of **grouping function** (Bruce 1985) in that each stressed syllable combines in some way with adjacent unstressed syllables to form a “stress-foot” or “stress group”. Among the arguments which are presented here in favour of such hierarchical structure are: a more satisfactory account of rhythmic characteristics (Hirst) as well as a simplification of the description of intonation patterns (Grønnum).

If we assume such a hierarchical structure, the recurrent pitch movements can be described as occurring not on the stressed syllable but on a higher-order prosodic unit. In British English and Danish this higher order unit is explicitly defined as consisting of one stressed syllable and any following unstressed syllables.

(11) (S s s...) (S s s...) (S s s...) ...

where **S** represents a stressed syllable and **s** an unstressed syllable. We shall refer to this using the terminology of metrical phonology (Selkirk 1984, Nespor and Vogel 1986, Halle and Vergnaud 1987) as a **left-headed** stress group. This is the rhythmic unit which Abercrombie (1964) called the **foot**, borrowing the term from traditional poetics, and which he used to describe the rhythm of English verse and speech. Abercrombie specifically claimed that the foot is “independent of word boundaries” (p. 17) so that a sentence like

(12) We reGREtted his eLEction.

would be analysed as:

(13) We re- (GREtted his e-)(LEction.)

where (...) corresponds to foot boundaries.

A more sophisticated model of prosodic structure in English had been proposed earlier by Jassem (1952) (cf. Hirst this volume) who makes a distinction between rhythmic structure and tonal structure. The **stress-group**

as described here corresponds more closely to the tonal structure of Jassem's description. A similar distinction is made by Grønnum (this volume) who suggests in fact that the prosodic stress group should be analysed as beginning with the stressed **vowel** rather than with the onset of the stressed **syllable**. This is obviously an area in which much work remains to be done. It remains to be seen, in particular, whether rhythm and melody can be accounted for on a cross-language basis by a single explanatory model of prosodic structure.

In French (Di Cristo) and Italian (Rossi), the stress group is described as consisting of one stressed syllable and any **preceding** unstressed syllables:

(14) (s s S) (s s S) (s s S) ...

We shall refer to this structure as a **right-headed** stress group. The typological distinction implied here is similar to that proposed by Wenk and Wioland (1982) between "trailer-timed" languages (left-headed) and "leader-timed" languages (right-headed). This distinction is obviously closely linked to that between "syllable-timed" and "stress-timed" languages, a distinction which has been commonly made since Pike (1945) but which has proved surprisingly recalcitrant to objective measurement (for a recent survey cf. Bertinetto 1989).

The case of Portuguese is particularly interesting. Neither Cruz-Ferreira nor Moraes makes explicit use of the concept of stress group. Something similar is implicit, however, in Cruz-Ferreira's analysis of an intonation pattern into Pre-head, Head and Nuclear tone along the lines of the English school of analysis:

(15)

pre-head	head	nuclear tone
┌ s s s ┐	┌ S s s S s s ┐	┌ S s s ┐

In particular, it should be noted that her analysis implies that unstressed syllables always belong to the same group as the preceding stressed syllable. Recent experimental work on the temporal organisation of stressed and unstressed syllables in Brazilian Portuguese (Reis 1995) seems to suggest a rhythmic organisation into right-headed stress groups. Whether European Portuguese also uses right-headed stress groups remains an open question: an analysis of the different phonological conditions on Portuguese clitic placement in European and Brazilian Portuguese (de Carvalho 1989) has in fact suggested quite independently that stress groups are left-headed in European Portuguese but right-headed in Brazilian Portuguese.

Several authors either do not mention any tendency to group unstressed syllables at all, or else assume that any such grouping will follow lexical and/or syntactic boundaries (cf. Alcoba and Murillo (this volume) for Spanish). It should, however, be noted that, as we mentioned above, similar proposals have also been made for English so that the analysis of rhythmic patterns by means of a single type of stress group in a given language is never uncontroversial. It would however be an extremely interesting result if it could be shown that

left/right-headedness is a language specific typological parameter. If we analyse the prosodic classification of Indo-European languages under this assumption we find a good correspondence with language families, since Germanic languages as a whole seem to make use of left-headed feet whereas Romance languages (except perhaps European Portuguese?) might all use right-headed feet, as in the following, extremely tentative, table:

Table 2. *Typology of non-lexical prosodic systems.*

pitch on stress group	GERMANIC (left-headed)	ROMANCE (right-headed)
Low High [□ ↑]	Dutch, German, Danish, Swedish	French, Italian, Portuguese (?)
High Low [↑ □]	US English	Spanish
Downstepping [>□]	GB English	Romanian

Note that the pitch movement described is that which occurs on the stress group as a whole rather than just on the stressed syllable.

The existence of such a language specific characteristic for stress groups seems a promising area for future research into objective criteria for typological prosodic parameters. It remains to be seen in particular whether reliable phonetic correlates of left/right-headedness, such as different durational patterns, can be established on a cross-language basis.

2.2 Mode and expressivity

It has often been assumed, although this assumption has of course been questioned, that one of the most uncontroversial functions of intonation is that of conveying different illocutionary aspects, or **modes**.⁵ Thus it is commonly maintained that a distinction between declarative and interrogative modes is one of the most universal characteristics of intonation systems and it is also often claimed that intonation contributes to the expression of other modes such as imperatives, vocatives, hortatives, optatives *etc.*, categories which gradually shade into what some authors see as a distinct category of expressive functions.

