THE UNIVERSITY OF HULL

Intonation and Second Language Acquisition: a study of the acquisition of English intonation by speakers of other languages

being a Thesis submitted for the Degree of

Doctor of Philosophy

by

John Stephen Barlow, B.A.(Hons.), M.A.

October 1998

Acknowledgements

There are many people whom I would like to thank. Firstly, Gertud Buscher has provided generous support from the very outset. I know that I am only one in a long line of students to benefit from her kindness and sensitivity.

Many thanks to Colin Richardson, who made sure that technical help was always available to me. Also, thanks to Steve Malcolm and David Wilson, who put right my mistakes, supplied me with innumerable items of hardware, software and whatever else was needed.

In terms of the research itself, I am grateful to forty-six informants who allowed me to record and analyse their voices. My gratitude also to Larry Duffy, Dawn Ebbrell, Nicolette Ellis, Piklu Gupta and Julian Halliwell, who proof-read far larger chunks of the thesis than they had expected.

Special thanks to Susana Z-R, whose support and encouragement never faltered. Neither did her patience.

Finally, my immense gratitude to Michael Lumsden, who, as the supervisor of this thesis, has been a continual source of sound advice and meticulous critique. The four years which I spent in Hull were enriched immeasurably by his presence, both as a linguist and as a friend.

Introduction

In the field of second language acquisition (SLA) research, the study of intonation, and prosodic systems generally, suffers from a considerable under-representation. This has far-reaching consequences. From the large body of empirical work on various aspects of SLA over the last three decades, a great deal has been turned to pedagogical use. Indeed, the field of SLA is closely linked to that of language pedagogy, as the dual acquisition-theoretical and pedagogical character of many current journals and conferences shows. However, the mutually nourishing relationship between SLA research and language teaching suffers if either component is inadequate.

In the case of intonation, this is exactly the case. At a time when the processes of SLA are under analysis from a wide range of linguistic, psychological and sociolinguistic perspectives, relatively little is known, even on a simple descriptive level, about the acquisition of intonation. There is no body of studies of L2 intonational form comparable, for example, to the 'morpheme studies' or to studies of 'developmental sequences' which informed much thinking in the field in the 1970s and 1980s (see Ellis 1994, Ch.3); no substantial body of work, that is, which might form the basis of further research.

The present study aims to contribute to current knowledge on the acquisition of intonational form in second languages. It seeks to provide a detailed account of how certain aspects of L2 English intonation develop, both in terms of their phonetics, and also in terms of the linguistic and discoursal ends to which they are put. The study is divided into two parts:

Part One: in which the theoretical and descriptive bases of the study are established. It deals first with aspects of intonational form in English, describing in detail the prosodic systems which are employed to mark various aspects of informational structure within the spoken language, and also considers briefly the current state of language teaching in these areas (Chapter One). Then a review of research into the acquisition of sound systems in second languages is presented, looking particularly at intonational form and other aspects of prosodic production and perception (Chapter Two).

Part Two: in which the experiments which have been undertaken as part of this study are presented. Firstly, the procedural and analytical aspects of these experiments will be described (Chapter Three). The findings will then be presented and discussed (Chapters Four to Seven). Finally, findings will be summarised and some general conclusions drawn (Chapter Eight).

Contents

PART ONE

PART TWO

Chapter Three : Experimental approach

Chapter Four : The production of prominence

Chapter Five : The use of tones to mark information

Chapter Six : Macro-prosodic structure

Chapter 7 : The role of L1 transfer in L2 intonation

Chapter Eight : Summary

Chapter One - Theoretical Foundations

1.0 Introduction

Intonation has been described as 'a collective label for a subset of prosodic systems' (Johns-Lewis 1986: xx). A study of the acquisition of 'intonation', then, might reasonably include any or all of these 'prosodic systems'. My aim in this chapter is to identify and describe a small number of such systems, preferring those which can most satisfactorily be incorporated into an analysis of intonational production by different speakers. I will concentrate on those elements which make a reasonably clear contribution to the semantic or discoursal structure of speech. By thus restricting the scope of the study to a small number of specific elements, a more detailed analysis of the productions of each of these at the phonetic level can be undertaken.

The first three sections of this chapter will involve discussions of the issues surrounding the linguistic description of the following forms: *prominence*, the sentence-level accentual system; *information marking*, by which new information is distinguished prosodically from what is already-established in the utterance; and *paratonic structure*, in which phonological 'paragraphs' are created prosodically.

In this study I will be looking at these aspects of intonational form in the speech of nonnative speakers of English. This analysis will involve making detailed phonetic analyses of these productions, and I will say more about this in Chapter Three. However, the very elicitation and analysis of prominences will also involve discussion of the linguistic systems which motivate these productions, in terms of how speakers react to and interpret particular contexts, and how information is subsequently mapped onto the suprasegmental level in production. So, whereas I will be interested primarily in the phonetic nature of production, interpreting the linguistic processes which underlie these productions is also a necessary part of such a study.

An understanding of the linguistic and discoursal character of these forms is necessary in the present study not only in the interpretation of data, but also because linguistic function is a key factor in devising adequate tests in the elicitation of speech data from informants. For this reason, the present chapter will deal with some of the ways in which these prosodic systems have been discussed, in terms of their contribution to linguistic structure and to the emergence of meaning in speech.

After this survey, I will assess ways in which these aspects of intonational form have so far been incorporated into pedagogical texts. Although this study is about acquisition rather than teaching, it is also pertinent to consider ways in which learners may (or may not) receive training in these areas of production. Also, the ways in which intonation and other aspects of prosodic form are dealt with by language professionals may tell us something about why certain forms are considered difficult to acquire.

1.1 Prominence

One or more parts of a string of spoken language are generally perceived as relatively prominent. This happens within multi-syllabic words, where one 'primary' stressed syllable usually stands out from the rest. Within longer sections of speech, such as single, unbroken 'units' of intonation, there also tends to be one or more recognisable loci of prominence. It is these latter prominences within intonation units which I will be discussing in this section.

In terms of articulation, the principal means of creating intonational prominence are through perturbations in the prevailing production of the three prosodic features of pitch, loudness and duration. However, the relative contribution which each of these prosodic features makes to the system of prominence in English is debated. I will discuss some differing accounts in this section.

Prominence has also been widely discussed in terms of its role within (suprasegmental) phonological structure, and in its relation to linguistic and semantic form. In this section I will assess briefly some of the main issues which such discussions have raised. My use of the term 'prominence' throughout this study serves as a convenient umbrella term, reflecting the general phenomenon; I will, though, use the terminology of different descriptions ('sentence stress', 'accent', 'tonicity' for example) within summaries of particular accounts throughout this chapter.

The review of different descriptions of prominence here will be a broad one, incorporating both phonological and phonetic accounts. Firstly, I will review work on 'sentence stress', looking at Trager and Smith's phonemic approach, and then moving on to the generative approach taken in *The Sound Pattern of English*. After this I will discuss the notion of 'accent', in relation to pitch systems and also to a contextual understanding of prominence. I will then look at 'contrastive' uses of prominence in English, and also at what is understood by the notion of 'information focus', before I draw comparisons with the 'nuclear' model of intonation which has been adopted by many British writers. Finally, I will discuss the role of prominence in pedagogic texts.

1.1.1 Syllabic loudness, pitch and sentence stress

In American analyses, stress (loudness) and pitch have often been considered to be separate systems. This can perhaps be seen most clearly in the work of Pike, for whom 'sentence stress', a phenomenon broadly compatible with our umbrella term 'prominence', occupies a separate layer of structure from pitch:

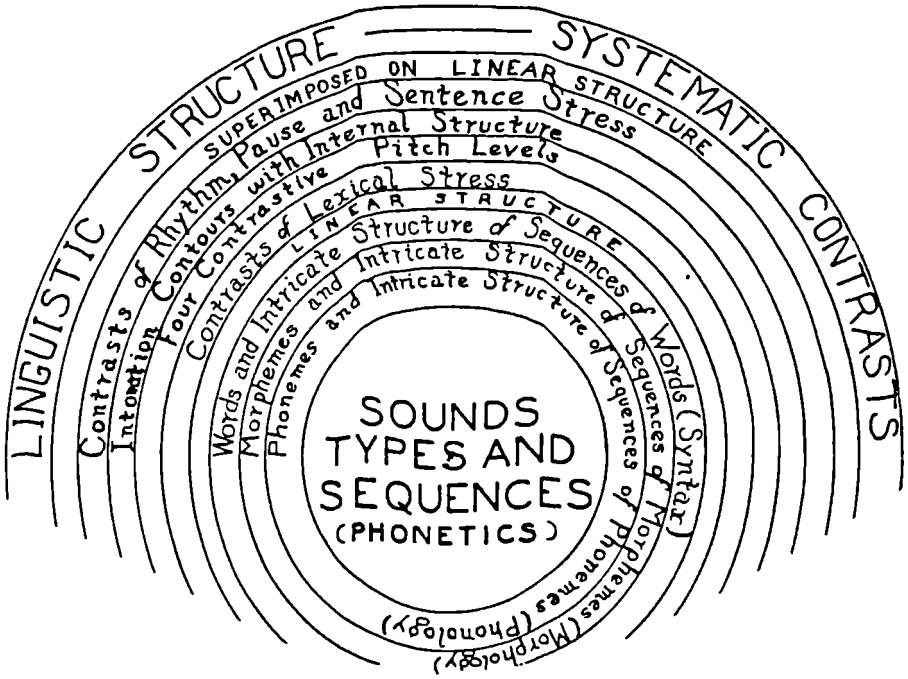


Figure 1.1 **Pike's *Layers of Structure of English***

In Pike's illustration (1945: 171), 'sentence stress' falls within a group of prosodic processes, including the use of pitch, loudness and duration (and pause), which are said to be superimposed onto 'linear' (core linguistic) structure. However, sentence stress is seen as being marked by loudness within a layer of structure separate from the pitch-related systems of both 'intonation contours' and 'contrastive pitch levels'.

The view that loudness is the principal phonetic marker of phonological stress is also seen in Trager and Smith's (1951; 1957) influential description of English structure. Here, stress and pitch exist as 'phonemes' within intonational form, but function independently and belong to different paradigmatic sets. Four degrees of loudness in utterance correspond to 'stress phonemes' (primary, secondary, tertiary, weak). Indeed, the four levels of stress function as markers of prominence on both the word and sentence levels. For example, adjustments in the relative strength (levels) of stress is said to be a feature of compounding:

stress level[1]:	2 1	3 1
	old maid (adj. + noun = *former servant*)	old maid (compound noun = *spinster*)

(Trager and Smith 1957: 39)

The string of word stresses in a sentence yields the stress pattern upon which the sentence stress pattern emerges. The pitch contour is overlaid onto this sentence stress pattern. However, pitch and stress remain independent systems, and pitch does not in itself contribute to the emergence of sentence stress. So, whereas variations in the three main prosodic features may occur at points of prominence (sentence stress), a hierarchical distinction between the contribution of loudness and pitch is made, in which only loudness is seen as the defining characteristic.

Ladd points out that Trager and Smith, in the original version of their description (1951), made no specific mention of the fact that '...what is perceived as sentence stress often coincides with the greatest pitch prominence of the intonational contour' (Ladd 1980: 7), although Trager and Smith did note this in a later (1964) version. Hockett (1958) also recognised this co-occurrence, using the term 'intonation center'[2] to describe the fact that one of the pitch 'positions' of a sentence will always occur on a primary stress, although he also maintains that pitch and stress systems remain independent (p. 47).

A similar notion of sentence stress is used in *The Sound Pattern of English* (Chomsky and Halle 1968) (hereafter *SPE*), where it is referred to as 'nuclear stress'. Virtually nothing is said about pitch in *SPE*. However, stress and pitch are again treated as separate systems, and the idea that pitch is overlaid onto the stress pattern is retained. In *SPE* stress patterns on both the word and sentence levels are seen to be predictable from a set of stress rules,[3] which together generate the stress pattern of individual words. From the string of word stresses, the Nuclear Stress Rule (NSR) then assigns a primary stress to the right-most primary word stress in the constituent under consideration. All other stresses in that constituent undergo a reduction of stress by one level. The NSR applies cyclically, beginning with the smallest constituent, and finally yielding a single sentence-level primary stress. The potentially infinite number of possible reductions in a word's stress level was seen as a major problem for the theory, although readjustment rules had earlier been proposed to avoid this (Chomsky, Halle and Lukoff 1956):

---

[1] Higher numbers here indicate stronger stress.
[2] Hockett's original term 'intonation center' was later used by Chomsky (1971).
[3] The Main Stress Rule, Alternating Stress Rule and Stress Adjustment Rule. There is also a Compound Rule, which assigns leftmost primary stress in compounds.

word stress[4]

$$[[\text{Peter}]_{NP}\ [[\text{baked}]_{V}\ [\text{a cake}]_{NP}]_{VP}]_{S}$$
(1 1 1)

NSR to be applied to NP and VP

$$[[\text{Peter}]_{NP}\ [\text{baked a cake}]_{VP}]_{S}$$
(1 2 1)

NSR to be applied to S

$$= [\text{Peter baked a cake}]_{S}$$
(2 3 1)

Figure 1.2 **The two cycles of the Nuclear Stress Rule, applied to the sentence *Peter baked a CAKE***

In *SPE*, then, the generation of a sentence stress pattern is governed by obligatory rules, which derive a stress contour on the basis of a combination of lexical and syntactic structure. The speaker has no conscious control over the shape of the contour, since its generation is as 'completely determined as, for example, the degree of aspiration' (1968: 6).

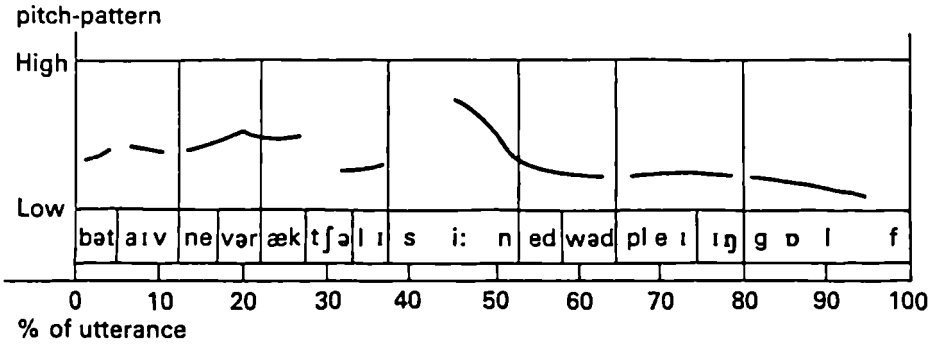
The nature of how pitch is seen to fit into the model of stress-generation is not covered in *SPE* (see p.15n). However, the traditional reading has been that the 'intonational center' (i.e. the major locus of significant pitch and primary stress) will fall on the primary stress as generated by the final cycle of the NSR. That is, intonational prominence in this sense is sentence stress *plus* the accompanying pitch phenomena. So, although pitch and stress (loudness) may maximise on the same place in a sentence, pitch is seen only to co-occur with stress and not to constitute a part of the generation process itself. That is, *SPE* retains the loudness-pitch separation of earlier American accounts.

1.1.2 Accent and the context of speech

Many writers have claimed that, rather than being the principal contributing factor in the articulation and perception of prominence, loudness may in fact be the least consistently

[4] Heaviest (primary) stresses here have the lowest number.

used of the three prosodic features in English. Fudge (1984) notes this in an example of the pitch-pattern of a sentence:



(Fudge 1984: 2)

Figure 1.3 **Typical pitch movements and the relative duration of syllables associated with the utterance of *But I've never actually seen Edward playing golf***

Fudge says that:

> the nuclear syllable **seen** is not necessarily much (if at all) louder than the rest, but it carries a very noticeable pitch movement, and is longer than any two-syllable stretch in the utterance (until the very last syllables, which tend to be prolonged in any utterance).
>
> (Fudge 1984: 2)

Halliday has noted something similar:

> What are often referred to as "four degrees of stress" are rather, at least in British English, structurally identified syllable classes whose exponents are marked by contrasts not only (if indeed at all) in intensity but also in pitch and duration.
>
> (Halliday 1967b: 14)

The interest in pitch and prominence (generally termed 'accent') runs counter to those descriptions of sentence stress reviewed above. Indeed, many linguists in the early 1970s responded to the debate surrounding the generation of sentence stress within the *SPE* framework by claiming that syntactic structure is simply not an adequate basis for explaining the phenomenon (Berman and Szamosi 1972; Bolinger 1972b). Bolinger had in fact long maintained that pitch is central to the perception and production of accent, rather than being merely overlaid onto the stress pattern (Bolinger 1958b). Since pitch prominence (or 'obtrusion') falls naturally (more probably) on the syllables of words which carry full lexical stress, he argues, pitch and stress will often co-occur at prominent points in the sentence. However, on their own the lexical stresses which form

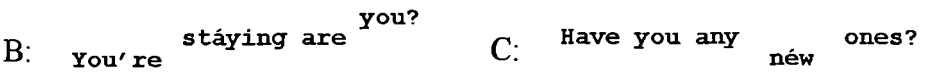
sentence-level stress patterns carry only the *potential* for pitch obtrusion. Indeed, he plays down the role of stress and claims, like Fudge, that the third of the three main prosodic features, duration (of a syllable), may be the more reliable indicator of accent than loudness:

> From the synchronic standpoint, duration is to a large extent a co-variable with accent. From the diachronic standpoint, duration accounts in part for how the unaccentables got that way.
>
> (Bolinger 1958a: 138)

Whereas some writers make a distinction between the specific system of prominence (sentence stress, accent etc.) and the broader domain of intonational contours, Bolinger's work seeks to unify the two. He specifies a set of three quasi-phonemic contours (*pitch accents* A, B and C) which have general profiles in terms of pitch movement, and which also make generalisable contributions to meaning:

Profile A involves '[a]ccent at a relatively high pitch followed by a jump down':

whilst profiles B is 'jumped up to', and C describes a rise somewhat as a mirror to A:



(Bolinger 1986: 142-78)

Profile A 'singles things out' or 'suggests contrast or newness', whereas B tends to express connectedness, and ties things together as opposed to the separateness associated with A. C 'plays down' the sense of the utterance or 'implies foreknowledge'.

These pitch accents, either alone or in combinations within the unit, provide the essential architecture of basic contours. The system of *intonation*, which involves the degree of pitch obtrusion and overall pitch width, serves to fine-tune that architecture, creating a gradient specification to the intrinsic meanings which are built-up from pitch accents. This is not a distinction generally seen in descriptions of English intonation, although Vanderslice and Ladefoged (1972) do use a similar approach.

In terms of accent placement there are no 'rules'; the placement of the pitch accents depends entirely on context and on the speaker's response to context. This is absolutely contrary to those accounts of sentence stress I have already discussed. For

Bolinger, accents mark important or highlighted information, or any other element which, for whatever reason, the speaker wishes to make stand out. The accentual system of prominence then, is seen as a product of context and of the speaker's assessment of information, and as such is not predictable at all. That is, accent is predictable only 'if you're a mind-reader' (Bolinger 1972b).

I will not employ Bolinger's system of pitch accents in this study. However, I do adopt a broadly context-based approach to prominence, rather than a syntactic one. I am interested here in the production of prominences from a phonetic point of view, and also in the production processes involved. I want to assess the ability of nonnative speakers to produce prominences which are nativelike in terms of the use of prosodic features. From a practical point of view, then, we need a broad experimental canvas which will admit all three principal prosodic features of pitch, loudness and duration into an analysis of prominence. I will also retain the idea that prominence relates to context and to speaker-decisions in response to context. Indeed, work into intonation and 'information focus', which I will refer to in following parts of this chapter, was directly influenced by the work of Bolinger.

1.1.3 Prominence as a device for creating contrast and emphasis

Contrastivity, says Cruttenden, 'is a term which most intonationists have to use but for which they find it difficult to give any precise definition' (1986: 90). The prosodic marking of a contrastive prominence is usually noticeably stronger than that of a non-contrastive one in the same context. However, this is not necessarily the case; the *position* of a prominence in the sentence may communicate the sense of contrast. I will return to this below. However, to begin with I will discuss contrastivity in relation to the relative prosodic weight with which it is often associated in utterances.

As opposed to the kind of prominence so far discussed, contrastive prominence/accentuation is interpreted in explicit relation to other elements within the context of its utterance. The contrastive marking of an element is intended to draw attention to how much *more* prominent that element is than the rest of the intonation unit. This sense of *more prominent* can be achieved through an expansion in the pitch, loudness or duration (or all) on the prominent element:

1. I didn't want VANILLA / I wanted RASPBERRY

With the extra prosodic weight of RASPBERRY (relative to the non-contrastive prominence on VANILLA), the second unit in 1 conveys a meaning which might be glossed as RASPBERRY AND NOT VANILLA, or RASPBERRY AND NOT ANY OTHER FLAVOUR, rather than simply as RASPBERRY.

The specific meaning of a contrastive prominence needs to be interpreted within the context of utterance:

2. The book is JOHN'S (normal accent on *John*)
3. No, the book is ALBERT'S (contrastive accent on *Albert*)

The two sentences from this fragment of discourse have the same accent placement. However, the use of greater prosodic weight (greater pitch height or greater amplitude, for example) on ALBERT'S than on JOHN'S indicates that something *more*, some kind of additional meaning, is being attached to ALBERT'S, based on its contrast with JOHN'S. By virtue of extra-strong prosodic marking, ALBERT'S carries not only its lexical meaning, but is seen to stand out from (to contrast with) the contextually known set of items to which it belongs. In this case the set includes contextually available names (referring to potential possessors of the book); ALBERT'S is used in contrastive preference to JOHN'S, so that possible glosses of 3 are ALBERT'S NOT JOHN'S and ALBERT'S NOT ANYONE ELSE'S, depending on the context.

The 'set' of items from which a contrastive 'stands out' may be available from the non-linguistic context, where the recognition of heightened prosodic marking entails the listener inferring that set:

4. Did you enjoy your tour of Austria, then ?
5. I liked VIENNA. (The rest of the country was awful.)

Here, the contextually known set against which VIENNA is interpreted (places to visit in Austria) is inferred by the hearer, and indeed such sets may not be 'known' in a recountable sense.

As I noted above, the *position* of prominences within the sentence can also contribute to their interpretation as contrastive. English prominence will tend to fall, other things being equal, on the last lexical item in a unit. In a general sense this is what Chomksy and Halle's Nuclear Stress Rule is said to generate; indeed, the same idea of right-most accent in the unit is central to very many different accounts of accentuation in English, with this default accentuation pattern being said to be the 'normal' accentuation

pattern in English. The notion of a normal pattern has informed many descriptions of prosodic form, and 'contrastive', and also 'expressive' and 'emphatic', accent have sometimes been excluded from discussions of the syntax and semantics of accentuation because they do not reflect such normal patterns. Chomsky, for example, excludes them from his discussion of the semantics of sentential accent, claiming that they potentially cloud the issue with extra kinds of information (1971).[5]

Contrastive accentuation in this sense is seen in terms of deviations from the normal pattern. Stockwell talks about normal, 'baseline' intonation, and states that it is exactly deviation from this prevailing pattern which characterises 'contrastive' intonations (1972: 87). Halliday talks about 'normal' intonation in a similar sense, although whereas for Stockwell the intonational form is determined by syntactic structure, for Halliday it reflects information structure (see section 1.2.1, below). Quirk and Greenbaum (1973) note that contrastivity is the process by which we can 'interfere with normal accentuation to highlight any word' (p. 453), although preferring the term 'marked focus' in later work (see Quirk *et al.* 1985: 399ff). So, contrastive accentuation often involves a prominence in non-final position.

Contrastivity alters the informational focus of the sentence. Either the unpredictability of the accent position, or its unexpectedly heavy prosodic marking (or both) indicate more strongly than otherwise that a particular item is the centre of attention, making it stand out very clearly, and effectively pushing other parts of the sentence further into the informational background.

Ladd (1980) has suggested that rather than there being a phonological system of contrastive accent, the whole spectrum of accent placement is constructed on a continuum. He notes, for example, that within a constituent which reflects the 'broadest' informational focus, that is, where the whole sentence comprises potentially focused information, the accent will fall towards the end of the sentence (a 'neutral' accentual system which conforms to the last-lexical-item norm). In cases where a prominence appears to be contrastive, however, it is also observed that the focused item is maximally 'narrow', in that the prosodically marked element alone occupies the informational

[5] A similar idea can be found in Trager and Smith (1951; 1957), where the highest of the four pitch phonemes is seen as fulfilling a differentiating role, and as such is said to have a restricted distribution, supplanting the more usual secondary stress on syllables which specifically require contrast or emphasis (1957: 52).

foreground or focus of the sentence. This is irrespective of where contrastive prominences fall within the sentence.[6]

Contrastivity is a major facet in a speaker's ability both to understand and to be understood. It serves as a clarification device and also as a means of correcting another speaker's speech:

6. Are they flying the UNION JACK today?
7. The Union FLAG's over there.

Emphatic prominence has also been seen as related to contrastivity. Intensifiers, for example, can carry contrastive-like prominence. In the following example, the prominence is shifted away from its position as marking the core proposition of the sentence in 8, and onto the intensifier in 9:

8. You're good at DANCING
9. You're SO good at dancing!

The proposition *You're good at dancing* becomes *de facto* given information in 9, since the prominence on SO indicates maximally narrow focus, both by virtue of its strong, contrastive-like prominence, and also because its non-final position is typical of narrow focus. So, the 'added meaning' in this case is related to the particularly overt marking of SO to intensify *good*, which is a gradable property. Hence, in 9 the strength of prosodic weight which SO carries will indicate how intensely the truth of the proposition *You're good at dancing* is asserted by the speaker.

The relation between emphatic and contrastive prominence can be seen in the fact that emphatic prominences on intensifiers often leave an implied contrast intact. Indeed, the expression of a gradable property itself implies a comparison with a norm:

10. It's SUCH a beautiful sculpture!

Here, the sculpture's beauty in contrast to an implied set of other beautiful works is still emphasised, in addition to degree of its beauty *per se*.

So, in terms of production contrastivity is an interesting phenomenon. It requires deliberate deviations from the norms of production of the prosodic features, often including notably more pitch movement or loudness than in normal prominences. In addition, contrastivity often entails an alteration in the sentence position of the

[6] Ladd does point out, though, that even the narrowest focus of this kind can yield differing interpretations of focus.

prominence relative to default patterns. Also, contrastivity is related to context in very explicit ways.

Contrastivity is particularly relevant to L2 speech, not only because such a facet may be lacking from production, but because its presence in the form of input from native speakers may not be perceived by learners, hence rendering input incomplete. These characteristics make it an interesting measure of prosodic performance. Contrastive prominence offers a particularly useful window on the intonational form of L2 speakers, both in terms of the bare production of prosodic features, and also as a gauge of their ability to use non-default accentuation patterns. Its analysis can contribute to the assessment of L2 speakers' overall fluency, their sensitivity to context, and their ability to manipulate (to 'interfere' with) the normal production of prosodic features. Also, contrastivity exhibits a very close relation to the context of utterance, which means that tests can be devised in which contrastivity is an obligatory component in an appropriate response.

1.1.4 Information focus and prominence

As we have seen, differing accounts of prominence share the assumption that the phenomenon has to do with indicating which part or parts of the sentence stand out from the rest. In a rather metaphorical sense, talk can be seen as occupying a conceptual 'ground', in which prominent sections are those parts which are brought into focus and given a foregrounded position.

The notion of focus, though, is problematical in its relation to prominence. The 'broad' focus of an all-new sentence may extend through the entire sentence, despite only one part of that sentence being marked prosodically.

Ladd (1980) observes that accent within broad focus seems to be determined in speech by normal stress rules, in which an end-focusing rule operates, leading to the accentuation pattern which has been described as 'normal'[7]:

> Instead of saying that normal stress goes on the rightmost accentable item in the sentence, we say that accent - in general - goes on the rightmost accentable item *of the focus constituent*. If the focus constituent is the whole sentence, we get 'normal stress'; if not, we get a narrower focus on the constituent identified by the placement of the accent.
>
> (1980:77)

[7] Ladd has discussed this issue at length more recently (1996: ch.5 and 6).

On the other hand, Ladd notes, the accentuation of a narrowly focused element relates to the idea of contrastive accent as expounded widely in the literature (1980: 75). So, he proposes an account in which a 'right-most accentable item' rule assigns accentuation in the default (i.e. normal) case, where there is broad focus, but that an information-based criterion of narrow focus takes over in the marked (i.e. contrastive) case. In the case of accent within broad focus, the criteria appear to reflect linguistic structure, in that an 'accentable item' involves a decision as to which is the right-most word with the appropriate word-class (i.e. is potentially accentable); the marking of narrow focus, however, involves strictly informational criteria. For example:

11. I think I'll have a BEER
12. Yes / why don't you have A beer? (but only one!)

Example 11, uttered 'out of the blue', has a broad focus, and thus receives accentuation on the right-most accentable (lexical class) word. 12, however, contains narrow focus on the (grammatical class) article A, and as such the marking conforms to an informational criterion, thus overriding the specification that only lexical class words can carry accent.

Bolinger's 'point of information focus'(1972b), which describes essentially the same phenomenon, is based on a concept of 'information' which has its history in the theory of information (see Hultzen 1956). Again, 'focus' is seen within a focus-presupposition framework, but as the 'focus' of information in terms of the relative weight of different parts of the message. Non-predictability of all kinds (lexical, contextual) contributes to the assignment of the accent on the item carrying most informational weight, leading to the assertion, noted above, that predicting accent would effectively entail knowing a speaker's mind. However, *explaining* accents is possible within this approach. For example, the location of accent in 11 and 12, above, might be said to reflect points of low probability in the utterance - that is, the least predictable elements in the informational structure, and hence those which become prominent.

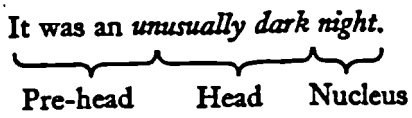
Schmerling (1976) proposes a distinction between 'topic-comment' sentences, and 'news' sentences. She claims that the topic and comment in a sentence are always potentially prominent, but that a tendency for right-most accentuation normally favours the 'comment' element; in 'news' sentences, on the other hand, the predicate is not 'strong', reflecting the relatively higher informational load on the actor (in Bolinger's terms, it has relatively low probability in the context), and thus prominence is retained on the 'comment' element.

Schmerling's distinction reflects a general claim in information-based accounts that intonational prominence falls on new information. Halliday also talks about recoverable and non-recoverable information (1967b), and indeed 'new' and 'given' are commonly used to refer to information which is focused and unfocused within the information structure. However, although 'new' and 'given' are clearly important terms within a discussion of information focus and prominence, I will review their use in the following part of this chapter, since the new/given opposition is also important in accounts of the tone system of English intonation. Nevertheless, we might bear in mind that, in the present chapter, all discussions of separate aspects of intonational form are to some extent related, with the potential for significant overlap, both in terms of linguistic function, and in terms of the phonetic characteristics of prosodic form.

1.1.5 The British tradition of the intonation nucleus

In the above review of prominence it was noted that a distinction between pitch and stress was made by many American analysts. The Trager-Smith pitch level system, for example, has exerted a strong influence on subsequent descriptions of pitch phenomena by American linguists (Liberman 1978 on intonation; Leben 1976 and Goldsmith 1976, from an auto-segmental perspective, for example).

The 'British tradition' of intonation analysis, on the contrary, has been framed largely in terms of the 'nuclear' model. The phenomenon of prominence is in many ways equivalent to what British analysts have often called *nuclear stress*, or simply *nucleus*:



(O'Connor and Arnold 1973: 18)

Figure 1.4 The constituent parts of the tone group in British nuclear analysis

In the *Primer*, Sweet (1901) observes that the principal 'synthetic' features, force and pitch and quantity, operate in 'intimate connection' (pp. 71-2). Sweet anticipates the widespread practice in British analysis of seeing the general phenomenon of prominence

as intrinsically part of the wider intonational structure, in which a combination of pitch and loudness makes up those parts of contours which stand out most prominently.

In contrast to American 'levels' analysis, British work has most commonly undertaken to describe whole intonation contours. This tradition sees 'curves' (Jones 1909) or 'tunes' (Armstrong and Ward 1931) as the foundation of an intonational structure, in which the nucleus (or tonic) carries the principal intonational movement. The nucleus will fall on a stressed syllable (typically, the last stressed lexical item in the intonation phrase), and the issue of loudness-as-stress is subsumed *a priori* into a statement about intonational structure in which pitch and loudness are both seen to be present.

British intonational models are often concerned at a descriptive level with various aspects of 'meaning'. Indeed, this was the main criticism of them made by American structuralists (Smith 1955, for example). In comparison to the accounts of suprasegmental phonology of Pike, Trager and Smith, and Chomsky and Halle, the work of Armstrong and Ward, O'Connor and Arnold, and Tench, looks positively holistic. Moreover, British analyses have often considered the relationship between intonation and emotion and other systems of 'meaning'.

British phoneticians have traditionally been more directly interested in language teaching than their American counterparts (Pike is an exception). Also, one result of taking an inclusive approach to intonational form, in which all 'systems' are incorporated into a single intonational fabric or architecture, is its availability as an accessible pedagogical description. Nevertheless, relatively little has been said about the specific system of prominence (nucleus) within an EFL context (see section 1.4.1).

1.1.6 Summary of different accounts of prominence

So far in this section I have reviewed some of the ways in which the phenomenon of prominence has been described in the literature. Indeed, there is an array of frameworks which can explain prominence within simple sentences. Brown *et al.* (1980) provide a summary of the shared claims of 'many scholars' in this field:

> ...the structure of information within the tone group will be such that given information precedes new information and that the focus of new information, hence the tonic, will fall on the last lexical item.
>
> (p.155)

In section 1.1.1, the simple sentence *Peter baked a cake* was used to illustrate how applications of the Nuclear Stress Rule lead to the prediction of sentence stress on the word *cake*. The various other paradigms for the description of 'prominence' discussed above can all be seen to explain this accentuation pattern:

Peter baked	a CAKE
given	NEW
background	FOREGROUND
high predictability	LOW PREDICTABILITY
(pre-head) head	NUCLEUS (tail)

In a very general sense, we might also characterise the function of prominence as marking the presupposition/FOCUS distinction, as well as the theme/RHEME paradigm (although see section 1.2.1).

1.2 The informational function of intonation

Discussions of prominence and of tone overlap considerably. In the present section I will continue the discussion of prominence, looking at how it reflects the structure of information, and dealing in some detail with the work of Halliday. After this I will discuss the concept of the intonation 'unit', in terms of its role as the phonological domain of prominence, and also how information is organised within these units. Finally I will look at the informativeness of tones, concentrating on the fall/rise distinction, in terms of its contribution to meaning and its uses in discourse.

1.2.1 Halliday's system of tonicity and its relation to information structure

Halliday's three systems of English intonation are: 'tonicity' (placement of prominence, which he calls the 'tonic'), 'tonality' (division of speech into intonation units, called 'tone groups'), and 'tone' (the kinetic movement of pitch on tonics) (see Halliday 1967b). Within the tone group, the system of 'tonicity' marks one or two points of informational focus, which Halliday refers to as 'points of prominence within the message' (1967a: 203). These tonic prominences are the intonational centres of the groups and carry the principle pitch movements (the tones) of the groups.

As I noted in section 1.1.3, Halliday claims that there is an unmarked, 'normal' accentuation pattern in English ('neutral tonicity'). In cases of neutral tonicity, the information presented by the speaker as prominent (through tonic placement) corresponds largely to the notion of prominence as described above, and 'falls on the last lexical item in the tone group' (1967b: 22).

For Halliday, '[W]hat is focal is 'new' information' (1967a: 204). The tonic position carries rhematic, or 'new' information in the default case, and that which falls outside the tonic domain is thematic, or 'given'. However, he also notes that the tonic will only fall *within* the rheme, 'not necessarily extending over the whole of it' (1967a: 205). I have mentioned this previously, in section 1.1.4, in relation to broad focus and prominence.

Halliday's concepts of new and given bear similarities to many of the other paradigms discussed in the previous section, most notably Bolinger's notion of 'point of information focus'. It also reflects the general Prague School claim that the end of a

linguistic unit, be it a clause, a sentence or a phonological unit, bears more informational weight than the beginning (see below, on Communicative Dynamism). As I noted in the previous section, whatever differences there are in different accounts of prominence, then, they do nevertheless predict similar things.

One of the methodological assumptions in Halliday's work is that default tonicity is naturally produced by speakers when sentences lack a context. This is the case, for example, when we read out a sentence 'cold'. Under such conditions neutral tonicity is employed, and the tonic will naturally tend to fall on the last lexical item.

However, tonicity within the Hallidayan approach is framed as a system of anaphora. Thus, material which occupies default tonic position (right-most) loses tonic marking in predictable ways in appropriate contexts:

13a. What's going on?
13b. Bill's made the DINNER!

13c. What about dinner?
13d. Bill's already MADE {it, dinner}

13e. Who made dinner?
13f. BILL {did, made it, made dinner}

Information in post-tonic position, as in examples 13d and 13f, conforms to Halliday and Hasan's (1976) rules of contextual recoverability, and is deaccented on the phonological level. In other words, it becomes given information, a phenomenon which has been discussed by various writers (Bierwisch 1968; Chafe 1974, 1976; Schmerling 1976), and which has also been termed 'predictable', 'repeated' and 'old' information. According to Halliday, focal information need not be new in a linguistic sense, but may instead be 'a feature of mood' (1967a: 204):

14a. I think he's done WELL (assertion)
14b Yes / he's done WELL (confirmation of assertion by another speaker)

In Halliday's given/new distinction, prosodic structure assists the listener in identifying that information which can be recovered from the broad context of utterance, and that which is newly introduced material, claiming that 'in English, information structure is expressed by intonation' (1970: 162).

Halliday's account of the role of tonicity within the 'message' of speech reflects Prague School notions of information. The way in which he discusses such notions as 'points of prominence', 'the message' and 'new' and 'given', may prove useful in the

present study. My aim here is to examine the acquisition of L2 intonational form, and this will involve the design of simple but reliable tests to assess speakers' ability to mark information status. Perhaps the adoption of rather intuitive notions of information will be a convenient way of framing experiments which are both rooted in the tradition of theoretical accounts of intonation, but which also appeal to an intuitive sense of information.

Applied linguistic research often involves a compromise between theory and potential applicability. This is evident here, where I want to assess L2 intonation, but also to make sure that my findings are of some potential use within the field of language pedagogy. As I will note in 1.4, below, there are currently very few L2 pedagogic texts which draw on and encourage learners' sensitivity to broad, contextual systems of information, despite its obvious attractiveness as a means of framing the difficult process of teaching L2 intonation.

1.2.2 Given and new information and prominence

Halliday's contextual interpretation of the informativeness of the tonic means that the prosodic structure of utterances can be understood in ways which incorporate diverse aspects of communication. This is a great advantage over syntax-based accounts of prominence, which are unable to draw on context as a possible factor in an explanatory account of the phenomenon (see Ladd 1996). In Chapter Three I will describe how the very broadest notion of context must be borne in mind when, rather than seeking an explanation of the system of prominence, one is attempting to elicit specific uses of it in the utterances of informants.

A central concern within the work of the Prague School has been to understand how elements in a spoken or written text are said to 'differ in the extent to which they contribute towards the further development of the communication' according to their degree of Communicative Dynamism (CD) (Firbas 1992: 7). The degree of CD is seen as the extent to which an element contributes to the dynamic process of advancing information. Thematic elements have a relatively low CD, since their presence simply iterates contextually known information, whereas rhematic elements typically advance the state of information by adding to what is known.

CD in the work of the Prague School tends to be seen as a continuum, and points of transition are identified in addition to the theme and rheme sides of the paradigm. Halliday's given/new distinction, on the other hand, cannot express such a variable; information is given or new on a binary scale (1967a). Although Halliday's 'given' category is enriched through a very inclusive notion of recoverability (anaphorically or situationally), there is still no talk of degrees of givenness. This may constitute an advantage in the present study. That is, we can use the straightforward new/given opposition as a means of making very simple distinctions in (elicited) speech data. That is, rather than concern ourselves with degrees of CD, we can simply elicit speech in which one item is very clearly new within an utterance which is otherwise composed of given or thematic information. Such utterances will be very explicit examples of what Halliday and Hasan (1976) call 'THEME and a remainder' (p. 325).