Even in the case of interrogative versus declarative, however, the nature of the distinction is far from uncontroversial since a question can be said to differ from a corresponding statement in its syntax, its semantics and its pragmatics, as well as any combination of the three. We shall not be concerned here directly with the different syntactic, semantic and/or pragmatic theories which have been proposed to account for these distinctions. While some authors in this volume claim that

there is no specific intonation pattern systematically associated with questions, ('t Hart, Botinis, Iivonen this volume), in a great number of languages, differences in the intonation contours of utterances are said to correspond more regularly to differences in the mode of those utterances, on whatever level this is described. In the following section we briefly compare the various prosodic features which are associated with various modes in the languages described in this volume.

a. Unemphatic yes/no questions

The distinctive properties of intonation patterns associated with non-declarative utterances, in particular with questions, can be classified, like the patterns associated with the corresponding declarative utterances, as **global**, **local** or **recurrent** characteristics, according to the part of the contour which is affected by these properties.

i. Global characteristics One of the most frequent characteristics described concerning interrogative utterances is a raising of the pitch of all or part of the utterance (US English, Swedish, Brazilian Portuguese, Finnish, Hungarian, Western Arabic, Vietnamese, Thai) whether or not the utterance finishes with a final rise.

In two languages (Danish, Vietnamese), questions are said to contrast with declaratives by an absence of global declination.

ii. Recurrent patterns In French, questions are said to contrast with declaratives through the choice of a different range of recurrent pitch patterns on the stress groups of the unit. The unmarked recurrent pattern is described as rising for declaratives but downstepping for questions, whether or not the question is marked with a final rise.

iii. Local characteristics. By far the most commonly described characteristic of questions is a high final pitch. This corresponds to the intuition that questions have rising intonation whereas declaratives have falling intonation. Bolinger (1978b) reported that about 70% of a sample of nearly 250 languages were said to use a rising terminal to signal questions and that the remaining 30% used a higher over-all pitch for questions than for non-questions.

In this volume a high final pitch for non-emphatic yes/no questions is said to be a common pattern for all the languages except Danish, Finnish, Hungarian, Bulgarian, Russian, Western Arabic and Brazilian Portuguese. In Danish, Finnish and Western Arabic, there is no local pitch characteristic for straightforward yes/no questions, and these are consequently distinguished from statements only by the global characteristics described above. In the case of Brazilian Portuguese, Romanian, Bulgarian, Russian and Hungarian, there is

said to be a distinctive pattern on questions which constitutes a final rise only when the last stressed syllable is in sentence final position. Otherwise the pattern consists of a rise on the last stressed syllable followed by a falling pattern on the subsequent unstressed syllables.

For the remaining languages a high final pitch is described as the unmarked pattern for yes/no questions although several authors note, as Bolinger (this volume) puts it:

There are both questions without rises and rises without questions.

When yes/no questions are pronounced without a final rise in these languages they are often said to be perceived as having added connotations.

b. WH-questions

In many languages (English, Spanish, Romanian, Russian, Greek) the intonation of WH-questions is described as being more similar to that of statements than that of yes-no questions. In French, unmarked WH-questions are pronounced with a falling final pitch like statements but the recurrent rising pattern found in statements is frequently replaced by the downstepping pattern observed in yes-no questions. As with imperatives, a rising intonation in these languages can be felt to mark the utterance as “friendly”. In Romanian and Greek, however, WH-questions are said to be more like emphatic declaratives and rising intonation is said to be rare. Finally, in British English at least (see §2.4), there seems to be a stronger constraint against breaking up utterances into smaller intonation units in the case of questions than in the case of statements.

c. Repeat questions

One particular type of question seems to be systematically produced with rising intonation in many languages. This is the type that either repeats or calls for a repetition. For example, in:

- (16) – How do you like it?
– How do I like it?

the repeat question echoes the original question and in:

- (17) – John came last week.
– When?
– Last week.

the question calls for a repetition of all or part of what was said. The latter may be compared with:

- (18) – John came last week.
– When?
– On Tuesday.

in which the question tends to fall because no repetition is involved. These repeat questions could probably be translated into a number of different languages with the same distinction.

Echo questions are said to be invariably rising in English, Swedish, French, Portuguese, Romanian and Finnish. The rise is said to begin and end higher than in ordinary questions in European Portuguese.

d. Tag questions

An utterance pronounced as a statement can be turned into a question in several languages by adding a question tag (...don't you?) which generally constitutes a separate intonation unit with the normal intonation for a yes-no question. These question tags are described for English, French, Romanian, Portuguese and Western Arabic but are certainly more widespread.

e. Unfinished utterances

Like questions, unfinished utterances or continuatives are commonly pronounced with rising intonation in many languages. In fact a raised final pitch is perhaps even more common for continuatives than for questions: thus both Finnish and Bulgarian are described as using a raised final pitch for continuatives but not for questions. Bolinger takes non-finality to be a universal criterion entailing rising or high pitch on both statements and questions.

In Greek and British English, it is said that the final rise in continuatives can be just as high (in spontaneous dialogue) as in questions. French, European Portuguese, Russian and Romanian are described as using a fuller rise for questions than for continuatives, Thai is said to show a "tense-ending rise" for questions and a "lax-ending rise" for continuatives, while in Japanese a question is said to end with a rise, and a continuative with a level pitch. In Danish unfinished statements are said to exhibit a slightly declining pattern, intermediate between that of statements and that of questions.

f. Other modes and expressivity

A number of other correspondences between intonation and mode or expressivity are described sporadically in various chapters. Thus for example some languages are said to use rising pitch to distinguish requests from commands (although Bolinger suggests that while a command usually goes down a repeated command may rise), others are said to use high pitch to express *surprise*, Vietnamese uses high register to express *obviousness*, Finnish uses a delayed peak to express a *hinting* connotation, Brazilian Portuguese is said to have specific patterns for "commands, requests, suggestions, advice and threats" *etc.* Intonation is not alone in conveying these expressive functions. As both Gårding and Cruz-Ferreira point out, we also need to mention voice quality and other "paralinguistic" features for some distinctions. There are, in fact, few cases

where the description of such correspondences seems to reflect any strikingly language specific characteristics. The expressiveness of intonation seems to be the area where universality is at its greatest although of course it is quite possible that this reflects more the state of the art than the state of the world.

2.3 Contextual effects and focalisation

One of the most important functions which has been ascribed to prosody is the part it plays in the way speakers organise and listeners identify information in an utterance. This has given rise to much recent work dealing with the pragmatics of accents, intonation and temporal organisation (cf. for example: Brown and Yule 1983, Eaddy and Cooper 1986, Terken 1985, papers in Johns-Lewis (ed.) 1986, Nooteboom and Kruyt 1987, Pierrehumbert and Hirschberg 1990).