Chafe (1974) doubts the validity of the notion of contextual recoverability, preferring the idea that given information emerges from consciousness rather than directly from context. He proposes that during spoken communication, information is retrieved from discrete levels of memory. In particular, he claims that the 'gamut' of degrees of CD (the extent to which information is understood to be already available to the interlocutor) is not in fact observed in languages. He makes reference to Bolinger (1972b) in claiming a broader definition of given/new based not on contextual recoverability, but on general recovery from memory. The mention of Bolinger's article is apposite at this point, since it once again underlines the fact that the discussion of information status briefly reviewed above is also by implication a discussion of prominence; the two (separate) phenomena coincide necessarily from the prosodic perspective. Both those looking at the psychological basis of given/new, and those looking at prominence ('tonic', 'accent' etc.) from a linguistic perspective (be it in a semantic or a syntactic framework), are talking about the same phenomenon.

Chafe also notes that the item of new information within a unit is marked by higher levels of pitch and amplitude. He is interested not only in the status and marking of information, but in the ways in which this is reflected in collaborative discourse, as speakers assess the probable reception of information by the hearer based on assumptions of shared information.[1]

[1] Far more detailed discussions of these issues have evolved in the areas of Relevance Theory (Sperber and Wilson 1986) and Representation Theory (Chapman 1994).

Information-based accounts of tonicity are not directly concerned with grammatical structure. Yet the debate over the relation between grammar and intonation is not always easy to deny. Crystal (1969) accepts that the tonic syllable will favour the informationally strong element(s) within the intonation group, and cites the findings of Quirk *et al.* (1964) extensively, which show significant parallels between the use of intonational forms (such as units) and grammatical components in language (such as clauses). Yet he also retains the idea of a grammatical function for tonic, in which the informational and grammatical systems are simultaneously marked in the prosodic structure.

However, the idea of a grammatical function has been denied by those for whom 'information' and 'intonation' are closely related terms. Tench (1990), for example, has suggested that whereas the large-scale, corpus-based findings of Quirk *et al.* do indeed show high correlations between constituent structure and tonic placement, the explanation lies in the encoding of information structure in the grammar. Because informationally rich constituents (lexical categories are usually rich) occupy predictable positions in the linear sequence, then so too does the tonic. Since there is a high probability in English of the object of a simple sentence following both the subject and the verb, and if the object has a high informational weight, then tonic will fall on the last lexical item in the sentence. According to those who oppose a syntactic interpretation of intonation, such 'encounters between intonation and grammar are casual, not causal' (Bolinger 1958c: 37).

1.2.3 Intonation units and their role in the structuring of information

Halliday's system of tonality describes how information is 'chunked' into phonological units (tone groups). Each tone group, he claims, carries one 'information unit' (1967a: 8). The idea of a phonological unit of this kind is adopted in very many descriptions of English intonation, and these units have been referred to as 'tone units' (Quirk *et al.* 1964; Crystal 1969; Brazil 1975), 'pause-defined units' (Brown *et al.* 1980), 'intonation units' (Chafe 1976) and, in the British tradition of intonation analysis, 'breath groups' or 'thought groups' (Jones 1956) and 'word groups' (O'Connor and Arnold 1973).[2]

[2] This list is by no means exhaustive.

These different terms do not all necessarily describe exactly the same units, which might differ in the specification of defining (phonetic) criteria, but they do all recognise that there is a phonological unit inside which structured intonation is built-up. For example, Tench notes, during the analysis of an extended piece of spontaneously occurring conversation, that:

> 'Information' may have to be interpreted fairly broadly but whether the speaker is reporting, expressing an opinion, making a comment, qualifying or elaborating on a previous assertion or whatever, he is deploying intonation in blocks that suit his purpose.
>
> Tench (1990: 153)

The division of an utterance into 'chunks' is to some extent under the control of the speaker, and a lexically identical sentence can be chunked in a variety of ways in utterance:

15a. / The way to a man's heart is through his STOMACH /
15b. / The way to a man's HEART / is through his STOMACH /
15c. / The WAY / to a man's HEART / is through his STOMACH / (etc.)

Given Halliday's claim of one informational unit per tone group, then the way in which a speaker chooses to divide his speech up will have important consequences on the effect of what he says. A larger number of (shorter) groups will result in the perception of there being more information, and this may be used as a rhetorical device, or simply as a means of increasing the overall force or perceived informativeness of what is being said:

16. WATCH / MY / LIPS / ... / NO / MORE / TAXES

Bolinger (1986) has made a similar observation regarding accent. He notes that accent can be used cumulatively, in that more and more accents can be loaded onto a sentence, creating an increasingly forceful effect. Such productions are typical of highly rhetorical or dramatic speech:

17a. What on earth is THAT!
17b. What on EARTH is THAT!
17c. WHAT on EARTH is THAT! etc.

This phenomenon is related to the size of intonation units in an interesting way, which illustrates the very inter-connected nature of intonational form here. Whereas Bolinger talks about increases in the number of accents, many approaches to intonation (but not

Bolinger's) allow for only one accent per unit. Therefore, the rhetorical use of greater numbers of units in example 16, above, in that it necessarily involves a concomitant increase in the number of accents, is in many ways analogous to the process of cumulative accentuation, as in example 17. Interestingly, both Hewings (1993) and Grosser (1993) have found a similar phenomenon in second language production, although one suspects that the motivation behind such cumulative accentuation in first and second language speakers is somewhat different. I will discuss these two studies in Chapter Two.

So, intonation units reflect speaker decisions as to the status of information uttered. As such, the nature of a speaker's unit production, in particular the length of units, may be a good indicator of his assessment of the information active within the discourse. Yet as an element in the measurement of intonational form within an SLA framework, intonation units do not in themselves constitute an unproblematic area.

1.2.4 Some problems with the prosodic marking of information

The supra-segmental level is multi-functional, yet has at its disposal a very limited set of prosodic features with which to mark a variety of systems. Sweet's analytic/synthetic distinction captures something of this. The three 'general factors of synthesis' in the structure of language are pitch, stress and quantity (Henderson 1971: 43), and are qualitatively different in their contribution to language than those elements which form the basis of the 'analytic' aspects of structure, such as units of grammatical form and discrete, individual sounds. The 'synthetic' nature of prosodic form, then, is inherently a system of competing speaker needs and limited resources.

'Synthetic' processes work to bind together discrete elements of structure into a meaningful whole, but as a consequence of this are more difficult to identify and describe in an adequate way. Since intonation falls squarely within the 'synthetic' area, it is difficult to assert the precise phonetic characteristics of any particular aspect of intonational form. For example, how do we account for those 'intonation units' which conform to a set of phonetic criteria but which intuitively we do not perceive as being 'units' in any satisfying sense? This is especially the case in spontaneous speech, and indeed, a distinction has been made between the intonation units as discussed above, and the elaborative and back-channelling devices (the 'softeners' and 'connectives')

commonly used in speech (see Crystal and Davy 1975). In addition, disfluencies, repairs and filled pauses may occupy their own intonation units, but these do not carry information in any substantive sense, although they perhaps can be said to contain information of some sort. Indeed, Crystal (1969) describes the division of utterances into intonation units as leaving a possible 'remainder' of interruptions, restarts and other utterance types which simply do not fit into the packaging model (p. 205). Nevertheless, such elements in speech must be dealt with in a principled way when speech data are assessed, particularly in L2 speech, where they are especially prevalent.

Another problem is in the widely held belief that informational prominence (focus, new) is marked prosodically by a predetermined set of prosodic devices. For example, higher pitch is often said to mark the new information of a unit. However, Yule (1980b) describes how *given* information can carry higher pitch than expected, noting that various interactional and topic-organisation systems use prosodic systems simultaneously with information-marking, and that these effectively distort any simple correlation between pitch and information status.

On the phonetic level, the direct relation between new information and high pitch within an intonational unit seems in fact to be a conflation of two separate systems. First, information which is new, or some element within a new expression which conveys the new information, is uttered as prominent by the speaker. In most varieties of English, prominence marking includes greater pitch height and increased loudness. The two systems - information marking and prominence - are separate, however. Evidence from non-standard varieties of English seems to support such a distinction; the prosodic marking of prominence in various dialects of English, such as Glaswegian and Belfast English, may be realised through significant drops in pitch (see Brown *et al.* 1980).

1.2.5 The role of pitch movement (tones) in marking information

There has long been interest in the contribution which specific pitch movements make to discourse. American analyses have traditionally looked at sentence-juncture phenomena. For example, the three terminal junctures of rising, level and falling tones[3] have been

[3] The normal transcription of these tones is: rising (//), level (/) and falling (#).

seen as marking the flow of speech in characteristic ways, and these remain in widespread use as transcription conventions (see Alvarez-Caacamo 1996, for example).

Something similar has been commonplace in the British 'nuclear' school of intonation analysis, although since the tonic syllable and the nuclear pitch contour are seen jointly to constitute nuclearity, no separation of prominence and juncture is maintained.

Although approaches differ, however, a general assumption is that a pitch movement carries 'meaning' in some sense. The widely reported sense of 'terminality' which a falling tone carries, not only in English but in many languages, seems to point to a universal basis for the meanings of pitch movements (see Bolinger 1978; 1986; 1989).

Within the study of English, interest in the contribution which different pitch movements make to the linguistic whole goes back at least as far as the seventeenth century; the observation that fall and rise function as a systematic part of the structure of English was noted by Butler in his *The English Grammar* of 1634:

> Tone is the natural and ordinary tune or tenor of the voice: which is to rise, or fall, as the Primary points shall require: and therefore it denominates the voice, High or low.
>
> (see Eichler 1910: 54)

Discourse-oriented models of intonation tend to make use of generalised categories of 'meaning' which can be attached to contrasting pitch movements, in particular the tones fall and rise. Halliday (1967b) draws attention to the different status of information which pitch structure can encode. For example, a falling tone is described as indicating 'major information' in the default case. He also describes certain informational dependencies between the tones of adjacent tone groups. After a falling tone, a low rise in the following tone group is said to carry 'minor information', which might be 'confirmatory' or 'dependent' (p. 27).

The way in which tones can be used to help create an informational structure can be seen in certain conventionalised intonation patterns in English. In listing, for example, there are three normal intonations patterns:

// R R R F // // F F R F // // F F F F //

(Couper-Kuhlen 1986: 150)

Of these, the first two patterns are the most common in English on lists which are informationally cohesive. O'Connor and Arnold discuss this within a description of their 'Low Drop' tone group:

> For example, if we say:
>
> He was \ tall, | dark | and \ handsome. ||
>
> we are treating each of these three attributes as being a complete and separately interesting feature of the man; but if we say
>
> He was / tall, | / dark | and handsome. ||
>
> we are linking the three together into a single, composite picture. In the same way the final item in a list which is taken to be complete is normally said with a Low drop, the other items in the list having a tone group with a rising nuclear tone.
>
> (O'Connor and Arnold 1973: 47)

Where the penultimate item takes a rise (or any non-fall), it draws attention to the dependency between itself and the final item, which is always a fall (terminates the list as a whole). Indeed, in the // R R R F // listing pattern, this notion of dependency is extended to all but the final tone, where the string of 'open' (non-fall) tones maintains the listeners' awareness of the fact that each item is part of a larger, continuing series, which is then terminated with a fall.

The contribution which the rise and fall make to listing patterns is part of a more general feature of their function as tones. Armstrong and Ward (1931) noted that incompleteness and completeness are typically communicated by rise and fall respectively, citing incompleteness as one of four 'meanings' of their rising Tone II (p.20). Indeed, this has been a common observation (Pike 1945; O'Connor and Arnold 1973; Gimson 1980).

However, the relationship between tones and proposed meanings is not straightforward. Tench notes that completeness is not co-terminous with fall:

> 18a. We saw the Prime Minister \YESTERDAY
> (*yesterday* is major, complete information)
>
> 18b. We saw the Prime minister /YESTERDAY
> (*yesterday* is major information, but rise indicates incompleteness - more to come...)
>
> 18c. We saw the Prime \MINISTER (unit break) /YESTERDAY
> (*yesterday* minor information but probably completes the proposition)
>
> (adapted from Tench 1990: 224-5)

Tench's example illustrates how the 'meanings' of tones are at best idealisations, and run the risk of distortion on many levels as the discourse is enriched with contextual information.

Halliday also points out that distinctions based on a simple association of pitch movement with a 'meaning' can be misleading. For example, a certain amount of prominence might be given to a piece of 'given' information, such as the low rise following a major information-marking fall, as a means of marking that given item with specific relevance:

19. \No | we don't often \GO to the ,theatre

(adapted from Halliday 1970: 44)

In the above example Halliday claims that, through adding a low rise on *theatre* (in a context where it is given information), the speaker wishes to convey that although 'this has already been referred to...it's a significant part of the message.' (1970: 44)

Tonal dependencies, then, can enrich the structure of information in speech. In particular, they bind units together, drawing attention to the flow of information, and marking the relationship of successive ideas in specific ways. In the field of Second Language Acquisition (SLA) such tonal dependencies have been used as a measure of developing L2 intonational form. Grosser (1993) notes that in two units which are syntactically co-ordinated, '...a sequence of "rising + falling" signals greater cohesion than a sequence "falling + falling"' (pp. 86-7). He reports that, in his data on L2 English acquisition, the suppliance of an appropriate rising tone in the first unit of such two-unit sequences increased as acquisition developed. In my own analysis I will also use measures of the appropriateness (nativelikeness) of tone sequences as a means of assessing acquisition of intonation.

1.2.6 Cruttenden's meanings of falling and rising tones

The fall and rise distinction has been associated with a variety of grammatical, semantic, pragmatic and discoursal systems. In a review of earlier work, Cruttenden (1981) concludes that the two basic movements represent a system of a higher level of abstraction than is usually claimed. This system, reflecting a broad 'closed' versus 'open' opposition, encapsulates abstract meaning of an order sufficiently high so that

grammatical, semantic and other distinctions can all be subsumed. There are, consequently, a variety of possible meanings associated with each tone:

Fall (closed)	Rise (open)
reinforcing	limiting
statement	question
finality	continuity
closed-listing	open-listing
conducive	non-conducive
statement	statement with reservation

(p. 195)

reinforcing	limiting
non-linking	linking
indubitative	dubitative
volitive	non-volitive
assertive	concessive
affirmative	detractive
dogmatic	conciliatory
unrelatedness	relatedness
informative	argumentative

(p. 333)

The polar meanings of the fall and rise pair is widely noted in accounts of tones in English:

20a. She doesn't lend her car to \ ANYone (definite statement)
20b. She doesn't lend her car to / ANYone? (querying the statement)

(Clark and Yallop 1995: 359)

Cruttenden claims that, although abstract in nature, such paired terms are no less precise than many others used to describe the characteristics of tones in the literature. He also goes some way to offering an alternative interpretation of the largely polarised debate over whether intonation is essentially a syntax or semantics based system:

> ...conflicts between discoursal and attitudinal meanings are not easily found, precisely because a particular use of a tune may be regarded as either discoursal or attitudinal depending on one's view-point. What remains clear, however, concerning the potential cases of conflict, is that grammatical considerations are overruled by lexical or attitudinal or discoursal considerations.
>
> (p. 199)

Cruttenden's list of paired terms can be used to make useful new distinctions between intonational 'meanings'. For example, the traditional generalisation equating fall with statement and rise with questions, which has in the past been used as the basis for the claim that intonation reflects the grammatical question form, can be refined using the

notion of *conduciveness*. There are two commonly held assumptions about question intonation: one is that questions carry rises; the other that polar-questions rise whereas wh-questions fall. Both these are disputable (see Svartvik and Quirk 1980) Whereas wh-questions do most commonly carry a falling tone in Svartvik and Quirk's data, polar questions are more varied, and 'meanings' such as *conducive* and *non-conducive* can help explain this variation. Polar questions in particular can be either conducive or non-conducive. In a highly conducive polar question the speaker has a high expectation that the proposition underlying his question is correct:

21. Would you like a drink? (as in *Have a drink!* uttered as you hand someone a glass of wine)

In example 21, the question is in fact similar to a statement in that its illocution is relatively assertive/statement-like, and the expected answer is largely an acknowledgement of the correctness of the proposition.[4] Polar questions may also be requests for verification, and these are also highly conducive. In both cases a falling tone is likely. On the other hand, polar questions may be relatively non-conducive, and in such a case may carry a rise:

22. Would you like a drink? (uttered speculatively, with a gesture in the direction of the drinks cabinet)

I will note below (section 1.4.2) that these insights have so far not been incorporated into pedagogical texts, where less adequate assumptions are still often preferred.

[4] Brown *et al.* (1980) use the notion of conduciveness. See p 178, for example.

1.3 Information staging

1.3.1 Phonological structure above the level of the intonation unit: the phonological paragraph or paratone

The question of how intonation units form dependencies, dealt with at length by Halliday, is also discussed by Crystal (1969; 1975). However, until quite recently there has been no detailed description of how larger units of discourse may be reflected in the behaviour of pitch. For Halliday (1967b), the existence of dependencies between the tones of different units does not alter the fact that the unit (tone group) itself represents the highest level in the phonological hierarchy. Crystal, on the other hand, proposes a system of tonal subordination, where the pitch contour of a superordinate intonation unit is repeated in an adjacent (subordinate) unit, but with a narrower (i.e. phonetically subordinate) pitch band width. Far earlier, Palmer (1922b) had described two-unit sequences, noting that a sequence of two like-tones is often found in the utterance of two co-ordinated clauses, whereas a sequence of two different tones often marks grammatical subordination (pp. 87-95). Yet, as Fox has pointed out (1973), few writers had, by the early 1970s, been willing to posit a phonological unit above that of the intonation unit.

The first suggestion that such a system might exist was made by Trim (1959). He recognises the dependent or non-final nature of certain types of 'tone groups', which demand a resolution in a subsequent (and, hence, connected) unit. He claims that there is clear phonetic evidence for a phonological level higher than the intonation unit. Firstly, in putative groupings of units, tonal sandhi brings the pitch of a unit's tail and the following unit's head closer than would otherwise be the case. In other words, pitch is used to emphasise cohesion across unit boundaries. The relative downward stepping in pitch of heads in succeeding units occurs to signal inter-unit cohesion, and contrasts with a greater rise in pitch to the head of the first unit of a new supra-unit. In a move away from the then predominant practice of seeing direct parallels between grammatical units and intonation units, for example, in the 'standard' work of the period, Schubinger's *English Intonation* (1958), Trim talks of 'sense groups', and has repeatedly been cited as the originator of work in this area.

Fox (1973) subsequently introduced the term 'paratone', and also provided early discussion of the possible characteristics of these proposed 'phonological paragraphs' (p.18). He suggests that it might be necessary to establish a system of 'paratonicity', the

paratonic equivalent of Halliday's 'tonicity'. He also tentatively advances the idea that whereas in Halliday's systemic description of English intonation the grammatical clause is comparable to the phonological tone-group, and the tone-group itself broadly comparable to a simple unit of information, a sentence might equate to a paratone, thus to a complex information unit.[1] As will be shown below, Fox's terminology has been retained, as has his suggestion that paratones relate to information structure. His suggestion of a sentence-paratone has not persisted, however, due to the development of discourse models which provide alternative frameworks.

Recent work in the area of supra-unit phonological structure has been conducted largely within a discoursal approach. Writers, most notably Brown (1977) and Yule (1980a, 1980b), have proposed that within the spoken text a 'phonological paragraph', or paratone, serves to structure the flow of the topic in ways analogous to the typographic conventions of the written paragraph. However, difficulties in specifying the exact nature of the phenomenon of 'topic' have been discussed widely in discourse analysis (see Brown and Yule 1983, Ch. 3). Indeed, we can see immediately that there is great potential here for discussions of the informational content of phonological paragraphs to be crowded with rather inexact terminology.

Brown and Yule have argued that instead of trying to define precisely the nature of 'topic' in discourse through exhaustive, proof-led analysis, we should concentrate on the exact nature of the 'packaging' of speech at this level of utterance. We intuitively recognise topic shifts, in that we recognise that an utterance will not always begin and end on the same 'topic' (again, used as an intuitive category). If two successive topics occur within one stream of utterance, then we naturally perceive that a topic shift has taken place. By locating this point of shift and noting how such shifts are marked, we can build up a picture of how a topic packaging system is created in speech. Although such analyses take into consideration lexical information (such as the repetition of a topic expression as an indication of the end of that topic), the central criteria for the marking of this kind of structure are seen to be prosodic.

[1] This he does in the spirit of Halliday's 'neutral case' approach.

1.3.2 Discoursal and phonetic properties of the paratone

Lehiste provided some early phonetic analysis to substantiate the claims of Fox and others in the 1970s. It was observed that in listener judgements the high pitch of a (whole) intonation unit was associated with paragraph-initiality (Lehiste 1975). The maintenance of high pitch throughout a unit, then, seemed to be perceived as denoting the establishment of a new topic. This high starting point for a paratone is interpreted by Lehiste as an articulatory planning device which seeks to accommodate the pitch declination which was observed to occur throughout the whole paratone, avoiding the need for both pitch re-setting and laryngealisation; resetting would involve a mid-paratone jump up in pitch between a tail and the following head above the line of downward stepping observed between the units of a paratone (see Trim 1959); arriving at a pitch low enough to result in laryngealisation before the tail of the final unit of a paratone is also undesirable, not least because it brings with it unwanted suggestions of finality (as well as being unsustainable). Lehiste (1979; 1980) also found that an intonational unit within a paratone tends to be uttered at a faster rate than the same unit spoken in isolation, and that the pauses preceding a paratone are significantly longer than either pauses which fall on a sentence boundary or pauses between units within a paratone. Finally, it was found that lengthening of the final stressed word of the last unit in a paratone is significantly longer than in cases of paratone-internal lengthening.

Within discourse-based accounts, the paratone is seen as the spoken realisation of a discourse paragraph. The discourse paragraph (as a constituent of a text) is explicitly distinguished from the orthographic paragraph. The latter is characterised by formal conventions and presentational enhancements (see Longacre 1979). The 'transcription' of either a spoken or written text paragraph excludes any such conventions (see Brown and Yule 1983: 97 for an example). Such a discourse oriented distinction has not always been made in the early literature on paratones, and perhaps this explains why some claims, such as Fox's suggestion of a sentence-paratone parallel, are unsatisfying. In this sense Chafe's suggestion that final extra-long pause is the spoken-text equivalent of the orthographic paragraph break (Chafe 1979) should perhaps also be seen only as an attempt to describe the superficial similarities between these two forms.

A discourse paragraph can be described as a block of language which maintains topic cohesion. Its formal structure is perhaps best identified with only superficial reference to discourse content - the term 'topic' is itself a fuzzy, intuitive category in

most approaches. However, this need not be an impediment to the interpretation of language data. Indeed, the identification of topic structure often constitutes a *post hoc* process in discourse analysis, being no more than a commentary on the content of the discourse paragraph. In written text, topic structure is usually marked by lexico-temporal devices, such as adverbial structure or macro-structural 'connectives' such as *furthermore, but, however* and *so* (see van Dijk 1977). In spoken paragraphs/paratones, the prosodic structure assumes some of these framing functions. There is an inevitable circularity in this methodology.

In a seminal article, Yule (1980a) attempts to clarify 'the vagueness which often accompanies discussions of 'topic'' (p. 33). He makes a distinction between the general 'discourse topic' - the general theme of the conversation - and the particular contribution to that general theme which a speaker makes. These contributing sections are referred to as 'speaker topics'. One of their identifying features is the prevalence of a lexicalised expression - the 'speaker topic expression' - which in spontaneous speech appears not only to introduce each speaker topic (usually as the opening unit), but also carries a high probability of being repeated towards the end of the speaker topic. A 'speaker topic' is most usually associated with a complete paratone. Yule notes that in this way the speaker can be seen as contributing discourse which conforms to 'the accumulation of linguistic contributions, together with their semantic and pragmatic presuppositions, from the preceding discourse' (Yule 1980a: 34). And Chafe (1976) also favours the idea of a sub-divided discourse topic as a domain or a 'spatial, temporal or individual framework within which the main predication holds' (p. 50).

A 'speaker topic', marked as one (major) paratone, can be shared between several speakers. Indeed, one of the strongest claims about the psychological reality of the paratone is the way in which its phonological shape can be maintained in collaborative discourse, where speakers are sensitive to the prosodic marking of a topic already in progress (Yule 1980a).

Brown *et al.* (1980) provide an extended analysis of the intonation of English (Edinburgh dialect). Having complained that the intonation unit, as expounded by Halliday and others, is simply not evident in their speech data, they establish pause-defined units (a workable equivalent), and find that groups of these units are indeed observed in their data to exhibit characteristics which appear to confirm the presence of paratonic structure. They take Yule's (1980a) mixed criteria for identifying paratone boundaries: extra-high pitch for major paratone onset (the pitch on the first stressed

syllable of the unit), often reaching the extreme high-point of the speaker's pitch range; a tendency for the baseline to be raised throughout the first intonation unit(s); end of paratone optionally (but in the majority of cases) marked by very low pitch, and low amplitude; long pause following a paratone. They also found that the 'speaker topic' was indeed often repeated towards the end of the paratone (especially in spontaneous speech). In Brown and Yule (1983) the authors claim that the whole of the 'introductory expression may be uttered with raised pitch', explaining that this extended phonological prominence (it might extend over the whole first clause or sentence) is used 'to announce what he [the speaker] specifically wants to talk about' (p. 101). They also point out that the phenomena which mark the conclusion of a paratone have much in common with turn-final marking in conversation.[2]

In addition to Yule's influential criteria for the identification of paratones, Couper-Kuhlen (1986) notes that the long pause following the end of a paratone might be filled.

1.3.3 'Minor' paratones

A 'major paratone', described above simply as 'paratone', can be considered as the phonological domain onto which a 'speaker topic' is mapped out, which itself has been described as a unit of discourse which contributes to the broader topic under general development in the discourse.

Yule (1980a) also claims the presence of 'minor paratones' in speech, although he gives only a very brief description of these. Essentially, groups of intonation units combine at a level beneath that of the (major) paratone and above that of single intonation unit. Brown *et al.* (1980) identify minor paratones as being particularly frequent in extended turns of spontaneous speech, lacking the significantly longer pauses of major paratones and having less extreme initial pitch peaks.

According to Brazil *et al.* 'in addition to everything else [prosodically] he does, the speaker marks the boundaries or 'chunks' of his utterance every so often' with the

[2] The intonational marking of paratonic structure is at best an abstract notion. In spontaneous speech, as has already been noted, prosodic features are utilized simultaneously by different systems of communication. Indeed, Yule notes (1980a:37) that the paratonic structuring in his spoken data is the 'least potent of the competing systems' and tends to disappear under pressure from contrast-marking, negation focus and given/new marking.

use of a low pitch tonic (1980: 61). This appears to correspond roughly to Yule's category of minor paratone, a number of which may exist within a major paratone. Couper-Kuhlen (1986) has developed Yule's specification of the role of minor paratones, utilising Brazil's (1975) notion of key.

Pitch at the end of a single intonation unit usually ends low in range. This entails a resetting of pitch to a higher level at the onset of a following unit. Couper-Kuhlen proposes that in English this jump in pitch to the onset can be to one of three levels. First, it can be to a very high onset, as would be the case if the new minor paratone happens to be the first in a new *major* paratone. However, within minor paratones, jumps to onset can be either medium or low:

> jump up to high onset - new topical orientation (new paratone)
> jump up to medium onset - (paratactic) additions or extensions to what precedes (i.e. the topic)
> jump up to low onset - (hypotactic) subordinations to what precedes
>
> (Couper-Kuhlen 1986: 193)

She provides an analysis of a transcription of one major paratone (p. 193-4). In Figure 1.5 I have reproduced this transcription, and annotated it with her interpretations of how pitch reflects paratonic structure, including the three types of jump to onset listed above:

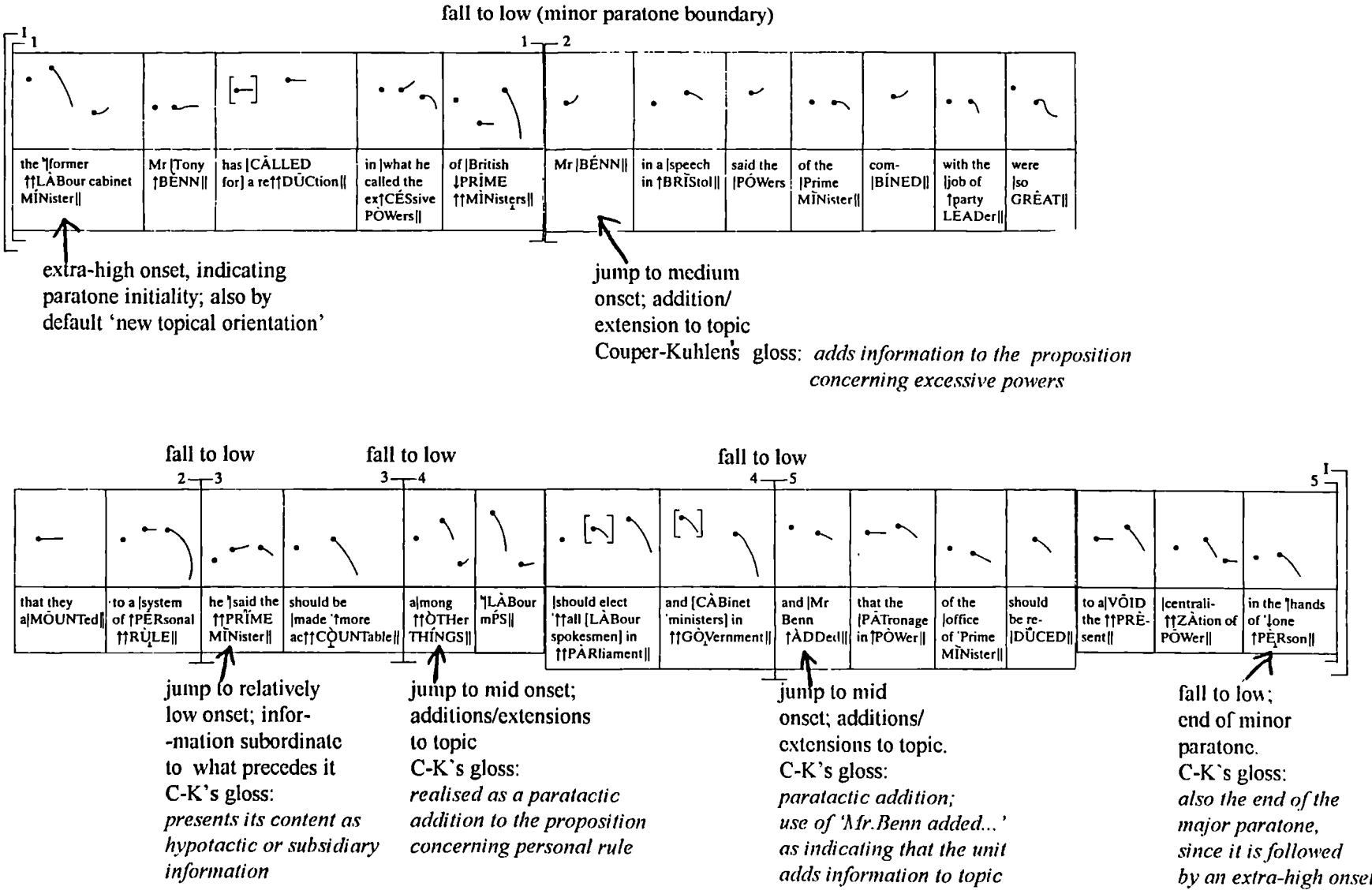


Figure 1.5 **The onsets and nuclei throughout a major paratone, annotated according to Couper-Kuhlen's description (1986: 193-4)**

In some cases, Couper-Kuhlen's classification of the type of information in each minor paratone seems somewhat arbitrary; I have already noted the difficulty dealing with the notion of topic.

However, from the point of view of the present study, what is important is that it does indeed appear that it is possible to see systematic use of onset and nuclear tone within a paratonic framework. In particular, it is interesting that onsets at different points in the structure can occur at significantly different pitch heights within a speaker's range. This appears to run contrary to the idea of onset as constituting the one point of relative stability in pitch production, by which pitch can be 'reset' as each new tone unit commences. Crystal (1969) notes that a speaker's onsets will be 'articulated at or around a stable pitch-level for the majority of his tone units' (p.143).[3] Perhaps these two views of onset are really not so different: the necessary resetting which Crystal observes perhaps equates to Couper-Kuhlen's jump to medium onset, with 'high' reserved for paratone initial units, and low onset to mark minor paratones which contain subordinate material, such as subordinate clauses, or other 'added-on' material and 'afterthoughts'. I will discuss the marking of subordination with pitch in Chapter Six.

In order to illustrate how variable onset levels can be, I have in Figure 1.6 re-transcribed the whole of the paratone from Figure 1.5, including only the pitch level of each *onset* (black squares), and also the minor paratone boundaries (vertical lines):

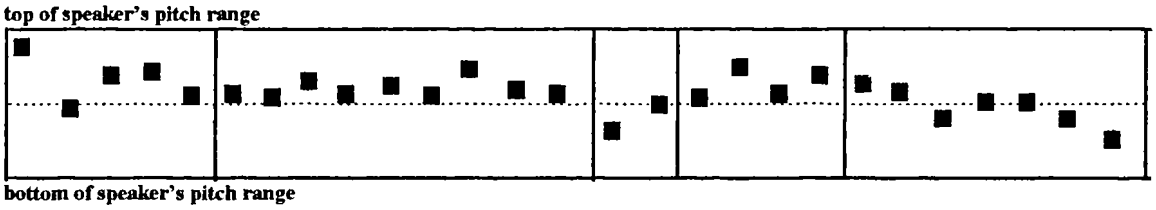


FIGURE 1.6 **Representation of the onsets through the five minor paratones in Couper-Kuhlen's transcribed major paratone**

From the point of view of bare pitch production, we can see that pitch at onset can be near to the highest point in pitch range, or significantly lower. As the notes on discourse structure in Figure 1.5 suggested, these variations in onset level have to do with the kind of orientation which the topic will be given during that unit.

[3] Crystal does note the extra-high pitch to mark the beginning of 'paragraphs' in BBC news readings (Crystal 1969: 144)

1.3.4 The organisation of spoken texts

'Every clause, sentence, paragraph, episode, and discourse is organised around a particular element that is taken as its point of departure' (Grimes 1975: 323). Grimes and others have noted that the linear sequence of items in spoken texts is 'staged', so that items referring to central ideas will precede those referring to subsidiary ideas, or to ideas which develop the initial one in some way. However, as I mentioned previously, the standard functional explanation of intonational prominence at the end of a unit is that *informational* prominence usually centres around the rheme, which typically follows the theme. Clearly, this interpretation conflicts with Grimes' statement that the initial item in a sequence (including a clause) will be most important in the structure of information. If prosodic structure were to reflect this staging requirement on the clausal level, then the accentuation patterns which I have reviewed earlier in this chapter would not be achieved. However, very many different paradigms attest to the tendency for a rightmost locus of information within the simple sentence, and also within the single intonation unit (prominence). Within higher levels of structure, elements in left position are not 'thematic' in the traditional sense. That is, a new topic expression announces itself as an addition to discourse, and will not tend to have the low communicative dynamism of thematic material. The nature of left-focus at this higher level of structure can be seen in Couper-Kuhlen's description of the marking of minor paratones by onset levels, above, and also in the fact that high (or extra-high) pitch has generally been noted to mark paratone initiality (i.e. leftmost position).

The notion of linearisation seems to relate fairly comfortably to that of paratonic form. Paratonic structure as I have outlined in this section can be thought of as a linearised text. Within the lowest level units (individual intonation units) prominence is by default at right-most position; the succeeding, higher levels of minor and major paratone exhibit a left-located prominence. As we have seen, this prominence is both informational, in that shifts in topic are marked here, and intonational, in that marking is realised largely through pitch. The following diagram is an abstract visualisation of the pitch structure of a major paratone. The individual intonation units show the usual pitch resetting between units. However, this resetting takes account of minor paratone structure after each three units, and the overall structure conforms to that of a major paratone:

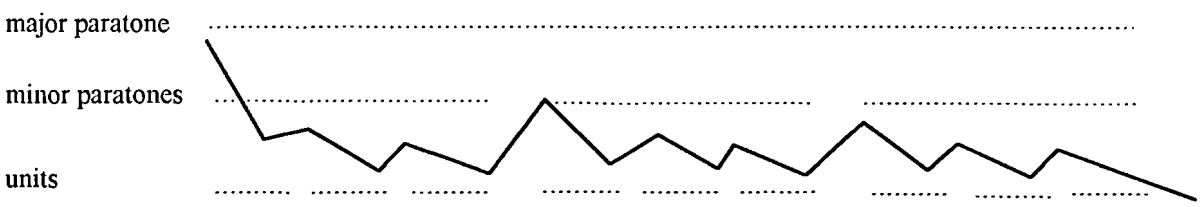


FIGURE. 1.7 **Abstract representation of the systematic relationship between major paratonic, minor paratonic and unit structure**

If pitch height is seen as a kind of gestalt of informational salience/prominence in speech (as Bolinger has generally asserted), then perhaps it functions simultaneously on different levels of structure. Hence, within separate intonation units the rightmost relative peaks on individual prominences indicate prominence to the listener, whereas at the level of paratonic structure it is relative height of the (left-most) onsets which indicates the relative informational salience of sections of the text. Such an idealised view is illustrated in the abstract representation above.

1.4 Intonational form and English language pedagogy

The *idea* that intonational form should be given a higher priority in second language teaching is not new (see Nash 1971; Wilkins 1974; Morley 1979). Indeed, some writers have proposed 'top-down' approaches to the teaching of pronunciation which favour intonation as a foundation rather than the more traditional, 'bottom-up' approach based on segmental work (Gilbert 1984; Pennington 1989; Jones and Evans 1995). However, Chun (1988) notes that, despite a growing interest in 'the pragmatic or functional rules...of utterances in spoken discourse' (p. 295), which informed many approaches to language pedagogy in the 1970's and 1980's, intonation was largely ignored. She describes intonation as 'a powerful and as yet untapped discourse tool which should be developed as part of the communicative competence of the foreign language student' (p. 298-9). The failure to consider intonational form within approaches to L2 pedagogy, she notes, is reflected in guidelines for proficiency testing. The American Council on the Teaching of Foreign Languages (ACTFL) *Proficiency Guidelines* (1986), for example, contained references to the use of intonation only in the top two of nine bands of proficiency, and 'the ability to distinguish main ideas from supporting information through syntactic, lexical and suprasegmental features (pitch, stress, intonation)' were only stated as a criterion for the very highest band (Chun 1988: 300).

1.4.1 Prominence

Recently, Ladd has summarised the different approaches to prominence, saying that '...[I]t is now generally accepted that sentence accentuation reflects - in some way - the intended focus of an utterance....' (Ladd 1996: 160). Such a system, which carries crucial information about how a sentence needs to be interpreted, will inevitably play an important role in both productive and receptive skills in an L2. Prominence marks focused information, creating fine semantic distinctions which, together with non-linguistic context, help to determine meaning. Also, prominences can be either contrastively or emphatically used, adding considerably to the speaker's ability to refine his intended message, as well as enabling corrections and clarifications to be requested or made. We might say, then, that improving the nativelikeness of an L2 learner's

production and reception of this system is very likely to result in significantly greater facility in the L2.

Taylor (1993) has provided a detailed discussion of prominence (accent) in an article which purports to explain what teachers of EFL 'need to know', concentrating on how information is structured (foregrounded, backgrounded etc.). However, treatments of prominence (or accent, sentence stress etc.) are seldom found in mainstream EFL pedagogical texts. Yet this dearth of any extensive inclusion of the system of prominence in teaching materials is curious, not least because it is fairly straightforward to conceptualise and produce. Indeed, Gilbert (1996) makes some very simple observations on how prominence might be developed in the speech of EFL learners: using whispering to draw attention away from other aspects of pronunciation when practising lengthening as a means of creating prominences; using body movements to emphasise prominences; the use of kazoos to imitate pitch movements of nuclei. Zawadski (1994) spends the first half of her book *An English Pronunciation Course* on word stress, prominence and rhythm, and pays particular attention to the message and the ways in which these specific prosodic systems help to create meaning.

In two widely used EFL pronunciation texts from the *Headway* series - *Intermediate Pronunciation* and *Upper-Intermediate Pronunciation* (Bowler and Cunningham, 1990 and 1991 respectively) - several of the main functions of prominence (referred to as 'sentence stress') are covered. Throughout the two books, 'sentence stress' is presented largely in relation to the character of the nuclear tone which it carries, as well as in its role as conveying speaker attitude (surprise etc.). There is, though, some brief mention in each book of emphatic and corrective prominence (1990: 26-7; 1991: 53), and of contrastive uses of prominence (1990: 32, 1991: 18). There is also some treatment of prominence-placement decisions and prominence-shifting in the more advanced of the two books (1991: 6, 38) Vaughan-Rees' *Rhythms and Sounds* (1994) also covers contrastive prominence (pp. 59-63), within a general approach which concentrates for the most part on building-up correct pronunciation from the basis of word stress patterns and speech rhythms.

The above texts from the *Headway* series include systems of prominence in only very brief and often indirect ways. Bradford's *Intonation in Context* (1988), on the other hand, provides an extensive treatment, with chapters on 'highlighting' (creating focus), and on emphasis and contrastivity (narrowing the focus). Bradford gives clear explanations of what happens to the prosodic features when prominences are produced,

and also includes some F_0 traces as examples of different tone movements (pp. 1, 12-13). Also, all her examples and subsequent practice exercises are contextualised, so that the relationship between prosodic form and context is maximised. Receptive skills, particularly in contexts where prominence is the key to accurate extraction of meaning, are given a central role in the presentation.

It seems, then, that prominence has gained some acceptance in EFL pedagogy, although this has begun to happen only quite recently.

1.4.2 Information structure and tones

Hewings (1993) claims that texts such as Halliday (1970) and O'Connor and Arnold (1973) contain 'complexities which make the material inaccessible for both teachers and students alike' (p. 6). Nevertheless, these texts provide clear and extensive accounts of both tonic placement and of tone. O'Connor and Arnold do not actually focus on tonicity as a system but on intonational patterns and their nuclei. However, the effect of their text as a whole is to present the student with a vast number of examples of utterances in English, all of which contain a transcription of where the nucleus falls. Yet Hewings' reservation about both these texts has some foundation; their descriptive rigour is a limitation in the EFL context.

Perhaps what has generally been lacking in pedagogical texts is a presentation of what O'Connor and Arnold call 'the anatomy of intonation', but one of a rather more manageable size than that of O'Connor and Arnold (pp. 1-45). We need a description which focuses on easily understood treatments of intonational form and their role in communication. As we saw in the previous section, some writers do include elements of intonational form in EFL texts. However, they tend to be very selective in what is covered, typically focusing on a small number of key areas: listing, questioning, expression of certain attitudes. It may be that the general move away from prescriptive approaches to language teaching has influenced this. What we do not see currently is any 'nuts and bolts' treatment. By this I mean some very basic starting points which might orientate a learner towards the broad forms of English intonation, but which in themselves cover the essential aspects of intonational form. For example, the idea of default tonicity is a simple and, in very general terms, an uncontroversial idea: the tonic

or nuclear element will tend to fall towards the end of the unit, often on the last lexical item. Yet, even such elementary advice is rarely found.

The given/new distinction is seen in two texts: Brazil *et al.* (1980) and Bradford (1988). The tone system which is used in both these (originally described in Brazil 1975) is very explicit, using a falling tone ('proclaiming') for new and a fall-rising tone ('referring') for given information. However, this immediately illustrates another problematical aspect of intonation and pedagogy, which has to do with the lack of a generally recognised description of English intonation. For example, if a pedagogical text includes this proclaiming/referring distinction in its exposition, it cannot easily go on to deal with 'tunes' in the way that O'Connor and Arnold do. The problem here is not just one of conflicting models, but also involves practical difficulties: the production of fall-rises on all 'given' information could hardly be incorporated into a 'tune' approach, since the prescription of using a fall-rise at certain points in an utterance may exclude the possibility of using different tune configurations as described by O'Connor and Arnold (as well as many others). This lack of a generally accepted description of intonation may have become a disincentive to writers of EFL texts, resulting in their selecting only the most basic aspects of intonational form for inclusion, which in turn impoverishes their treatments.

Of the small number of assumed ideas which generally comprise EFL accounts of intonation, the most notorious is question intonation. In EFL texts a simplistic assumption has often been made whereby wh-questions receive a falling tone and polar questions a rising one (see section 1.2.6 above). We can see this in *Headway Intermediate Pronunciation* (Cunningham and Bowler 1990), where such a distinction is made explicitly in two short exercises, which cover polar questions (p. 3) and wh-questions (p. 4) separately. In the more advanced *Headway Upper-Intermediate Pronunciation* (Cunningham and Bowler 1991), the same distinction is made, again very early in the book (p. 6), and again very briefly. Such accounts of questioning intonation are, as we have seen, substantially misleading; the degree of conduciveness informs tone choice on questions. Thompson (1995) makes this point, citing the findings of Svartvik and Quirk (1980), in which wh-questions were indeed seen to carry a falling tone, but polar questions were observed to carry either a rise or a fall. She argues for an approach to teaching question forms which in some way incorporates the assessment of how conducive yes/no questions are. It might be added, however, that such an approach is not easy to imagine.

This illustrates yet another problem for applied linguistics in this area. 'Applied' linguistics, where by 'applied' we mean the application of linguistic knowledge to practical areas such as SLA research and language pedagogy, suffers from a considerable delay from the time when new linguistic approaches and models gain recognition within the field of linguistics, to the incorporation of application of them in the 'applied' field. Moreover, this is exacerbated when descriptive models are in themselves complicated in their mechanics, and when, as in the case of intonation, they are seen to touch on a wide variety of different linguistic and discoursal systems.

1.4.3 Information staging

Bearing in mind the delay in bringing new linguistic models into use in 'applied' linguistics, it is no surprise that paratonic structure has yet to be incorporated into EFL texts. The only partial exception is Bradford (1988) who includes material on 'key' (see Brazil 1975), and one of the ways in which she describes the use of low key is as a subordinating marker used on an element when it has the same meaning '...as the piece of information which comes before.' (p. 35)

A major problem in the teaching of intonation is that it involves making the student aware of a system which is naturally planned at a subconscious level by speakers. Thus, we might ask whether paratonic form, as well as perhaps other prosodic devices at the utterance-level, can ever realistically be 'taught'. On the other hand, there is no reason to suppose that the inappropriate (nonnativelike) use of pitch, speech rate, and pause in extended turns of speech is any less intrusive in either the perception of foreign accents, or indeed in the ability of nonnative speakers to communicate successfully in the L2. Again, what is needed is a simplified account, although with paratonic structure perhaps we need to be rather more selective in what we include in a pedagogical account. Some things, though, are straightforward. For example, we know that pauses of different duration are used in English to differentiate between units of information which carry different amounts of informational importance within the spoken text (see 1.3.2., above). Yet, no EFL texts mention pause in this way. Why is this so?

1.4.4 Intonation and language pedagogy: the role of second language acquisition research

It might be the case that advances in L2 proficiency are generally greater in those areas of language structure with which writers of texts and teachers are most willing to engage. Vocabulary, grammar, and even the pronunciation of segments might thus advance more rapidly than intonation, simply because they are taught more often and more systematically. Also, inappropriate teaching, such as advice to produce a rise on polar questions, may exert a negative influence over a learner's acquisition. The inadequate treatment of intonational form in most EFL teaching and learning material should thus be borne in mind when data on the SLA of intonation are considered. Learners have a generally impoverished experience of intonation in terms of formal contact with the L2, and this may affect overall acquisition. Language teaching, of course, is only one possible source of L2 input and I will discuss the various ways in which learners receive input during acquisition in Chapter Two.

In the preceding sections I have suggested that several factors may have led to a relative absence of intonation in English pedagogical texts. These are:

- descriptive accounts of intonation are seen as complex and inaccessible[1]
- descriptive approaches to the teaching of target language structure are not currently favoured
- the delay in transfer of knowledge from the theoretical to the applied domains
- the lack of a commonly recognised description of intonational form

These factors, I suggested, might explain the following:

- the absence of basic guidance on many fundamental aspects of the structure of intonational form in EFL texts
- the inclusion of a small number of simplistic and sometimes inaccurate aspects of intonational form in many pedagogical texts

[1] A point also made by Hewings (1993: 6)

Woolard (1993) has described the situation from the perspective of language teachers:

> The average teacher is uncomfortable with intonation, treating it as a difficult subject: difficult to isolate, difficult to describe, and difficult to formulate rules for - rules which will allow students to generate appropriate examples for themselves. As such, it tends to receive little explicit focus in the classroom.
>
> (p. 24)

And Faerch, Haastrup and Phillipson (1984) make a similar point:

> ...we suspect that teachers tend to spend a good deal of time on the individual sounds of English and hope that rhythm and intonation will look after themselves.
>
> (p. 122)

Palmer (1922a) was aware of the problem of teaching L2 intonation, observing that in the case of pitch contours, although the patterns themselves can be taught, the assigning of meaning to these abstract contours is difficult to teach.

This situation has consequences for SLA research. Over the last thirty years, research in SLA has often had a direct influence on the writing of pedagogical texts. Krashen and Terrell (1983) is one such case, where a large body of empirical work (mainly into the acquisition of grammatical morphology) led to the formulation of a model of SLA, which in turn spawned new pedagogical methods.

There has been very little research into L2 intonation, and this itself may have exacerbated the situation with regard to the writing of appropriate texts. Also, the lack of a commonly recognised descriptive model may cause researchers themselves to be wary of the subject, just as Woolard suggests teachers are.

In framing the present study, then, several key issues were borne in mind. In terms of descriptive models, I have attempted to avoid some of the problems by concentrating on features of intonational form which are not seen as exclusively part of one descriptive framework. For example, in the discussion of prominence, above, several different descriptions were referred to. I argued that, at least in a very broad sense, they all include an essentially comparable notion of 'prominence'. It is indeed this rather broad notion of 'prominence' which I will adopt in Part Two of this thesis. The selection of other aspects of intonational form also reflects this approach.

Regarding applicability, the aspects of intonation under investigation here all relate in a fairly clear and transparent way to the informational structure of speech. That is, they represent some of the more straightforward things which speakers 'do' with intonational form: make contrasts and create broad or narrow focus, enumerate, mark

new and given information, signal subordinate information, structure longer stretches of speech. Although this study is concerned with the acquisition of intonational form, it is also undertaken with the aim of making an (indirect) contribution to the state of L2 pedagogy in this area. So, whereas I will be considering aspects of L2 acquisition of intonation here in considerable phonetic detail, and also discussing what my observations reveal about SLA processes, the findings will, nevertheless, be of potential interest as a source for applied work for those involved directly with language pedagogy.

In section 1.4.3, above, I asked why it is that a parameter of speech such as pause is not taught to EFL students, given that it is potentially a very productive element in terms of the communicability of speech. Perhaps part of the problem is a general lack of familiarity with research in these areas within applied linguistics. If this is so, then perhaps this study might go some way towards providing the field of EFL pedagogy with a partial account of, a) some aspects of intonational form and how they operate in spoken discourse, b) how these aspects develop in the acquisition of English as a second language.

Chapter Two - Second Language Acquisition

2.0 Introduction

This chapter will begin with a broad overview of research into second language acquisition (SLA) over the past thirty years, and will then move on to a selective review of work on sound systems in SLA. The third section will consider some linguistic issues related to both first and second language acquisition, concentrating in particular on the role of prosody. Finally, a more comprehensive review of research into SLA and intonation will be presented.

2.1 Approaches to Second Language Acquisition Research

A language teaching methodology is always informed by a theory of acquisition in some way. The structure of a pedagogical text implies decisions by the writer as to exactly what processes of acquisition are likely to be enhanced by teaching. As such, the awareness of SLA as a linguistic phenomenon goes back at least to the time of the first published language learning texts. Duwe's *An Introductorie for to learne to rede, to prononce, and to speke French trewly* (1534), for example, emphasises the learning of verbs using transformational drills, and contains little reference to rules of grammar. This itself tells us something about Duwe's beliefs regarding the naturalness of acquisition and the contextual nature of good language learning.

Over the last thirty years the field of SLA has become increasingly influenced by the desire to develop theoretical models of acquisition. In some ways a separation has occurred between language teaching methodology and the study of SLA as a theoretical endeavour, so that the two are often now seen to exist independently. As James and Leather point out, '[l]earning an L2 (or 'acquiring' it...)...has been scientifically differentiated from teaching it....' (1987b: 1).

The increasingly 'scientific' field of SLA research has followed a number of key shifts in approaches to research, and in the general parameters of the framing of models of acquisition. In this section I will examine these briefly, and consider their possible implications for the study of L2 intonation.

2.1.1 Contrastive Analysis and the transfer of L1 in SLA

Contrastive Analysis (CA) seeks to identify differences between the structures of different language systems. Within the study of SLA, CA has been used as a way of describing how specific L2 forms are acquired from the starting point of a particular L1, where similarities and dissimilarities between L1 and L2 are seen to affect the nature and rate of acquisition. Such an analysis effectively becomes a predictive model for the acquisition of the L2 by speakers of a particular L1.

Contrastive accounts of SLA have often been concerned with sound systems. Wolff (1950), for example, compares phoneme sets of English and Puerto-Rican Spanish, and Stockwell and Bowen (1965) provide a far more extensive CA of English-

Spanish phonology. A comparative L1-L2 approach is potentially very useful in my present study. It might be possible to explain certain nonnativelike aspects of L2 English intonation in terms of informants' L1. For example, the absence of a particular tone (fall-rise, rise-fall etc.) from an informant's L2 speech may reflect its absence in the L1 of the learner. The question in this case, then, would be to what extent does this tone emerge in the L2 over time as the influence of L1-transfer diminishes.

However, in the present study there are several problems related to a wholly CA-based approach. Firstly, there is a dearth of adequate descriptions of the intonational systems of many languages. This is particularly so in the case of studies of the phonetics of intonation. Hence, experimentation would entail the elicitation and analysis of data on both L1 and L2 production of informants. This would be not only a very time-consuming approach, but would also involve considerable care in the elicitation of data of a comparable nature across languages.

There are also more theoretical arguments against the wholescale adoption of a CA approach, which have been voiced repeatedly in the field of SLA as a whole. One common argument has been that CA is blind to the *nature* of acquisition, and cannot provide a qualitative account of exactly how the transition from L1 to L2 form develops. An L2 feature might be novel for a learner, yet easy to acquire, for reasons of conceptual clarity or functional transparency. Conversely, an L2 item might prove unexpectedly resistant to acquisition, despite putative similarities with L1 forms.

Contrastive accounts of SLA tend to suppose that differences between the two systems constitute a source of difficulty, or interference. In terms of language teaching, such predicted difficulties can be given particular attention in order that this L1-interference, or negative 'transfer', be overcome. CA is in this sense founded upon a view of acquisition as the 'constant warfare' which results from the difficult accommodation of two language systems in the mind of the learner (see Marton 1981). Indeed, the very idea of constantly referring back to L1 forms in a description of the acquisition of L2 forms itself suggests that acquisition constitutes something of an unnatural co-existence of conflicting systems. As we will see, this approach has been challenged.

During the 1960s and 1970s, dissatisfaction with CA prompted a shift towards new research orientations in SLA. At about this time also, interest in SLA and sound systems (which had been central to many CA studies) began to diminish. As I will note below, SLA research since then has most often used evidence from phonetics and

phonology in only a peripheral way. L1-interference, so central to accounts of the acquisition of L2 sound systems, has become regarded by many as only a minor aspect of SLA. This was expressed unequivocally in 1980 by Felix:

> ...our data on L2 acquisition of syntactic structures in a natural environment suggest that interference does not constitute a major strategy in this area... it seems necessary to me to abandon the notion of interference as a natural and inevitable phenomenon in L2 acquisition.
>
> (Felix 1980: 107)

For Felix, and indeed many others, the phenomenon of SLA became synonymous with that of the acquisition of L2 syntax.

However, there is no doubt that the L1 of a speaker influences substantially the acquisition of the phonology of a second language. For this reason, current work in the area of the phonetics and phonology of SLA is perhaps most clearly distinguishable from other areas of SLA research in that a central role is still given to the process of transfer.

In my own analysis, I will make specific reference to informants' L1s (Chapter Seven). Also, I will conduct some limited analysis of comparable speech productions made by informants in both their first and second languages.

2.1.2 The study of errors in SLA

In an influential study, Dulay and Burt (1974) found that only 4.7% of grammatical errors in an L2 corpus (of L2 acquisition by children) were explicable in terms of L1 transfer. The idea that many errors are *not* explicable in terms of interference generated interest in processes of acquisition unrelated to the L1. Errors, instead of being seen as the debris of L1-L2 'warfare', began to be seen as evidence of other acquisition mechanisms.

Error analysis (EA) involves the collection of data on L2 production. However, the construction of a reliable, representative corpus of learner-errors is itself not straightforward. Significant deviation from normal production may occur during data collection; informants may deliberately avoid using certain forms, and, conversely, they

may overuse those forms which they know are correct, either as a production strategy or due to the reinforcing effects of particular teaching practices.[1]

Data interpretation is also problematical in EA. Corder (1967) noted that not all examples of incorrect production were errors. 'Errors' are seen as evidence of an unacquired or incorrectly acquired form, whereas 'mistakes' are simply performance-related slips. In later publications Corder proposed a number of assessment criteria for the classification of errors and mistakes, including judgements of how an intended meaning is expressed, and decisions as to whether 'a plausible interpretation' can be put on the sentence in context (1974: 129). L2 production, he claimed, also includes 'covert' errors, in which a correct surface form is used inappropriately. Corder's stipulations for the description of errors are seen at their most complex in Corder (1981). However, Ellis notes that, despite their apparent scientific character, such procedures depend on two unreliable elements: the researcher's own (subjective) assessments, and subject-as-informant questioning (Ellis 1994).

Whatever its merits and drawbacks, EA seems immediately unsatisfying as an approach to the assessment of L2 intonation. We can of course analyse intonation in a number of *quantitative* ways. For example, we can ask whether the average pitch range of certain tones, or of speech generally, is broader or narrower in the production of nonnative speakers than in the case of native-speaker norms, and also whether L2 production develops in the direction of nativelike norms over time; we can note which tones are present in a learner's repertoire, and we can look for significant differences in the frequency of use of different tones in comparison to nativelike norms. These approaches are not concerned with *errors*, but rather with the quantification of the prosodic parameters of speech. I will indeed look at my own data from this perspective.

However, in addition to the phonetics and phonology of intonation, we can also look at the functions of intonational form in discourse, expressing this as the ability of a speaker to use the appropriate ('correct'?) intonational device in a given context. For this reason, in Chapter One I specified a number of relatively discrete semantic and discoursal functions of intonation - normal and contrastive prominence, listing intonation, the marking of paratonic structure - in an attempt to arrive at a set of specific intonational forms to be used as a means of assessing acquisition here. Given suitable data, then, we can judge whether or not an informant is capable of employing the correct intonational

[1] Indeed, Selinker sees the 'transfer of training' as one the five cognitive processes in SLA (see Selinker 1972).

device in appropriate, contextualised production. Again, inappropriate production here will not constitute *erroneous* production in any strict sense.

The elicitation of suitable data is an issue which has interested researchers into SLA, although once again we see the preoccupation with the investigation of grammar here. The Bilingual Syntax Measure (Burt, Dulay, and Hernandez-Chavez 1975) is an elicitation procedure which involves speakers describing a series of pictures, which is designed to oblige the speaker to supply specific grammatical forms at certain points (irregular past tenses, plurals, continuous forms etc.). Much research in the late 1970s and 1980s was based on such procedures; they allow a clear and precise picture of the current stage of acquisition of the learner to be obtained, and also let the researcher state with a greater degree of certainty whether an informant has or has not acquired a specific form.

Eliciting data on L2 intonation in this way is not without problems. The constituent parts of the intonational whole occur within the prevailing prosodic layer of speech. So, it is not always easy to assess whether particular forms (a systematic high-fall/low-fall distinction, for example) have been attempted at all. Nevertheless, some form of controlled elicitation of speech may help here, in that production can at least be guided towards the probable use of predicted forms, whilst remaining spontaneous speech. In Chapter Three I will describe how my own experiments include tests which bear some similarities to The Bilingual Syntax Measure, although incorporating a looser approach, and more detailed subsequent interpretation of data.

2.1.3 Acquisition processes in SLA

One result of EA-based research in the 1970s was that a number of processes of acquisition were suggested to account for different kinds of errors in SLA. Richards, for example, provides the following error categories:

1. overgeneralisation
2. ignorance of rule restrictions
3. incomplete application of rules
4. false concepts hypothesised

(1974: 174-8)

Such an approach involves the description of how particular cognitive processes (including error-generating processes) influence the acquisition of L2 rules. However, whereas data on the production of L2 syntax and lexis can be understood in terms of these cognitive processes and their influence on the development of rules, the study of L2 intonation does not fit comfortably within this approach.

This perhaps relates to Sweet's broad distinction between the 'analytic' and 'synthetic' layers of speech. In terms of SLA, at least some of the elements in the 'analytic' layer (syntax, lexis, segments) are present from the earliest stage of acquisition; the first stumbling, error-strewn attempts at simple sentences contain the beginnings of an L2 lexis and some attempt at target-like syntax probably follows soon after. Errors in the production of these systems are the result of processes which promote the emergence of nonnativelike forms during acquisition. The 'synthetic' layer, however, is perhaps not a necessary element in L2 structure in the same way as the 'analytic' layer is. Learners are likely to be understood by native-speakers even if their prosodic production has not begun to take on L2 characteristics; they may use a pitch range identical to that of their L1 production, and employ intonational tones and contours exactly as they do in their L1 (they may also have no awareness of L2 word-stress patterns). As L2 prosody emerges, especially intonation, the very nature of the meaning-enriching and discoursal functions of many elements of these prosodic systems are difficult to express in terms of acquisition processes. As Færch *et al.* (1984) have noted, '[C]orrect or incorrect prosody is far more a question of degree than of being right or wrong' (p. 127).

Processes of acquisition such as 'overgeneralisation' and 'ignorance of rule restrictions', then, do not seem to offer a satisfactory framework for the analysis of L2 intonation. Indeed, how could 'overgeneralisation' be measured here? To take an example, a speaker of L2 English may produce a tone such as O'Connor and Arnold's *Low Drop* (see O'Connor and Arnold 1973: 47-57), with an (inappropriately) wide pitch width, leading to possible interpretation of this as a *High Drop*, since the tonic movement would describe a high rather than low fall. The reserved sense of definiteness of a *Low Drop* would in this case be lost, and the more strongly assertive *High Drop* supplanted. However, this cannot easily be seen as a discrete error, or in any sense as evidence of overgeneralisation (of a rule), although communication may well have been impaired. Indeed, depending on one's model of intonation, such a case might not be seen in terms of an error at all, but as a paralinguistic variant on one (falling) tone (see Ladd 1996: 39).

So, whilst we recognise that intonation comprises distinct forms (different tones, for example), in production the prosodic features are flexible parameters. In such cases as an inappropriately wide pitch on the falling tone of the *Low Drop*, above, we must be rather careful in interpreting this as deviant, since the appropriate pitch movement has been used, and whereas the production might be seen as contravening nativelike norms of production, there may be a host of possible reasons for this which would form an acceptable explanation of such a production in the speech of a native-speaker in the same context.

These issues will be borne in mind during the analysis of the data in Part Two. Whilst I will use a form of elicitation which aims at collecting data where the use of certain aspects of intonational form is highly probable, I will also recognise that the relatively fluid nature of intonational form cannot be dealt with simply in terms of whether a particular form is or is not used by a speaker. Instead, I will look for *tendencies* in production, in which it will be particularly important to have data from native speakers, against which to make comparisons.

2.1.4 Developmental approaches to the study of acquisition

Early in the 1970s several studies of First Language Acquisition (FLA) proposed common developmental patterns in the acquisition of a language, in which grammatical morphemes were observed to have been acquired by different children in the same order (Brown 1973; de Villiers and de Villiers 1973). Brown, for example, found that the acquisition of L1 English by different children correlated with scales of 'cumulative complexity', in which grammatical morphemes carrying a more complex set of sense relations (agent, object, locative, active etc.) were acquired later than less complex ones (p. 185).[2] Brown also discovered identifiable stages in the acquisition of particular morphemes, in which processes such as overgeneralisation are seen to typify the developing acquisition of individual items (p. 325).

A developmental approach was also adopted by many writers in SLA at this time, and 'developmental orders' of the acquisition of grammatical features have been widely reported in the literature (see Ellis 1994, Ch. 3 for an extensive discussion). Others have

[2] Brown discusses the notion of 'tansformational complexity' in similar terms (p. 377).

described similarities in the specific stages of the acquisition of single items ('developmental sequences'). Such studies tend to claim that developmental orders are similar for learners of an L2 with different L1s. Again, research has concentrated largely on studies of grammatical morphology.

The idea that acquisition is in some way systematic at its various stages of development along an L1-to-L2 path led to the view that L2 speech can be seen as a series of emerging, transitional languages, with characteristic acquisition processes and patterns. The term 'interlanguage' (IL) was coined by Selinker (1972)[3] to describe the concept of a continuum of progressively more target-like language systems which develop as acquisition progresses.

ILs are said to exhibit many of the characteristics of emerging first languages (such as the learner's use of hypothesis-testing as a means of discovering new rules). However, ILs emerge under different conditions than L1s, and hence the processes and strategies of acquisition are said also to be different. A set of SLA-specific strategies is proposed by Selinker:

1. Language transfer
2. 'Transfer of training' (influence of a 'learned' L2 rule)
3. L2 learning strategies
4. Strategies of communication in an L2 (with native speakers)
5. Overgeneralisation of L2 linguistic material

(from Selinker 1972)

Selinker also describes the process of 'fossilisation', where the acquisition of L2 forms halts some way short of full nativelikeness, due to the persistent negative transfer of the L1. Corder (1967) also noted such a phenomenon, and his well-known claim that ninety-five per cent of learners don't achieve nativelike proficiency might be restated as the observation that in most cases ILs undergo complete fossilisation some way short of nativelike levels. Causes of fossilisation have been greatly debated in the literature; the learner's attitude to L2, communicative environment, and maturation have all been discussed.

The above list of strategies appears to offer a useful set of terms for the discussion of intonation. They imply that SLA is typified by a set of general learning principles which are naturally employed during SLA, without the suggestion that these principles relate only to the development of 'rules' of 'analytic' linguistic systems. For

[3] The term was first used in 1935 by American creolist John Reinecke, to describe non-standard varieties of a language (see Larsen-Freeman and Long 1991: 74, n4).

example, Selinker's 'transfer of training' involves the 'influence of a 'learned' L2 rule'. However, a 'rule' here refers to an item of conscious linguistic knowledge - what has sometimes been called 'metalinguistic knowledge'. The example of teaching students to use a rising tone on polar questions, which I discussed in section 1.2.7, would be a case of 'transfer of training' in this sense. As another example, I will note in Chapter Five that in my data L2 speakers overuse rising and level tones relative to native-speaker norms. It seems clear that this represents either 'language transfer' or some kind of 'strategy of communication in an L2' (perhaps a strategy to mark one's speech with doubtful, non-committal open tones). Of course this does not imply that any linguistic 'rule' has been broken or inaccurately acquired; it is an aspect of verbal interaction, and we can comfortably talk about how such a 'strategy' diminishes as L2 proficiency increases, rather than about the suppression of 'IL rising tone rules' or any such processes.

2.1.5 Production styles and their influence on L2 production

Labov (1972) observes that in careful, formal speaking styles speakers will tend to employ prestige norm variants which they do not use during casual speech, and notes that whereas social conditioning leads speakers to monitor speech towards the prestige norm in formal speaking situations, in casual, vernacular speech, processing constraints disallow for this kind of monitoring.

Tarone has noted that in L2s, production reflects a continuum of different styles (Tarone 1979, 1982, 1983, 1985). A more careful style will lead to high levels of accuracy, involving the use of the most advanced, newly acquired target forms, along with conscious attention to linguistic form. On the other hand, vernacular, casual styles will not be form-focused in this way.

Again, this relates to the nature of speech data in SLA. What speaking style most adequately reflects a learner's current stage of acquisition? A form may be present as metalinguistic knowledge in careful production, where a conscious use of taught forms is possible. But is this conscious awareness of linguistic structure a useful measure of acquisition in any convincing sense? The principle of Tarone's continuum paradigm is that conscious attention to linguistic form during careful production actually distorts the picture, since in studying acquisition we are interested in what has been internalised as

linguistic competence, and not in how metalinguistic knowledge can be used in careful, form-focused productions. This approach has much in common with Krashen's Monitor Theory (see below).

Some evidence from phonology, however, does not support Tarone's claims. Sato (1985), for example, does not find the continuum paradigm model to be reflected in her data. She notes that, in consonant clusters in Vietnamese-English IL, more accurate, nativelike productions are found in vernacular than in careful style. Oyama notes a similarly 'interesting' finding in her study of immigrants' L2 English pronunciation:

> For the nonnatives, then, the increased attention to pronunciation ... did not necessarily allow closer approximation of the norm. For those whose command of English phonology was shaky, increased attention seemed to have a deteriorative effect on performance.
>
> (1976: 270)

Burmeister and Ufert (1980) also found more L1 interference, backsliding and avoidance strategies under conditions of formal production. The interpretation of this must have something to do with the intrusive nature of the conditions under which formal styles are adopted; such styles may draw to the attention of the learner that his knowledge of L2 rules is incomplete, with this awareness creating negative pressures on production.

Although these issues have a direct bearing on methods of data elicitation in my study, prosodic production is a largely unmonitored aspect of even careful speech. Indeed, it is not clear whether a highly monitored style would lead to significantly differing production than in casual speech. For example, we can imagine a French speaker's highly monitored L2 English containing non-velarised (or partially de-velarised) /r/, and some attempt at correct word stress patterns which in casual speech would be very much less target-like. With intonation, though, these issues are far less clear-cut, as well as remaining largely untouched by research. As Leather (1987) has pointed out, work in this area does not have the advantage of a sizeable research literature upon which to base predictions or to frame hypotheses. I will say more about this throughout the remaining sections of the present chapter.

2.1.6 Monitor Theory and pronunciation as a monitored skill

An influential interpretation from evidence of the many studies of developmental patterns in SLA has been the Monitor Model. Krashen (1977) claims 'amazing' levels of

uniformity of orders of acquisition from a large number of findings, with this uniformity strongest in data from vernacular, spontaneous production. As the most comprehensive theory of SLA of the 1970s/1980s, Krashen's Monitor Theory (MT) has been under constant criticism throughout its development (McLaughlin 1978; Sharwood Smith 1981; Gregg 1984, for example). MT has also been discussed widely by language teaching practitioners, and indeed has led to detailed proposals concerning the principles of language teaching (Krashen and Terrell 1983).

The theory comprises five central hypotheses: Acquisition/learning, a distinction between what is acquired as competence and what is merely 'learned' as metalinguistic knowledge; Natural Orders, in which the patterns of acquisition are said to be predictable; Affective Filter, where the effects of states of high anxiety or low motivation are seen to block the reception of useful target-language input; Input, the quality of the L2 which the learner receives; Monitor, the extent to which L2 production involves editing and revision.[4]

For the purposes of the present study, the last two of these hypotheses are particularly interesting. Input, Krashen claims, must be comprehensible but should also be one step ahead of the current level of acquisition. Linguistic and non-linguistic context will provide overall meaning in such cases, thus allowing new forms to be understood in context. Useful input is therefore expressed as comprehensible input (*i*) plus one unacquired element (*i*+1). Crucial to the model is the claim that acquisition results from input. Immersion programmes, for example, provide large amounts of such comprehensible input. The Monitor Hypothesis describes how the 'learned' system of linguistic rules can only be used as a planning and editing device during production when the rule is known explicitly and when time constraints allow. The conditions necessary for this 'Monitoring' (with an upper case 'm') of speech do not often pertain in the normal contexts of language use. On the other hand, 'monitoring' (with a lower case 'm'), in which 'acquired' linguistic competence can be employed both to initiate utterances and to make intuitive editing decisions, is available in all production conditions.

What implications for pronunciation do notions of Input and Monitoring have? Few learners ever learn any formal rules of phonology or 'rules' of articulation in general. Krashen recognises this; pronunciation in MT is seen to fall wholly within the

[4] For a detailed account of the five central hypotheses see Krashen 1981b, 1985.

scope of the processes of 'acquisition', and is not available as a learnable system. This is reflected in MT-based approaches to language teaching, which discourage the teaching of pronunciation:

> Pronunciation ability, or good accent, may be nearly completely dependent on what has been acquired, not on rules which have been learned...
>
> ...we do not recommend any specific activities for pronunciation, especially in the early stages.
>
> (Krashen and Terrel 1983: 89)

There is conflicting evidence in the literature regarding the effects of careful styles on pronunciation. Flege notes a lack of clear evidence of any benefit (1985), but elsewhere has suggested (1987a) that speakers of L2s may be able to regulate (and modify) broad 'bases of articulation' when high levels of monitoring obtain. Oyama finds that careful reading in an L2 was adjudged by NSs to be more heavily accented than spontaneous speech (1976).

With regard to the conscious use of phonological or articulatory rules in L2 speech, Dickerson (1987) has pointed out that there have been few attempts to make 'rules' available to learners. He notes that little research had (by 1987) been carried out into pronunciation and Monitoring. In an extended classroom experiment, he taught specific phonological 'rules' to learners of L2 English, such as using spelling as a guide to the prediction of stress position and vowel quality. The 'learning' of these guidelines was seen to have made a clear, positive difference to the nativelikeness of pronunciation. In keeping with the general MT model, such explicit pronunciation guidelines did indeed result in better pronunciation in highly careful productions - that is, when used during Monitoring. However, longer term improvements were also observed in un-Monitored pronunciation, specifically in the realisation of those elements of pronunciation which had been 'learned' at an earlier stage using explicit guidelines. This is counter-evidence to the acquisition/learning distinction in MT, which emphasises the non-transferability of the learned to the acquired system.

Dickerson points out that his subjects were particularly encouraged to practise pronunciation using these explicit guidelines, Monitoring and repeating their speech in what he calls 'covert rehearsal'. Covert rehearsal has been observed as a productive process in FLA (Weir 1962) and may be a transferable language acquisition skill. Simply, it entails highly form-focused practice of linguistic forms in private, perhaps even *sotto*

voce. Dickerson claims that the learner's very careful rehearsals become input for the learner himself, constituting comprehensible input.

Whilst Dickerson's experiments support the MT claim that learned rules do not transfer directly to spontaneous production, they also suggest that by guiding students through a learning process which includes much (non-communicative) covert rehearsal, learned phonological rules can positively effect medium to long-term increases in overall performance. Thus, Dickerson proposes that rather than being a relatively superficial process in SLA, Monitoring is both an important presentational device in public situations, when the learner wishes (and is able) to improve the quality of his output, and is also a means of generating comprehensible input during private rehearsal of the L2. This latter is 'a decidedly fundamental role, because it involves shaping a grammar' (Dickerson 1987: 137).

2.2 The acquisition of L2 sound systems

In recent discussions of SLA, sound systems have often been relegated to a position of secondary importance. Weinert notes that 'a system of generative rules still motivates the majority of SLA studies' (1995: 180). Also, transfer has itself become 'unfashionable' as an approach to SLA analysis in recent years (see Gass 1984). However, as I have noted, transfer is undoubtedly a central process in the acquisition of L2 sounds. Indeed, whereas the notion of an IL as a series of approximating linguistic systems has provided research into the phonology of SLA with a workable investigatory framework, the notion of transfer still informs much work in the study of SLA sound systems:

> Perhaps the most general question in the study of second or foreign language (L2) phonological acquisition is: how, in the development of schemata for the sound systems of L2, are general capacities for auditory pattern processing constrained by the established sound system of L1?
>
> (Leather 1987: 59)

In this section I will review three areas of interest in accounts of interphonology: the role of the L1, the nature of IL phonological rules, and universal IL processes.

2.2.1 The role of L1: 'new' and 'similar' sounds

Trubetzkoy (1939) claimed that L2 sounds which are phonemically indistinct from L1 ones ('similar' sounds) are 'filtered' by L2 speakers, so that phonetic differences remain unperceived. The distinction between L2 sounds which are 'new' for the learner, and those which have a broad L1 counterpart, has proved a fruitful framework for the discussion of issues relating to the acquisition of L2 sounds.

Wolff (1950), in his contrastive description of the phoneme sets of English and Puerto-Rican Spanish, notes that phonemes in an L2 may be familiar from the L1 but may have different allophones or distribution in L1 and L2. Interestingly, he states that these phonemes are likely to present more difficulty for the L2 learner than completely unfamiliar phonemes, and goes on to indicate some of the ways in which L1 forms constitute complex kinds of interference, such as confusion over phonotactic behaviour and the mastery of new articulations. This qualitative approach to contrastivity differs

from Lado's then influential claim regarding L2 acquisition generally, that what is new in an L2 is more difficult to acquire (see Lado 1964).

The question of whether all-new or L1-similar sounds are more resistant to acquisition in SLA raises interesting issues. In terms of neurology, the decreased plasticity of a mature, lateralised brain may mean that the establishment of new perceptual models and motor plans constitutes significant, possibly insurmountable, difficulties in adult L2 acquisition. This would suggest that new sounds might be harder to acquire than ones which are in some way familiar. It has been noted by speech pathologists that in the treatment of sound production disorders, the remediation of distorted sounds is easier than the teaching of sounds new to a speaker (see Flege 1987a).

However, a great deal of attention has been paid in the literature to the process of acquiring L1-similar sounds, and evidence here has tended to disconfirm claims that 'new is harder'. On the contrary, it is the production of L1-similar sounds which seems to be typified by a shortfall in the attainment of nativelikeness. Acquisition of such sounds is seen as describing a classic intralingual progression, where the developmental IL system is neither L1 nor L2. Flege has explained this in terms of 'equivalence classification' (see Flege 1981, 1985; Flege and Hillenbrand 1984).[1] The development of accurate perceptual and articulatory models of an L2 sound, when the target sound has a close counterpart in the L1, results in a mediated, compromise sound. For example, Flege and Port (1981) report that in the L2 English of speakers of Arabic, productions of voiceless plosives were observed to have voice onset time (VOT) at a mid-way point between (longer) L2 English and (shorter) L1 Arabic norms.

The 'gravitational pull' of both L1 and L2, claims Flege (1987a), occurs due to our natural instinct to establish sound categories from acoustically differing tokens of a speech sound (input). Whereas in FLA this adaptive mechanism allows the child to evolve sound categories from the input of one language, in SLA input comes from two separate languages. Where L1 and L2 sounds are similar, then, the mechanism involves meditation, especially at early stages of acquisition. And since our ability to categorise sound input is not known (or supposed) to decline in the adult, then the principal difference between FLA and SLA here 'may be the existence of well-defined phonetic representations' (p. 16) at the time when L2 acquisition begins.

[1] Weinreich (1953) observed a similar phenomenon, terming it 'interlingual identification'.

So, L1-similar sounds may be more difficult to acquire to nativelike levels than all-new ones. This is the case in Flege's (1987a) comparison of the formant frequencies of vowel production in speakers' English-French IL.[2] Whereas English speakers produced French /u/ (L1-similar) with persistently higher second formant frequencies than target (French) norms, their production of French /y/, which is an all-new sound for anglophones, was realised with accurate, target-like first and second formant frequencies. These differences are observed to persist into late stages of acquisition, where overall levels of L2 proficiency are very high.

The persistence of approximated L1-similar sounds is demonstrated in Bohn and Flege (1992). In a cross-sectional study of German-English IL, the L1-similar sounds / i / and / ɪ / were observed to approximate between L1 and L2 values in terms of vowel quality and duration; these approximated sounds remained stable in productions by learners who had spent anything between six months and over seven years in English-speaking environments.

The concept of equivalence classification supposes that an accurate perceptual model of the L2 sound is substantially intact in the mind of the learner (see Flege 1987a) so that medi ation can take place. Ioup (1984) reports that native speakers of English can identify foreign accents in their L2 Spanish, reflecting a highly developed conceptual model of the L2 sound system, and suggesting also that perceptual models of L2s may be more highly developed than corresponding levels of production. Indeed, there are considerable difficulties in assessing to what extent L2 sound production reflects articulatory or underlying phonological deficits. Flege admits (1987a) that differences in vowel production in his study might have arisen due to purely articulatory factors.

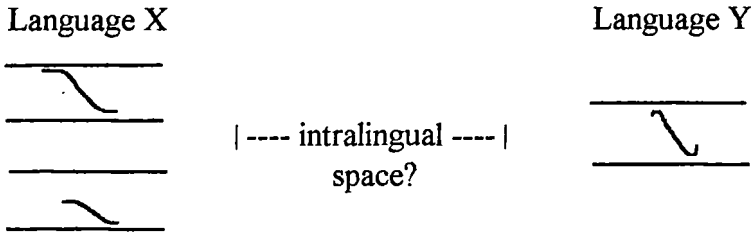
Recently Flege (1997) has found that in Dutch-English IL, the new L2 English sound /æ/ can be fully acquired by Dutch speakers,[3] with formant frequencies comparable to target norms. With L1-similar vowels, though, production remains nonnativelike in terms of their formant frequencies, even at high overall levels of proficiency. Through his model of the Speech Learning Module (SLM), Flege proposes that there is no decline in the ability to acquire speech sounds at all, and that 'basic speech learning abilities' remain into adulthood. However, only new sounds will evade

[2] It is noteworthy that although the field has generally been referred to as 'interphonology', the most effective methodology tends to involve the close instrumental analysis of speech data, yielding quantitative data. Perhaps 'intralingual experimental phonology' more accurately encapsulates the approach.