These studies make use of a number of related concepts such as theme/rheme, topic/comment, given/new, presupposition/focus, psychological subject /psychological predicate, background/foreground *etc.* which have been proposed at various times to describe the complex nature of the pragmatic organisation of an utterance and its relation to the discourse context. There is, however, no overall pragmatic framework generally accepted today which is capable of taking into account all the different aspects of this organisation.

Sperber and Wilson, in a recent attempt (1986) to provide such a general pragmatic framework remark that:

there is a huge descriptive literature in this area, but nothing approaching an explanatory theory of the relation between linguistic structure and pragmatic effects. (p. 203)

a. Ground and figure

The basic idea behind all work in this area is that communication takes place against a background of shared knowledge so that the way a listener interprets an utterance will be partly dependent on the (situational) context in which the utterance occurs.

This idea follows a more general principle which had been proposed in the beginning of the century by the Czech psychologist Wertheimer [1886–1943], one of the founders of the Gestalt School of psychology, according to which the perception of a stimulus, particularly in the case of vision, generally consists in attributing a structure in which one part of the stimulus, called the **figure**, seems to stand out against the rest of the stimulus, called the **ground** (cf. Thomson 1968 chapter 13). Under different conditions, the same stimulus can be structured differently into figure and ground, giving rise to a number of familiar optical illusions.

The first person to apply a similar idea to language was another Czech, the linguist Mathesius [1882–1945], founder of the Prague Linguistic Circle, whose theory of **Functional Sentence Perspective** was taken up and developed by other linguists of the Prague School such as Daneš and Firbas. According to this theory, as expounded by Daneš (1967), utterances generally have a bipartite structure consisting of, on the one hand, what the utterance is about: its subject, theme or topic, and, on the other hand, what the utterance says about it. According to Daneš, the original Czech terms **základ** (or **téma**) and **jádro** were first translated by “some Czech scholars” as respectively **theme** and **rheme**, and then later by Y.R. Chao, followed by Hockett (1963) and Daneš himself, as **topic** and **comment**.

The idea that utterances contain a third level of organisation besides syntax and semantics and that one of the main exponents of this organisation is rhythm and intonation (cf. Daneš 1964) was taken up by Halliday in an influential series of publications (1963a,b,c, 1967a,b, 1970) in which he argued, following Mathesius, that besides the thematic structure of an utterance **theme/rheme**, there is a parallel **contextual structure**, according to which information is classified into on the one hand what is **old, already known** or **given** and on the other hand what is **new**.

Halliday also introduced the expression **information focus** (1963b), (although the term “focal accent” had been used by Sharp 1958) to refer to the part of an utterance which is highlighted by pitch prominence. The term focus was taken up later by generative linguists after Chomsky (1970) suggested that the semantic interpretation of an utterance needs to refer to the focus and presupposition, where:

The focus is the phrase containing the intonation center and the presupposition is determined by replacing the focus by a variable. (p. 91)

The link between presupposition/focus, theme/rheme and given/new was made explicit by Jackendoff (1972):

The focus and presupposition designate what information in the sentence is intended to be new and what is intended to be old. (p. 3)

It would obviously take us too far afield to discuss the more recent history of these concepts in any detail but for general discussion from various points of view cf. Firbas 1964, Chafe 1970, 1976, Haviland and Clark 1974, Prince 1981, Reinhart 1981, Culicover and Rochemont 1983. For discussion of their relevance to the study of intonation cf. Ladd 1980, Rossi *et al.* 1981, Brown 1983, Fuchs 1984, Gussenhoven 1984, Schaffer 1984, Selkirk 1984, Rossi 1985, Couper-Kuhlen 1986, Cruttenden 1986, Bolinger 1986b, 1989.

There seem to be two basic uses which have been made of these dichotomies and which are reflected in the contributions to this book. The basic distinction is obviously a **cognitive** one. As Chafe puts it:

Given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance.

While it seems difficult to dispute the cognitive relevance of such distinctions, opinions differ as to whether it is necessary to set up formal linguistic categories as exponents of them.

Hirst (1989, this volume) following Sperber and Wilson (1986) claims explicitly that these categories are not necessary:

Once we find a way of accounting for these phenomena along the pragmatic lines Sperber and Wilson suggest, we are free to assume that neither focus, nor theme, nor topic have any place in the linguistic description of sentences. (p. 141)

A similar position seems to be implicit in the usage of some authors in this volume (Gibbon, Cruz-Ferreira, Moraes) who make use of the distinction old/new or theme/rheme only in a cognitive sense. In other cases, it is felt necessary to use a different term so that, for example, given/new is used as a cognitive label and theme/rheme as a linguistic category defining a constituent of the utterance (Rossi). For Bolinger “theme/rheme” is defined both cognitively (“what the utterance is about” vs. “what is said about the theme”) and to refer to constituents of an utterance: “the rheme contains the most important – new, contrastive, interesting – information.”

Halliday’s original claim was that intonation provides a way of breaking up an utterance into “information units” and that the nucleus (or “tonic” in Halliday’s terminology) of each unit marks the end of the new information.

This claim that there is a one-one correspondence between prosody and information has been the subject of much controversy. Thus it has been argued that in sentences like *The TELEvision has just exploded*, the nucleus does not necessarily mark the end of new information (Schmerling 1976), that an Intonation Unit can have more than one nucleus (Brown 1983), that given information can be just as accented as new information (Nooteboom and Krut 1987) and finally that the distinction between given and new is scalar rather than binary, reflecting different degrees of *givenness* for different items (Terken 1981).

Some authors in this book claim that there is a specific type of accent which acts as a marker of a rhematic element (Gibbon, Alcoba and Murillo, Di Cristo, Svetozarova, Iivonen) or that the distinction corresponds to different levels of accentuation with a higher accent for rheme and a lower one for theme (Bolinger, Iivonen, Nikov and Misheva, Svetozarova). There is, however, fairly general agreement on the fact that the specific pitch patterns of theme and rheme depend on the mode of the utterance (statement or question) as well as, for the theme, on its position in the utterance.