[3] Although he doubts whether this sound is absolutely 'new' for Dutch-speaking informants (see p. 39).

equivalence classification, and will therefore be the only sounds acquired without the distorting influence of L1. Additional evidence of the successful acquisition of English /æ/ as an L2 new sound is found in Bohn and Flege (1997), whose data on German-English IL again show production at nativelike levels in learners as old as 30 (see also Flege, Bohn and Jang 1997).

The question of how intonational form can be considered within this framework of new and L1-similar sounds is not at all clear. We might talk about 'forms' with reference to certain tones which are either new or L1-similar in terms of a particular description of two languages, but any such contrastive account would need to be very sensitive to the overall parameters of speech in both *L1* and *L2* simply to express how phonetic differences correspond to phonological differences. For example, language Y has only one simple falling tone, whereas language X has two. In such a case do we take into account aspects of (phonetic) realisation, such as pitch range, in determining where the similarities of forms occur?



Cruz-Ferreira (1987) deals with these difficult issues in the case of the contrastive analysis of two intonation languages (see section 2.4.3 for a review), in which some aspects of linguistic function have been incorporated into an assessment of the acquisition of several tones and tone patterns. However, the more fundamental issues remain: how do we designate and predict patterns of the acquisition of intonation within an IL where the following conditions prevail:

a) where prosodic features are an ever-present part of production at all stages on the IL continuum, so that there are no unequivocal points at which L2 intonational 'forms' might be observed to have been acquired;

b) when there is an array of functions of prosodic structure which are to be accounted for simultaneously;

c) when many aspects of prosodic production are instinctive and may be subject to imprecise and hardly-explicable contextual and psychological influences?

Even if we assume that what is phonetically familiar (for example, pitch movement) can be phonologically new (for example, complex tones for many speakers of L2 English), it is not clear in what way this phonological 'newness' equates to the above discussion of 'new' in terms of segmental forms.

Leather (1997) notes these difficulties in his study of the acquisition of Chinese lexical tone by speakers of Dutch (see section 2.4.2 for a review). In particular he notes the possible influence of intonational universals in both the FLA and SLA of Chinese, where in both cases the rise and fall-rise tones are acquired later than the fall and level tones. Cruz-Ferreira also suggests that the universal nature of intonation informs the acquisition of L2 intonation. I will explore these issues in more detail in parts 2.3 and 2.4 of this chapter, below.

2.2.2 The nature of IL phonology: universals

Bannert (1980) found that, in a study of a group of non-native speakers of Swedish with different L1s, one strategy in the production of all-new L2 vowels was to replace these with vowels which were relatively more common in terms of universal hierarchies. Eckman (1977) has also argued that a contrastive account of interphonology can be predictive (rather than just explanatory) if the universal markedness of L2 items is taken into consideration.

CA, within either a taxonomic phonemic or a standard generative phonology framework, fails to account for some evidence on the L2 acquisition of segmental phonology. For example, German-English and English-German ILs differ in the presence or absence of voiceless obstruents in final position. A standard generative phonological representation, in which underlying representation of the final position obstruent includes voice in both languages, will predict the following patterns of acquisition: English-German IL will involve the acquisition of the German terminal devoicing rule of obstruents; German-English will involve the dropping of this L1 rule. Hence, we might say that the *dropping* of the devoicing rule for speakers of L1 German would be cognitively simpler than the acquisition of that rule by anglophones. However, final obstruents are well known to take far longer to reach target-like norms in German-English IL (i.e. the acquisition of final voicing) than the contrary case (the English-German IL acquisition of final devoicing). In other words, there is a definite

directionality in the degree of difficulty running through this pair of ILs with regard to final voicing and devoicing.

This directionality can be accounted for by assessing the universal criteria of typological markedness of items in a particular IL frame. Eckman's Markedness Differential Hypothesis (MDH) predicts that acquisition of forms or rules is influenced by how marked (universally) the form or rule to be acquired is, with markedness equating broadly to relative difficulty. Such a model explains the generally unproblematic acquisition of L2 German final devoicing by anglophones, since final devoicing is relatively less marked (universally) than English voicing.[4] The greater difficulty which German speakers experience in the acquisition of final voicing in English is similarly explained.

The claim that typological markedness can be used to explain the relationship between IL and L2, Eckman notes (1987), 'is to say, simply, that the IL and TL are related in the same way that other genetically unrelated and geographically non-contiguous languages are related' (p. 144). In this study, consonant cluster reduction rules for Cantonese/Japanese/Korean-English ILs appear to be unsystematic, yet are subsequently shown to be explicable in terms of the relative markedness of possible reduced clusters. For example, a stop-fricative-stop cluster will reduce to fricative-stop rather than a stop-stop, based on observed universal preferences for such combinations in natural languages (see Greenberg 1978b). A second finding here is that no subject adheres to the rule all the time; stop-stop reductions do occur. ILs are, by their very nature, highly variable (see following section). However, Eckman also notes that, whilst some stop-stop reductions do indeed occur, they are produced only by speakers who also produce the predicted, less marked reductions, whereas no speaker only produces stop-stop clusters. In other words, the rule, like other IL rules, is observed to vary in its use, but even here, that variability is orientated towards the principle of markedness.

The MDH has important implications for the present study, bearing in mind the prominent role of universalist claims regarding intonation, in terms of its phonological form, its semantics, and its broader communicative functions. Indeed, Sato (1985) makes the point that universal processes may be relatively more influential in IL phonology if aspects such as stress, rhythm and intonation are considered. However, such an approach is also difficult here. As I mentioned earlier, there are no detailed accounts of the

[4] See Dinnsen and Eckman (1978) for universal hierarchies of the markedness of voicing in different positions.

universal nature of intonation, partly because such universality is to some extent assumed. However, even those few strong universalist claims made - the terminality marking of a fall, for example - are necessarily associated with meaning within the discourse. Indeed, the more universalist the discussion of form, the vaguer that description tends to become (see Cruttenden 1981 on the abstract meanings of fall and rise, for example). So, we cannot talk simply of phonological markedness of intonational form, as in the case of vowel hierarchies or scales of sonority, for example. On the contrary, discussions of universal processes will entail the cross-linguistic assessment of function. I will say more about problems associated with this approach in Chapter Three.

2.2.3 The nature of IL phonology: variability

Much (synchronic) variability in ILs can be explained in terms of context, monitoring and production constraints. However, others have identified a further source of variability. Tropf (1987), for example, reports the following forms for <nicht>in the production of one subject's Spanish-German IL over a period of one hour, where all predictable sources of variability were held constant:

[niçt], [niʃt], [niç], [niʃ], [niʒ], [nis], [nik], [ni], [net][5]

(Tropf 1987: 174)

Ellis (1985) compares this free variation to Gatbonton's similar notion of 'diffusion' (Gatbonton 1978). New forms emerge 'spontaneously' in the acquisition of all natural languages, and this high degree of free variation occurs even more so in ILs. A form is underspecified in terms of its exact function when it enters the learner's repertoire, leading to its employment in free variation with other forms. Free variation then gives way to processes of 'reorganisation', in which the use of forms is gradually refined and narrowed down (although not always to nativelike levels). His model is, like Gatbonton's, exemplified with reference to the acquisition of speech sounds. One clearly problematic issue which Ellis does not resolve, however, is that his model implies the existence of some random selection operation within the acquisition process. Whereas free variation in SLA is never likely to be explained exhaustively in terms of learner-

[5] Some of these are themselves acceptable dialectal forms in German.

internal processes, the positing of a 'diffusion' model does little more than explain currently inexplicable variation.

Tropf proposes that universal patterns and variability work *together* in interphonological development. In his data, a universal sonority hierarchy scale (vowel-glide-liquid etc.) is seen as informing choice of L2 variant consonant cluster reduction. For example, in the L2 German cluster /pf/ (final position), where the plosive precedes a (more sonorant) fricative, a preference for /f/ is reported in L2 production. This bears similarities to Eckman's MDH.

However, Tropf denies that, as claimed by Eckman, these IL forms are based on L2-like underlying phonological representations. He argues that the notion of (L2-like) underlying representations being simplified on output implies a complex (and psychologically unreal) account of IL process. In other words, there is an inherent paradox in claiming that phonological form is established to L2 norms and that subsequent output rules somehow simplify that form during production. Apart from being cognitively costly, it is unnecessary. One might add that the idea of establishing a well-formed L2 representation simply in order to reduce it during production is implausible. The same argument was often put forward by writers such as Corder as a general counterclaim to the notion that ILs are no more than simplified L2s: learners cannot simplify what they have still not acquired. Tropf's claim is that, rather than being evidence of simplification processes, free-variation shows that underlying representations are constantly evolving. Progress in an IL involves adding to the underlying representations in such a way that variation in realisation is gradually reduced. The number of variants for a learner is therefore in inverse proportion to the level of acquisition of a form. Sonority is an example of one determining parameter in a developing underlying representation for each form, where those forms start out as very basic and underdetermined (at least where transfer does not intervene), and are gradually constrained.

Such models of variability are difficult to apply to studies of L2 intonation. For example, we might look at the distribution of different tones used by speakers at different stages in acquisition, and express this as a kind of variability in the tonal system. Any significant differences from target-like norms would in this sense represent 'variation', which might (or might not) reduce over time. However, this would be an expression of *variation from L2 norms*, and not *variation from an L2 form*. In order to study exactly

how specific intonational forms may be subject to greater variability in L2s, we need to assess their production phonetically. Yet even here things are not straightforward. For example, we might look at the phonetic characteristics of a falling tone in L2 data and compare it to target-like norms, looking at pitch width and steepness of fall. However, we also know that many contextual and psychological factors contribute to the production of prosody in ways which they do not in segmental production, and these factors are hard to control for.

What is needed, then, is an approach which recognises that a detailed study of L2 intonation is indeed an attempt to express *variation from L2 norms*, and that L2 prosodic production represents a complex speaker-response to linguistic and non-linguistic context. So, each datum does not simply represent a variation on a target form, as in Tropf's / niç /→ [niçt], [niʃt] etc., but a far more nebulous phenomenon. For this reason we need to account for production in as much (phonetic) detail as possible whilst retaining a close interest in context and linguistic use during interpretation of data. This, unfortunately, means that we cannot adopt wholescale any existing model of L2 phonological acquisition. We cannot express our findings purely in terms of the phonetics and phonology of intonational systems; context must inform our analysis of production here.

2.2.4 Universal IL processes

I have already referred to several studies which cite the influence of universal processes in the development of L2 phonology. The idea that language acquisition is guided by a set of language-independent processes has obvious attractions for the study of both FLA and SLA and particularly for the study of L2 intonation.

Major (1987) proposes how, in a broader sense, some SLA phonological processes may be universal. Acquisition, he argues, begins with a period in which the main changes involve the influence of L1 phonology being eliminated through new constraining rules. After this, developmental processes kick in, leading to increasingly nativelike production, and then decrease as either nativelike pronunciation or (more likely) fossilisation is reached. The following two examples are of Brazilian Portuguese (Carioca)-English and English-Brazilian Portuguese ILs (p. 216):

	BP-E :	E-BP :
Target	Eng. /r/ (prevocalic)	BP /r/ (postvocalic)
early stage	[x]	[r] (Eng.)
-	[x] ~ [χ]	
-	[x] ~ [χ] ~ [r] (a 'partial /r/')	[-] (deletion of sound)
-	[χ]~ [r]	
late stage	[r]	[r] (Por.)

In the early stages, both cases see L1 interference (Portuguese /r/ → [x]). The BP-E case then shows evidence of approximation with a developmental form, a 'partial /r/', whilst the E-BP case shows the reintroduction of a natural phonological constraining rule (postvocalic deletion). Later stages of the BP-E show progressive developmental forms, characterised first by the dropping of L1 'interference' forms and then by the dropping of less target-like forms as the developmental forms increasingly approximate to the L2 target. The symbols in the latest stage may represent somewhat fossilised forms.

However, in all-new sounds this chronology does not pertain. As with all the examples of new sounds previously discussed in the present chapter, developmental processes take over immediately here.

So, in Major's proposed chronology L1 transfer gradually gives way to developmental processes (which follow natural phonological patterns). After this, where acquisition continues at all, the number of developmental processes declines as target norms are approached. Pronunciation may fossilise early, resulting in a heavy and identifiable accent. The elimination of deviant sounds progresses through the introduction of new constraining rules. In SLA, some of these deviant sounds will be L1-related. On the other hand, the starting point of an IL may well provide much shared ground, the 'positive transfer' which often facilitates rapid progress in SLA but which, since natural languages commonly share at best only very few identical sounds, rarely assists with the attainment of nativelike levels of pronunciation.

In this sense SLA and FLA can be seen to share the basic mechanism of progressive phonological constraining. However, rather than the starting point comprising only the universal 'implicit phonetic forces' of FLA, in SLA the starting point also involves a well-formed L1 phonology. The permutations of coinciding and conflicting forms of constraining rules in L1 and L2, then, will constitute qualitatively differing conditions for the L2 acquisition of different forms.

Major notes the usefulness of Stampe's Natural Phonology (NP) framework in his discussion. NP is based around the notion of phonological 'processes', where a process is defined as:

> a mental operation that applies in speech to substitute, for a class of sounds or sequences presenting a specific common difficulty to the speech capacity of the individual, an alternative class identical but lacking the difficult property.
>
> (Stampe 1979: 1)

In FLA, these processes reflect the child's tendency towards simple sound production. Innate, 'natural' instincts are said to lead to preferences for certain physiologically easy sound productions. Stops, for example, are simply easier to produce as voiceless than voiced, and consequently a common NP process in early speech is the prevalence of the devoicing of voiced stops.

According to Stampe (1979), natural processes are most evident in the earliest stages of acquisition, where utterances typically consist of 'sequences of simple consonant-vowel syllables' which lack 'positive articulatory properties' (p.2). During acquisition, many natural processes are suppressed as the child gravitates towards a specific L1 phonology. Phonological acquisition, then, involves the gradual *unlearning* of some innate tendencies, rather than the acquisition of L1 rules. Some processes, though, may be compatible with the L1, and these will not require suppression. In addition, the child may need to refine a process in keeping with a restricted or variable occurrence in the L1, and here such 'rules' must indeed be learned by the child. Donegan and Stampe (1979) claim that, unlike rule-based accounts of phonological acquisition, NP is sensitive to the emergence of language in individuals and communities as a reflection of 'the forces implicit in human vocalisation and perception' (p. 126).

NP has been seen as particularly applicable to models of interphonology. For example, if we return to the example of final devoicing in German and English, NP offers an elegant explanation: given that obstruent devoicing is an NP process (Stampe 1979: xv), then the relative ease of acquiring final devoicing in the case of English-German IL is explained as the reversion to an innate process; the difficulty of acquiring final voicing in German-English IL, though, entails the suppression of a process which has been reinforced as an active element in the L1 phonology.

So, whereas interest in NP in its own right has waned recently (Fudge 1994), it may offer a useful mechanism for the discussion of SLA. However, Wieden (1993) points out that although NP has interesting implications for SLA, within the field of

interphonology it remains necessary to develop a model in which the role of transfer is still accounted for explicitly. For example, are processes which have been rendered latent or suppressed during FLA available for re-activation during adult SLA? Dziubalska-Kolaczyk (1987) suggests that there is perhaps a distinction to be drawn between processes which become latent during FLA but which can be reactivated in SLA, and processes which are not thus available, suggesting that processes with 'a 'universal' phonetic conditioning, i.e. which originate from process types which are not totally suppressed in the two languages in question' (p. 203) may be available for reactivation.

Processes in NP are said to fall into three categories: *fortification* and *lenition*, concerned with strengthening and weakening of sounds respectively, and *prosodic* processes.[6] So, what kind of natural processes are involved in the emergence of intonational form? Can we talk about suppressing such processes in the same way as a process such as obstruent devoicing? As I have already mentioned, intonation has often been seen in very strongly universalist terms. Perhaps, then, the nature of L1 acquisition here is somewhat different, leaving relatively more of the fundamental parameters of prosodic production intact as the child progresses from 'natural' to L1-based speech. This in turn would suggest that we retain a relatively greater 'innate' basis for prosodic than for other aspects of production, and this would have profound implications for the acquisition of intonational form in a second language.

In section 2.3 I will explore these issues, looking at how intonation and prosodic form generally develops in FLA, especially at the earliest stages.

[6] See Fudge (1994: 2761). Stampe (1979), in the original statement of NP, does not mention intonation.

2.3 Issues relating to the acquisition of prosody in FLA and SLA

The 'cognitivist' accounts of SLA (see Lightbown and Spada 1993, Ch.4) which have formed the bulk of recent work in SLA are based largely on evidence of the acquisition of grammatical morphology. One of the unique features of SLA is that learners tend to be older, hence cognitively and psycholinguistically more advanced than L1-learning children. However, advanced cognition may well not be a necessary condition for the acquisition of intonational form in either first or second language acquisition. The notion that intonation may be partly or wholly informed by factors not directly related to cognitive processes has important implications for our understanding of its development in L2s.

2.3.1 FLA: early instinct for prosody in babbling and one-word stage

The acquisition of L1 phonology is often described in terms of the infant's gradual development of a set of contrasting speech sounds. For Jakobson (1968), the development of *Kindersprache* entails the transition from babble to meaningful speech as the gradual emergence of *oppositions* within an expanding set of sounds. The sound set at its earliest stage typically comprises no more than a single vowel and consonant. Within this general framework, the acquisition of phonology is expressed largely in terms of the development of a segmental system.

However, the emergence of prosody pre-dates that of the segmental system. Prosodic features are present from the moment of the very earliest vocalisations, although the questions of what exactly constitutes prosodic *form*, and what emergent systems of *meaning* might be involved, add considerably to the difficulties inherent in the study of infant prosody.

Neonates are known to respond to changes in all three principal prosodic features (Morse 1974; Cruttenden 1979). Lieberman (1967) has also observed that the crying of neonates and young infants' babbling has a rise-fall intonation contour, which constitutes the basic pattern of vocal production at this stage. He argues that since the rise-fall is the basic unmarked contour for statements in English, as well as in many other languages, then this early pattern is itself a proto-intonation contour, and forms the foundation for

the development of intonation (see Cruttenden 1979 for a critique of this strong assertion).

Bare production and reception, of course, do not necessarily imply meaningful or systematic use. The early use of the vocal apparatus (for crying etc.) and the adult's production of meaningful intonational form share the same prosodic features. However, the extent to which these two modes of production share any broad systems of meaning is not immediately clear, although assumptions here are easy to make. For example, in a study of the intonation of Russian infants, Tonkova-Yampol'skaya (1973) finds 'intonation' signifying such states as *happiness*, *placidity* and *discontent* in the speech of neonates:

> ...even a newborn infant has a primary intonational structure....This fact testifies to the biological nature of this primary intonational pattern...
>
> (p. 135)

This 'fact' does indeed testify to the patterning of prosodic features in the earliest vocalisations, but the attribution of an 'intonational structure' here is perhaps somewhat unsatisfactory.

Children acquiring tone languages have, though, been observed to develop correct use of tones earlier than segments (Li and Thompson 1977; Kirk 1973), and this at least supports the general claim that tonal acuity emerges early. Nakazima (1962) found that at eight months infants are good mimics of pitch patterns, can repeat a pitch contour, and can also superimpose it onto a series of other (nonsense) words.

Evidence of an emerging system of tones is generally noted in one-word stage,[1] although it is perhaps not surprising that that there is disagreement over the question of possible functions of these tones. Cruttenden's data (child aged 1;1) contain uses of high falling and some rising-falling contours, but he notes no evidence of discriminate use of these. Halliday (1975), on the other hand, has claimed that intonation is present in the earliest stage of acquisition, and in his own data (age 0;9 - 1;6) he finds a developing system of tones, including one complex tone (fall-rise), with evidence that the child has begun to match certain speech acts with specific tones. He also observes discrimination between high, mid and low pitch levels in utterances at this early stage. Dore (1974) has also reported early 'primitive speech acts' in one-word stage, and Menyul and Bernholtz

[1] One-word stage typically begins at around 0;9 years.

(1969) observe falling, rising and rising-falling contours on naming, asking and emphatic functions respectively.

The claim, then, is made that intonation can be demonstrated to have an ontogenetic foundation based in the earliest crying of infants and traceable through to adult forms, and that prosodically realised 'proto-performatives' such as requests are amongst the earliest acts of meaning (Bates, Camaioni and Volterra 1975; Bates 1979; Bruner, Roy and Ratner 1982). The infant, having discovered the ability to produce a rising contour, soon finds that this can be used to elicit specific responses in his caretakers (attention, feeding).

Sensitivity to an early system of word stress also appears to begin to emerge in one-word stage. In Cruttenden's data, his 1;1 year-old repeatedly discriminated between ['dada] ('downstairs/upstairs') and [da'da] ('bath water').

Infants also develop the ability to modulate their utterances with gestural and facial expressions early. In terms of Bolinger's view of intonation, they are discovering the expressive, gestural possibilities of communication generally; the 'symptomology of up and down' (1986: 194) emerges as part of this gestural system. The fact that intonational forms tend to develop to some extent 'ready made' in early vocalisations may support the notion that they can be understood as gestural responses to the world. Crystal (1979) has referred to 'prosodic idioms' to describe relatively fixed intonation patterns which accompany the production of early linguistic forms. It is certainly not the case that adult-like patterns are mastered before any real words are learned, as was sometimes asserted in the past (for example Bever, Fodor and Weksel 1965). However, the existence of communicatively significant pitch patterns in the first two years has been widely reported (see Weeks, 1980; Abecassis and Tardy 1978).

2.3.2 FLA: the development of intonational form from two-word stage onwards

Halliday (1975) reports a clear fall/rise distinction for statement-like and command-like utterances respectively early in two-word stage,[2] and also that all utterances have whole contours from this point on. This systematic distinction of tones, it might be remembered, predates any corresponding ability on the segmental (or indeed the

[2] Two-word stage typically begins between 1;6 and 1;9 years.

syntactic) level, and Halliday interprets the phenomenon to be the earliest evidence of the emergence of two functional categories in meaningful speech: 'mathematic', a vague statement-like category, and a 'pragmatic' category, involving requests and demands. This, he suggests, is amongst the first linguistic evidence of the gradual process of the socialisation of meaning in the child.

Marcos (1987) has also noted the use of pitch direction and pitch range in two-word stage French to mark acts such as labelling, showing, giving and requesting. Pitch range in this study was consistently higher for requests than for labelling; higher pitch was also noted on the initial request than on subsequent repetitions, suggesting that the urgency of the request appeared to affect the pitch production. Pitch contours were produced systematically with communicative function early in year two, with falling and rising tones for labelling and requesting respectively. Cruttenden (1979) has also found 'functional' use of prosody in two-word stage within the emerging system of accent placement. For example, principal accent in the early 'sentences' at this stage will prefer the goal to the agent of an expression, regardless of surface position. Cruttenden also observes that intonational form at this stage 'may be the most consistent criterion for dividing child speech into sentences' (1979: 34).

During late two-word stage and beyond, prosodic form exhibits some curious developmental aspects. Cutler and Swinney (1987) express this as an apparent paradox. They note that at three years the child's verbal production is very advanced in terms of utterance-level intonational form (Allen and Hawkins 1978), in the prosodic marking of such semantic/discoursal categories as new and given (MacWhinney and Bates 1978) and in the use of stress for compounding (Wieman 1976). However, related comprehension of these aspects of intonation develops considerably more slowly. In separate studies, MacWhinney, Pleh and Bates (1985), Cowin, Mann, Schoenheimer and Berman (1984), and Solan (1980) all found that during comprehension children between the ages of 4 and 6 were unable to use correctly such prosodic cues as stress and sentence juncture to assign thematic role, to disambiguate syntax and to assign pronominal referents. The source of this paradox, Cutler and Swinney claim, lies in the misinterpretation of the child's productive abilities. Drawing on Bolinger's (1983) notion of intonation as a gestural system, they propose that an element such as 'accentual focus' occurs as a result of a primitive physiological mechanism in which bodily tension produces rises in pitch, and that speaker-excitation will usually coincide with parts of the utterance which are central to meaning. It is supposed that prosodic marking of this type will develop just as

soon as utterances become adequate to carry such vocal effects; that is, before the child has acquired complex underlying semantic representations of the world. As such representations do emerge, prosodic systems are incorporated into a more complex, 'socialised' competence, including accurate perception of this kind.

So, adult-like production can occur in children as a result of a physiological mechanisms, rather than as a result of adult-like semantic competence. This may have a bearing on our understanding of SLA in this area. Does the 'socialising' of these physiological mechanisms into adult-like linguistic competence bring with it a reduction in or depletion of the availability of the early, innate mechanism itself? There is perhaps no obvious reason to suppose that this does happen. For example, the physiological consequences of excitation naturally lead to strong points in prosodic production, with linguistic 'prominence' developing later than the prosodic phenomenon itself. However, the fact that complex linguistic uses of physical prominence do develop later in childhood does not appear to constitute a process of constraining in any way; nothing has been *unlearned* in the NP sense.

In SLA, learners do not have the opportunity to 'start from scratch', and the conditions of SLA include the existing L1. However, there may be less reason to suppose that phonological constraints have to be overcome in the case of prosody. That is, rather than hypothesise that basic physiological mechanisms of prosodic form are relatively unavailable to change in SLA because they are so early-learned and so fundamental to language development and instinct, it might be argued that their very nature as the earliest building blocks of a gestural language renders them more available to alteration during L2 acquisition.

A major difficulty in approaching these issues is the relative lack of work in this area. Indeed, part of the aim of the current study is to provide some empirical evidence of how intonational systems develop in SLA.

2.3.3 Caretaker talk and input in FLA

The input which a child receives as caretaker-talk (CT) is generally thought to stimulate the early emergence of intonational form. Compared to the norms of adult-adult speech, the CTs of various languages have been observed to contain higher F_0, greater F_0-variability, shorter utterances, and longer pauses (see Fernald *et al.* 1989; also, section

3.3.4, below). The particular characteristics of prosodic production in CTs , it has been suggested, are a result of caretakers recognising the child's evident sensitivity towards prosody at this age. This sensitivity is manifested in production (Fernald and Simon 1984; Morgan 1986), and also in reception; infants exhibit greater interest in CT, with its exaggerated prosody, than in normal adult-adult talk (Fernald 1982). Slobin (1985) and Gleitman and Wanner (1982) suggest that certain 'operating principles' in the child's early perceptual system mean that prosody may be the optimal (perhaps the only) means which a caretaker has of directing meaningful language to the child. The prosodic organisation of CT also appears to facilitate awareness of syntax; Broen (1972) found that pauses were more reliable cues to clause boundaries in CT than normal adult-adult talk, and Garnica (1977) found that pitch contours also cue clauses more overtly in CT. Perhaps most tellingly, Fernald (1984) has interpreted the prosodic structure of CT utterances as constituting particularly well-formed *gestalten*.

There appears to be a strong element of universality in CT prosody (Garnica 1977; Sachs 1977). In a study of CT in French, Italian, German, Japanese, British English and American English, higher mean F_0, higher F_0 maximum and minimum, and greater F_0 variability were all observed, as were longer pauses (Fernald *et al.* 1989).[3]

We might ask, then, whether in the language directed at L2 learners' prosodic form plays as great a role in SLA as it does in FLA.

2.3.4 Caretaker talk and foreigner talk

Foreigner talk (FT) bears similarities to the phenomenon of caretaker talk (CT), although the existence of significant differences was pointed out early (Hatch *et al.* 1978, for example).

Some close analysis of the relation between the prosodic characteristics of CT and FT has been undertaken. Mechthild and Hwang (1991) suggest that there exists 'a universal didactic competence' in all caretakers. In an interesting study of CT and FT in Mandarin Chinese, they note an expansion of prosodic features both in talk to pre-syllabic infants (CT) and in speech to non-native speakers (FT).

[3] The extent to which these prosodic features were employed differed from language to language, however.

Close analysis of prosodic structure, such as Mechthild and Hwang's, has not often featured in the study of FT, where interest has most normally centred around the use and distribution of syntactic and morphological structure. Where prosodic/intonational systems are alluded to in the FT literature, it is often without principled observational criteria. Ellis (1994), for example, mentions that intonation and slower speech rate are typical features of FT, but without referring to any specific studies or data (pp. 248-54).

Tarone (1980) observes all the following amendments to speech (taken from Hatch 1979) in FT:

slow speech rate
clearer articulation:
- final stop release
- heavier voicing on final voiced stops
- use of glottal stops before words beginning with vowel
- fewer reduced vowels
- fewer contractions

longer pauses
extra loudness
exaggerated intonation

Elsewhere, Hatch (1978) notes that longer pauses are used in FT as a means of making more salient the (following) introduction of a topic expression, and Long (1983) has noted that stress is used in the same way. In terms of L2 production, the modified L2 which nonnative-speakers (NNSs) produce when communication difficulties arise also exhibits some of these characteristics of FT (see Pica 1988). However, Pica's study exemplifies the generally inadequate approach which has often been taken in this area; the 'phonological' criteria here are not explicitly laid out, and phonological adjustments are assessed in terms of a yes/no occurrence. This is perhaps not an appropriate means of analysis for intonational form, with its degrees of realisation.

2.3.5 Formulaic language in SLA

In *A First Language*, Brown (1973) notes linguistic forms which precede the first recognisable stage of linguistic production ('*Stage 1*') generation in FLA, suggesting that they must be the product of an 'inflexible', non-generative production routine. The prefabricated or memorised nature of certain elements or 'routines' in language production has long been noted: in linguistics generally (Bolinger 1976; Fergusen 1976);

in FLA (Fillmore 1976; Hickey 1993; Peters 1977); in SLA (Bahms *et al.* 1986; Dechert *et al.* 1984; Weinert 1995)[4]; and in second language teaching methodology (Nattinger and DeCarrico 1992).

One of the defining characteristics of formulas is the manner of their realisation in speech. Raupach (1984) specifies a procedure for the identification of such 'performance units', including intonational marking, length of unit, and pause delimitation.[5] Peters (1977) claims that formulas are the result of 'gestalt' rather than 'analytical' production, and notes similar phonological criteria ('phonological coherence'): there should be complete fluency ('phonological integration'), with no planning hesitations, and an unbroken intonation contour. She notes that, in her (L1) data, particular intonation contours may form part of the lexical chunk, so that in repetitions of this unified chunk of language the intonation contour remains the same. In L2 also, production of both segmental and suprasegmental phonology is observed to be more nativelike on formulas than on normal linguistic production (Raupach 1984; Weinert 1995).

It is interesting that many of the phonological criteria used in the identification of formulas are non-segmental. The production of prosodic systems has often been associated with right-hemisphere activity; Bryan (1989) found that patients with right-side cerebral damage performed significantly worse than those with left-side damage on a whole battery of prosody-related linguistic tests, including lexical stress, emphatic stress, intonation, prosody in discourse, and the use of prosodic cues in language identification. If the right side does play a greater role in the processing of prosodic systems than in other aspects of language generation,[6] then Krashen's assertion that formulas are processed in the right hemisphere, hence are not part of central linguistic processes, may support the idea that prosodics play a major role in formulas (Krashen 1981b). This itself may have serious implications for an explanatory description of the acquisition of L2 intonation.

---

[4] The literature tends to distinguish 'routines', which are prefabricated and allow no internal variation, from 'patterns', in which one or more component may be variable in a pattern-with-slot(s) form. The transition from routine to pattern has been seen as a process in the naturalistic acquisition of both first and second languages (see Krashen and Scarcella 1978; Wode 1981). This process involves a gradual analysis of the structure of the routine within the developing linguistic competence of the learner. The discovery that one can replace certain elements in the routine with others of the same category is seen as evidence of one of the acquisition strategies used by learners.

[5] The issue of whether hesitations constitute delimiting pauses is discussed by Raupach along with the issue of how hesitation phenomena are different in first and second languages, given the different planning constraints in each.

[6] Bearing in mind the relative nature of this assertion (see Danesi 1988).

2.3.6 Critical period

Oyama notes that it has long been 'part of folk wisdom' that early training has a positive effect on second language learning (Oyama 1976: 261). This has been a particularly intrusive notion in the case of the acquisition of phonology, where the idea that post-pubertal students lose their ability to master an L2 phonology and its attendant articulatory systems has long been cited as a justification for 'starting early'. Penfield and Roberts (1959), for example, advocated early L2 teaching, citing the greater 'plasticity' of the young brain.

The idea that there is a critical period for the acquisition of language (see Lenneberg 1967) provided researchers in SLA with a useful explanatory concept for investigations of qualitative shortfalls in attainment in L2s. Indeed, Lenneberg himself mentions foreign languages in *Biological Foundations of Language*, suggesting that whereas second languages can be learnt after puberty because the 'basic language skills' have already been acquired, foreign accents are rarely overcome (p. 176). Evidence from aphasiology in general supports claims of a critical period. In aphasias resulting from left-side lesions, whereas the right-side can take over the (usually left-side) language functions of a damaged pre-pubertal brain, this ability declines steeply through puberty (Obler 1981). Such a decline is usually taken as explanatory neurological evidence for the inverse relationship between age and eventual attainment in an L2, especially its sound system, and both critical age, and evidence from aphasiology, have often been used in discussions of SLA (Neufeld 1979; Patkowski 1980; Taylor 1986; Johnson and Newport 1989).

However, research into the acquisition of L2 sound systems and age has shown contradictory findings. Obler (1981) reports on research (Gaziel *et al.* 1977) in which right-handed Hebrew-speaking children appeared to be more extensively lateralised (left hemisphere-reliant) in their L1 than in their L2, as attested by observed responses to both written and audible language. Increased reliance on the left-side developed more slowly over time. Snow and Hoefnagel-Hohle (1977) claim that the ability to mimic L2 Dutch nonsense words shows no critical upper age limit, but on the contrary increases over time (their oldest subject, aged 31, was the best mimic). Looking at pre-pubertal development, Ervin-Tripp (1974) and Ekstand (1976) claim evidence that older children

acquire nativelike pronunciation faster than younger children, Ervin-Tripp comparing 4-6 and 7-9 year-olds and Ekstand 9-11 year-olds. On the other hand, Fathman (1975) claimed superior pronunciation in 6-10 than 11-16 year-olds, and in a later study (1982) noted that, in naturalistically acquired L2s, levels of accent retention are higher for those who began acquisition at a younger age.

Krashen *et al.* (1979) explain these early findings as evidence that whereas younger learners may take some considerable time to 'catch-up' (and often to surpass) older learners in acquisition of syntax, 'phonological competence' in younger learners develops very much more quickly (p. 579, n. 4), a point which Snow and Hoefnagel-Hohle also make. Indeed, Krashen has claimed that lateralisation may occur at around 5 years (Krashen 1973), and claims also that in terms of L2 pronunciation the negative effects of early lateralisation are clear in the literature. In a more recent survey of research, Johnson and Newport (1989) accept that most research has confirmed a childhood advantage in the acquisition of phonology.[7]

Whereas Krashen's claim that older is faster but younger is better (see Krashen 1981a; 1981b) has been roundly challenged as a general claim, few deny that it does appear to hold for pronunciation. Walsh and Diller (1981), for example, claim 'overwhelming behavioral evidence' that no critical period intervenes in SLA, although they concede that the exception is phonology/pronunciation. They explain this in terms of the advancing consolidation of neural architecture by six to eight years, making the acquisition of new phonological systems, and especially the new motor-articulatory processes which this entails, difficult. They claim that, neurologically, SLA in early childhood is qualitatively different from pubertal/post-pubertal acquisition, and that the earliest starting ages are indeed the most effective for mastering pronunciation; post-puberty, they suggest, is a more 'pertinent' time for the 'integration of higher order linguistic processes' (p.14). By proposing a lower/higher distinction for linguistic processes, Walsh and Diller can explain why pre-pubertal learners excel in L2 pronunciation, whereas post-pubertal ones appear (according to their evidence) to acquire syntactic and (particularly) semantic aspects of L2 structure more effectively;

[7] Johnson and Newport's own experiment confirms that younger learners also have an advantage in terms of ultimate performance of various grammatical rules, with native-like performance eventually occurring only in the English L2 of immigrants who had arrived in the USA on or before their seventh birthday. More pertinently for their study, maturation is seen to be reflected in a lowering of performance only up until puberty, after which further decline is only very gradual, possibly describing a general decline in human learning ability.

> Lower order processes such as pronunciation are dependent on the early maturing and less adaptive macroneural circuits, which makes foreign accent difficult to overcome after childhood. Higher-order language functions, such as semantic relations, are more dependent on the late maturing neural circuits, which may explain why college students can learn many times the amount of grammar and vocabulary that elementary school students can learn in a given period....
>
> (p. 18)

Scovel (1988) has proposed that a critical period exists for phonology only, and that there is a specific evolutionary reason for this. With the onset of puberty the individual becomes a potential contributor to the gene pool of his or her population. Characteristic vocal production, it is argued, is one key means of group-identification, and the solidification of phonological ability at around this time may thus be explained in evolutionary terms.

Another recent development in the literature on critical period is Major's suggestion (1997) that what cannot be fully acquired after puberty is a second language *per se*. He proposes that, whereas this usually results in an L2 being imperfectly acquired, two further scenarios can emerge: replacement of L1 by L2, with the original L1 deteriorating to nonnative levels, the old L2 becoming the new 'first' language of the speaker; L1 deteriorating but L2 not becoming nativelike. The influence of an L2 upon the L1 phonology has been reported previously in the literature (Flege 1987b). Such a scenario of L1 deterioration may be seen in the speech of long-term immigrants. The role of the critical period, Major argues, is in representing a deadline for the acquisition of two or more languages to nativelike levels, after which the adult mind can only retain one such competence.

2.4 SLA and prosody

In the literature, non-segmental phonology is often seen to be inadequately acquired in L2s. Johanssohn (1978) found that errors of prosody represented a greater degree of observed deviance than segmental ones in a study which attempted to rank the 'gravity' of errors in L2 English. Backman (1979) found L2 intonation in Venezuelan Spanish-English IL to be consistently too narrow in pitch range. There seems to be a widespread belief that L2 intonation is highly resistant to acquisition (see Grosser 1993). This belief may itself contribute to the widespread opinion amongst language teachers that intonation is 'a difficult subject' (see section 1.4). In the present section I will review in some detail the recent literature in this area.

2.4.1 Problems in the assessment of L1 transfer

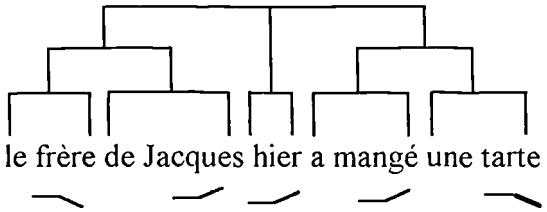
Voice-setting has been seen to transfer from L1 in L2 production (Esling and Wong 1983). Færch, Haastrup and Phillipson (1984) note that the flat intonational contours of Copenhagen Danish transfer to L2 English. They also noted that their Danish-speaking informants produced higher than normal levels of pitch on unstressed syllables in L2 English, again reflecting Danish prosodic structure. However, they also comment that overall reduction in pitch range seems to be the result of insecurity or tentativeness in L2 speech (p. 125). So, L1 transfer is not identified unequivocally here.

Kelm (1987) looks at the production of 'contrastive emphasis' (what I have referred to as contrastive prominence) in American English-Mexican Spanish IL, noting that, compared to English, in Spanish there is significantly less expansion in pitch height on contrastively emphasised elements (relative to that used in normal, declarative utterances). In his data, contrastive marking in L1 Spanish is likely to be realised using a mixture of prosodic and syntactic or lexical means, whereas in L1 English, prosodic devices are the principle means of marking contrastivity in comparable tasks. When the English speakers, who were advanced speakers of Spanish and had been immersed in a Spanish-speaking environment for over a year, completed similar tests in Spanish (their L2), their native English levels of prosodic marking of contrastivity were transferred to L2 production. Kelm concludes that in the case of English-Spanish IL, intonational form will not automatically be acquired through long and continued exposure alone. He also

notes that, although many writers have commented on the importance to effective communication of nativelike intonation, this area is left largely untouched by language teachers.