For statements with neutral word order, *i.e.* theme followed by rheme, a continuative pattern for the theme (rising movement or final rise) is described for several different languages (American English, Brazilian Portuguese, French,

Italian, Russian, Bulgarian). In French and Italian, an initial element pronounced as a question is described as constituting an emphatic theme. When the theme follows the rheme, the pitch pattern is variously described as rising or accentless (American English), low falling (Brazilian Portuguese and Bulgarian) or low flat (French and Italian).

The only languages for which the treatment of the theme in questions is discussed are Brazilian Portuguese, Romanian and French. When the theme comes first it has a final fall in Brazilian Portuguese and a globally falling (downstepped) pattern in French. When the theme is postposed it is pronounced with a rise in Brazilian Portuguese and with a high flat pattern or a rise in French.

The rheme is generally said to be characterised by a terminal pitch pattern regardless of its position in the utterance. This pattern is described as falling or rising-falling for statements (French, Italian, Brazilian Portuguese) and rising for questions (French, Brazilian Portuguese, Russian).

b. Focalisation and emphasis

So far we have presented the notion of focus/presupposition as equivalent to that of theme/rheme, given/new or topic/comment. In fact focus is a more general notion than the others which attempts to account by means of a single mechanism both for the given/new aspects of an utterance and at the same time for what has traditionally been treated under a separate heading as **emphasis** or **contrast**. **Contrast** has traditionally been held to be a subset of **emphasis**.

Chafe (1976) remarks that:

Contrastiveness is different from other statuses, and particularly the status of focus of contrast is different from the status of new. There are confusions of contrastiveness with the given-new distinction. There is an unfortunate tendency of both linguists and psychologists to pick foci of contrast as paradigm examples of new information.

Both focalisation and emphasis may be simply defined as the speaker's highlighting of part of an utterance. This is sometimes called **narrow focus** as opposed to **broad focus** in which all the parts of an utterance are given equal prominence (Ladd 1980). A similar distinction is made by Luksaneeyanawin (expressive focus/end focus) and by Nikov and Misheva (emphatic focus/neutral focus). Most contributors to this book, however, use the term **focus** or **focal accent** in its narrow sense.

Some authors prefer the term **emphasis** (Hirst, Dascălu-Jinga, Abe), **pragmatic accent** (Rossi) or **special prominence** (Svetozarova) to that of **focus**. Not much depends on this terminological choice, except for the fact that whereas emphasis is basically a paradigmatic notion, *i.e.* a given item is either emphatic or non-emphatic, focus is basically syntagmatic: focus is on one element of a sequence (more strictly it applies to one node of a tree). This implies that while it is possible to refer to emphatic and non-emphatic versions

of an utterance consisting of a single word, (such as *no.* vs. *NO!*) it is not possible within standard treatments of focus to distinguish broad and narrow focus on a single item (for discussion cf. Hirst this volume).

The traditional distinction between **emphasis for intensity** and **emphasis for contrast**, which Jones (1918) attributes to Coleman (1914), is adopted by a number of authors under various different names including, respectively: **expressive / informative focus** (Gårding): **focus for intensification / focus for contrast** (Di Cristo), **emphasis / contrast** (Dascălu-Jinga, Svetozarova, Nikov and Misheva, Iivonen).

Most languages provide both textual means and prosodic means for emphasising an element of an utterance: very often the textual means is not sufficient by itself. Thus in Western Arabic (Benkirane) an emphasised word is likely to be followed by the morpheme /lli/ and accompanied by an emphatic accent.

It is interesting that emphasis **can** be manifested by intonation alone, without any modification of the syntactic structure. This seems to be the case for all the languages described in this book. It is often claimed that any word of an utterance, including functional words, can receive contrastive emphasis (as for example in French and Chinese).

In European Portuguese (Cruz-Ferreira), a relatively free word order, combined with a principle of **end focus**, has the effect of creating a strong pressure for the word with the highest informational content to be placed at the end of the intonation group, especially in the case of contrast.

(19) Eu prefiro que ela VENHA (I prefer her to come).

(20) Eu prefiro que venha ELA. (I prefer HER to come).

There are however limits to the freedom of word order and intonational devices take over in cases where grammatical constraints forbid the shifting of a contrasted word to final position. In this case, in Portuguese too, any word within an utterance may be given prominence and take the nucleus.

In the majority of languages described in this volume, focalisation and/or emphasis is said to be manifested by an extra pitch prominence, giving rise to larger F_0 movements often accompanied by extra intensity and duration.

In some languages, it is said that focalisation does not imply the use of a specific pitch pattern. This is the case for Dutch ('t Hart) where focalisation for contrast may be achieved by simply flattening out the surrounding pitch accents. According to 't Hart, contrast is not always deducible from the shape of the pitch contour alone and the interpretation of an accent as contrastive in Dutch needs to refer to the context. Other languages which are described as not having a specific emphatic pitch pattern are Swedish (southern dialects) where the rising-falling pattern of the focused word is similar to that of the sentence accent, and Copenhagen Danish which has no specific emphatic form as opposed to the

dialect of Bornholm which uses a rising-falling pattern, similar to a final sentence accent.

When focalisation is accompanied by extra prominence this may be **aligned** with a specific syllable of the focused word: this can be the main stressed syllable as in American English, Romanian, Spanish, Bulgarian *etc.* or an earlier unstressed or secondary stressed syllable as in Brazilian Portuguese, Finnish, French and Chinese. In some cases, (French, Japanese, Chinese) the extra-pitch prominence may spread over the entire word. In a few languages, focalisation is manifested by a rising pitch movement, which may spread over the emphasised word, as in Hungarian and Arabic, or which is more locally defined, as in the dialect of Stockholm (Bruce 1977, quoted by Gårding this volume) where the rise follows the fall of the preceding word accent, with different timing depending on the accent. The last emphatic pitch pattern described by contributors to this volume is a rising-falling one (French, Italian, Russian, and Greek). The timing of the rise and the fall is not specified for Russian, but in Italian and Greek the rise is aligned with the stressed syllable and the fall takes place on the post-nuclear syllable. In French (focalisation for contrast), the rise can take place on any syllable of the focused word, and the last full (stressed) syllable receives a terminal fall, delimiting the focused item as an independent intonation unit. In the case of a focalisation for expressive contrast, both the rise and the fall are compressed onto the final stressed syllable.

The focalisation of part of an utterance leads to modifications of the global pitch pattern which are more or less important depending on the language and/or the dialect. In some languages, focalisation is accompanied by a reduction of surrounding pitch accents both before and after the nucleus. In Danish the flattening of surrounding accents is said to be more important than the upward boosting of the emphasised syllable, while in Dutch, as mentioned above, the surrounding accents may be completely deleted. Western Arabic is said to have no accentual prominence outside the focused word and in Chinese the underlying prominence pattern in the focus domain is occasionally overridden.

After focus, practically all languages in this volume are described as exhibiting an accentless pattern (*i.e.* a low or parenthetic F_0 plateau): French, Italian, Brazilian Portuguese, Danish (non-contrastive focalisation in Copenhagen) and Greek. The only exception described in this collection is Swedish (Southern dialects) where the accented syllables (word accents) often retain their pitch prominence after focus although both phrase and sentence accents are said to be deleted. In general, the pre-focal part does not seem to show such drastic modifications: the prefocal pitch pattern is described as unmodified in Swedish, Italian and Brazilian Portuguese and simply reduced in Danish.

2.4 Phrasing and textual organisation

a. Levels of representation

A very serious problem facing any description of intonation is, as we saw above in section §1.2, that of representation. Specifically we need to decide how we are to represent the fact that the phonetic events that constitute an intonation pattern (however these are themselves represented *i.e.* as tones, turning points, pitch accents *etc.*) seem to have a certain autonomy with respect to the segmental phonematic material, just as in a song the words and the music each have a certain autonomy.

One of the first formal solutions proposed for this problem was to think, as in the representation of a song, in terms of two more or less independent lines where the constituents of intonation were referred to as “prosodemes” (Bloch and Trager 1942), “suprasegmentals” (Hockett 1942), “long components” (Harris 1944) or “prosodies” (Firth 1948). Since the two lines are obviously not entirely independent there must be some way of linking them at various points just as the bar-lines in musical notation provide a point of synchronisation between the text and the music. This was done by postulating a certain number of “secondary phonemes” (Bloomfield 1933) or “juncture phonemes” (Trager and Bloch 1942, Trager and Smith 1951). These junctures were later taken over into the theory of generative phonology (Halle 1959, Chomsky and Halle 1968) and have left their mark on much subsequent work on intonation within a generative framework (Leben 1973, Goldsmith 1976, Pierrehumbert 1980 *etc.*)

Another way of synchronising the segmental and the suprasegmental elements of a representation was to assume that intonation has a hierarchical structure and that it is this structure which is linked to the segmental material. This approach has been characteristic of the British school of analysis⁶ from Palmer (1922) to Cruttenden (1986) and Couper-Kuhlen (1986) although the first person to have proposed such an analysis in print seems to have been Klinghardt in his analysis of the intonation of English (Klinghardt and Klemm 1920), where he breaks down a “speech measure” (Sprechtakt) into component parts:

- (21) **Auftakt** **Taktkopf** **Neuhebung** **Tiefton** **Abtakt**
 (Pre-measure) (Head) (Re-raising) (Low tone) (Post-measure)

This analysis, with only slight modification, was to become the familiar structure as described (with some slight variation) for example by Palmer (1922), Kingdon (1958), O'Connor and Arnold (1961), Crystal (1969):

- (22) **Pre-head** **Head** **Nucleus** **Tail**

A systematically hierarchical analysis of intonation patterns was introduced by Jassem (1952). This model proposed a two-level structure by which a complete utterance can be segmented into a discrete sequence of **tone-groups**, each tone group being itself segmented into a discrete sequence of **tonal-units**. Jassem's

tonal unit is in fact identical to the **stress-foot** introduced later by Abercrombie (1964) and adopted for describing intonation by Halliday (1967a). Abercrombie and Halliday, as we saw earlier, however, adopted the same unit for modelling rhythm and intonation whereas Jassem describes distinct structures: the rhythm unit and the tonal unit.

In this volume, a two-level structure is used by several authors under different names as in table 3. A similar binary hierarchy is implied in the use of major and minor continuatives (Dascălu-Jinga) or intonation morphemes and accent morphemes (Rossi).

It remains an empirical question whether it can be shown that more than two levels are necessary for an adequate theory of intonation. Liberman and Prince (1977) proposed a theory of binary branching phonological structure which contains a potentially infinite number of levels. Martin's (1977, 1981) model of intonation also allows the same sort of infinite recursivity. A more highly articulated model of phonological structure has been proposed by Selkirk (1978) and developed by Nespor and Vogel (1986). According to this model a phonological representation consists of a nested structure of phonological categories ranging from Utterance to Segment, which include Intonation Phrase, Phonological Phrase, Clitic Group, Phonological Word, Foot and Syllable, each category consisting of a discrete sequence of categories of the next lowest level. Other authors have suggested that a more satisfactory approach might be to use a smaller number of constituent types but with a certain degree of recursivity (Hirst 1983b, 1987, Ladd 1986). There is obviously a great deal of work which remains to be done in this area in order to examine how the constituent structures defined by these authors correspond to the generally simpler structures used in describing intonation patterns.

In this volume several authors feel the need for at least one intermediate level between the intonation unit and the stress-foot.