There are several aspects of Kelm's study worthy of note. First, it concentrates on a single element of production, which allows him to describe in some detail the comparable systems of production of that element in both languages. Secondly, contrastive prominence, as I have noted, is reasonably straightforward to elicit, and seems to function in comparable ways in the two (intonation) languages concerned (see Chapter Seven, below). However, Kelm's conclusions are too broad. Ideally, we need to examine production of a number of different prosodic systems in order to build up a picture of the acquisition of intonation.

In section 2.2 I noted that the influence of the L1 is complex; data which are initially puzzling may in fact be explained if one's model of L1 interference is sufficiently detailed. Lepetit's (1989) findings on the acquisition of L2 French intonation by speakers of Canadian English and Japanese support this assertion. He uses groups representing different stages of overall L2 acquisition, and analyses the production of French 'melodic contours', which are essentially series of falling and rising tones within an utterance, but described at three levels of syntactic structure (see Martin 1981). The lowest prosodic level (the rhythmic foot) carries tones which reflect how the corresponding syntactic unit is to be interpreted relative to the following unit. For example, rising tones mark syntactic units on the same hierarchical level, whereas where two units correspond to a subdivisions of a larger syntactic unit, this is marked in the first such unit by a 'melodic slope inversion'. This will usually be a fall. However, when a syntactic unit is divided into two sub-units of which the second takes a terminal fall, the inversion results in a rising tone in the first unit of the two:



(p. 401)

This framework is interesting for several reasons. Firstly, Lepetit can elicit data which conform to a predetermined model of target-like production. Secondly, the nativelikeness of L2 production can effectively be quantified in terms of appropriate tone selection, which has great advantages for acquisitional research since the broad fall/rise opposition is not likely to result in inaccurate analysis due to 'borderline' decisions. Finally, the use of tone in linguistically sensitive ways at various levels of structure provides evidence of a very complex planning in the production of intonation. I will discuss this in relation to Grosser's work (below), and indeed I will use a similar approach in my own analysis of tone.

In Lepetit's findings, neither the English-speaking nor Japanese-speaking informant groups showed any increase in accurate production of these contours throughout the ability-ranked groups. Like Kelm, Lepetit comments that his informants had never been taught these aspects of French. He also provides interesting findings concerning transfer. For example, on the first (lowest) level of structure of melodic contours, a systematic difference in typical patterns in English and French (in the same linguistic conditions) does indeed appear to interfere with the English speakers' L2 French intonation,[1] with incorrect productions reflecting the L1 of the speakers: the French rise-fall on two lowest-level units (indicating their membership as equal constituents at the immediately higher level) was produced by the English-speaking informants as fall-fall. This corresponds to Martin's own description of such a sequence in English within the same dependency context. However, such evidence of transfer can be misleading. For example, Martin specifies that both English and French have the same rise-(terminal) fall pattern, yet in Lepetit's study the L2 French of English speakers still shows an average of 26.1% incorrect patterns here. This and other evidence prompts Lepetit to note that whereas cross-linguistic influences on L2 intonation are of 'central

[1] The study is made simpler because Martin's original description of this theory of melodic contours was presented for both French and English (Martin 1981).

importance', the phenomenon is not a unitary one, and that we 'should not underestimate its degree of complexity' (p. 407). In particular, his findings suggest that experimental data in this area should be interpreted in the first instance as evidence of nonnativelike production, and only as evidence of *possible* transfer, even when L1-like elements are identified.

2.4.2 Perception

The perception of sounds must to some extent precede production in L2 acquisition. Gimson bases his EFL pronunciation manual on this idea (Gimson 1975), and Cruttenden (1994) has recently reiterated the point with particular reference to the acquisition of intonation. Nonnativelike perception of L2 intonational form has been seen as a key factor in the poor acquisition of productive skills. Scuffil (1981), for example, found that German speakers of L2 English had difficulty matching recordings of English sentences to their appropriate contexts. However, as de Bot has pointed out (1986), the very task of matching uttered speech to a specified context (or disambiguation tasks incorporating intonational meaning) is fraught with methodological problems; informants often need to possess a reasonably advanced level of L2 simply to perform such tasks, which usually involve paraphrased meanings or other linguistic interpretation. Also, it is not always clear what perception tests tell us about the linguistic or communicative ability. In Scuffil's tests, for example, native English speakers also performed the tests, and showed inaccuracies in their perception of their own L1.

On the other hand, Luthy (1983) shows how a group of highly proficient (>TOEFL 500) English L2 speakers were significantly less accurate than native English speakers in matching attitudinal and discoursal meanings to short, decontextualised phrases where segmental information was not available, mis-perceiving items more than one quarter of the time, whereas the native speakers showed consistently high levels of correct perception (97.33%).

Lehiste and Fox (1992) make a highly controlled assessment of how Estonian and English native speakers perceived the relative prominence of a series of nonsense syllables, where one syllable in a series carried either increased duration or amplitude, with F_0 held constant throughout the series. For Estonian speakers the durational cues were stronger, whereas the English speakers responded more often to amplitude cues.

Also, in a similar test using 'nonspeech' sounds (noise), where F_0 was again constant, all speakers showed a higher degree of sensitivity to amplitude and durational cues, which is predicted by the authors, and is seen as a result of the fact that F_0 is expected to be a cue of prominence in speech and, hence, its absence has a lesser effect on nonspeech stimuli. So, although speakers show similar levels of sensitivity to the perception of relative prominence, L1 is seen here to influence the nature of perception in a very fundamental way.

The exact nature of how these L1 perceptions inform 'perceptual acquisition' was addressed by Leather with reference to tone perception (Leather 1987), looking at speakers of non-tone languages (English and Dutch) and their ability to perceive linguistically significant tonal contrasts in an unknown tone language (Mandarin). Leather notes that the development of acquisition, particularly of productive ability, is founded upon a prior perception of phonetic contrasts. As I have noted previously, one difficulty in assessing varying aspects of the acquisition of non-segmental phonology is that the central prosodic features are always present in both L1 and L2 in some configuration. Leather, too, notes that whereas the Chinese tone distinction under observation (Chinese tones one and two, level and rising respectively) were 'new' as lexical tones to English and Dutch speakers, they are only new phonologically, and 'would not be unfamiliar in either articulatory or acoustic terms' to the informants. He observes that whereas it is reasonable to expect previously established L1 tonal perceptions to influence the ability to deal with these new tonal contrasts, 'there is no theoretical basis for detailed prediction of the nature or extent of such transfer' (p. 64).[2] All informants' perceptions were less accurate than those of a Chinese-speaking control group, but did nevertheless show evidence of perceptions which were broadly the same as those of the Chinese speakers. No L1 transfer was positively established.

In a more recent experiment, Leather (1997) assesses the acquisition of Chinese lexical tones by two groups of Dutch speakers undergoing one of two different learning treatments: one group followed a programme of 'perceptual learning', in which digitised speech and computer graphic representations were used to train students to perceive tonal contrasts successfully; the other group learned the same tone set 'cold', in which repetitions and pitch-line visualisations (of their own productions) were used to foster learning through production only. Both groups were subsequently tested for accuracy of

[2] A point which I made above in the discussion of Lepetit's findings

both perception and production. About half of the 'perceptual' learners, when tested for production, were able to produce a tone set with the correct tonal contrasts, and those learners most successful in the production of tones also tended to be highly successful at perception. However, the 'productive' learners also developed perceptual ability, despite having received no perceptual training. So, although the interrelationship between perception and production is a close one, it appears that it is not necessary to receive training in perception in order to develop accurate perceptual models of L2 tones, and neither is training in the production of tones necessary for a perceptive ability to develop. Leather's findings also suggest that tonal universals may affect SLA in the same way as they do FLA: his data reveal that Chinese tones 2 and 3 (rise and fall-rise respectively) constitute the most difficult contrast to acquire. He notes that this relative difficulty has been observed before in both the SLA (Kiriloff 1969) and the FLA (Clumeck 1980) of Chinese. We will see, below, that universal tendencies have been noted in several studies.

2.4.3 Cruz-Ferreira

Cruz-Ferreira's (1983; 1987) work addresses several of the issues so far raised in the current discussion on SLA and intonation. Her approach to the question of perception and L1 is a more inclusive, linguistic one than much other work. She sees intonational form as determining meaning in utterance through its interaction with lexico-grammatical information. Indeed, she comments that L1-L2 differences may lead to a 'mismatch of syntactical and intonational forms', which renders L2 speech nonnativelike even if intonational patterns themselves have been mastered (1987: 103).

In Cruz-Ferreira (1987), speakers of L1-Portuguese/L2-English and L1-English/L2-Portuguese were tested for L2 comprehension, using a series of double paired tests in which sentence pairs were superficially identical, with differences in meaning conveyed by intonational form alone:

sentences:	a)	they've left the \children
	b)	they've \left / the children
glosses to be paired to sentences:	a)	somebody has left the children
	b)	the children have left

(Cruz-Ferreira 1987: 108)

Here, sentence a) is paired with gloss a). Sentence pairs with differences of meaning based on different illocutionary force and different attitude were also used. Cruz-Ferreira proposes three central nonnative interpretative strategies for intonation: transfer, pitch height, and lexico-semantic strategies.

Transfer can be positive. For example, in the sentence pair above, unit divisions distinguish different syntactic structure. This is a feature of both English and Portuguese. Transfer can be negative, as in the use in Portuguese of low-fall or high-fall to distinguish positive and negative presupposition (respectively), which was not comprehended by English speakers, because although the intonational pattern was familiar, its linguistic use was not.

Pitch Height strategy is a rather broad notion. She suggests that 'pitch-dependent' decisions based on L1 norms can inform comprehension, even when particular forms are all-new in an L2. For example, 'command' versus 'warning' is marked in both languages with tones which differ in the extent of their highest point, with the 'warning' tone typical of each language being relatively higher than the 'command' tone. Even when the particular tones used do not occur in the L1 (for example, the low-low tone which typically marks commands in Portuguese is all-new to native speakers of English) a 'pitch-dependent' decision based on the association of pitch height with relatively more 'open' or 'closed' meanings can be construed in an L2. In this case, the typical fall-rise of a warning in English is higher than a fall. The interpretative use of 'gross phonetic shape', she notes, may reflect the operation of intonational universals here.

Lexico-syntactic strategy involves the association of less marked linguistic meanings, based on 'first interpretation' (or 'high bias' interpretation) of linguistic form, with changes to the intonational form being ignored. This occurs where no comparable L1 intonational form exists. Portuguese speakers of L2 English, for example, do not comprehend the shifting of a nucleus from rightmost position to a prior epistemic verb as a means of switching polarity in a sentence:

> I thought she was married (but in fact she isn't)
> I thought she was married (and indeed she is)
>
> (p. 113)

In such cases, either the least marked lexico-syntactic meaning is taken, or, where no such high bias interpretation is evident, a random interpretation is made from possible meanings.

Central to Cruz-Ferreira's study is the role given to linguistic decisions in assessing intonational form, in particular the association made between judgements of familiarity of 'structures' (lexico-grammatical form) and familiarity of 'patterns' (intonational form).

The comprehension of L2 forms tells us something about how production may develop. Cruz-Ferreira's findings suggest that L1 may be involved in the comprehension of contextualised speech where two conditions obtain: when there is a direct similarity in L2 and L1 between intonational form and linguistic function; where an unfamiliar linguistic structure nevertheless carries a familiar overall intonational pattern, in which case any gross phonetic similarities with L1 (possibly intonational universals) will be sought, and meaning through this slightly abstract L2-L1 comparison arrived at.

One recurrent comment in the literature is that intonational form is not taught to learners systematically. Cruz Ferreira's study begins the task of specifying what needs to be taught, by describing the conditions under which L2 intonation will not be comprehended correctly.

2.4.4 Studies of German-English

Wieden's (1993) quasi-longitudinal study of L2 English prosody is particularly relevant to the present study in that in it he attempts to look at evidence from the developing acquisition of the use of the separate prosodic features. His data on prosody are taken from an earlier, large-scale study of 400 Austrian-German speaking schoolchildren learning English (reported in Wieden and Nemser 1991). He notes that the data appear to show three distinct stages in the acquisition of intonational form.

At the earliest stage, the use of amplitude is observed not to have 'linguistically motivated modulation' but rather to be used in a kind of extreme syllable timing, with each successive syllable marked with similar levels of amplitude. F_0 was also thought to lack linguistic motivation, with only a very small number of fixed contours - predominantly a 'steep rise-moderate fall', which Wieden sees as a possible default pattern at this stage. Although occurring at an early stage in acquisition, these

phenomena do not reflect L1 norms. On the contrary, Wieden suggests that inherent physiological constraints motivate their use. He suggests that early L2 acquisition, rather than relying heavily on L1 models of production, might in fact include the (re)activation of those natural processes identified in Natural Phonology as innate production processes.[3]

During the intermediate stage, pitch contours exhibit 'pitch time rising', a series of short, regular rises with a stable highest point. Wieden cedes that this may be evidence of the marking of doubt, or the transfer of L1 enumerative intonations, but, alternatively, may be evidence of a gradual increase of control of pitch reminiscent of child speech. Accent (prominence) also develops in this stage; gradual differentiation in levels of amplitude creates utterance-level prominences. The processes seen as typical of intermediate stage are also generally non-target like in nature, and Wieden explains them as either reflecting L1 norms, or as evidence that approximations to the L2 prosodic structure are being made through attempts to overcome inherent constraints.

At the advanced stage, accent and intonation develop. Such features as the maintenance of the same pitch over a series of syllables (evidence of deliberate manipulation rather than natural, physiologically-determined production) and evidence of use of nuclear prominence emerge.

One issue which Wieden cannot address, given the nature of his informant base, is the question of whether the putative 'inherent constraints' on early stage productions (such as the use of the rise-gradual fall contour) are processes which diminish or disappear into adulthood (his early stage informants were in the range of 7-9 year olds), or whether they remain available to adult learners. If not, then perhaps adult acquisition here may side-step the need to revert to these kinds of (non-targetlike) productions in adult L2 acquisition. On the other hand, if natural prosodic processes, in the Natural Phonology sense of 'process', do indeed become constrained by the L1 over time, perhaps they block (L2) acquisition here, in that acquisition entails passing through such processes as a necessary part in the acquisition of L2 forms themselves. Such questions have so far not been addressed in the literature.

Grosser (1993) assesses the production of accent, and also one aspect of tonal development, in a longitudinal study of 10-12 year-old Austrian-German speaking

[3] He also cites anecdotal evidence that the recorded speech of his youngest (and least experienced) speakers was adjudged by native English speakers to sound like utterances of native English-speaking children.

students of L2 English. Noting the very similar accent assignment rules for German and English, he describes patterns of accent placement which are neither target-like nor related to L1 patterns. The first of these is 'cumulative accentuation', in which most or all words receive an accent. This is clearly not a process transferred from L1, and he in fact cites Wieden and Nemser's similar findings (1991) in concluding that this reflects some sort of non-linguistic, temporal patterning, going on to suggest that it may reflect a 'quasi-mechanical, word-oriented production principle' (p. 84). Hewings (1993) has discussed a similar production strategy in nonnative speech, and suggests that a process of *word selection* is preferred by NNSs to the normal prosodic planning based on *sense selection* of whole units (p. 158).

Grosser finds that 'backshifting' of accents occurs, where rightmost elements in a unit are accented despite the contextual inappropriateness of such accentuation. Conversely, a pattern of inappropriate 'fronting' of accents is also observed. In the first year of acquisition, inappropriate 'backshifting' (too far right) accounts for 70.7% of errors whereas fronting only 29.3%. He claims that this might result from general L2 variation here, although noting such frequencies might also reflect the influence of teaching input. I have previously discussed both the 'transfer of training' and effect of variation in this chapter.

Grosser claims that, despite similarities between L1 and L2, such as their accentuation rules, these similarities are not immediately drawn upon in L2 production. Rather, 'basic production constraints' may lead to such phenomena as accent flooding or accent alternation, especially at early stages of acquisition, although unlike Wieden (1993) he does not go on to examine this claim in any detail.

Grosser's study is also one of the few studies to examine tonal production developmentally. Since I will take a very similar approach in my own analysis of aspects of tonal production in Chapter Five, I will outline Grosser's findings here. He looks at points in a reading where two clauses (uttered as two intonation units) have a co-ordinate relationship. Where the second clause carries a falling (terminal) tone in utterance, the first may have a fall or a rise, although the 'open' sense of a rise more clearly anticipates the following information, and hence represents the marking of a more cohesive sequence of clauses (see also section 1.3, where I discussed tonal dependencies of this kind). In data from learners of L2 English with between one and two years' learning experience, a very clear preference for a falling first tone (at one year's learning, 76% falling, n=120) gradually declined until, after two years, this preference had

disappeared (at two years' learning, 54% falling, n=120). Learners' productions of two-unit utterances, in that they were increasingly produced with a rise-then-fall sequence, were evidence of a developing ability to use tone in textually cohesive ways across more than one unit.

However, the roughly equal preference for fall and rise in the more advanced learners was also seen to reflect very similar preferences by same-aged speakers of both English and German. So, this development in the use of tones may represent an approximation to the L1 model, where this move towards L1 follows the earlier influence of innate (i.e. non-L1 specific) processes. Perhaps, after some early 'natural' processes, it is the L1 which kicks in.

To examine this further, falling tones were analysed as either 'downstep' (a typical German, slightly flattened fall), and a 'peak type' (a typical higher English fall). In L1 English speech, the preference for a 'peak type' fall is very pronounced (75% in Grosser's data). Taking German informants with between one and two years' L2 English learning, their production of falls in similar contexts in both their L1 German and L2 English was examined. It was found that in both L1 and L2, a preference for downstep falls decreased, at roughly the same rate. So, although the L2 model seems to have been approximated to, we in fact see that the change in frequency of use of these two variants of the fall reflects very closely similar changes which occur in the L1 at this age (11-12 years). Also, the fall-rise variant of the rise, which was observed in the L1 English data, was observed to no significant extent in either L1 German or L2 English data, providing negative evidence for approximation towards L2 of this particular tone.

So, we once again see that apparent L1 transfer may have alternative explanations when sufficiently detailed analysis is conducted.

2.4.5 'Discourse intonation'

The two studies which I will discuss here, Wennerstrom (1994) and Hewings (1993), adopt very different approaches, but are both concerned with the analysis of a variety of systems of intonation. Both take a broadly discoursal interpretation of the functions of intonation which, as I have said, will be the approach which I adopt.

Wennerstrom provides a detailed and highly specific account of several aspects of intonational form in her study of Thai, Japanese and Spanish speaking learners of L2

English. She uses a clearly defined discourse model, which allows her to identify discoursal uses of intonational forms with some certainty. I will discuss her methodology in some detail in Chapter Three. Her analysis involves data from both text readings and free speech, and considers production of the following aspects of intonational form, largely using Pierrehumbert's (1980) model of intonation:

1. high phrase accent[4]
2. high/low pitch accent[5]
3. paratones (I have discussed these in Chapter 1)
4. high boundary tone[6]

Instrumental readings provided data on group tendencies in the prosodic marking of these forms. For example, F_0 and dB readings were taken on a word conforming to the discoursal function of high phrase accent (marking a unit as topic-establishing). Readings were then compared to a word which did not have this discourse-marking function. Where significant differences in F_0 and dB were noted throughout the native-speaking group, these differences were taken as indicative of nativelike tendencies. The other categories, above, were measured in similar ways. For example, a newly-introduced word, bearing the high pitch accent marking of 'new' information, was compared to a subsequent repetition of that word, bearing the low pitch accent of 'given' information. Thus, the *English* column in Table 2.1, below, shows whether in each case F_0, or dB, or both, were consistently used throughout the NS group to mark particular features in this context:

[4] In Pierrehumbert (1980) the 'phrase' level represents an intermediate level of tone structure between the tone ('pitch accent') level and the utterance ('boundary tone') level.
[5] high/low tone used to mark information as new/given in the discourse.
[6] A high tone signalling incompletion of discourse, such as rising boundary tone, inviting further discussion of the topic.

	English	Spanish	Japanese	Thai
high phrase accent (F_0)	p=.024	p=.001	p=.003	-
high pitch accent (F_0)	p=.043	-	-	~~p=.012~~ *
low pitch accent (F_0, dB)	p<.001 (F_0) p<.001 (dB)	- p=.030 (dB)	p=.024 (F_0) p=.045 (dB)	-
paratones[7] (F_0)	p=.005	p=.003	-	-
high boundary tone[8] (F_0)	p=.002	p=.035	p=.024	-

TABLE 2.1 **Wennerstrom's findings in tabular form: presence of significant pitch and amplitude marking in text reading by different speaker groups**

* ~~p~~ indicates that pitch here was significantly lower, rather than significantly higher (as in the target production)

The 'overwhelming consistency' of the native speakers of English in realising these phonetically marked intonational contrasts is said by Wennerstrom to be evidence that such aspects of intonational form have 'morpheme status' in English speech. The range of target intonational forms, then, can be seen as a set of items, which lends the study some similarities to the developmental approaches to the acquisition of grammatical morphemes discussed earlier. Wennerstrom then, can state which of these 'morphemes' has been acquired.

In Wennerstrom's analysis of free speech, however, significant problems arose in the native English-speaker data, where intonational 'morphemes' such as pitch accent were not seen to be realised as contrasts. The analysis of high and low boundary tones in free speech, though, is very interesting. Sequences of phrase accent and boundary tones were considered together as yielding quasi-contours of 'high' (high phrase accent plus high boundary tone), 'plateau' (high phrase accent plus low boundary tone), or 'low' (low phrase accent plus low boundary tone). Overall frequencies of the use of these contours were calculated. It was found that, whereas both native English-speaking and Spanish-speaking informants produced very similar percentages, the Japanese and Thai speakers significantly overused low tones and underused high tones, the Thai speakers' percentages being most deviant from target norms:

[7] Mean ranges of f_0 and dB were compared for a paragraph-initial and a paragraph-medial sentence which was identical apart from the final word, and different means of f_0 seen to reflect different levels of paratonic structure (see Wennerstrom 1994: 405, 412)

[8] Measured as extent of significantly higher f_0 on a questioning discourse-final boundary tone than on other boundary tones within the text.

	'High'	'Plateau'	'Low'
English NS	40	48	12
Spanish	43	44	13
Japanese	25	31	44
Thai	14	39	47

TABLE 2.2 **Percentage distribution of the three quasi-contours in the free speech of informants in Wennerstrom (1994)**

Wennerstrom notes that such nonnativelike overall production of tones might have serious consequences for the comprehensibility of L2 speech, where inappropriate cues are given regarding the status of information in utterance. Such findings also suggest that speakers of tone languages encounter greater problems in the perception and production of intonation languages, with speakers of pitch-accented languages (such as Japanese) encountering an intermediate level of difficulty, and those speaking other intonation languages(such as Spanish) facing least difficulty. Such a conclusion, though, is speculative, although it does appear to be supported in general by Wennerstrom's findings. In Table 2.2, above, the Thai speakers perform considerably less well than the Spanish and Japanese speakers, although in these latter two groups neither is clearly more nativelike in their production than the other.

Wennerstrom's findings afford no insight into developmental features of acquisition. Also, the selection of informants was based on a placement test carried out for a University English programme, and although this ensured that all informants were of a comparable overall level of English proficiency (intermediate), there were significant differences between the three groups in terms of language background and length of stay in the USA at the time of the experiments. Nevertheless, the study provides a principled experimental framework in an area where such approaches have not previously been employed, and I will draw substantially from Wennerstrom's approach in several areas of my own experiments.

The final study to be reviewed in this chapter is that of Hewings (1993). Again, I will refer to the methodological approach in the next chapter, and will here discuss only the findings.

Adopting a descriptive approach based in Brazil's 'discourse intonation' framework (see Brazil 1975), Hewings assesses production of L2 English intonation by

native speakers of Greek, Indonesian and Korean. His data are collected from a series of readings of scripted English dialogues, performed in pairs by informants, and also from the free speech of the same informants. Under consideration are tone unit length, number and distribution of prominences, tone choices, use of key and termination. His findings indicate that, in comparison with native speakers, nonnative speakers:

- use shorter tone units
- flood speech with prominences
- tend to place prominence later in the unit than is appropriate
- produce relatively fewer fall-rise tones and more rises than NSs
- produce more level tones than NSs

Hewings accounts for his findings in terms of five main influences:

- lower capacity of NNSs to plan ahead during speech
- tendency for NNSs to 'disengage' more frequently
- influence of language classroom procedures
- NNSs' lack of familiarity with L2 lexis
- NNSs' lack of familiarity with exploitation of intonational devices in social conventions

(Hewings 1993: 302-4)

2.5 Summary

At the beginning of this chapter I noted that the discussion of three central issues has dominated recent work in interphonology: the nature of transfer, the nature of IL phonological rules, and universal IL processes.

It has been claimed that L2 sounds are not in themselves unavailable to the adult learner. Where the presence of an L1 does not impede a new L2 representation, there appears to be no problem in either the perception or the production of such sounds. Where L1 sounds are similar, a phonological conflict arises which, as we have seen, may result not only in an approximated L2 sound, but also in some disturbance in the original L1 sound. In an extension of this idea of L1-L2 reciprocity, Major (1997) has gone on to suggest that any L2 can become nativelike, but will necessarily involve deterioration of

the L1. Flege (1997), taking a modular view of speech skills, has also claimed that the potential to acquire language does not in itself deteriorate over time, as evidenced by new sound acquisition.

In terms of intonational form, the implications of research are more difficult to interpret. How, for example, does 'equivalence classification' take place in the case of the phonological representation and phonetic realisation of prominence, or in the use of tones? Do different uses of prosodic form for linguistic and discoursal purposes exhibit predictable developmental patterns in SLA? Are these patterns influenced by L1 or by intonation universals? Prosodic features are present in all languages (unlike, for example, certain segmental features), but does this constitute an advantage or a hindrance to the acquisition of correct L2 intonational form? We have seen that purposeful use of basic intonation develops very early in infants. Does this mean that L1 intonational form is more resistant to alteration in a late-acquired L2, given that such acquisition constitutes a very difficult uprooting of the earliest-learned communicative/productive vocal instincts?

As the above review of work on L2 intonation illustrates, there are a variety of issues relating to experimental design in this area which simply do not arise in many other areas of acquisitional research. For example, we can claim to be interested in 'intonational form', yet there are a number of 'systems' within the intonational whole which can be studied. In Chapter One I described those intonational systems which make a reasonably clear contribution to the semantic and discoursal structure of speech, and which consequently will form the basis of my own experiments here. In Chapter Three, below, I will discuss this in terms of what I will *not* investigate here. Not least is the problem with the nature of data themselves. Intonation is very awkward (and time-consuming) to analyse. Wieden (1993), for example, analyses the recorded speech of 400 informants in terms of segmental production, yet assesses the intonation of only four of these informants. Other issues also distort any possible general conclusions: informant profiles vary considerably across studies; target-language control groups are not always used in empirical research; decisions as to what type of study (longitudinal etc.), form of analysis (auditory or instrumental), and indeed form of data collection, have all contributed to a general lack of comparability in the literature. Table 2.3 represents the more recent studies of production (not perception) of L2 intonation reviewed in this chapter, together with a break-down of the nature of the experimental approach in each case.

Study	What is under observation	Informants			Languages		Type of study			Type of analysis			Data elicitation		
		Total number*	Age-range	Use of an L2 control group	L1 of informants	Target language (L2)	Single group analysis	Cross-sectional	Longi-tudinal	Auditory	Instru-mental	Both	Free-speech	Elicited speech	Both
Færch, Haastrup and Phillipson 1984	Pitch patterns; tones; pitch range; prominence placement; word stress	123	12-25	NO	Danish	English	-	✓	-	✓	-	-	-	-	✓
Kelm 1987	'Contrastive emphasis'	18	17-27	YES	English	Spanish	✓	-	-	-	✓	-	-	✓	-
Lepetit 1989	'Melodic contours' - use of intonational cues (tones) to mark syntactic structure	75	c. 19-23	NO	English Japanese	French	-	✓	-	-	✓	-	-	✓	-
Leather 1997	Acquisition of tones in Mandarin under different learning conditions	c. 50 (not stated)	19-28	YES	Dutch	Chinese (Mandarin)	-	-	✓	-	✓	-	-	✓	-
Wieden 1993		4	7-15	NO	German	English	-	-	✓	-	✓	-	-	✓	-
Grosser 1993	prominence placement; tone choices in multi-unit utterances	8	10-12	YES	German	English	-	-	✓	-	-	✓	-	-	✓
Wennerstrom 1994	Use of pitch/loudness to mark elements of discourse structure	30	adult	YES	Spanish Japanese Thai	English	✓	-	-	-	✓	-	-	-	✓
Hewings 1993	Prominence, tone, key and termination according to Brazil *et al.* (1980)	12	adult	YES	Korean Greek Indonesian	English	✓	-	-	✓	-	-	-	-	✓

* not including control groups

TABLE 2.3 **Summary of recent research into the acquisition of L2 intonation**

Chapter Three - Experimental approach

3.0 Introduction

In Part One of this study, some of the issues involved in research into the acquisition of L2 intonation were discussed. Firstly, there are many differing descriptions of intonational systems in English. Secondly, within most descriptive frameworks a variety of different constituent parts together constitute the intonational whole, and identifying which of these parts can most profitably be used in acquisitional research is itself not easy. Thirdly, intonation is closely associated with context, but it also emerges spontaneously in speech, and may not be under the same kind of control as other aspects of linguistic structure. Paradoxically, we need speech data produced spontaneously, but which exhibits very clear linguistic function in context. A compromise between the need to elicit highly controlled data and the need for spontaneous speech is a feature of research in this area.

In the present chapter I will first discuss some of the general issues surrounding approaches to research. I will argue for the exclusion of certain kinds of data on experimental grounds, and in doing so re-establish the selection of those intonational systems which I described in Chapter One as being particularly amenable to this kind of study. Then I will make some broader observations on the nature of second languages and the status of intonation here. Bearing in mind that this study is cross-sectional, I will then discuss the nature of testing pronunciation, and the validity of different ways of banding informant groups. Finally, I will describe in detail the particular experimental procedures used in the elicitation of those data which will form the basis of discussion in Chapters Four to Seven.

3.1 Issues relating to research into the acquisition of L2 intonation

3.1.1 What will not be examined

In Chapter One, reference was made to a number of intonational systems: prominence, the marking of information status, and paratonic structure. These are only three of the many ways in which prosodic form contributes to the meaning of the spoken text; others include the signalling of turn-taking, the delineation of illocutionary force, the communication of affective states, and reflections of surface grammar.

Recently there has been growing interest in the ways in which the systems of intonation contribute to the structure of collaborative discourse. It has been noted that the overall phonetic shape of a paratone can be maintained by several contributing speakers (see Yule 1980a); Lehiste (1979; 1980) has also described how speakers manipulate paratonic structure as a conversation-management strategy. In general we might see these interactional uses of intonation as a further exploitation of the basic prosodic fabric - so that another layer of meaningful discourse emerges above the level of basic intonational form. For example, in the notion of 'key' as described by Brazil *et al.* (1980), different pitch levels are used to mark aspects of the discourse of an individual's speech: low key marks calmness, and high key marks excitability; low key also marks 'equivalence' relationships, whereas high key marks 'contrastive' ones. However, this same system of key operates as a system of pitch concord *between* speakers, and is utilised to encode speech in terms of the dynamics of its interactional 'discourse' (transactions, moves etc.). Other analyses have also considered in detail the contribution of intonation in the structure of conversation (see French and Local 1983; Couper-Kuhlen and Selting 1996).

Collaborative/multi-speaker uses of intonation will not be included in the present study. Here, I will concentrate entirely on the non-collaborative speech of informants. There are a number of reasons for this. Within multi-speaker discourse, the status of information is potentially very fluid; information can be presented as new or given, depending on speaker decisions, so that the issue of exactly what 'information' exists as shared knowledge between speaker and hearer may be complex. Hence, new and given in a conversation are far from straightforward categories (Brazil *et al.* 1980: 17). Such information marking may also be used for rhetorical or other stylistic purposes which deliberately misrepresent the state of mutual knowledge (see Gunter 1966). Speakers can also manipulate accent placement, to hold the floor or to take charge of the information

flow of the exchange in some other intentioned way (Lehman 1977; Abdul-Ghani 1978; Yule 1980a). These choices may not be at all clear during analysis.

The prosodic organisation of speech in multi-speaker discourse is therefore significantly more complex than single-speaker speech, and thus the 'meaning' of a particular (intonational) form in discourse is always potentially hard to pin down. In terms of the elicitation of speech data, the avoidance of multi-speaker contexts helps to minimise the intrusion of these largely uncontrollable aspects of interaction. On the other hand, I still want to assess the ways in which intonation is used in realistic speaking contexts, rather than simply noting its bare form in the uncontextualised reading of single, isolated sentences.

Those aspects of intonational form described in Chapter One - prominence, information marking, paratonic structure - can all be elicited in non-collaborative speech, and also bear a relatively direct relation to context. Hence, they provide a clear window on production. Also, the behaviour of prosodic features in the realisation of these systems is reasonably straightforward to assess: prominence may be marked by perturbations in any (or all) of the prosodic features; the marking of information status necessitates uses of distinctive pitch tones; paratonic structure requires systematic use of onsets at different pitch as well as other specific uses of relative pitch height, and also a systematic use of pause as a framing device.

The incompleteness of analysing only part of the intonational whole is recognised here, especially the absence of any analysis of whole-tune contours. Applied linguistic research continually suffers from the opposed pressures of being maximally broad, so as to render its findings 'applicable', and maximally narrow, so as to be able to claim methodological rigour and thoroughness. The approach here is one of a workable compromise between on the one hand attempting to encompass all intonational phenomena and, on the other, concentrating on only one aspect. This allows for an analysis of considerable phonetic detail but also for findings of sufficient breadth to be interesting as a study of the acquisition of intonation.

3.1.2 Quantitative and qualitative approaches to data

Ladd (1996) makes the distinction between the quantal/categorial nature of linguistically encoded intonation phenomena and the scalar/gradient nature of paralanguage. For

example, a rising tone has certain categorial functions - in Cruttenden's exposition these include question, continuity, open-listing, non-conducive, statement with reservation etc. However, the *scalar* nature of the rise in realisation - the width of the pitch - is also of paralinguistic significance, relating to the intensity of the message and reflecting the speaker's involvement or emotional state. For Ladd, the linguistic and paralinguistic complexion of tone are distinct systems. Thus, a rise has a linguistic 'meaning', and in addition the nature of its realisation reflects the paralinguistic conditions of utterance. However, Ladd notes that some writers feel that the breadth of pitch of a tone, broad and narrow pitch on a rise, for example, is evidence for two separate tone categories. O'Connor and Arnold (1973) distinguish between high-rise and low-rise (pp. 8-9) in this way. Ladd cedes that the distinction between the 'linguistic' categories of intonation and the paralinguistic nature of utterance is a fuzzy one, not least because the very same prosodic features are used in both linguistic and paralinguistic aspects of tone production, and notes that there is no easy way of resolving this 'paralinguistic stalemate' (p. 39).

In terms of second languages, it would seem logical to look first at the way in which prosodic features are used in the production of linguistic categories in this sense, rather than at the ways in which prosodic features contribute to paralinguistic meaning. One way of achieving this might be to concentrate on aspects of intonational form which have a direct relation to some aspect of linguistic form. However, the distinction between linguistic and paralinguistic intonation here is still problematic. Firstly, the paralinguistic constraints upon second language speech are probably relatively more pervasive than in L1 speech; levels of anxiety, timidity and frustration all have a potential effect on prosodic production, and such states of mind are generally held to be more pervasive in L2 than L1 speech (see MacIntyre and Gardner 1991 on 'foreign language anxiety'). Specific second language speech strategies, such as conspicuously high levels of acquiescence or doubtfulness, might also lead to deviant production, for example in a proliferation of concessive, non-volitive and conciliatory rising tones.

In order to avoid these problems, the elicitation of speech here is designed with the aim of collecting data which is easily expressible in terms of its function in speech. Secondly, the tests themselves will be designed so that anxiety and other intrusive states of mind are kept to a minimum.

Even so, we are still left with the problem of accommodating two different kinds of analysis within the experiments. The first of these assesses the use of *forms* by

speakers, such as the selection of tones in a certain context, or the ability of an NNS to mark appropriate prominences as contrastive. The other kind of analysis looks at the *degree* to which prosodic features are used in the realisation of particular forms. For example, pitch may be used as a marker of prominence by a NNS, but by looking at the degree of marking we can ask the following questions: a) does the NNS employ prosodic marking in a target-like way? b) does NNS production become more native-like as global levels of speech increase?

So, there is indeed a potential problem in the relation between intonational forms and the gradient nature of the realisation of prosodic features. In the present study, then, both quantitative and qualitative evaluations of the production of intonation will be used.

3.1.3 A note on cross-sectional studies

There is little work on the development of L2 intonation in the literature. Wieden (1993) and Grosser (1993) are both exceptions here, although their longitudinal studies both involve data from children, and there is a clear interplay here between the developing L2 intonation and ongoing developments in the L1.

In the present study, I intend to consider acquisition of intonation using a cross-sectional approach to data from a group of adult speakers. The relative advantages of cross-sectional (as opposed to longitudinal) studies in SLA have been discussed widely, and neither approach has been seen as inherently preferable (see Reichardt and Cook 1979; Larsen-Freeman and Long 1991). I will discuss this further in section 3.1.5, below.

A principled account of exactly *what* ability is an appropriate criterion for the framing of a cross-sectional spread of informants is also needed. Studies which address particular aspects of L2 acquisition sometimes use existing EFL tests of proficiency as a general measure of acquisition. Jamieson (1995), for example, uses TOEFL scores. However, there are pressing arguments against the use of such tests as yardsticks in research into areas of speech such as intonation. Such tests typically under-value productive speech skills, and are usually unsophisticated (both in design and application) in their sensitivity to aspects of pronunciation.

In the present study, we need cross-sectional groups which allow us to assess how intonational form contributes to general perceived nativelikeness of pronunciation over time. What is under investigation is the way in which intonational form in the

speech of NNSs develops as the overall perceived quality of pronunciation becomes more nativelike. In this sense, general measures of proficiency, subsuming a spread of linguistic competencies, are not useful; the various different skills of an L2 speaker often develop at different rates, and pronunciation may fossilise at a very early stage in acquisition relative to other skills. Informant subgroups will, therefore, be based on global levels of perceived nativelikeness of speech production (see section 3.1.5, below).

Also of interest, of course, is the extent to which NNSs can attain nativelike norms in their intonational production. The highest-ability subgroups will serve as a representative example of advanced pronunciation, and will allow assessment of how, in speakers perceived to have the most nearly nativelike pronunciation overall, production of intonation either matches or falls short of target norms. It will not provide evidence of the upper limits of acquisition, as in the case of Neufeld's work. Neufeld (1977, 1978) took great care in selecting informants with the very highest motivation and disposition to acquire. In my study, on the other hand, no such criteria were employed.

3.1.4 Observations on the assessment of L2 pronunciation

At the opening plenary session of a recent conference on L2 pronunciation testing, recordings of the English speech of four NNSs were played. Members of the audience were asked various questions, including, *On a scale from 1 to 10, how would you assess the learner's pronunciation [in each case]?* Bradford (1997) notes the high level of 'discussion and disagreement about the relative merits of the four learners' which this and similar questions provoked (p. 6). She goes on to raise the question of 'whether, in fact, it is realistic (or possible) to separate out pronunciation from a more global assessment of oral communication'. If those with specialist, professional interest in such issues could arrive at no agreement as to the assessment of four NNSs' pronunciation,[1] one might ask whether such an undertaking is viable at all. At the very least, one must seek to clarify exactly what is under assessment.

Within the broadly 'communicative' approach which characterises much current foreign language testing, 'pronunciation' components are rarely framed in terms of how near to nativelikeness the examinee comes. For example, examiners' guidelines for

[1] The conference was attended by members of IATEFL's Pronunciation and Testing special interest groups, as well as by phoneticians and those working in the field of EFL.

UCLES[2] English examinations specify only that there be 'little L1 accent' or that the L1 is not 'intrusive' in order that a positive assessment of pronunciation be made (see Jenkins 1997: 7). Another EFL examination body, IELTS[3], has descriptors for nine ability bands in its guidelines; pronunciation is mentioned only fleetingly in the top two bands, is not mentioned at all in the middle bands, and only very generally at the lower ability range; 'pronunciation is likely to be strongly influenced by the first language and to significantly impede communication' constitutes the entire relevant descriptor for band 3 (see Tench 1997: 31).

Current pronunciation testing, then, is framed with relatively more weight placed on communicability than on the precise phonetic character of realisation (a point which both Jenkins and Tench make). Tolerance of an 'unobtrusive' foreign accent is widespread, and indeed Gimson (1980) had earlier suggested that NNSs should aim at being *intelligible*, rather than accurate pronouncers of the L2 (p. 320).

A common theme running through recent discussions of pronunciation testing, which is perhaps related to Gimson's notions of intelligibility, is that assessment should ideally be based on judgements made by non-specialists (Kenworthy 1987; Tench 1997; Porter and Weir 1997). If indeed pronunciation is to be judged in terms of how it contributes to broad intelligibility, and not in terms of phonetic accuracy, then the most appropriate judges would seem to be those who have no specialised ability to adopt detailed judgement criteria.

3.1.5 NNS subgroups

The subgroups in my cross-sectional study here were required to reflect global levels of pronunciation. That said, whereas we require judgement criteria which allude to global levels of pronunciation, we need also to avoid forcing judges to be too specific. Hinofotis (1976) found that when judges were asked to rate particular skills, such as vocabulary, accent and grammar in L2 speech, they were influenced by the 'overall communicative effect' of that speech to such an extent that the assessment of the particular skills turned out to be more or less baseless.

[2] University of Cambridge Local Examinations Syndicate.
[3] International English Language Testing System.

In the literature on native-speaker reactions to nonnative speech it has been noted that certain mispronounced sounds create a stronger feeling of 'foreignness' than others (see Cunningham-Andersson 1997; Morris-Wilson 1997). In an earlier study of Spanish-English L2, it was found that native English speakers judged such speech as unfriendly, unpleasant, and as associated with low social class (Ryan, Carranza and Moffie 1977). Indeed, it has been claimed that English speakers in general harbour unconscious, negative feelings about a variety of different non-standard accents (see Edwards 1982). Whereas the intuitive reaction of judges cannot be avoided, we can at least keep the linguistic form of the sample constant in each case, so that samples are all linguistically well-formed. In this way, we can avoid the possibility that judges, in assessing 'pronunciation', respond unconsciously to aspects of deviant production unrelated to pronunciation.

On the other hand, we also don't want judges to be aware that they are assessing pronunciation, lest they adopt *ad hoc* criteria of their own, listening out for a few features which they associate with poor pronunciation.

For these reasons it was decided to take as a sample one paragraph from each informant's reading of the paratone text (see 3.2.2, below). These samples were then played to native-speaking judges in random order, along with the same samples of speech from an English native-speaking control group. Judgements of the nativelikeness of NNSs' pronunciation are (presumably) influenced by both segmental and non-segmental aspects of pronunciation. Hence, there is a certain circularity in seeking to express putative developmental patterns in the acquisition of intonational form against these global assessments of pronunciation. However, there is no easy alternative, other than ranking informants' pronunciation in terms of close phonetic criteria (of the production of particular segments, for example). This would seem to take us further away from the basic issue of global pronunciation.

One further justification for using a set text in the judgements was that in free speech NNSs may (either consciously or unconsciously) use suppression strategies, managing to reduce the extent of use of certain difficult sounds. Jenkins (1995) has found that this happens in discourse between NNSs, especially where there is a common L1, where an attempt is made to suppress use of sounds particularly prone to negative L1 transfer.[4]

[4] Avoidance strategies have been identified in other areas of L2 linguistic production (Schachter (1974), although Kellerman (1992) has pointed out some of the difficulties in even expressing logically how

A total of eight different randomised series of samples from the speakers (25 NNSs and 8 NSs) were prepared. Using the same paratone reading meant that the context was the same for each speaker, and whereas the judges were requested to rate the overall quality of the speech (they were not aware of the nature of the research), the fact that each informant was saying the same thing made it probable that the overall quality of the pronunciation was the main focus of their attention. Eight judges rated all speakers (including NSs) on a scale from 1 to 5, in response to the general question *How well does this person speak English?* (see Appendix C)

A high degree of agreement was found between the scores of the judges. From these judgements, outlying scores[5] were discarded, and the remaining scores were then averaged for each speaker.

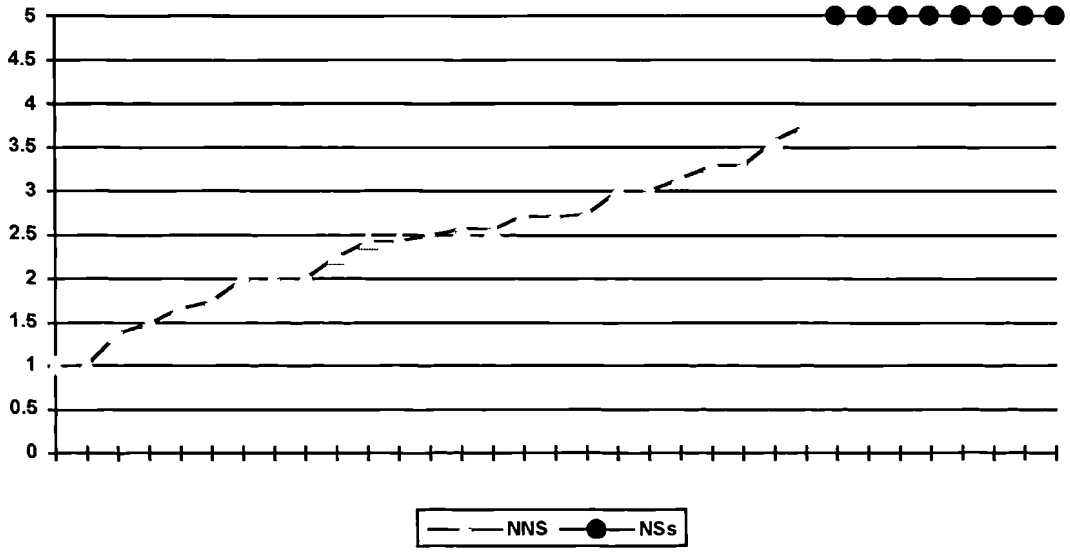


FIGURE 3.1 **Ranked average scores of the pronunciation of the 25 Spanish-speaking NNSs and 8 NSs**

On the basis of these averaged, modified scores five subgroups were established, with scores in the following ranges:

Subgroups	score band
group 1 (n=6):	1 - 1.99
group 2 (n=6):	2 - 2.49
group 3 (n=6):	2.5 - 2.99
group 4 (n=7):	3 - 4
group 5 (n=8):	5

avoidance could function; do you have to 'know' the form in some way before you can avoid it? He does cede that avoidance may be possible in certain cases, such as when a form is known but where circumstances (such as fluent, conversational speech) make its accurate realisation difficult.

[5] Any score given to an informant by only one judge.

Group 5 contained all the NSs and none of the NNSs.

There is an inevitable degree of arbitrariness in the division, whereby speakers with very similar scores are nevertheless put into adjacent groups. A cross-sectional (rather than a longitudinal) study of this type is always likely to be in some ways unsatisfactory. It does, nevertheless, allow generalisations about tendencies in acquisition over time to be expressed.

We might remember, though, that at this stage the very assumption of *development* is baseless. We simply do not know to what extent (if at all) the production of intonational form changes during L2 acquisition. Indeed, a lack of any change would itself be interesting here.

It seems likely that any development in L2 acquisition would occur rather slowly. After all, we are looking not for newly-acquired forms (as in the case of grammatical morphemes, for example), but rather for modifications to existing norms of prosodic production in gradually more target-like ways. Such changes might be very difficult to track continually in an individual informant's development, and indeed one justification for the use of a cross-sectional approach here is that it allows for significant group generalisations to be made. Also, we do not know to what extent production in this area is variable between speakers, and whereas findings within a longitudinal framework constitute a more direct expression of the 'facts' of the data, variation within findings cannot so easily be 'ironed-out' in an attempt to make generalised claims about acquisition. So, rather than follow the development of production in one or even several speakers over time, it is more appropriate here to take a larger number of informants and express generalisations about any *tendency* towards development in speech within their ability-group.

As I noted in the previous chapter, it is indeed the *variation from L2 norms* which I am interested in here, and any strong group tendencies will reflect exactly that notion of a gradual movement towards target norms.

In Chapter Two I discussed 'developmental' approaches to the study of both first and second language acquisition. In both these fields, work has built-up to provide an increasingly complex and refined picture of how various aspects of language are acquired over time. Hence, studies here can rely on a broad base of assumed ideas, and their findings can be compared to existing work. In the case of L2 intonation, there is no such

bank of existing work, a situation which de Bot (1986) has referred to as 'the missing database'. Since 1986 some work in this area has been done, yet those studies which most closely reflect the approach to intonational form which I am taking here (Cruz-Ferreira 1987; Hewings 1993; and Wennerstrom 1994) are non-developmental.

3.2 Methods of data elicitation

The main body of data which I will use is from a group of 25 speakers of L1 Peninsular Spanish, plus a group of 8 native English speakers. Two smaller groups, speakers of Italian and Mandarin Chinese, provided additional data, which I will consider only in Chapter Seven. All informants were between 21 and 30 years old, and were all studying at the University of Hull, either as exchange or postgraduate students. More than half were students of subjects other than English, and almost half had not studied English formally since the age of 14.

The use of a relatively large number of same-L1 informants means that any cross-cultural transfer of intonation, such as that observed by Loveday (1981) in his study of Japanese-English speech, will be constant throughout the group. Also, it avoids the influence of different L1s on judgements leading to the sub-groups. L1 was found by Purcell and Suter (1980) to correlate with judgements of quality of L2 English pronunciation and, whilst I will consider the influence of L1 in Chapter Seven, it is useful also to have data which does not need to be analysed with constant reference to the variable influence of different L1s.

The tests themselves were carried out in a normal office environment. The microphone used [6] was unobtrusive, due to its flat design, and for the most part went unnoticed. The tasks were simple, and were preceded by a series of (unrecorded) practice tasks, which familiarised informants with the general nature of the responses required.

The elicitation procedure was designed to minimise the stressfulness of speaking in a formal situation. Informants were all recorded individually, since there is some evidence (Giles, Coupland and Coupland 1991; Jenkins 1997) that NNS's will 'accommodate' the phonological variation (including errors) of fellow speakers in a

[6] Yamaha PZM.

given situation. So, whereas far more data could reasonably have been gathered in language classrooms - for example, by using a language laboratory to record simultaneously a whole group of NNSs - it was felt that a language classroom may not have led to optimum conditions for pronunciation.

3.2.1 Screen-question tests

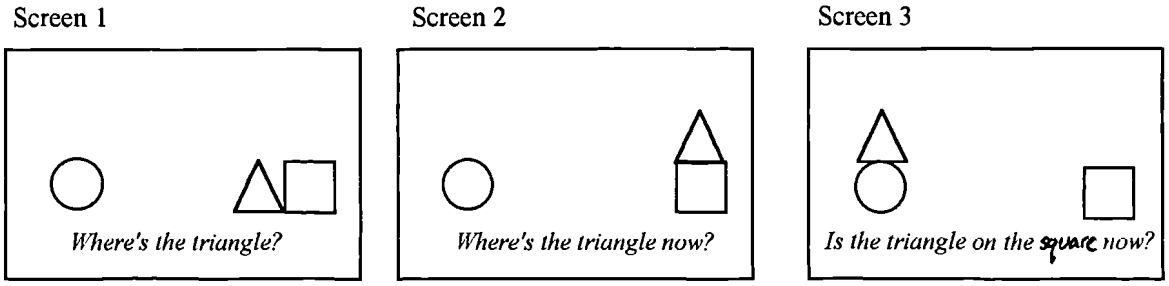
Screen-question tests consisted of a series of grey-background screens presented to each informant on a PC. Each screen had a number of coloured shapes arranged in various configurations.[7] In successive pictures one thing had changed relative to the preceding screen - either the position of one shape or the introduction of a new shape. At the bottom of each picture a simple question was written, which was read out by the experimenter as the picture was shown. Informants were asked to reply simply and briefly to each question. In an attempt to guide informants towards simple responses, several practice screens were first shown which, in addition to the question, had a sample answer added.

Screen-questions were designed to elicit responses containing one of the following forms:

- normal (default) prominence
- contrastive prominence (in medial sentence position)
- contrastive prominence (in right-most sentence position)
- double contrastive prominence (two elements in unit obligatorily contrastive)
- listing intonation
- listing intonation with one element marked as 'new' information
- responses which typically contained four points of information (prominences), within a two-unit utterance

The following sequence of screen-questions illustrates the elicitation of the first three of these:

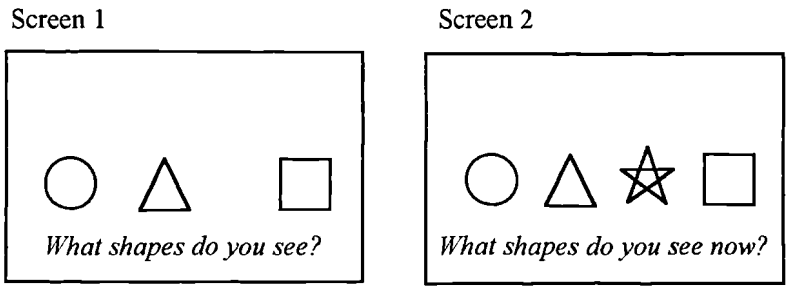
[7] This procedure was similar to that of Swerts and Geluykens (1994). However, the element of moving shapes, which was central to the notion of describing the process in their experiments, was avoided here, since this may have prompted more varied responses than were required.



Typical response:		
The triangle's next to the square	*Now it's on top of the square*	*No, it's on top of the circle*
Gloss:		
Default (normal) prominence	Shift of focus leftwards onto new preposition within the otherwise repeated PP, creating a contrastive prominence on *top*	Contrastivity here corrects the proposition of the question, creating contrastive prominence in right-most position

FIGURE 3.2 Screen-questions for the elicitation of normal and contrastive prominences

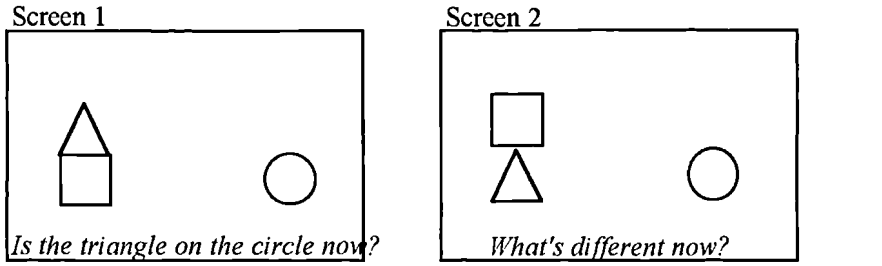
Listing intonations were elicited in the same way:



Typical response:	
A circle, a triangle and a square	*A circle, a triangle, a star and a square*
Gloss:	
All items are 'new' here (although familiar from previous tests)	The same list, but with the addition of a totally new item, which is predicted to be marked as such relative to the other list items

FIGURE 3.3 Screen-tests for the elicitation of listing intonations

Responses containing 'double contrastive prominences' were elicited from screen-questions such as the following:



Typical response:
No, it's on top of the square

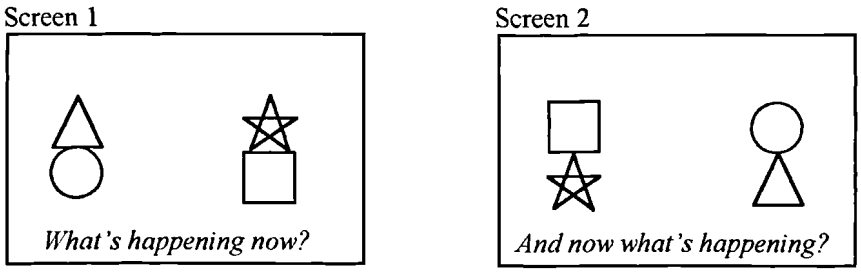
Gloss:
Contrastive prominence on final element

Now the square is on top of the triangle

Reversal of functional roles of two items active in previous screen

FIGURE 3.4 Screen-questions for the elicitation of double contrastive prominences

Responses containing two units, each with two points of information, were elicited with sequences of screen-questions such as the following:



Typical response:
The triangle's on the circle /
and the star's on the square

Gloss:
Points of prominence on two elements in each unit

Now the square's on the star /
and the circle's on the triangle

Switching of roles of the four shapes, with prominences on each

FIGURE 3.5 Screen-question sequence for the elicitation of two-unit utterances with two prominences in each unit

Appendix A contains all screen-questions used in the experiments.

The use of these screen-questions was designed to yield maximally similar responses. For example, the first screen in Figure 3.2 asks *Where's the triangle?* Responses here, although occasionally of more than one intonation unit, in most cases contained a central unit which, not surprisingly, contained *The triangle's on the square* or *It's on the square*. All such units, for both NSs and NNSs, were of a very stable average length, being generally between 3 and 5 words long. Hence, the data elicited were of a very similar structure. At the same time, these data are all spontaneously produced. Research into intonation has recently moved towards a greater recognition of the need to elicit spontaneous speech from informants (see Swerts and Geluykens 1994, for example). Indeed, one of the major criticisms of earlier work has been that it is based on context-less readings of sentences. Halliday's work on intonation adopted this approach, and Bolinger made repeated criticism of context-less analysis of intonational form, complaining that where an informant is requested to recite an isolate form he will simply invent a context; that context will, according to Bolinger, remain unknown to the analyst, so that the exact relationship between intonational form and the linguistic and non-linguistic context of utterance cannot be fully addressed (see Bolinger 1972b).

Wennerstrom notes that free-speech data is preferable in the analysis of intonation because here informants speak 'from meaning' whereas reading tasks tend to promote form-attended production (1994: 413). However, Wennerstrom's free-speech data are, due to the unguided nature of their production, less useful to her than the text-reading data used in the same study. In my elicitation, I have tried to maintain the ethos of free-speech, whilst aiming at very similar responses in all informants. In the screen-question data here, both linguistic and non-linguistic context was the same for each speaker. Hence, the questions prompt responses of a very similar linguistic form in most cases, and the simple relationships between shapes on the screen means that there is little scope for misinterpretation of the non-linguistic context. It is an attempt, in Bolinger's words, to be a 'mind-reader', through carefully designed contexts.

A very controlled context has other benefits. In the data on prominence, the difficult issue of the bi-directional/non bi-directional relationship between (phonetic) prominence and (informational) focus is largely avoided, since the interpretation of what information needs focusing is straightforward. The question of where within a focused element prominence falls is not avoided entirely. However, given the nature of the

contexts and the linguistic forms of the responses, the scope of a focus is generally predictable.

The high degree of similarity in linguistic form (especially the length) of the responses also minimises the danger of skewing of subsequent analysis due to the effects of declination on different utterances. For example, a very long and a very short unit might differ in terms of the presence or absence of maximum pitch or amplitude (instrumentally identified) at a comparable sentence position due only to the greater effect of declination on the longer unit. For this reason, only responses to questions which were of a very similar linguistic form and length throughout the data were pre-selected for analysis. Also, within a body of responses to a particular screen-question, all responses considered aberrant within these very narrow parameters were excluded. So, for example, the data on 'normal' prominences were all one unit long, had no obvious breaks, and were all clearly judged to be non-contrastive in terms of their content. This reduced the amount of available data, although these were far more uniform in structure as a result.

Three series of screen-questions led to a total of nineteen responses from each informant. After these tests, each informant was asked to provide a brief description of four similar screens in his/her own L1. These four screens did not contain written questions, and were intended to elicit data on L1 production, which was used subsequently in a discussion of L1 transfer (see Chapter Seven).

3.2.2 Paratone text-reading

After the screen-question tests, informants were asked to read out a printed text in English. All informants had been asked to read and understand this text prior to the start of the session. Here, particular care was taken that all NNSs understood the whole text, and that they were given as much time as necessary to familiarise themselves with its content. A bilingual dictionary was provided. The text contained only words from the 2000 most common words in English, using *The Longman Defining Vocabulary* (Longman 1995: B12-18).

Reading aloud from a text in experimental conditions creates stressful speaking conditions, and as well as having an effect on articulation, this may also inhibit the normal levels of information processing which the speaker must carry out before post-

lexical phonological decisions are made. The above preparations were intended to minimise speaker-anxiety. These readings are clearly not spontaneously produced, yet there seems no more authentic way of eliciting comparable data on macro-prosodic production from a large informant group with the aim of group comparisons.

The text itself, which was written for this experiment, was intended to conform to the characteristics of paratonic structure as discussed in section 1.3, above. Such an approach has been used before in the study of SLA and intonation (Wennerstrom 1994). One 'block' of language was needed, which was both typical of a contribution to a broader topic of discourse, and also informationally cohesive in a way which could be related to the notion of paratonic structure.

The following text was read out by all informants, in the reading conditions described above:

> These days many students come to the United Kingdom to study, often on exchange programmes. If you want to study in the UK, there are nearly one hundred universities. And many towns and cities have two or more universities to choose from.
> Many institutions have a modular system, in which you can study separate units (called modules) from different departments.
> A modular system allows you to choose the things you want to study, depending on your own particular interests.
> If you want to study in Hull, there are plenty of modules to choose from. Modern languages are a popular choice; there are modules in French, German, and many other languages. So, you should never get bored if you do come here to study.

What we might regard as the opening topic (*coming to the UK to study*) is reintroduced at the end of the text, following Yule's observation regarding 'speaker topics' (1980a: 33). Tentatively, then, this whole text reading might be seen as potentially yielding a series of paratones. More importantly, the text is created as informationally layered, and prosodic devices are required to mark these layers during reading. For example, within the text three sections represent three shifts in sub-topic. They are distinguished by line breaks, although these were intended to be suggestive only, and are not followed by paragraph-initial indentations. These three sections were subsequently found to contain several units which were each comparable in phonetic profile to the notion of a 'minor paratone'.

Hence, this text is an attempt to prompt the use of certain devices in speech. In subsequent analysis of the readings, it was borne in mind that a pre-determined

informational structure which is 'written-into' the text might well fail to promote readings which exhibit the predicted paratonic structure. The readings of the native-speaking informants were used as a model of nativelike production, even when these (as a group) deviated from predicted patterns. In Appendix B the text is annotated with suggestions as to some of the ways in which it might be structured during reading.

3.2.3 Technical note

High quality tape recordings were made with a Yamaha PZM microphone and a Tandberg Audio Tutor cassette recorder. These were then digitised as .wav sound-files (22050 samples/sec., 16-bit PCM). These files were then converted into .sig format, the sound file format used by the Macintosh phonetics application Signalyze (version 3.0). Subsequent analysis was performed using Signalyze.

3.3 Analysis of data

3.3.1 Prominence: the use of prosodic features

As described above, the responses selected for analysis had very predictable points of prominence. However, all data were transcribed, including intonation units and prominences, without reference to the predicted 'correct' response. In order to double-check that transcriptions were accurate, a selection of utterances was played (in randomised sequences) to naive native-speaking informants, who were asked to *identify the word or words which the speaker is focusing on as particularly important*, following a procedure used by Wells (1986). This served as a verification test for the data, and proved to support the transcriptions almost without exception.

The phonetic analysis carried out on each example of prominence was also based on Wells' (1986) study of focus [8] in spoken English. He identifies seven distinct prosodic devices (from a list of sixteen) which are observed to contribute to the marking of various types of 'focus'. From this he is able to posit four distinctive degrees of

[8] Wells' 'focus' is more or less coterminous with 'prominence' as used in Chapter One.

prominence-marking, based on the total number of prosodic devices ('features') observed to be located at the prominent point. Degree of prominence is therefore defined in terms of the number (or score) of the following devices present:

pitch peak
maximum pitch range
kinetic tone
loudness peak
decrescendo
tempo marking
pause/drawl

In effect Wells' work establishes a prototypical framework for the analysis of prominence. No single device itself needs to be present for the unequivocal marking of prominence. However, strongly prominent elements (such as 'contrastive' prominences) will tend to be marked by a greater number of these devices. This number of devices can be expressed as a score, where a higher score indicates stronger prosodic marking.

However, the data in the present experiment were for various reasons not fully amenable to analysis within Wells' original framework. So, several alterations were made to the model. *Pitch peak* as a single category was expanded to include *maximum* F_0 and *minimum* F_0, although the latter, it turned out, was seldom observed in the data. *Maximum pitch range* was used, but, given the relatively short units which were typical of most speakers' responses, could not always unequivocally be said to be present. It was noticed in the data that significant use was made of what has been called 'pitch obtrusion' (the term reflects something of the IPO/Eindhoven notion of 'pitch excursion'). Therefore, two categories which Wells had first adopted but then dropped from his model, *obtrusion to* and *obtrusion from* [prominence], were reintroduced. *Decrescendo* is not used here; its is a dubious category in Wells' own study, since he observes that the phenomenon occurs with equal frequency in elements with low and high scores, suggesting that amplitude declination, rather than intentioned use, might explain its presence (p. 61). Durational marking in Wells' category includes *allegro* both prior to and after the prominent element. Since the sentences in the present experiment were short, post-prominence allegro could not always be assessed. Therefore, *prior allegro* alone was retained. Also, *slight pause* before the prominence, another of Wells' original seventeen categories, was used as a separate category here. This effectively spreads the scoring on durational devices in a way which compensates for the absence of scoring for post-prominence allegro. This leaves Wells' final category, *pause and/or drawl* on prominent element, reduced to drawl only, which is interpreted here as any

significant expansion in the duration of the prominent syllable or syllables relative to those in the rest of the sentence. *Kinetic tone* and *maximum loudness* (here maximum dB) were both retained in the model. However, in the latter *relative loudness obtrusion* was included as an additional criterion, so that obvious jumps in loudness on prominences were taken into account, even if they did not reach the same level of amplitude as preceding material in the unit.

The final list of ten devices for the scoring of the realisation of prominence is:

maximum F_0 on prominent element
minimum F_0 on prominent element
maximum pitch range within the focused constituent (in which prominence falls)
obtrusion to prominence
obtrusion from prominence
kinetic tone within the prominent element
maximum amplitude or relative obtrusion of loudness
allegro immediately prior to the prominent element
slight pause prior to the prominent element
drawl/slow-down on prominent element

The issue of whether auditory or instrumental analysis should be used in the measurement of intonation is another 'stalemate' in the field of intonational studies. The most cogent approaches in the literature seem to be those which establish the mixture of auditory and instrumental analysis which is most appropriate to the particular aspect of form under investigation (Crystal 1969; Brown *et al.* 1980). The present study seeks to do likewise. The benefit of combining auditory and instrumental methods of analysis is that where the analysis is founded on auditory discrimination, the use of instrumental readings to check auditory judgements acts as a clarification of the acoustic reality of auditory analysis. Conversely, auditory assessments can be used to double-check the perceptual reality of instrumentally extracted measurements. In addition, mechanical readings provide useful statistical data on the gradual changes in the use of prosodic features during acquisition of an L2.

There follows a brief discussion of the different analytical techniques used to assess each of the prosodic devices used in the 'scoring' of prominences:

Maximum F_0 on prominent element (MXF)
Minimum F_0 on prominent element (MNF)

Of interest here is whether NNSs use these maxima to mark the prominent element within a unit, and whether the group-wide disposition to do so changes as global

levels of spoken English do. The acoustic character of these maxima (whether, for example, physical pitch range expands or contracts as ability increases) is not itself under observation.[9]

Maximum pitch range within focused constituent (MXR)

The measurement here was instrumental only insofar as an F_0 trace was examined to check the extent of the F_0 range within the focused element (not just the prominence itself).

Obtrusion to prominence (OBT)
Obtrusion from prominence (OBF)

Both these categories were used in an attempt to capture elements typical of pitch movements within short utterances. Assessment here was again auditory, but inspection of the F_0 trace was used to check that obtrusion, and not simply a movement within the prevailing contour of the unit (such as a stepped contour), had taken place.

Kinetic tone within prominent element (KIN)

Given the short utterances under observation, this proved to be a difficult category to distinguish from that of maximum pitch range. Kinetic movements often begin within a prominent element and spread over subsequent material not strictly within the 'salient' element as recognised by naive informants, and it may not be until the 'tail' of a short intonation unit that a movement is safely identified, despite the stipulation that a nuclear tone must, by definition, begin its characteristic movement on the salient element (the 'nucleus'). Assessment was auditory.

Maximum amplitude/ relative obtrusion of loudness (MXA)

dB peaks for each response were located mechanically. Obtrusions were checked manually against the dB trace.

Allegro immediately prior to the prominent element prominent (PAL)
Slight pause immediately prior to the prominent element prominent (PAU)

Auditory assessment, backed up with inspection of the wave where necessary.

[9] The pitch range of speech will be discussed in terms of its relation to ability in Chapter Seven.

Drawl/slow-down on prominent element (DSO)

Auditory analysis. However, where a drawl/slow-down was suspected but unclear, duration measurements (100ths/sec) were added to the transcription and average syllable duration on the prominence compared to that of the average for the whole unit.

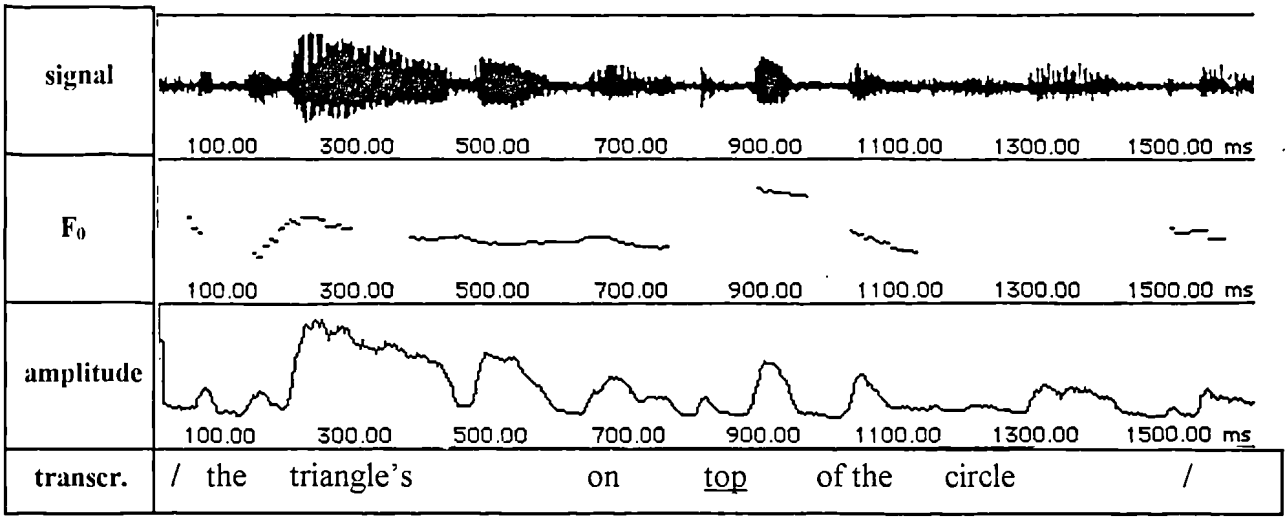
The analysis thus comprises a mixture of instrumental and auditory techniques.

	instrumental measurement	instrumental to confirm auditory judgement	auditory judgement
max. F_0 (MXF)	✓		
min. F_0 (MNF)	✓		
max. pitch range (MXR)		✓	
obtrusion to (OBT)		✓	
obtrusion from (OBF)		✓	
kinetic tone (KIN)			✓
max. loudness (MXA)	✓		
allegro (PAL)		✓	
slight pause (PAU)		✓	
drawl/slow-down (DSO)		✓	

TABLE 3.1 **Schema of analysis of prominence**

3.3.2 Examples of the scoring system

In the following section, a number of responses are discussed with reference to how they were scored using the system described above. All responses are from NS informants, and each is presented as a raw oscillogram, with corresponding F_0 and dB traces below.



scored: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 4

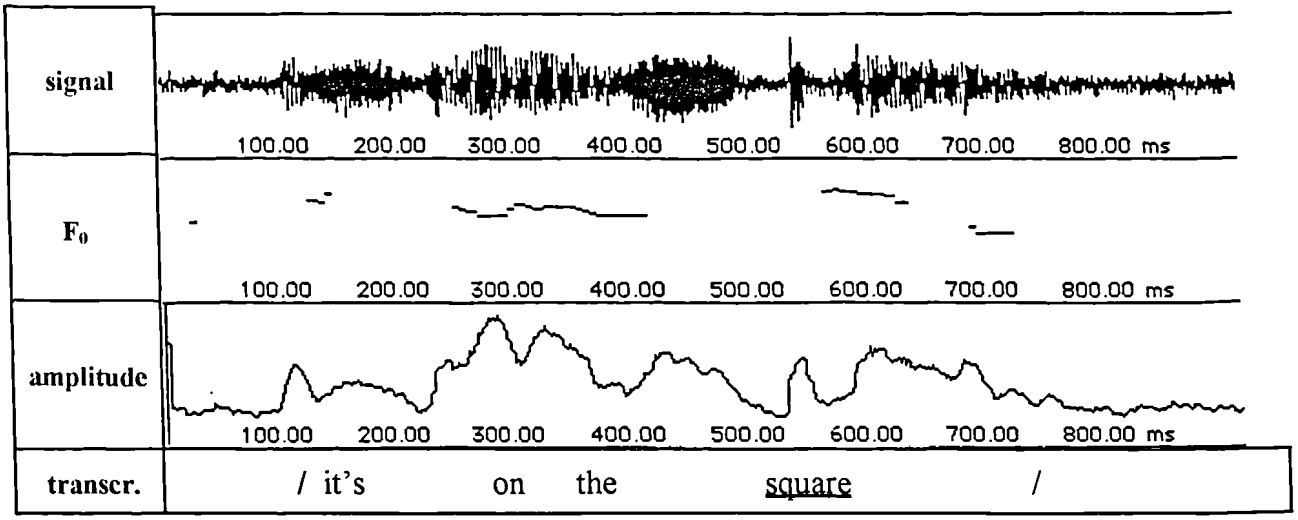
FIGURE 3.6 A unit containing a contrastive prominence, spoken by a native speaker

MXF: the contrastive element, *top* (*of*), is marked here with the highest F_0 in the unit.

OBT: F_0 jumps up from the prevailing unit contour on the prominent element.

OBF: When an upwards-jumping pitch obtrusion is made some way before the end of a unit in English, the remaining portion of the unit may fall steadily throughout the remainder of the unit. However, here the F_0 falls on the word *of*, so that *top* (the prominent element) is effectively isolated at a high pitch level in relation to both preceding and succeeding parts of the pitch structure. The two categories of obtrusion are useful when a prominence is very overtly pitch-marked in either or both of these ways, but nevertheless does not carry a kinetic movement salient enough to be perceived auditorily.

MXA: the dB peak on *top* in this (contrastive) response is not the maximum in the unit, yet obtrudes from the expected line of loudness declination which otherwise characterises the response. Additional evidence of a deliberate use of amplitude marking here was that, in the speaker's immediately preceding response (*It's next to the square*), there is no increased amplitude on the comparable section of the PP.

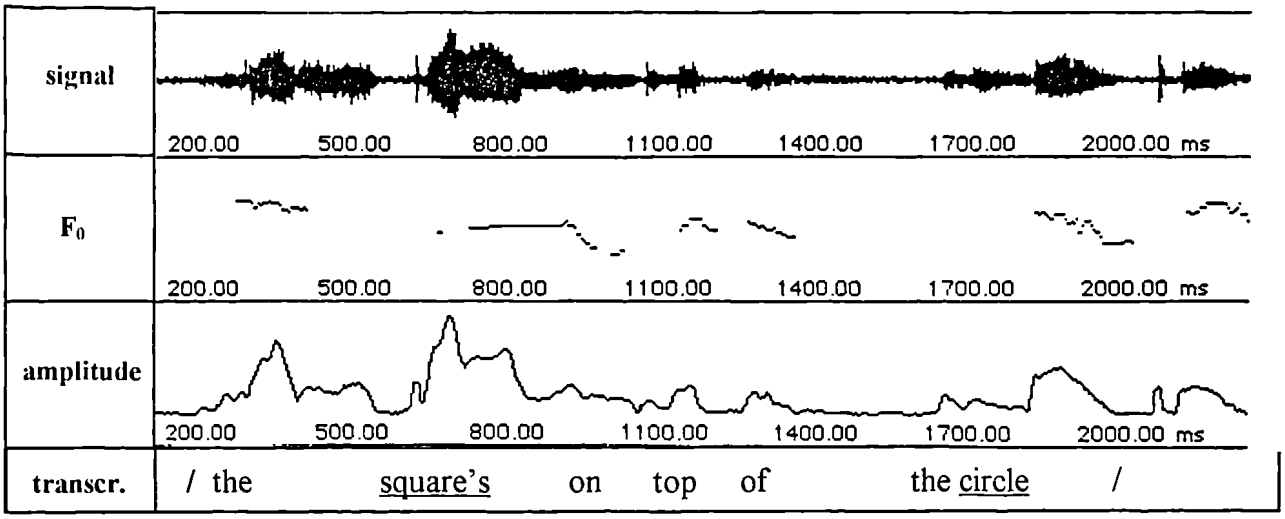


scored: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 3

FIGURE 3.7 **A unit containing a contrastive prominence (in final position), spoken by a native speaker**

MXF: it is noticeable that, in the marking of a contrastive prominence in final position, it has been necessary to distinguish this element from a non-contrastive prominence, which would usually be in final position. Thus the final element (*the*) *square* is marked with pitch equally high as at the high onset here, which is quite uncommon in the data.
MXR: The full pitch range occurs on the focused element.
KIN: The kinetic movement here is a (high) fall.

The following three examples are used to illustrate the use of the scoring system with single intonation units which contain two prominences. All three are the responses to the same screen-question, by three different NSs.



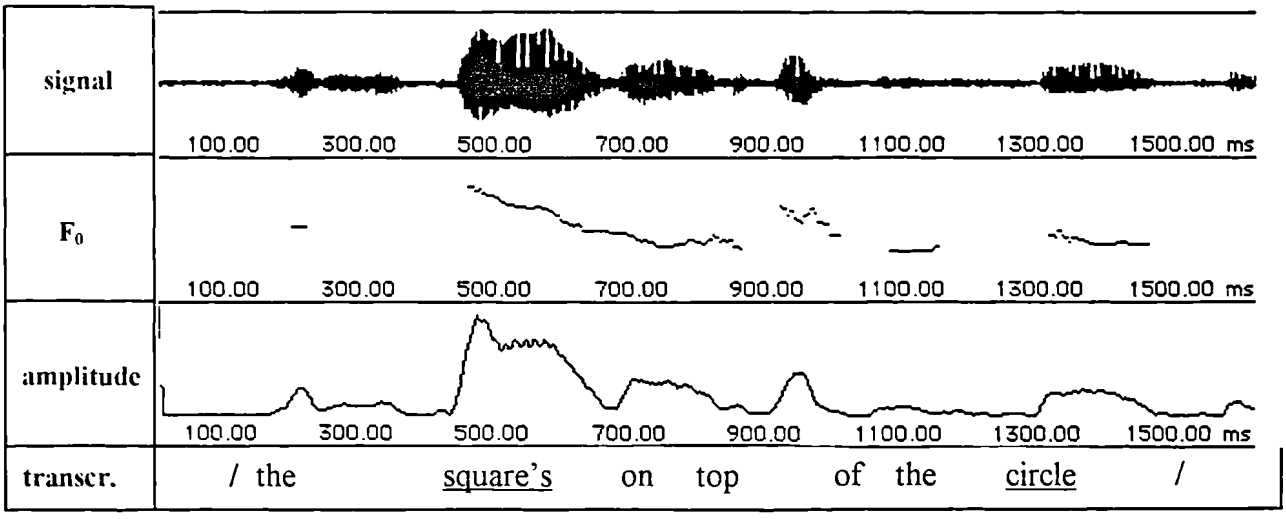
scored: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 2
scored: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 4

FIGURE 3.8 **A unit containing two prominences, spoken by a native speaker**

The integrity of the response's single-unit status might, in terms of traditional analysis, be evinced by the fact that there is only one kinetic movement. Indeed, all preceding material can be interpreted in terms of a high falling head. The unit pattern also conforms to Bolinger's 'suspension bridge' prototype. The point here is that the first of two informationally strong elements is contextually 'contrastive' and is hence given prosodic prominence. Although indeed rather like a high falling head, there is here a noticeable obtrusion down from *the square 's* onto the following word (*on*), against the prevailing contour.

MXA: because this category is interpreted to include relative obtrusions from a prevailing line of amplitude (normally a declining line), two prominences can both be scored, although clearly both cannot be marked with maximum amplitude. In this example, there is such an obtrusion on the final element *the circle*. This analysis is hardly conclusive, and inevitably involves a subjective assessment. In this case, further evidence included a comparison with responses from the same speaker which contained single, non-final contrastive prominences; here it was observed that the final elements in such cases generally received far less of an amplitude peak.