It has sometimes been suggested (e.g. Trim 1959) that there is additional structure **above** the level of the Intonation Unit. Fox (1973) proposed the term "paratone" for a higher order level of this type. Misheva and Nikov refer to "phono-paragraphs". Several authors in this volume (Gibbon, Gårding, Grønnum, Botinis, Iivonen) mention the problem of intonational structuring in continuous discourse and dialogue. Botinis uses the terms "turn-unit" and "sub-turn-unit" for units of dialogue structure.

b. Acoustic marks of intonation boundaries

Most authors assume that the boundary of an Intonation Unit is signalled by a terminal pitch movement (English, German, Dutch, Spanish, Romanian, Italian, Japanese, Finnish, Western Arabic) or by the presence of a final sentence accent or nucleus (Portuguese, Greek, Thai). Other cues which are mentioned are silent or filled pauses (Swedish, European Portuguese, French, Russian,

Table 3. *Three-level hierarchical structures used by authors in this volume.*

Level 1	Level 2	Level 3	authors
prosodic sentence	prosodic phrase	prosodic word	Gårding
prosodic sentence	prosodic phrase	stress group	Botinis Grønnum
intonation unit	prosodic word	tonal unit	Di Cristo
phrase	syntagma	phonetic word	Misheva & Nikov
phrase	speech measure	foot	Iivonen

Finnish, Thai), final lengthening (Spanish, French, Italian), rhythmic cohesion (German), change of tempo (European Portuguese) and pitch resetting (Dutch).

c. Syntactic and pragmatic constraints

Since the earliest studies of intonation, there has been a debate over to what extent Intonation Units are syntactically, semantically or pragmatically determined. Extreme positions have been vigorously defended at both ends of the spectrum. A reasonably general consensus (which is however somewhat short of unanimity), seems to have been reached to the conclusion that while pragmatic reasons for dividing an utterance ultimately prevail there are nonetheless a number of syntactic constraints which need to be respected.

The idea that intonation units somehow constitute units of information or sense units is agreed on by most contributors (Bolinger, Hirst, Gårding, Grønnum, Cruz-Ferreira, Moraes, Di Cristo, Rossi, Misheva and Nikov, Svetozarova, Luksaneeyanawin). They also generally agree that while there is no one-one correspondence between syntax and prosody, there is a tendency for intonation units to correspond to clauses (German, Dutch, Portuguese).

It will be clear from the above that there appears to be no evidence that different languages make use of a different number of levels of intonation structure. Nor do the different languages seem to show much difference in the way pragmatic and syntactic constraints determine the division of an utterance into intonation groups: rather the data from each language needs to be weighed in order to establish empirical evidence for an appropriate model for representing intonation which would be valid for the description of the intonation system of all languages.

2.5 Stereotyped patterns

An extremely interesting variety of intonation patterns called “stereotyped patterns”, “stylised patterns” or “intonation clichés” have been described for a

number of different languages (Pike 1945, Abe 1962, Gibbon 1975, Ladd 1978, Fónagy and Fónagy 1983). These patterns are characterised by the fact that they give the impression of being halfway between speech and song. In this volume, the most frequently cited use of these chants is for calling: a pattern of this type is mentioned for English, German, Dutch, French, Romanian, Japanese and Thai. Another fairly common usage is in jeering chants particularly (but not exclusively) in children's speech (English, German, French). Other uses mentioned include greetings (German), warnings (Thai), enumerations (Spanish, French, German). Stereotyped patterns are certainly more widespread than this but have not yet been studied in several languages. There is some indication of language specific use of these patterns which seem to have wider use in German (Gibbon) and French (Di Cristo) than in English (Bolinger, Hirst) for example.

3. Comparisons with other systems

3.1 Comparisons within languages

Theoretically, there is no reason why differences observed between different languages should not also be observed between dialects of the same language (assuming of course that a distinction between dialects and languages can be founded on linguistic rather than political or sociological criteria). After all, differences between Indo-European languages presumably derive at least in part from differences between earlier dialects of Proto-Indo-European. We can, however, reasonably suppose that, on average, differences between distinct languages will be greater than differences between dialects of the same language, and we should thus expect to find that dialectal variations are statistically more often minor or superficial differences rather than major fundamental ones. In so far as the prosodic characteristics we have described in this chapter are concerned, it seems reasonable to assume that lexical characteristics are more fundamental than non-lexical ones. There does not, however, appear to be any *a priori* reason to classify other prosodic characteristics as fundamental or superficial. The study of dialectal variation could consequently provide a useful tool for this task.

a. Lexical characteristics

Dialectal variations can affect the lexical characteristics of a prosodic system although in this volume there are far fewer cases of this type than those concerning non-lexical characteristics.

In respect of the typology of lexical prosodic systems we proposed above (§1.1, table 3) there is only one case where a dialect would need to be classified under a different category from that of the standard language. In the dialect of Swedish spoken in Finland (Gårding) the distinction between Accent 1 and Accent 2 is neutralised so that this dialect would be classified as a free stress system rather than as a tonal accent system.

In other cases, lexical characteristics vary between dialects but without changing the typological category of the system. The Emilian dialect of Italian (Rossi) is characterised by a tendency for final word stress rather than penultimate stress which is predominant in most other dialects. In Japanese (Abe) the tonal systems are manifested differently in different dialects, the same morpheme being pronounced low-high-high in Tokyo, high-high-high in Kyoto and low-low-high in Kyushu for example. Dialects of South Vietnam differ from those of the North by a neutralisation of one of the tonal contrasts observed in the standard dialect. Dialects of central Vietnam are said to show greater variety than in either the North or the South. In some central dialects a certain degree of neutralisation of tonal contrasts occurs so that speakers from these regions are said to speak “with a level voice”.

b. Non-lexical characteristics

These characteristics fall into three main categories: rhythm and temporal organisation, sentence accent and pitch patterns.

i. Rhythm and temporal organisation. We suggested above (§2.1) that there is some evidence that stress groups in European and Brazilian dialects of Portuguese may be organised into left-headed and right-headed structures respectively. If this proves to be the case it would be the only example in this volume of a dialectal variation of this type although this is an area which has yet to be explored for many languages.

European Portuguese is also said to be characterised by a faster speech rate than Brazilian Portuguese (Moraes), an effect which is partly due to the fact that, contrary to Brazilian Portuguese, pre-nuclear central vowels are frequently dropped in European Portuguese.

Dialectal variations of rhythm are briefly mentioned as typical of some dialects of Finnish (Iivonen), Russian (Svetozarova) and French (Di Cristo) where a number of dialects from the North of France are characterised by a strongly marked stress on the initial or final syllable of words, which sounds like an emphatic stress to speakers of General French.

ii. Sentence accent. In Danish (Grønnum), the dialect of Bornholm can be distinguished from Standard (Copenhagen) Danish and other dialects by the fact that it possesses a sentence accent. Furthermore, whereas in Standard Danish, final lengthening is said to be “slight and non compulsory”, the dialect of Bornholm exhibits a shortening of final syllables. This means that in Bornholm dialect a final syllable bearing a default or sentence accent is generally shorter than a non-final one, despite large pitch variations.

Final lengthening also has a different effect in two Southern French dialects (Provence and Languedoc), where schwa is maintained in the final syllable of an

utterance. Final lengthening affects the last full syllable in Provence but the final unstressed syllable in Languedoc. These differences give the impression that the final contour in statements and yes/no questions is mainly realised on the nuclear syllable in Provence and on the post-nuclear syllable in Languedoc.

In Swedish (Gårding), dialectal variations are characterised by different manifestations of sentence accent: a two-peaked Accent 2 is found in central dialects in which sentence accent and word accent remain distinct, while a single-peaked Accent 2 is observable in Bergstaden and Gotland where sentence accent and word accent coincide.

The form of the focal accent in yes/no questions is different in dialects of Romanian (Dascălu-Jinga), being manifested by a positive pitch-obtrusion in North-Western dialects but with a negative one in Standard Romanian.

iii. Pitch patterns. The choice of a particular type of recurrent pitch pattern (§2.1 table 5) is described as a dialectal variation in English (Hirst), where the unemphatic pre-nuclear pattern is described as a sequence of falling pitch accents in several American and Scottish dialects but as a downstepping sequence in Standard British and as a sequence of rising accents in Welsh English. A similar variability is said to occur in German dialects (Gibbon). In Western Arabic (Benkirane) the pre-nuclear pattern is described as generally flat or rising whereas Syrian and Egyptian varieties of Eastern Arabic are said to show more melodic variation together with a distinct declination effect.

Qualitative pitch differences relative to the distinction between statements and yes/no questions are observed in dialects of Danish, Portuguese, Spanish, Romanian, Greek and Arabic. Sentence mode is said to be signalled globally (by declination slope) in Standard Danish and a number of other dialects, but locally in Bornholm and Sonderberg by a characteristic pitch rise located within the last stress group. In European Portuguese, yes/no questions are usually pronounced with a falling pitch on the nucleus followed by a rise on the final unstressed syllables (Cruz-Ferreira), while in Brazilian Portuguese, pitch rises on the nucleus and drops on the final unstressed syllables (Moraes). In the Muscel (South Carpatian) dialect of Romanian (Dascălu-Jinga) yes/no questions have a rising terminal contour in all cases, whereas in standard Romanian, the final rise is determined both by the placement of the question focus and by the stress pattern of the final word. Questions in Eastern Arabic dialects are said to be commonly pronounced with a final rise whereas this is described as very uncommon in Western Arabic (Benkirane).

Contours with a final high pitch are found in declarative utterances in a number of Northern dialects of British English (Hirst) as well as in the Extramadura and Tucuman dialects of Spanish (Alcoba and Murillo). A similar observation is made for the Greek dialect of Corfu where statements may sound like questions to speakers of standard Greek (Botinis). Contours of this type are

also said to be more frequent in Southern dialects of American English, particularly among younger speakers (Bolinger).

A rising-falling variant of the falling nuclear pitch pattern is described for dialects of three romance languages: Portuguese, Spanish and French. This pattern is frequently used in European Portuguese (Cruz-Ferreira), and is interpreted as attitudinal in Brazilian Portuguese (Moraes). Puerto Rican Spanish also has a pitch pattern of this type in declaratives and questions with interrogative pronouns. In French, a similar pattern is frequently observed in the final contour of Intonation Units in Corsica and Auvergne. A similar dialectal variant is described for German (Gibbon) where the neutral nuclear pitch pattern is described as being typically high with glide to low in the same syllable for Coastal dialects but with a delayed high pitch and slow post-nuclear fall to mid for Rhenish dialects.

In the light of the observations made by the authors of the chapters in this volume, it seems, as might be expected, that lexical prosodic characteristics are those which are the most resistant to dialectal variation, together possibly with the left/right-headed nature of the stress group. Among the parameters which seem more susceptible to dialectal variation are the recurrent pattern typical of unemphatic declaratives, the presence or absence of a final high pitch on yes/no questions as well as the presence or absence of a final high pitch on unemphatic declaratives. This list is of course extremely tentative but may perhaps suggest an extremely interesting direction for future research.

3.2 Comparisons across languages

As we mentioned in the beginning of this survey, (§1.2), the description of an intonation system is the result of a complex interaction between a language and an explicit or implicit theory. This fact is particularly apparent when we attempt to establish comparisons across languages since it is very often difficult to decide whether differences between descriptions are due to different theoretical and/or methodological approaches or whether they are to be put down to genuine differences between the systems constituting language specific prosodic parameters. In this section we briefly list a number of possible candidates for such parameters suggested either by the individual descriptions or by direct comparisons made by the different authors. It will be obvious that in the present state of our knowledge such a list can be neither exhaustive nor well-established: it constitutes rather a number of questions which we feel need to be addressed in future research.

a. Lexical characteristics

These are probably the most heavily theory-dependent of all prosodic characteristics. We suggested (§1.1 table 3) a typology defined by two

independent parameters: accent and tone. It is not at all evident what objective criteria might be used for establishing these parameters since, as we emphasised, the fact that stress or tone is lexically distinctive in a given language does not necessarily imply any easily identifiable acoustic pattern.

b. Non-lexical characteristics

As with dialectal variations, these can be grouped under three headings: rhythm and temporal organisation, sentence accent (and emphasis) and pitch patterns.

i. Rhythm and temporal organisation. There are few direct comparisons between the rhythmic characteristics of different languages but the fact that such characteristics are mentioned as typical of dialectal variations (Portuguese, French, Russian, Finnish) suggests that contrastive studies might well be able to identify systematic cross-language differences in this area. One suggestion (Botinis) is that the average ratio between stressed and unstressed syllables in languages forms a scale along which different languages can be situated.