PAU: the scoring of a pause before the element *the square* is evinced by the markedly long duration of *the*. There is clearly some differentiation between speakers as to whether the realisation of a prominence in these simple contexts includes potentially unstressed parts of a constituent, such as an article. Here the speaker appears to lengthen the duration of this word, including some silence, in order to make more prominent the following word *square*. (Notice that here the lack of any F_0 trace on *the* is simply a result of the recording conditions, and this was not used as evidence of pause. Rather, an auditory judgement was made). This example suggests that the categories PAU and DSO are sometimes interchangeable. If, in this case, the word *the* is taken as part of the prominent element, then over the whole element a significant slow-down does occur. Two things argue against a DSO marking in this case: even taking syllable duration for *the square*, the element is only 1.3 times average syllable duration for the whole response; this extent of slowing down can in any case be expected on a final element. In those cases where DSO is marked in the analysis, an average syllable duration of about twice average syllable duration was measured. An example of this, using another NS's rendition of the same response, follows.



scored: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 4
scored: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 2

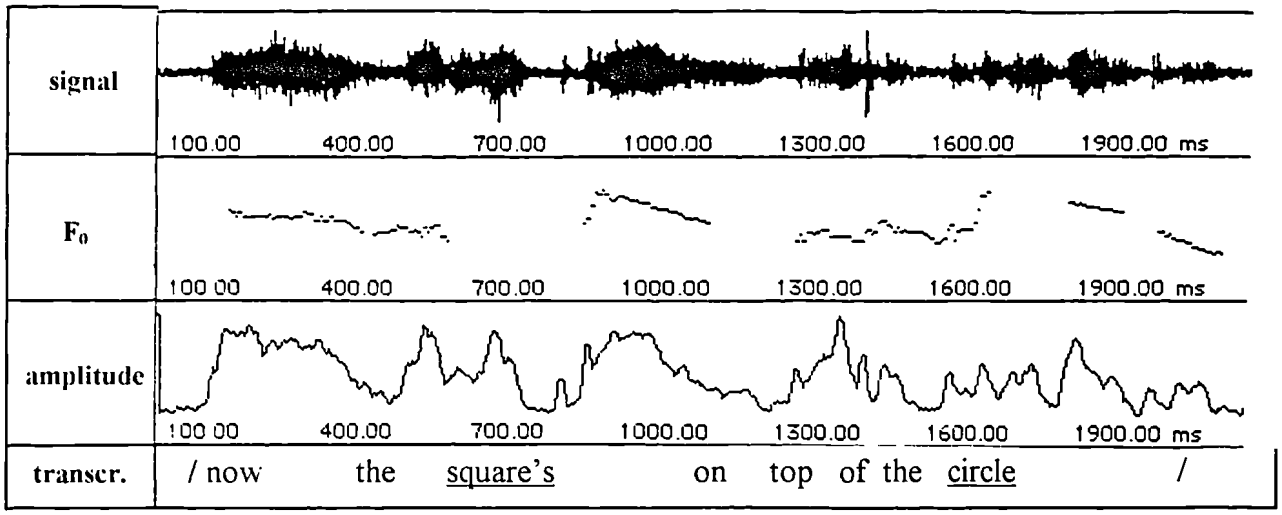
FIGURE 3.9 **A unit containing two prominences, spoken by a native speaker**

MXF: The first prominent element (*the square's*) is marked by a considerable jump up to maximum F_0 for the unit, but, since this high pitch simply corresponds to the normal jump from an unstressed, pre-head section onto the first stressed element (the 'head') no

OBT score was given, in that there is no convincing evidence that the jump here was used purely to delineate a prominence. Indeed, it reflects normal pitch behaviour associated with high heads (especially in the case of this particular speaker).

MXR: maximum pitch range is observed, because the F_0 range within the prominent element is far greater than on any other element in the unit. Clearly, marking of MXR will usually coincide with KIN (as in this case). However, MXR can occur where there is no perceptible kinetic movement, for example in a 'double contrastive unit' where the greatest pitch range occurs on a stepped pitch line. In such a case, a kinetic movement might fall on the second prominence. Indeed, such units would be considered to comprise a stepped high falling head plus nucleus in more traditional analysis. In the example above, though, both prominent elements have kinetic movement. This illustrates why a strict 'nuclear' analysis has not been adopted in the present research. As Brown *et al.* (1980) note, a strict descriptive model of intonation is often inapplicable to real spoken data, resulting in convoluted and awkward explanations of why the data is at variance with the model. Here, although reference is made to heads, nuclei and the attendant terminology of traditional British description, the analysis is far more flexible, and allows for such phenomena as two 'prominences' within a single intonation unit, without the necessary identification of one as a head (i.e. subsidiary, pre-nuclear).

DSO: this element was uttered at more than two times the average syllable duration for the whole unit, and was thus scored for DSO.



score: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 4
score: MXF MNF MXR OBT OBF KIN MXA PAL PAU DSO = 3

FIGURE 3.10 **A unit containing two prominences, spoken by a native speaker**

MXF: both contrastive prominences here have amplitude peaks either maximal in the unit or significantly greater than predictable from their position.
KIN/OBT/OBF: the first element appears on the F_0 trace to have a clear kinetic movement. However, auditory analysis overrides this in the first case. Instead, the word *square* seems to be obtruded to and from, with a marked jump down to baseline following it. For this reason, this element is scored for both OBT and OBF, but not KIN. There is no unequivocal means of distinguishing conceptually between what Bolinger called 'obtrusion from' and a kinetic tone which effectively takes the pitch line away from the prevailing level within the central portion of the prominent element. Like many aspects of intonational analysis, something of a compromise is needed, and in the present study that compromise is built into the scoring system. The second of the two (contrastive) prominences in the above example, (*the*) *circle* exhibits 'obtrusion to', which in this case is an upwards obtrusion.

3.3.3 Elicitation and analysis of paratones

As described above, an attempt was made to write a text which followed a predetermined informational structure, and which might be expected to be read in a way which required the use of paratonic devices. Great care was taken in the elicitation to make sure that all informants had unlimited time to understand the text, and to familiarise themselves with it thoroughly. Nevertheless, important issues relating to the exact nature of read versus spontaneous speech remain unresolved here; whereas it is the case that work on paratonic structure generally takes spontaneous speech as data,[10] an analysis of paratonic structure across a large informant group poses insurmountable problems. Spontaneous data are very difficult to handle in subsequent analysis. In Brown *et al.* (1980), for example, some of the most detailed observations on paratonic structure were made based on data from only a very small number of recordings from their original data. Indeed, they base their most detailed claims on the speech of only one informant, explaining this decision in terms of the very time-consuming nature of analysis (p. 65). Wennerstrom (1994) assesses the production of a 'paratone' by measuring whether there

[10] Indeed, I have argued for the use of spontaneously elicited data, above.

are differences in pitch and loudness at only two points in a read-out text, where only one of these is seen as 'paratone initial' (p. 412).

In the present study, I have preferred to specify a model of intonation which, in terms of both elicitation and analysis, remains manageable in the analysis of a large informant group, but which is nevertheless broad enough so that substantial claims can be made. From the literature on paratonic structure mentioned above (and discussed at length in 1.3) I will look at the ways in which pitch and pause are used as framing devices within these readings. In particular, I will use the NS data as either confirming or rejecting the presence of clear prosodic marking of shifts in topic structure within the text. Where NS production is regularly sensitive to the informational structure of the text, I will take these regularities as representing nativelike production.

Several difficulties arise in the comparison on group-wide data on pause (duration) and pitch (height). Readings have to be normalised across the group. For pause, the average word duration was calculated and pause durations expressed in terms of number of word-lengths. Hence, across the group we have a measurement of what proportion of speaking time is devoted to pauses by each speaker, expressed in a comparable form (although see section 6.2 on difficulties with this approach). Pitch was converted to semitone scores and then normalised, using a standard procedure (see t'Hart, Collier and Cohen 1990, p.24). I will discuss these issues at greater length in the next three chapters, along with the results of analysis.

4.0 Introduction

This chapter describes the production of prominences in the data from the screen-question tests. Firstly, findings on the marking of 'normal' prominence will be presented and discussed, including differences between native and nonnative speech. A similar account of 'contrastive' prominence is then presented. Finally, some comments will be made on the nature of intonation unit. Only the data from the Spanish-speaking informants and the native English speakers will be discussed in this chapter.

4.1 The prosodic marking of 'normal' prominence

There were a total of 37 NS and 112 NNS responses in which the principal intonational unit was unequivocally 'normal'. NNS data were arranged into four subgroups, reflecting overall spoken ability, as described in Chapter Three.

An amended version of Wells' prominence marking model (see section 3.3.1, above) was used to generate a 'score' for the prominence in each unit, indicating how many of a possible ten prosodic devices were observed to be present on the prominent element. Scores were then averaged for the NS group and for the four NNS subgroups, giving an average score per normal prominence for speakers within each group:

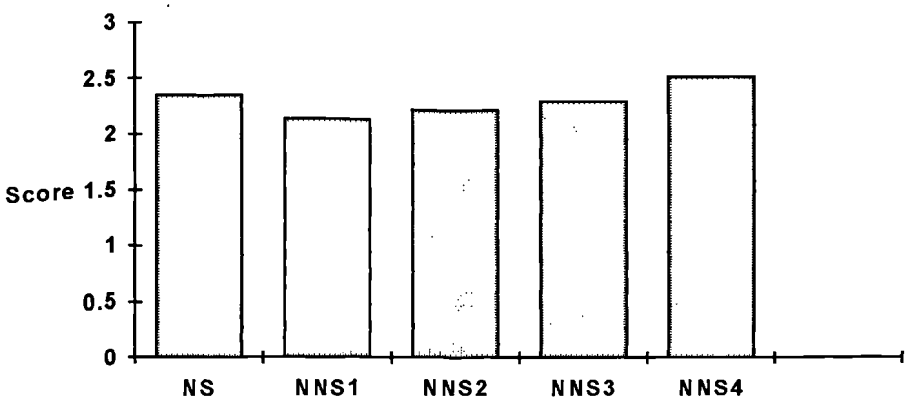


FIGURE 4.1 **Average group scores for 'normal' prominence**

As I outlined in Chapter Three, scores on each prominence analysed reflect how *many* of the ten possible marking devices are observed to be present. Thus, the group scores here serve as a rough indication of how much of speakers' available prosodic means are employed on average to mark these points of prominence. The group average score for NSs here is just under 2.5 for the production of normal prominences.

With the NNS groups, the lowest-ability group (NNS1) has the lowest average score, and there is a gradual increase in average scores throughout the four ability NNS groups. Indeed, in group NNS4, nativelike levels are surpassed. Tentatively, then, we might interpret this as evidence of developing acquisition.

However, this increase in scores may be explicable in terms of increasing confidence in L2 speech, rather than in the 'acquisition' of target-like prosodic systems. That is, as speech becomes more proficient, it becomes easier for speakers to exploit the natural prosodic resources of production. In this sense, such findings may not reflect processes of L2 acquisition at all, but rather the particular conditions of L2 speech production, in which the natural parameters of prosody are to some extent narrowed or constrained in less proficient speakers, with a gradual falling-off of these production constraints as proficiency increases.

As well as looking at these rather abstract scores representing total prosodic marking, we can also look more closely at the development of the use of the three prosodic features in these data. The analysis used here permits us to sub-categorise scores according to use of pitch, loudness and duration. For this purpose, the items in the scoring model are grouped separately within their broad prosodic feature category:

pitch-related items:	maximum F_0 minimum F_0 maximum F_0 range obtrusion to obtrusion from kinetic movement
loudness-related item:	maximum amplitude/obtrusion
duration-related items:	pre-allegro pre-pause drawl/slow-down

Hence, in Figure 4.2 a more detailed account is presented of the extent to which speakers rely on each of the three prosodic features in the production of prominences:

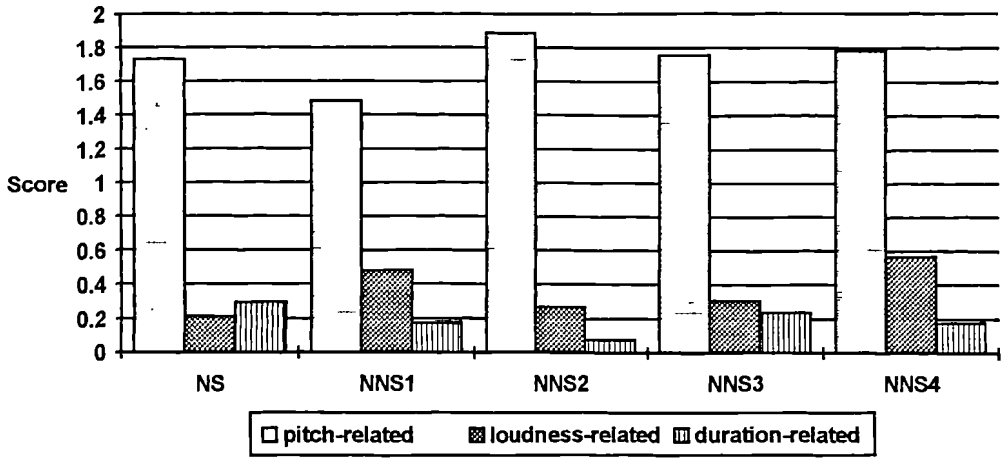


FIGURE 4.2 Average use of prosodic features per 'normal' prominence

In all but NNS1, pitch-related marking is at about nativelike norms. Loudness, on the other hand, appears to be a consistently over-used feature in all NNS groups. Loudness perhaps relates most directly to notions of excitability and the physiological nature of gesture, and for this reason may be the most natural means of creating points of prominence in a language whose exact prosodic configurations have not been mastered by a learner. I have already noted that increased loudness is a general feature of production in which maximum communicability is aimed at (caretaker talk and foreigner talk), and thus we might understand these findings as evidence that a comparable strategy operates also in SLA. Certainly, it appears not to be an effect of L1 transfer, since there is no decline in this overuse of loudness, which we might otherwise expect. Durational marking is consistently under-used in all NNS groups, although the generally very low scores in all groups make strong claims difficult to support here.

These scores, as I have noted, are simply rough measures of the nature of average, typical productions, based on real spoken data. They are useful in that tendencies can be expressed numerically. One such tendency in these data is the far greater variability in NNS than NS production. Taking an average score for all normal prominences by each individual speaker, whereas the NSs are very consistent in the spread of speaker-average scores (2.00-2.67), the entire NNS group (which is, of course, a far larger group) varied considerably more (1.33-3.75).

High variability is observed to be a characteristic feature of L2 production, as I discussed in section 2.2.4, above. However, of greater interest here is the production of specific prosodic features and how this production may be more or less variable in NNSs. Of the 37 NS responses, only one (2.7%) was marked by neither a kinetic pitch

movement nor a pitch obtrusion. This reliance on pitch *movement* to indicate prominence appears to predominate in the NS group; maximum amplitude and maximum F_0, for example, are often found outside the scope of a unit's prominence (for example in unit-initial, or 'head', position), whereas the salient movement of pitch - its principal kinetic movement, its clearest point of obtrusion - will almost always coincide with the prominent element. On the other hand, 8 NNS responses (14.1%) contained no pitch-related marking of normal prominence. So, although NNS group averages for pitch marking were generally comparable to that of the NSs, production was more variable.

4.2 The prosodic marking of 'contrastive' prominence

Contrastive prominence is an interesting aspect of production for several reasons. Firstly, it is associated with a deviation from the norms of prosodic production, in which speakers tend to make overt, heavily marked prominences, rather than those typical of 'normal' prominence marking. It is a measure, in this sense, of how successfully they can optimise the clarity of a narrowly focused utterance using prosodic means. Also, contrastivity often entails shifts in the placement of prominence in the sentence, in which the end-focusing of normal production is disturbed, and the position of the prominence is shifted leftwards in the unit.

There were 32 NS and 102 NNS units containing clear, unequivocal use of contrastive prominence. Some subjects responded to a 'contrastive' question in a way which was clearly inappropriate, so that the predicted contrastive element was not marked at all; some made an attempt at contrastive marking which left two elements partially marked.

4.2.1 'Inappropriate'[1] contrastive prominence

Failure to mark a predicted contrastive element at all was observed only in the NNS subjects. This might be seen as evidence of an inability to recognise which item required narrow focus in the context of a screen-question. However, two factors argue against this. In

[1] The term 'inappropriate' is somewhat misleading; these data are simply responses which did not conform to expectations within the highly controlled context in which they were elicited.

the data there are only six examples in total, spread between five NNS speakers, with at least one speaker from each of the four subgroups. This suggests that the phenomenon is not related to any shortfall in productive ability if we assume that influences would diminish as proficiency increased. Also, all speakers who in one or two 'contrastive' responses failed to mark the predicted item nevertheless did so successfully in all other responses. These six examples, then, appear rather to represent aberrant data, explicable in terms outside the scope of this study. Especially bearing in mind that the NSs produced no such responses, we can assume with a fair degree of certainty that the nature of the experiment, and any number of consequent influences on production, might have led to these particular responses. These six examples are not included in the subsequent analysis.

4.2.2 'Appropriate'[2] contrastive prominence

NSs produced unique, contrastive marking on predicted, narrowly focused elements 25/32 times and the NNSs on 76/102 occasions. That is, there was strong agreement amongst all speakers that a certain element (typically not rightmost in the sentence) required contrastive prominence in the response. From this we can assume that the elicitation was successful in its aims.

In addition, about 20 percent of responses, both by NS and NNS, contained some 'sharing' of prosodic weight on two separate points in the unit. This phenomenon will be discussed below (section 4.2.4, below). This sharing occurred only in response to questions which were designed to elicit contrastive prominences, and from these responses the score on the predicted point of (contrastive) prominence will be taken here, given that there is, in all cases, an observed attempt to create a prominence at this point. The data on contrastive prominence, then, are drawn from a total of 32 NS and 96 NNS responses.

Group average scores for contrastive prominence production were compared to scores for normal prominences:

[2] See previous footnote.

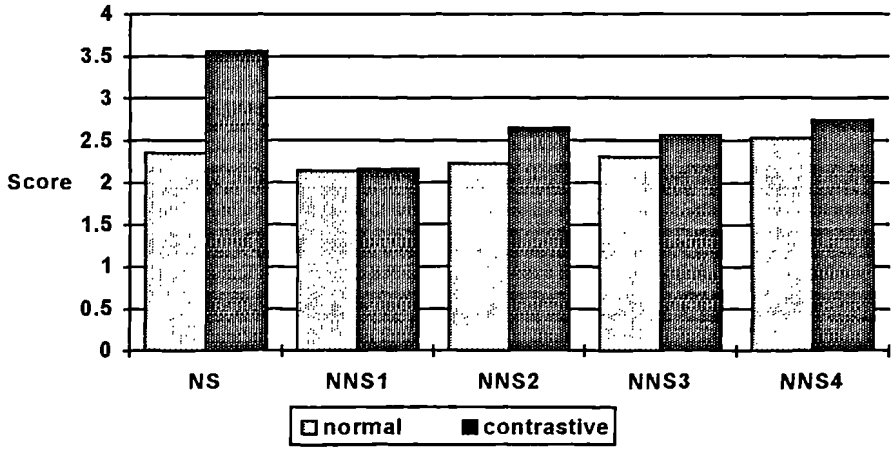


FIGURE 4.3 Group-average scores for 'normal' and 'contrastive' prominences

There is a striking increase in the NSs' scores for contrastive prominence. However, the NNSs do not produce any such increase and, indeed, their scores are only slightly (although consistently) above those for normal prominence in each subgroup. There is only a very slight increase from groups NNS1 to NNS4 in the extent of contrastive marking, indicating that the nativelike production of contrastive prominences does not significantly increase as proficiency does in these data.

Again, these data can be broken down to show the use of the three prosodic features:

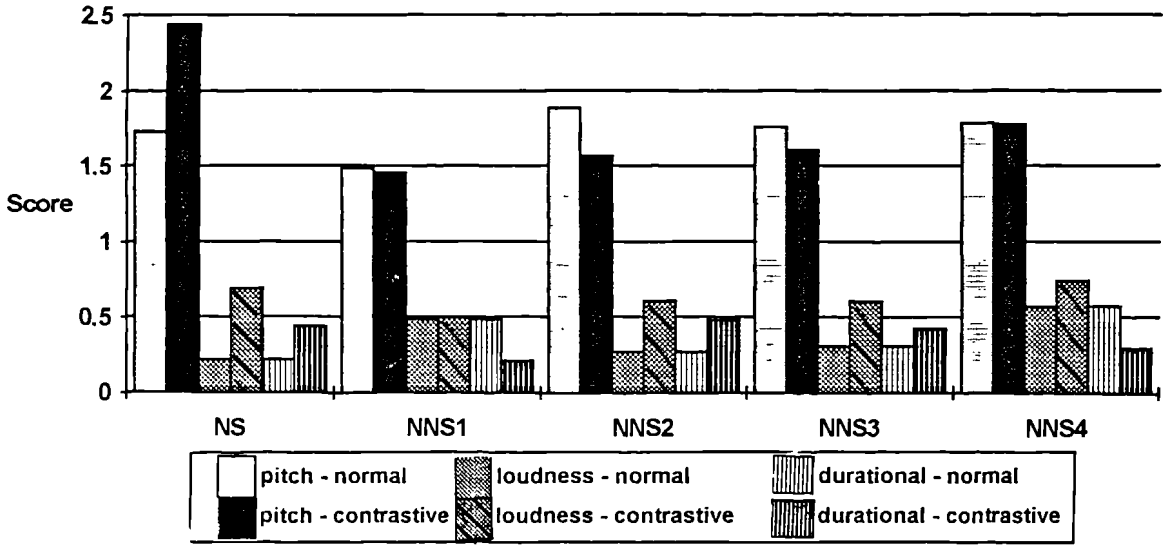


FIGURE 4.4 Use of prosodic features in normal and contrastive prominence

The NSs use pitch in the marking of contrastive prominences significantly more than in normal prominences. On the other hand, the NNS data shows no such increase, and the use of pitch marking is actually lower in contrastive than normal prominences in groups

NNS2 and NNS3. This is a very clear finding; NNSs recognise the need to make prominent an element which would not otherwise stand out (most contrastive prominences were sentence-medial), yet the degree of pitch-marking, as expressed in the scores here, does not exceed that for normal prominence. So, the NNSs seem not to show any sensitivity to the use of pitch as a means of creating particularly heavy points of prominence.

Loudness is three times as likely to be used as a marker of contrastive than of normal prominence in the NS data. This degree of increase is not repeated in the NNS data, largely, one suspects, because loudness is already an over-used resource in the marking of normal prominences here. The use of loudness does, however, increase somewhat in the NNSs' contrastive prominences, and this reaches nativelike levels in group NNS4. Throughout groups NNS2 and NNS3 there is actually a reduction in the (excessive) loudness used to mark normal prominences, and this may be connected with the increase in loudness as a marker of contrastive prominence throughout the groups. That is, NNSs may initially rely on a simplified kind of production in which all types of prominences are flooded with loudness. The ability to distinguish normal from contrastive prominences then evolves over time in terms of a refinement in two directions, with more marking on contrastive and less on normal ones. This would appear to reflect the staged acquisition of the use of loudness noted by Wieden (1993) in a quasi-longitudinal study of L2 English (see section 2.4). He finds a gradual differentiation in levels of amplitude to create prominence over time.

Scores for durational marking are, again, below significant levels, although there appears to be a greater (more nativelike) use of them in contrastive than normal prominence marking in the NNS data.

The majority of responses with contrastive prominence follow a very straightforward pattern here, in which the contrastive element is in direct contrast to a same-class element in the previous response. Typically this narrowly focused element is either a shape name or a preposition, and occurs in a non-final position in the sentence, although the response to one question yielded a final-position contrastive prominence, which I will discuss separately in section 4.2.3, below. In several of the NS responses the contrastive prominence is followed by a clear kinetic movement in the final (non-prominent) element in the sentence, perhaps because the contrastive element is too far from the end of the sentence to carry the whole of the principal pitch movement, and speakers feel the need to terminate their sentence in some overt way. In traditional

British analysis this phenomenon is a recognised feature of post-nuclear sections, especially where the nucleus falls some way before the end of the unit, in which cases the nuclear tone is seen to describe its movement from the nucleus and stretching into the tail. In the present study, because scores for contrastive prominence reflect only marking on the element itself, this is a problem, given that a kinetic movement which may *commence* its characteristic movement on the contrastive element may in fact not receive a 'KIN' score, due to the location of the *perceived* movement lying outside the scope of the prominent element itself. Indeed, NSs were observed to mark contrastive prominence with kinetic tone marginally less often than they did normal prominences (and average of 0.56 instead of 0.59 per prominence), and the increased use of overall pitch marking for contrastive prominences in the NS is explained by the increased use of pitch height, which for prominences in medial position is realised either by maintaining the high pitch of a high onset, or is accompanied by a sudden obtrusion in pitch, usually upwards, but occasionally downwards.

4.2.3 Contrastive prominence in right-most position

Further evidence of the failure of NNSs to mark contrastive prominences with sufficient prosodic weight can be seen in the one screen-question in which a contrastive prominence was required in final position to communicate correctly the narrow focus on that element. This is the context in which it was elicited:

Question (2.4):	Where's the triangle now?
Typical response:	/ it's on the <u>circle</u> / (normal)
Question (2.5):	Is the triangle on the circle now?
Typical response:	/ no / it's on the <u>square</u> / (contrastive)

If the response to 2.5 is to make a communicatively appropriate contrastive prominence on *square*, we might expect that the score for this contrastive prominence will be higher than other, non-final position contrastive prominences, given the narrow focus and also the need to differentiate it from default 'normal' prominence placement on the last lexical item. Group average scores for this particular prominence are here compared to averaged scores on all normal and contrastive prominences in the data:

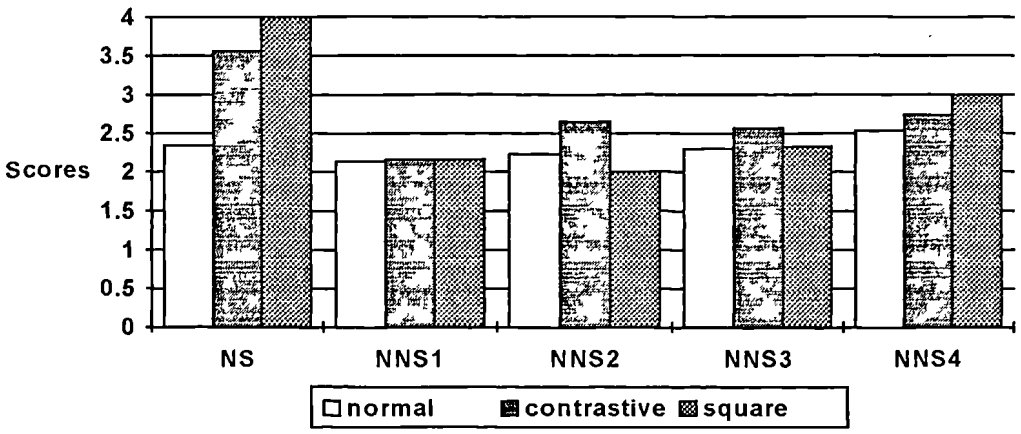


FIGURE 4.5 Strength of marking on 'normal' and 'contrastive' prominences and on the word *square*

The NSs do indeed mark *square* most heavily. The NNS groups show that only in the final group, where spoken ability is highest, do speakers manage to make this distinction through prosodic means, which is highly suggestive of late development of this aspect of production. The NNS data as a whole here may indicate that contrastive prominences are difficult to realise for two reasons: the sentence position of such prominences goes against predictable (contrastive) accentuation patterns, as also does the need to create extra-strong prosodic marking in this position. We have already seen that contrastive marking is itself slow to emerge in the NNSs, yet here even heavier marking than a nativelike contrastive prominence is required. This may explain the fact that in all but group NNS4, *square* received prosodic marking which only matched group averages for normal prominences, and fell short even of this in group NNS2.

4.2.4 Prominence shared by two elements

In some responses an appropriate contrastive prominence is accompanied by a second point of marking in the same unit which to some extent shares the perceived prominence. 7/32 (21.9%) of the NS and 20/102 (19.6%) of the NNS responses carry such marking.

In the NSs, this appears to represent a gradual process of the shifting of the point of prominence, from final to (contrastive) non-final position, and seems to be linked to the nature of the context. Of the seven examples in the NS group, three speakers produce such shared-marking on responses to the same two questions (questions 1.3 and 2.4, see Appendix A). Both of these questions were the *first* in sequences which were

designed to promote contrastive responses. So, the non-prominent elements of the contexts of these two questions - that is, the 'given' elements - were evidently *less given* than in the case of the following question in each case, since here the context is exactly the same, and hence the 'given' status of backgrounded information is reinforced.

In these NS data, the shift was from a unit-final normal prominence position to a non-final contrastive one. The 'sharing' of prosodic marking, then, appears to reflect production in which speakers produce units which retain some of the phonetic profile of 'normal' production, whilst some shifting of the available marking leftwards onto the contrastive element also occurs. That is, relative to the intonational form of the previous utterance, speakers shift some (but not all) of the available prosodic means onto the contrastive element. Evidence for this can be seen in the responses to the *following* questions in each case (questions 1.4 and 2.5), in which all three of these (native) speakers produce fully contrastive intonations, with no perceived prominence remaining on the final element.

In the seven cases of shared prominence in the NS data, only in one case was there actually a higher score on the element which was *not* predicted as (contrastively) prominent; one other case had the same score on the two points of prominence. Of the remaining five responses the predicted contrastive element had considerably heavier prosodic marking. Indeed, in three of these cases the presence of a second point of prominence (identified by native-speaking judges) could not be supported using mechanical analysis at all, and was identified only in auditory judgements. The marking of prominence by maximum amplitude peak (or by amplitude obtrusion) was located on the predicted contrastive element in 5/7 of these responses, was shared (both elements marked) in one further response, and showed no clear marking function (on either element) in one other.

These particular questions, then, seem to have provided examples of the process whereby prosodic resources are shifted gradually leftwards, as the degree of 'givenness' of other elements within a developing context increases. This phenomenon perhaps corresponds very broadly to Ladd's 'deaccentuation' (Ladd 1980), and in the present data the simplified context used results in the process being seen in operation over several responses.

The majority of the 20 NNS responses of this type were, like those of the NSs, to questions 1.3 (6 responses) and 2.4 (8 responses), but with three examples each from the *following* question (questions 1.4 and 2.5), which in the case of the NS group were

always fully contrastive, with no evidence of shared prominence. If the increasing extent of contextual givenness throughout these series of questions had a direct effect on the extent of the shifting of prosodic means leftwards onto the contrastive element, then perhaps the fact that this process of shifting spilt over into the subsequent response only in the NNSs' productions suggests that the context is handled in the same way by the NNSs. However, changes in contextually derived knowledge relating to the relative informational weight of potential points of information takes longer to be encoded prosodically in speech production. This may be due to the relatively high cognitive load required simply to speak in an L2, leaving the speaker with less capacity to deal with contextual changes than in the case of L1 production. Hence, appropriate changes to the intonational form of speech are observed to take longer to effect. This again illustrates the importance of interpreting prosodic production with reference to context. In this case, it appears that cognitive factors may influence L2 production.

Of the twenty NNS responses with shared prominence marking, eight had higher scores on the non-predicted element, whilst only one had a clearly higher score on the predicted contrastive element. So, although the predicted contrastive element is marked, the shift of prosodic resources tends to be smaller than in the NS responses. In general NNSs' responses display a far greater degree of spreading of prosodic resources than those of the NSs. For example, NS units containing non-final contrastive prominence where this prominence is not marked by increased amplitude typically have a steadily declining series of amplitude peaks throughout the unit, so that at no point is loudness an obtrusive characteristic of the unit; in other words, obtrusions of amplitude, where not used to mark a non-final (contrastive) prominence, are not subsequently used in the unit. However, the converse is often true in NNS units; there is less marking on contrastive prominences, and at the same time there is also un-nativelike maintenance of other markers, especially amplitude, on other elements. In the example below it can be seen that the speaker maintains an amplitude peak on the final element, despite marking a non-final element as prominent (and being judged to have done so):

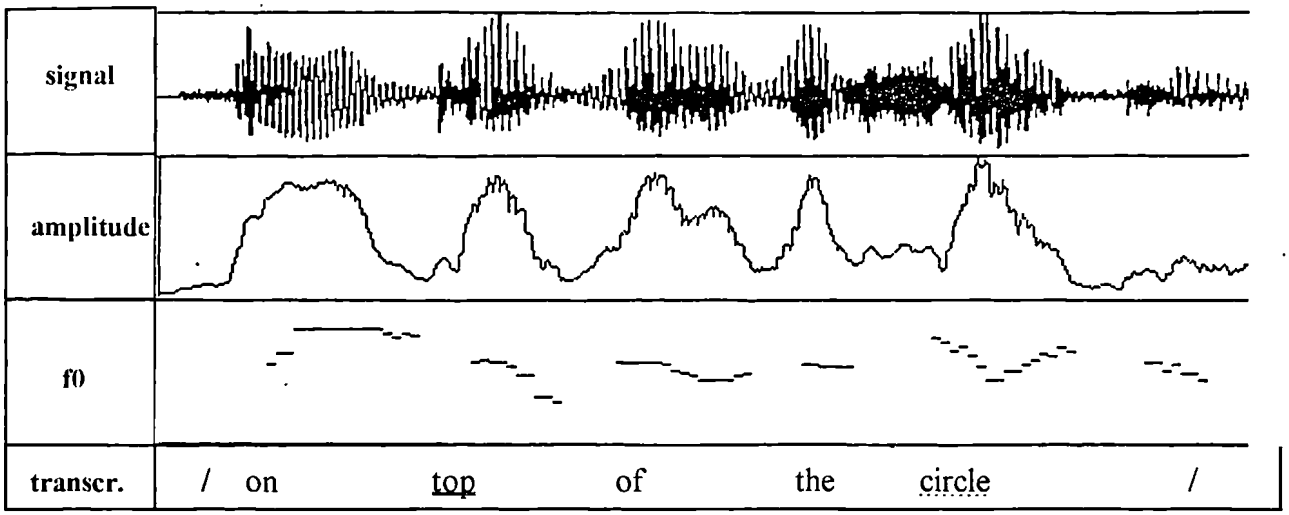


FIGURE 4.6 A unit containing a (non-final) contrastive prominence but with clear amplitude obtrusion on the final element in the unit, spoken by a nonnative speaker

This suggests that, having recognised the need for contrastive prominence marking, the speaker here shifts only some of the available resources from their 'natural' final position, and indeed makes a considerable articulatory effort in maintaining a final-position amplitude peak. There were eleven NNS responses in which equal amplitude peaks were noted on the predicted, non-final contrastive element and on the final element in the unit.

In 13/20 of the NNS 'shared-marking' responses, the predicted contrastive element was marked by maximum amplitude/amplitude obtrusion, which represents a less consistent use than in the NS responses here. However, many of the NNS responses had low marking scores on the predicted contrastive element in general, and in half the responses amplitude was either the sole marker or one from a total score of two, using the ten-marker scoring model. So, although the overall process of prominence shifting appears to take longer to effect, amplitude is the prosodic feature most easily shifted from one part of the intonation unit to another by NNSs. If we recall that, in 'normal' prominences, NNSs over-used amplitude marking, then their ability to shift this marking onto contrastive elements seems to suggest that loudness is the feature which speakers most readily associate with creating points of prominence in utterance, and which they are most easily able to employ in the marking of prominence at different sentence positions.

4.2.5 Units containing two points of prominence

There were two screen-questions that elicited responses in which two (potentially contrastive) prominences were necessary in order to communicate successfully the full informational import of the response relative to the previous screen (1.5 and 2.6, see Appendix A). Below is the second of these questions, together with the previous screen-question:

Question (2.5): Is the triangle on the circle now?
Typical response: / no / it's on the square /

Question (2.6): What's different now?
Typical response: / now the square's on the circle /

Both nouns in 2.6 are predicted to be contrastive, given that they have changed roles within a sentence frame which is the same (NP *be* PP[loc]), the preposition being a given element here. Throughout the speaker groups, the responses to these questions were surprisingly close to the desired frame in most cases. Below is a summary of where speakers placed prominences:

placement of prominence(s) (underlined)	NS	NNS
NP *be* prep. NP	8	30
NP *be* prep. NP	7	4
NP *be* prep. NP	1	2
NP *be* prep. NP	-	4
NP *be* prep. NP	-	6
NP *be* prep. NP	-	2

TABLE 4.1 **Number of examples of different prominence placement in NS and NNS**

I take the first two patterns of prominence placement here to be context-appropriate. Although in the second (NP *be* prep. NP) the preposition is unnecessarily prominent, this was in all NS responses part of a unit in which all three elements were marked as informationally strong (I discount here the possibility of prominence in the copula), and I take this to be a pattern which reflects the speaker's perception that 'everything has changed' relative to the preceding context, leading to a tendency to mark all available

elements. Crucially, only one case of prominence marking *not* falling on both NPs occurred in the NS data.

From table 4.1 it can be seen that the NNSs in general succeed in marking the predicted appropriate prominences. However, 12 NNS responses (one quarter) are, within this analysis, contextually inappropriate. I have already noted, above, that NNSs may take longer to shift prominence onto sentence-medial loci of narrow prominence. In the case of 'double prominences' here, the context was very explicit, yet the necessary pattern of marking was itself a complex one. It seems unlikely that the contextual 'newness' of both NPs would *not* be appreciated by NNS subjects, the context here being not only very simple but also language-independent. On the contrary, it seems that failure to place prominences appropriately in the NNS data can be attributed to problems in the production of such an accentuation pattern during speech.

Data were also classified according to the four NNS subgroups. The spread of both appropriate and inappropriate prominence placement was not seen to be related to level of ability. Indeed, the lowest of the four ability groups produced the highest percentage of responses which had 'appropriate' double prominence on the two NPs. Moreover, fifteen out of twenty-five NNSs produced at least one response which fell into one of the four 'inappropriate' categories. So, not only is inappropriate placement common to NNSs of all levels, it is not confined to a few speakers but common to the majority. The production of such sentences, with two obligatorily prominent elements, appears to pose serious problems for even highly proficient NNSs, and this may have to do with the upper limits on informational load which can be incorporated into the planning and realisation of these sentences.

4.2.6 Prosodic marking of two contrastive prominences in a unit

How strong is the prosodic marking within these units containing two points of prominence? Data from responses to questions 1.5 and 2.6 are here taken together, since their contexts, and hence the predicted response, were intended to be identical. Figure 4.7 compares group average scores for normal and contrastive prominences cited earlier in this chapter, and the average scores on the prominences in the two-prominence units:

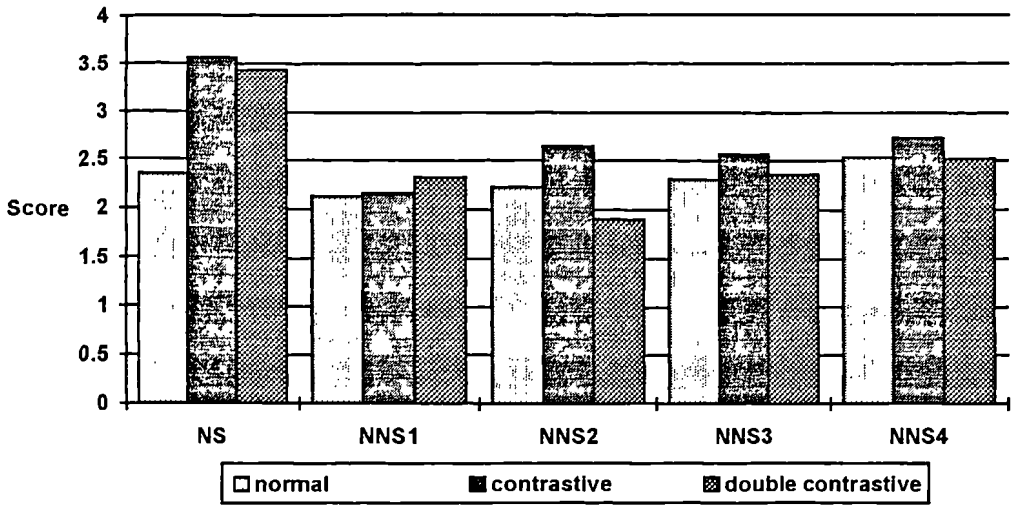


FIGURE 4.7 Group-average scores for marking of normal prominences, contrastive prominences and for the marking of prominences in the 'double contrastive' responses

Despite the fact that *both* loci in the two-prominence units demand heavy (contrastive) marking, the NSs here employ prosodic marking on both these two points in the unit at a level almost equalling that of their production of single contrastive prominences. NNSs clearly fail to mark these double prominences, despite the extremely overt nature of their (contextual) contrastivity in the elicitation. Indeed, group scores here remain at 'normal' prominence levels, just as they do in the case of single contrastive prominence. It appears, then, that whereas the NSs' intonational competence includes the ability to mark two contrastive prominences, both with strong prosodic marking, the NNSs show no ability to do this, even in the most proficient group.