We proposed a tentative prosodic parameter distinguishing **left-headed** languages, those in which unstressed syllables are grouped with a preceding stressed syllable, and **right-headed** languages where unstressed syllables are grouped with a following stressed syllable. We also suggested that if such structuring is in fact part of the mental representation of speech, we should expect this to give rise to different durational patternings in these languages. It remains, however, to be seen whether all languages can be classified according to this parameter, and whether a simple prosodic structure such as the stress group is sufficient or whether more complex structures need to be posited.

ii. Sentence accent and emphasis. One obvious candidate for a language specific prosodic parameter is the presence or absence of a specific type of pitch pattern for the final stress of unemphatic utterances (Grønnum). As mentioned above, it remains an open question whether the observed differences are best interpreted as different types of prominence or as different ways of signalling prosodic boundaries. Final lengthening also seems to apply differently across languages. Interestingly, Grønnum points out that this phenomenon is not an automatic correlate of a specific nuclear accent. There is also evidence from Grønnum's discussion of Germanic languages that the degree of final lengthening varies from one language to another. Another possible parameter concerns the existence of a specific emphatic pattern signalled other than simply by the reduction or suppression of pitch accents before and/or after the nucleus.

iii. Pitch patterns. We have distinguished three types of characteristics of pitch patterns which we refer to as **global**, **recurrent** or **local**.

Global characteristics. For these, an obvious question which springs to mind is whether some languages are spoken on a higher overall pitch than others. Such a possibility is not mentioned in any of the chapters of this book but there is, as far as we know, little or no empirical data bearing on this issue. A similar question concerns the range of pitch variations used in different languages. In this respect, 't Hart notes that while two pitch levels were sufficient to model Dutch intonation patterns, three levels were needed to model English patterns adequately. It remains to be seen, however, whether an empirical procedure could be defined which would bring this characteristic to light.

Another global characteristic concerns the presence or absence of declination in an utterance. Botinis notes that little or no declination is to be observed in spontaneous Greek utterances but ascribes this rather to the nature of the speech sample than to a language specific characteristic. It is not known whether the difference of declination depending on sentence mode described for Danish is also observed in any other languages.

Recurrent characteristics. Recurrent pitch patterns coinciding with the stress group can be characterised as typically rising, falling or downstepping. Although this is a strikingly obvious characteristic of intonation patterns, considerable variability is found within individual languages depending on such factors as sentence modality (French), expressivity or dialectal variation (English, German).

Local characteristics. In all standard dialects of the languages described in this book, unemphatic statements are said to finish typically on a low pitch. In a few non-standard dialects (English, Spanish and Greek and no doubt other languages), however, the unmarked pattern appears to be one which finishes on a high pitch.

One of the most obvious local characteristics is the use or not of final high pitch in questions. Final high pitch is said to be a common way of signalling questions in most of the languages described here but is described as uncommon in a number of languages from different phyla (*e.g.* Danish, Finnish and Western Arabic). It is worth noting, however, that some of these languages commonly use final rising pitch for unfinished utterances. In a number of languages, two different types of rises, one for unfinished utterances and another for questions, are described although it is not known how widespread or how systematic such a distinction is.

4. Conclusions

We have emphasised over and over again during this introduction the fact that it is extremely difficult to factor out the language specific prosodic characteristics of a language from the theoretical assumptions and background of the author. In

order to establish a typology of prosodic parameters more firmly, it will, in our view, be necessary to undertake systematic co-operative research on an international basis. The most promising direction would perhaps be to attempt to establish a list of objective procedures making as few theoretical assumptions as possible with a view to establishing empirical evidence for the existence of language specific prosodic parameters. It is to be expected that the wider availability of modern technology, including in particular high quality synthesis applied to a large number of languages, will prove an extremely effective tool for this type of research.

One interesting, and to our mind extremely promising, line of research for coming years will be into dialectal variations of intonation systems. It is, after all, arguably easier for one linguist to familiarise himself with three or four dialects of his native language than it is for him to attain an equivalent degree of competence in three or four different languages. As we mentioned above, the variability observed in the intonation systems of different languages is also reflected to a great extent in the variability between dialects of the same language. The study of dialectal variation may provide a valuable source of information as to which parameters of an intonation system are deeply anchored in the system and which are more superficial and hence more susceptible to variation from one dialect to another.

Another aspect which will probably prove important for future research concerns the type of corpus submitted to analysis. The majority of the work reported in this volume is based on the analysis of a form of speech which has come to be known, sometimes rather disparagingly, as “laboratory speech”, consisting of isolated sentences pronounced out of context, usually read rather than produced spontaneously (but cf. in particular the chapters by Grønnum, Botinis and Kratochvil). An obvious question which needs to be answered is how far does variability in the situations in which speech is produced influence the results obtained under these conditions? To what degree do generalisations obtained from isolated sentences apply to more spontaneous situations of communication? Do speakers make use of the same prosodic competence in all cases or are completely different processes at work? There is obviously still a great deal of work to be done in this area before we can even begin to answer these questions.

Finally, it is to be hoped that progress will be made in the next decade towards what one of us has referred to elsewhere as a “Third Generation” model of intonation (Hirst 1991), which would go beyond single language descriptions (first generation) and multi-language descriptions (second generation) by defining a number of independent levels of representation determined by more general linguistic principles. Such a model can only come about as the result of a concentration of collaborative work. It is our hope that this volume will be a first step in that direction.