Therefore, despite NNSs' recognition of the *requirement* to mark as prominent two elements in a unit, we again see that their production does not include the use of prosodic means to mark contrastivity here, as reflected in the scores.

We might see this as potential evidence of a general lack of flexibility in prosodic production in response to context. Such nativelike flexibility appears here to be resistant to acquisition by NNSs. L2 production, perhaps, remains relatively static in this area, as evinced also by the data on (single) contrastive prominences, above. This may have important implications not just for models of SLA, but also for approaches to the teaching of L2 speech skills.

4.3 Unit-divisions and the use of pause

Relatively little has been said in the SLA literature about the production of intonation units, although as I noted in section 1.2.3, they play an important role in the way that information is structured in spoken language. Several writers have noted the general tendency for nonnative speakers to produce intonation units of a shorter average length than in comparable target-language production (Jenner 1984; Brazil 1986; Hewings 1993). This really reflects two features of production; the tendency to produce shorter utterances in general, and the tendency to say less within a single unit.

From all the recorded data discussed above, all responses which were perceived as being one unit long, and which contained no internal pauses, were assessed for length:

NSs (n = 68)	6.47 words per unit
NNSs (n = 153)	4.83 w.p.u.

In the calculation of unit length in the present study, all units containing pauses were removed from the data. It is these units containing pauses which I will now discuss.

Pause is not solely a means of marking unit divisions, especially in nonnative speech. The intrusive nature of processing limits, which often results in linguistically unmotivated pausing, may seriously complicate any analysis of unit divisions, not least because the perception of there being more units may lead to the identification of prominences which, rather than being linguistically motivated, are merely the result of our native expectations that a unit will, in normal cases, contain at least one prominence.

In the case of nonnative speech, then, some generalisations which serve to illustrate the relationship between pause and intonation units are required. For example, in the response in Figure 4.8, below, the speaker appears on first analysis to produce a three-unit utterance. However, the pauses here in fact do little other than divide what is, in other respects, a downwards-stepped contour with a single nucleus on *circle*:

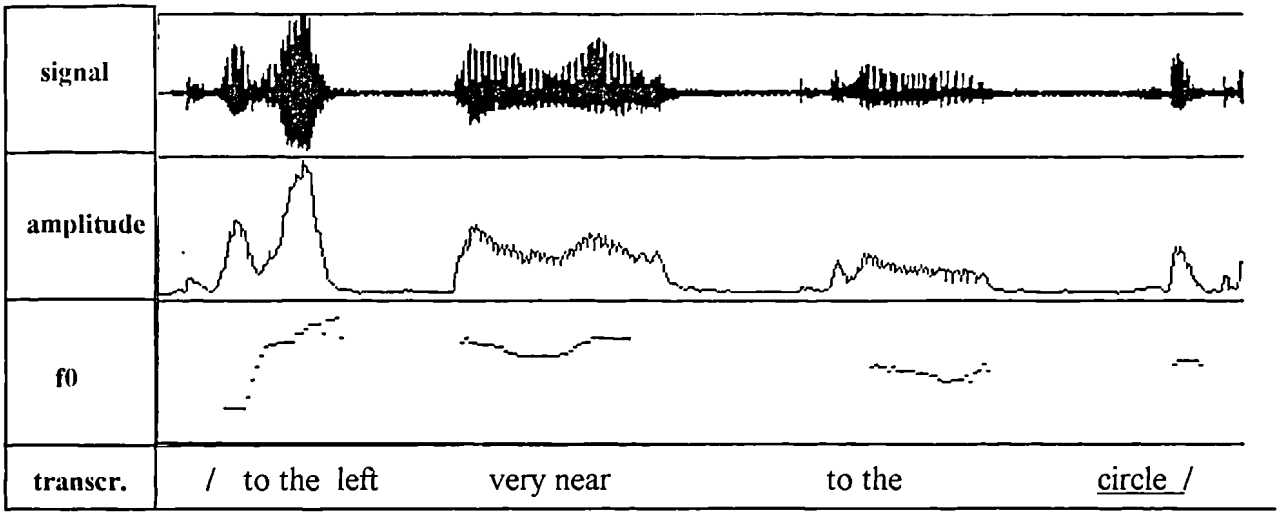


FIGURE 4.8 A unit containing a normal prominence, spoken by a nonnative speaker

Pausing here appears to divide what might otherwise be a single unit; that is, an utterance which was *planned* by the speaker as one unit long.

In 214 NNS responses from the screen-question tests, 61 contained substantial pausing. In 12 cases these pauses were clause-internal and filled. Such filled pauses may be the result of delayed word-retrieval or of a deliberate attempt to maintain fluency (and, hence, audience attention) through 'buying' extra production time. Pausing strategies of this kind have been discussed before (see Raupach 1984). On the other hand, unfilled pauses are potential markers of unit divisions.

In order to look more closely at the relationship between pauses and intonation units in the NNS speech, the remaining 49 NNS responses, containing unfilled pauses (or a combination of filled and unfilled), were divided into two groups. All responses to the screen-question tests had previously been transcribed manually; responses which had been transcribed as containing two units (18 tokens) were grouped together, and units which had been transcribed as one unit, but containing a pause, formed the other group (31 tokens). The two groups were then considered in the light of both their linguistic and their intonational characteristics.

4.3.1 Two-unit utterances

Although the elicitation of prominences was designed to yield responses which were of predictable intonational structure (one intonational unit was expected), the NNSs in

many cases produced speech less easily categorisable as such than in the case of the NSs, who produced no two-unit utterances here. Out of 18 NNS utterances which were clearly identified as containing two units, only two are divided at a point which splits an XP constituent. The other 16 are composed of a syntactically well-formed first unit, followed by a second unit which in all cases contains additional material (i.e. is not a tag, or a conversational device of any kind). The first units of these two-unit utterances are not shorter but in fact *longer* (average 6.53 words per unit) than the average word count for NNSs' single-unit responses (4.83 w.p.u.).

A possible explanation of the use of a greater number of units by L2 speakers here is as a strategy for avoiding the heavy processing load involved in longer, informationally richer units. A strategy such as this appears to reflect similar characteristics of careful speech (in L1s), in which shorter units free up processing time for planning. However, a slightly different explanation is that these divisions indicate points where, although processing is a delaying factor in production to the extent that a break in production occurs, the speaker nevertheless attempts to maintain supra-unit cohesion to mitigate the intrusion of the unwanted pause. The significantly longer duration of these first units in NNSs' two-unit utterances appears to suggest that an upper limit on processing has indeed been reached, the speaker then providing whatever remaining information is perceived as requisite in a following unit; the division is not motivated by the context but is forced upon the speaker, who attempts to maintain intonational cohesion in the following unit.

In the above cases there is evidence that the NNSs can use intonational devices to maintain cohesion in this way. Of the 18 responses, in nine cases the speaker used a rising (or high-level tone or a fall-rise) towards the end of the first unit and a fall to baseline at the end of the second; such a rise and fall sequence in two adjacent units has often been seen as the most cohesive two-tone sequence (see section 1.2, above). In another, a fall to mid-range in the first was followed by a fall to baseline in the second. In three responses, the same-type kinetic movement described a narrower pitch range in the second than the first unit (a form of pitch-marked subordination). The five remaining responses contained a rise or a fall in both units. So, the majority of responses appear to contain some attempt at inter-unit cohesion. A *non*-cohesive unit pair, on the other hand, might have an unresolved tone sequence (fall and rise, or fall and fall-rise), or might have pitch which falls to baseline in the first but not in the second, indicating a point of finality

in the first unit, followed by a second tone which is (presumably) resolved in a subsequent unit.

4.3.2 Intra-unit pause

The elicitation of very short responses was intended to reduce the occurrence of mid-unit hesitations. From 69 NS responses, only one clear hesitation (of 290 milliseconds) was observed. Hesitation is a far more prevalent feature of nonnative speech. There were 31 NNS responses containing clear pause within a single unit.

Care must be taken when assessing the status of pause within the production of NNSs, as Figure 4.9 illustrates:

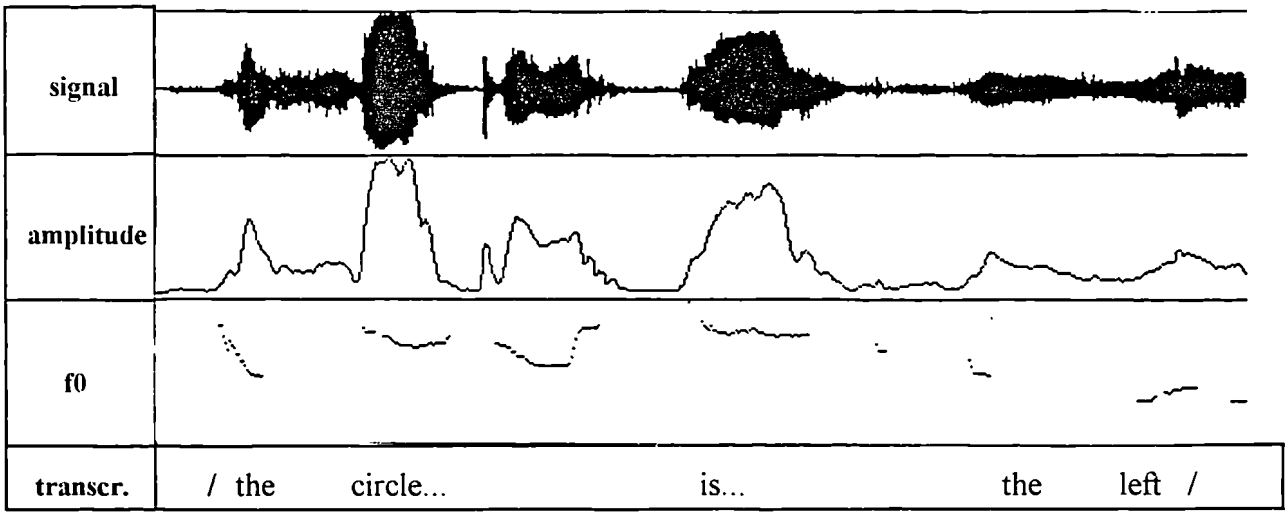


FIGURE 4.9 A unit containing a normal prominence, spoken by a nonnative speaker

In the example here there is considerable hesitation both before and after *is*. Here the speaker fails to produce a syntactically well-formed sentence. There is considerable pause, which appears to create three intonation units. So, the proliferation of units here might be seen as the result of breakdowns in linguistic processing. However, these three units are not simply reflections of how linguistic processing has broken down; they also show how the speaker has attempted to maintain a single-unit prosodic structure, given that the pausing is not intentional. Pitch and amplitude peaks on all three informationally strong elements of this utterance follow lines of declination, with a discernible kinetic

pitch movement at the end of the whole utterance, that is, in predictable right-most position. So, in Figure 4.9, whereas pause appears to detract from the overall impression of well-formedness, certain aspects of its phonological form still correspond to a single-unit pattern.

Of the 31 cases of intra-unit pause in the NNS data, many pauses are long enough to constitute unit division markers within the normal analysis of English intonation. However, according to the transcriptions, none of these responses was perceived to have a second unit. Data from responses to questions 1.2, 2.3 and 3.3 accounted for 30 out of the 31 examples. I will now discuss these in detail, since they all follow the same syntactic pattern:

NP *be* PP[loc] (e.g. *The X is* [prep.] *the Y*)

The analysis here seeks to identify whether, despite being perceived as single-unit utterances during transcription, there is any link between use of unit-pause and any observed tendency towards a greater number of unit divisions in nonnative speech. The location of pauses within the syntactic structure of the responses is summarised here:

- Pause before copula: there is one such example (that illustrated in Figure 4.9), in which two pauses effectively isolate the NP, copula and PP of the sentence;
- Pause between copula and PP: of only four examples here, three also contain some filled pausing, suggesting a delay in the processing of the following PP, especially the retrieval of the head preposition;
- Pause within the PP, but leaving NP of the PP in tact: all five examples here conform to the patter *The X is on top* <pause> *of the Y*;
- Pause immediately prior to NP of the PP: two cases;
- Pause within the NP: nine teen cases, all of which fall between determiner and head noun.

So, the majority of pauses come at positions in the linguistic structure least appropriate as potential unit divisions. On the contrary, the prevalence for pauses between determiner and noun strongly suggests that processing factors (word-retrieval) can explain these pauses. The most appropriate place for a unit division within the above

structure would perhaps be between the NP and the copula, yet here there is only one such case. Below is a representation of how many pauses fell at which points in the sentence:

the X 1 *is* 4 [prep. 5] 2 *the* 19 *Y*

Some extra evidence for the assertion that these utterances are planned as single units comes from their internal intonational structure. The presence of kinetic movement is often used as a means of identifying intonation units. None of the parts of the above responses *not* carrying the prominence contain perceived kinetic movement, although some of course contain obtrusion or maximum F_0, especially where they contain the head of a putative unit. More tellingly, there was no clear evidence of fall to baseline prior to the final section of any of these utterances.

4.4 Summary

In this chapter I have discussed the screen-question data in relation to the production of normal and contrastive prominences. The nonnative speakers in these data were all speakers of L1 Spanish. I have suggested that:

- The NNSs make less use of prosodic features as markers of normal prominence than native speakers of English, but average use increases slightly as proficiency does.
- Nativelike use of the three prosodic features of pitch, loudness and duration each show distinct developmental sequences. Generalising for both normal and contrastive prominence:
  - pitch marking is under-used by NNSs but increases towards nativelike levels in normal prominences, but shows no increased use (or development) in contrastive marking;
  - loudness is constantly over-used on normal prominences, with an overall increased use on contrastive prominences. However, this relative increase is smaller than that in NSs' normal and contrastive prominences, and NNSs approach nativelike extent of use in contrastive contexts;
  - durational phenomena appear to be consistently under-used, but were not present at significant levels in either the NS or NNS data.

- NNSs place contrastive prominences correctly when responding to simple contexts but consistently fail to mark them with sufficient prosodic weight.
- NNSs do not significantly increase the extent to which they mark contrastive prominence as their global levels of spoken English increase.
- Both NSs and NNSs can 'shift' prominence marking from default right-most position onto a contrastive element in the sentence in response to context. NNSs appear to 'shift' less of the available prosodic means in the same context.
- NSs increase contrastive marking to create emphatic prominences where the linguistic context might otherwise be unclear, such as in the case of final-position contrastive prominence. NNSs do not do this.

From the data on 'double contrastive prominences':

- NNSs are capable of creating two points of prominences within a single unit. However, in cases where the context makes this necessary, their ability to do so is neither consistent nor does it increase in more proficient speakers.
- As with single contrastive prominences, NNSs mark double contrastive prominences with no greater degree of marking than they do normal prominences.

In terms of the relationship between pausing and intonation unit divisions:

- NNSs tend to speak in shorter intonation units.
- Many pauses are in fact likely to be simple breakdowns in fluency and not intonationally significant in terms of a strategy of using shorter units.
- NNSs use intonational means to make two units cohere where such units are likely to be the result of processing limits forcing otherwise unwarranted unit divisions to be made.

Chapter Five - The use of tones to mark information

5.0 Introduction

In Chapter Two (section 1.2) an account of different kinds of 'information' was presented, including notions such as the new/given distinction and communicative dynamism. These kinds of categories of information will often receive recognisable pitch marking in English. It was also noted that the use of pitch to mark information status is related to the global 'meanings' of tone, in particular the tonal opposition of fall/rise.

In this chapter, I am interested in establishing ways of expressing with some precision the nature of nativelike tone choices in the 'packaging' of specific kinds of information in speech. This will provide a basis for the assessment of comparable NNS data, in a wider discussion of tone in L2 production. Firstly, I will consider listing intonations and their tonal cohesion, and also how this cohesion can be disturbed to accommodate other aspects of information marking. Then, I will analyse data from the screen-question tests which elicited utterances which were two units long, looking particularly at the multi-unit cohesion which can be built-up through the use of tone.

5.1 Listing intonation patterns in English

Halliday's assertion that falling tones encode major information and rising tones subsidiary information (1967b) corresponds to many similar descriptions of the basic roles of fall and rise in the prosodic marking of information. Many of Cruttenden's (1981) glosses for the meanings of fall and rise can also, very broadly, be understood as reflecting this general polarity. The intonation of lists in English exploits these basic meanings of falls and rises.

English lists are not normal sentences, and their prosodic organisation is similarly idiosyncratic. Each item in a list will usually be given its own intonation unit, and will hence carry its own prominence, including a pitch movement. Lists have a clear informational cohesion, and the structure of a listing pattern serves to bind items together to create a composite, unified structure.

The two most common listing patterns in English, expressed in terms of a series of rising and falling tones, are:

//RRRF// //FFRF//

(Couper-Kuhlen 1986: 150)

The contribution which falls and rises make to lists can also be expressed in terms of the opposition of 'closed' (fall) and 'open' (rise). Where the penultimate item takes an open tone, it draws attention to the dependency between itself and the final item; the unresolved nature of an open tone creates a sense of 'more to come', which is resolved by the subsequent closed tone. This final closed tone is most often to baseline, whereas other falls within the pattern may not be so. The pitch pattern of a list can take the form of a series of level tones, with a relatively high penultimate tone and a low final tone.

5.1.1 Production of lists by NSs and NNSs

There were two pairs of questions in the screen-question tests (questions 2.1, 2.2 and 3.1, 3.2, see Appendix A) which elicited lists as responses. In both pairs, the first question elicited the description of shapes which, whilst new in the context of the particular test, were nevertheless familiar from the previous test(s). In the second question of each pair, a 'new' shape was included as the penultimate item in the series.

By looking at the tones used to mark the last two items in lists, we can assess the speaker's use of open tones in penultimate position. Also, we can assess whether a speaker's selection of tone on 'given' and 'new' items in this penultimate position differs (i.e. whether new items are marked with tones which disturb the sequence).

The sequence of tones used on the two final items in each listing response was analysed in terms of the four possible configurations of tones, adopting the open/closed opposition:

open / closed
closed / closed
closed / open
open / open

Transcription of tones was auditory, and some decisions were marginal. In general, level tones were taken as 'open' when in high or high-medium range, and as 'closed' when

close to baseline or, for non-final levels, when medium-low or low in range. Below are the results for both types of list:

sequence of final two tones in list	lists with no 'new' item	lists with a 'new' item	lists with no 'new' item	lists with a 'new' item
	NSs (n=16)	NSs (n=15)	NNSs (n=50)	NNSs (n=46)
open / closed	5 (31%)	3 (20%)	33 (66%)	29 (63%)
closed / closed	7 (44%)	8 (53%)	10 (20%)	11 (24%)
closed / open	3 (20%)	4 (27%)	3 (6%)	1 (2%)
open / open	1 (6%)	-	4 (8%)	5 (11%)

TABLE 5.1 **Tone choices on last two items in lists**

In Table 5.1 it can be seen that in the lists with no new items, the NNSs rather than the NSs favour more heavily the configuration which corresponds to the pattern open / closed. The small NS sample makes it impossible to extrapolate too much from this finding; the NNS group, on the other hand, is large enough to allow us to say that, from the evidence here, the construction of normal listing intonation patterns here exhibits a high degree of nativelikeness. Perhaps we shouldn't find this too surprising; enumerative intonation comparable to English listing patterns is reported for many languages, not least for Spanish (see Chapter Seven, section 7.3.1).

When a new item is introduced as the penultimate item of an otherwise identical list, the overall use of a closed tone at this position by the NSs increases from 64% to 80%. However, the NNSs make no increase at all in their use of a closed tone (26%) here. This does appear to be a significant tendency. Whereas NSs' use of an open tone on 'new' information overrides normal tone selection at this position in the list, this doesn't happen in the NNSs' productions.

There are two possible explanations of this. Firstly, perhaps NNSs have not developed a target-like sensitivity for the need to mark new information prosodically. Secondly, NNSs may be unable to disturb normal patterns of production such as a listing tone sequence (a sequence which itself may have been transferred from L1). Cruttenden's (1981) suggestion that certain discoursal requirements 'override' others in tone selection seems to be true here for the NS but not the NNS data. Such an ability to override established patterns may be an aspect of L2 prosody which is resistant to acquisition.

Finally, a more traditional tone inventory was used to quantify all tone selection on *all* items in all the lists:

	fall ↘	rise ↗	fall-rise ∨	rise-fall ∧	level —
NS n=90	63.33	8.89	7.78	5.56	14.44
NNS n=398	28.14	32.16	1.51	0.0025	56.67

TABLE 5.2 Distribution (%) of tones throughout all the lists

The spread of tones indicates that NNSs consistently use rises and levels more than the NSs. Such differences in tone selection will be explored further in the next section.

5.2 The use of tone in multi-unit utterances

In the screen-question tests, questions 3.4, 3.5 and 3.6 all elicited utterances which were of a predictable form: two units each, both of which contained *two* informationally strong points (shape names in all cases):

The A is on the B / and the C is on the D

In each of the two units of an utterance, the first prominence was arguably a head rather than a nucleus. However, I take a looser interpretation here and interpret each strong point as a 'prominence', not least because typically here both units have two very clear informationally strong points. Also, there is a tendency for the first point in each unit (A and C in the above example) to carry a rather more marked movement than would be expected in a head, and neither does each unit show obvious signs of being internally divided. These issues of description are not strictly of interest here; I want only to identify any patterns within the pitch structure of utterances which are longer than a single unit. The tone used at each point of prominence was noted.

Several of these responses deviated from the predicted *sequence* in which the shapes were actually described.[1] What is under observation here is the way in which two-

[1] *The square's on the circle and the triangle's on the star* instead of the usual response *The triangle's on the star and the square's on the circle*, etc.

unit utterances are structured prosodically, and not the mapping of specific kinds of information. So, these few differently-sequenced responses were analysed along with the other data here.

5.2.1 Tone distribution

Measuring the distribution of different tones has been a common feature of the intonational analysis of spoken corpora. Several well-known studies, which all use a similar tone repertoire, are presented in Table 5.3. One distinction which they make is between tones with a complex kinetic movement (fall-rise, rise-fall) and compound or 'binuclear' tones, which have two distinct movements within one complex pitch form (fall + rise, rise + fall, see Crystal 1969: 217-220):

	fall ↘	rise ↗	fall-rise ↘↗	fall + rise ↘...↗	rise-fall ↗↘	rise + fall ↗...↘	level →
Crystal (1969)	51.2	20.8	8.5	7.7	5.2	1.7	4.9
Quirk et al.(1964)	52.5	24.7	6.9	9.3	3.9	0.6	-
Davy (1968) reading	50.2	24.6	11.1	5.5	2.1	0.6	5.5
Davy (1968) conversation	58.7	24.6	7.4	5.1	4.2	0.4	8.0

TABLE 5.3 Summary of % distribution of tones from several corpus studies[2]

My present data can be presented in a similar way, although in my analysis it was not necessary to distinguish between complex and compound tones, perhaps because the nature of the tasks made the use of complex tones unlikely. So, Table 5.4 represents tone production in three two-unit responses, classifying tones within a five-tone set:

[2] Compound tones with very small distribution are omitted.

	fall ↘	rise ↗	fall-rise ∨	rise-fall ∧	level —
NSs n=95	60.42	14.58	9.38	10.42	5.21
NNS (all) n=280	40.36	17.86	2.14	1.07	38.57
NNS 4 n=77	51.95	15.58	3.90	1.30	27.27
NNS 3 n=68	40.30	17.91	4.48	-	38.26
NNS 2 n=71	30.99	19.72	-	2.82	46.48
NNS 1 n=64	37.50	18.75	-	-	43.76

TABLE 5.4 **Tone repertoires for all groups**[3]

In the NS data here, there is a greater preference for falls than in the norms established in the studies cited in Table 5.3, and also a slight increase in the proportion of rise-falls as compared to the sum of both fall-rise and fall + rise in the cited studies. There is a corresponding lower proportion in the use of rise and also in the overall use of rise and rise-related tones. Levels are at similar percentage levels.

We might explain the greater use of fall and fall-related tones here in terms of the very factual, declarative nature of the utterances elicited in the tests, leading to a predominance of 'proclaiming', declarative falling tones. Because the context appears to have influenced the distribution of tones in the NS production here, I will compare the NNS data with the NS group data, since the context was the same in both cases.

Taking the NS data as an appropriate model of nativelike distribution, we see that, for the simple tones fall, rise and level, the NNS groups all differ from native percentages, but that in each case distribution progresses towards the NS norm as proficiency increases. Whereas the gradual reduction in the over-suppliance of rises is only slight, arriving at nativelike levels in group NNS4, the increase in the under-suppliance of falls is very significant, although still well below nativelike levels in group NNS4.

However, this highest-proficiency NNS subgroup does prefer falls at a percentage comparable with data cited from the corpus studies in Table 5.3. How should we interpret this? Perhaps these data reflect a general *tendency* towards target norms, which might not generally lead to full nativelike production. As I have previously noted, the study of L2 intonation, and of prosodic form generally, is more a matter of degree

[3] The compound tones fall + rise and rise + fall do not appear in my data.

than of right and wrong. Significant in these data, then, is that the nativelikeness in the tone repertoires of the NNSs does indeed graduate by degrees towards the target.

In Table 5.4, the most striking feature of the data is the over-suppliance by NNSs of level tones. There is a developmental reduction in these un-nativelike high percentages, which again attests to the fact that acquisition of intonation here does appear to encompass a continuing process of adjustment of tone distribution. However, the percentage of levels for group NNS4 is still far higher than in either the NS group, or in any of the corpus studies.

Why is this? We are familiar with the idea that level tones in English can be used as markers of non-commitment, or where the speaker seeks to convey a lack of personal involvement in the topic. Perhaps such uses of level tones may be a general feature of L2 production. Færch, Haastrup and Phillipson (1984) suggest that flattened and generally narrower pitch range in L2 speech might be a consequence of speaker-insecurity or tentativeness (p. 125). Also, we know that Spanish, despite being an intonation language, operates on a relatively narrow pitch band (see Chapter Seven), and this may contribute to negative transfer in our data. Both these questions can be tested to some extent: we can assess how widespread the over-use of levels is by looking at speakers of different L1s performing the same screen-question tests; we can assess the role of L1 transfer by looking at the tone repertoires of the L1s of our subjects. Both these questions will be addressed in Chapter Seven.

5.2.2 Open and closed tones

One of the ways in which these data can be interpreted is as reflecting binary speaker-choices between open/closed. In this way, a simple, percentage ratio can be stated. In the analysis here, this ratio is expressed as *open : closed.* The four tones (one on each of the two informationally strong elements in both units) for utterances 3.4-6 were summed and percentage ratios calculated, for the NS group and for both the NNS subgroups, and also for the NNS group as a whole.

In the present study, the *open : closed* binary distinction subsumes all rise-related and fall-related under the two categories open and closed. Once again, level tones were interpreted as either open or closed depending on their relative height in a speaker's range:

'open' tones	'closed' tones
rise	fall
fall-rise	rise-fall
(fall + rise)	(rise + fall)
low level	high level

The data from the large-scale corpus studies cited above can also be re-analysed to yield statistics which reflect this binary distinction. So, in addition to our NS data, we have other data on nativelike distribution:

	open	closed	remaining level-tones[4]
Crystal (1969)	37.0	58.1	4.9
Quirk et al.(1964)	40.9	57.0	-
Davy (1968) - reading	41.2	52.9	5.5
Davy (1968) - conversation	37.1	63.3	8.0

TABLE 5.5 Summary of several studies of tonal repertoires of native speakers of English, with data re-analysed to reflect choice of open and closed tones (% ratio)

In the present experiment, the distributions were as follows:

	open tones	closed tones
NS n=95	26	74
NNS n=280	51	49
NNS4 n=77	40	60
NNS3 n=68	51	49
NNS2 n=71	58	42
NNS1 n=64	56	44

TABLE 5.6 Tone choices as percentage of open and closed tones for the whole NS and NNS groups, and also for the four NNS subgroups

The NNSs, taken as one group, chose open and closed tones at a ratio of about 1:1. This represents a significant deviation from the target-like distribution of the NS group, and also from the re-analysed data of other studies, presented in Table 5.5.

The open tone category subsumes all those tones associated with rising pitch, and closed tones with falling pitch. Hence, an open/closed opposition captures something of the essence of the universal tonal meanings often attributed to intonation (see Bolinger 1978; Hewings 1993). There may be psycho-physiological justification for claiming that general 'meanings' emerge in natural languages through the use of distinctive open and

[4] Levels are omitted from this re-analysis, since no evidence of relative pitch height is available.

closed tone types. It is these very broad kinds of meaning which will prove useful in the interpretation of the current data.

It has been noted widely that certain kinds of open tone phenomena may be associated with speaker-anxiety. For example, Hirschberg and Ward (1995), summarising the literature on high-rising contours in English, report that writers have sought to explain its use as related to 'hesitancy and deference' (Lakoff 1975), 'tentativeness' and 'uncertainty' (Ladd 1980), and as a type of trying-out of the mutuality of information (Gussenhoven 1983, Sacks and Schegloff 1979; Clark 1992). Hirschberg and Ward themselves propose an account based on the speaker marking information as 'to be added to mutual beliefs' where, in addition, the speaker has some doubts as to whether this information conforms to (i.e. can simply be added to) the hearer's own (unshared) beliefs. Writers often go on to suggest that these kinds of general meanings of tones may reflect a universal aspect of natural languages.

If pitch movements which are characteristically 'open' have a broad sense-relationship to notions of hesitancy and uncertainty (amongst others), then we might expect to see a larger than target-like preference for open tones in nonnative speech. Anxiety as a factor in L2 speech is widely attested and, indeed, 'language-anxiety' has been seen as a specific, and very pervasive, phenomenon in of SLA (Horowitz, Horowitz and Cope 1986; MacIntyre and Gardner 1991). In the findings here, a drop in over-use of open tones occurs as proficiency increases. Since language-anxiety perhaps also declines as proficiency increases, these data might be interpreted as evidence of a clear developmental pattern in acquisition which is partly explicable in terms of the gradual reduction in language-anxiety.

However, psychological influences are only one possible explanation of nonnativelike production here. Of interest also is the speaker's ability to respond to the structure of the information being uttered, and to see whether native tendencies are reflected in the NNS data here. To explore this question in greater depth it is necessary to look at tone choices at each of the four separate tone positions within the utterances, and to typify nativelike production norms here relative to context.

5.2.3 Tone sequences across two-unit utterances

In our data, above, the NSs show a greater preference for closed tones than in the (re-analysed) corpus studies cited in Table 5.5. The NSs do, nonetheless, conform to the general tendency of using a significantly greater proportion of closed than open tones. On the other hand, the NNSs (taken as one group) show no such preference, with a significant over-preference for open tones, reflected in an open:closed ratio of 51:49.

These data are from the tones used within two-unit utterances. Before moving on to talk about the NNS data, then, I will briefly discuss NS preferences here. Table 5.7 gives distribution of tones at the four different tone-locations within the two-unit utterances, taking data from all three utterances together (responses to questions 3.4, 3.5 and 3.6):

	two-unit utterance			
example	*The A*	*is on the B*	*/ and the C*	*is on the D*
	unit 1		unit 2	
	tone 1	tone 2	tone 1	tone 2
open : closed tone choices NSs (n=24)	12 : 88	50 : 50	38 : 62	4 : 96

TABLE 5.7 **Average choices (% ratio) between open and closed tones throughout all two-unit utterances**

The overall average of tones use by the NSs was 26% open and 74% closed.

It has been noted earlier that native speakers of English are aware of tonal relations between two adjacent units. In the screen-question tests, the visual contexts of the utterances were fully available as processing began;[5] also, speakers were immediately aware of the nature of the language which was expected.[6] Thus, utterances here were produced under conditions which made forward-planning possible.

The overall average open:closed ratio for tone choices in these responses in the NS data is 26:74. Against this average, then, we can compare ratios at specific points throughout the utterance. The second tone (unit 2, tone 2) is almost always a fall, and is marked more consistently as a closed tone (96%) than any of the other three tone-positions. On the contrary, the second tone of the *first* unit is marked in fifty per cent of

[5] This differs from Swerts Geluykens's (1994) procedure, which uses moving shapes in an emerging context.

[6] At this stage in the elicitation, speakers had become familiarised with the expected linguistic pattern of responses.

cases with an open tone, the highest NS group average for an open tone. We might interpret both these tones as reflecting prosodic structure at the multi-unit level, in which relative shifts away from the prevailing average open:closed percentage distribution mark the two units of the utterance as cohesive and informationally bound. This cohesion is created in that the open tone which terminates the first unit requires resolution in a subsequent (closed) tone; open tones, it was noted earlier, carry a general sense of 'more to come', whereas subsequent closed tones are potentially markers of termination. Grosser (1993) in fact uses a similar measurement of the nativelikeness of his L2 English-speaking subjects, in which he sees a rise and fall sequence in the two units used to utter two syntactically co-ordinated clauses as the most cohesive prosodic pattern.

Against this level of structure, the first tone of each unit in the present data appears to cohere to tonal dependencies within its *own* unit. The tendency in the data is for first tones to switch polarity with the following tone in that unit. This corresponds roughly to Lepetit's (1989) 'melodic slope inversion', discussed in Chapter Two (section 2.4). So, tone 1, unit 1 is predominantly closed (88% closed tones against a prevailing average of 74%). This reflects a tendency to mark it as a point of informational prominence in its own right (i.e. separate, significant information), whilst also creating a tonal dependency with the following tone, indicating the connectedness of the two points of information within the unit. In unit two a similar tendency to switch tonal polarity occurs, with 38% instead of 26% open tones in the first tone, and 96% instead of 74% closed tones in the second. It is perhaps worth pointing out that the above comments on deviations from the prevailing speaker norms in these data can all be made by comparing the same deviations with the average of open:closed tones in the corpus studies mentioned above (39:58), with the exception of tone 2,1.

Tone 1, unit 2 is marked by the NSs with an open tone in half the cases. This is, as I have said, a significant deviation from normal tone distribution. However, even in those twelve NS utterances where a closed tone occurred at this point, other elements in the prosody tended to contribute to the sense of a dependency with the following terminal closed tone of the second unit. In six cases the subsequent closed tone in unit 2, tone 2 (the utterance terminal) was to a lower pitch than the closed tone in unit 1; in another two utterances the fall was to an equal pitch; indeed, the only case of a higher pitch to end the second rather than the first unit was in the one case of an open tone in utterance-terminal position. Also, only in two utterances did a closed tone in position 1,2 end on a *lower* pitch than a subsequent position 2,2 tone.

So, we have some evidence that, where tone choice does not conform to our expectations, NSs do, in the majority of cases, use other prosodic devices (here pitch) to mark some form of dependency between these two tone positions.

I will now discuss the NNS data for the same tests.

	utterance			
	unit 1		unit 2	
	unit 1, tone 1	unit 1, tone 2	unit 2, tone 1	unit 2, tone 2
NSs (n=24)	12 : 88	50 : 50	38 : 62	4 : 96
NNSs (all) (n=72)	59 : 41	73 : 27	52 : 48	19 : 81
NNS 4 (n=27)	42 : 58	55 : 45	40 : 60	22 : 78
NNS 3 (n=23)	59 : 41	77 : 23	47 : 53	24 : 76
NNS 2 (n=24)	72 : 28	72 : 28	56 : 44	24 : 76
NNS 1 (n=23)	62 : 38	88 : 12	69 : 31	6 : 94

TABLE 5.8 **A summary of the use of open and closed tones (% ratio) at each tone position in two-unit utterances**

The majority of the NNS data suggests (with the exception of group NNS 1, tone 1, unit 1) that there is significant development through the NNS subgroups 1-4 towards more nativelike tone choices. However, the choice of tone 2,2 differs from this trend in three out of four NNS groups by being relatively stable at a ratio of about 1:4 for open and closed tones. A closed tone in this position (utterance terminal) is the most highly predictable choice, not least because falling tone is a natural consequence of bringing normal speech to a close, and its presence as the preferred terminal tone to most NNSs' whole utterances is, perhaps, of little surprise. However, the persistent use of an open tone in a quarter of utterances (discounting groups NNS1) is perhaps further evidence of the general prevalence of open tones in NNS speech, noted above.

Some aspects of the NNS data are confusing. Against a consistently high preference for a closed tone at utterance-terminal position, the first tone in unit two sees in all groups a *greater* preference for open tones even than in the NS data, this preference lowering to levels similar to the NS ones in group NNS4. Yet the open tone is indeed appropriate here, and its over-use is hard to interpret. Perhaps the *reduction* in the degree of preference for it throughout the groups represents a reduction in the over-use of open tones as a general feature of NNS production, an over-use which I have already observed in other productions, above.

We also recall that, where NSs used a closed tone in the second tone position of *both* units, they often used pitch to mark the two unit-terminal closed tones differently, making the first of these as subordinate (generally less 'terminal') to the second. Looking at the NNS group as a whole, a comparable analysis reveals that of thirteen utterances with two closed tones in these positions, eight had lower F_0 minimum in the second (utterance-terminal) closed tone. Moreover, the distribution was evenly distributed through the groups. Comparing these to the NS data, there is no tendency within this very small number of responses to suggest any NS/NNS differences. On the other hand, the proportion of open tones at the end of unit one shows a developmental increase through the different ability sub-groups, and this may indicate that, given that NNSs most often use closed tones in utterance-terminal position (even NNS1 speakers), there is a gradual development of the sensitivity to creating contrasting terminals (in this case, pre-empting a terminal utterance closed tone with an open tone to end the first unit) which reaches nativelike levels in group NNS4 here.

In the analysis of the NS data, I suggested that the first tone in each unit was marked with reference not to the structure of the utterance as a whole, but to the *unit*-terminal tone which they precede, showing some tendency to switch polarity with the following tone in their unit. If this is true, any evidence for a development in NNSs' ability to reproduce such a tonal system can be seen more easily in the second unit, where a relatively stable 1:4 ratio of tone choices in the tone 2,2 can be seen as exerting an equal influence on the planning of tone-choice of the preceding tone throughout the data. The preference for open tones in first tone position here, which in NNS1 is 69:31, diminishes throughout the sub-groups, until in NNS4 nativelike levels are attained. Given the stable utterance-terminal tone here, this appears to be strong evidence for an emerging nativelike ability to plan an intonation unit in which pitch movement reflects two points of information in a unit-cohesive way.

Subjects here were required to plan rather short and linguistically straightforward utterances, and hence these data suggest that, in the utterances of NNS4 speakers, the linguistic planning necessary is sufficiently straightforward so that the ability to plan tone choices can be employed fully. So, although the production of these dual-prominence units may not in itself constitute a major element in intonational competence, the sensitivity to the mutual effect of sequences of tones within intonational structure does.

5.3 Summary

The findings on listing intonations suggest that:

- Spanish-speaking NNSs of English can use tones to construct well-formed English listing intonations in terms of the final two tones.
- NNSs seem not to be able to disturb the tonal sequence of a list in order to accommodate the marking of 'new' information within that listing pattern.
- NNSs use level and rising tones during the production of listing intonations far more than NSs do.

In multi-unit utterances:

- NNSs significantly over-use open tones and under-use closed tones, but these (non-target-like) uses reduce as proficiency increases.
- The prevalence of open tones may be a feature of nonnative speech in general.
- In the prosodic structure of multi-unit utterances, changes in tone choices by NNSs appear to reflect a developmental sequence towards more nativelike levels. This occurs in the ability to create tonal dependencies over sequences of two units, and also within units containing two tones. This is taken as a developing ability to use tones to reflect complex informational structure.

6.0 Introduction

In this chapter the text readings (see Appendix B) will be discussed. The textual properties of this text, and the macro-prosodic structure which it was intended to elicit during readings, were described in Chapter Three (sections 3.2.2 and 3.3.3).

6.1 Temporal organisation of readings

Reading speed was lower in the NNSs than the NSs. Also, even though the text was heavily punctuated, hence giving strong hints as to where intonation unit boundaries might begin and end, the NNSs used on average more units than the NSs; consequently, their units were, on average, shorter. NS/NNS differences were small and did not show any clear development over the four subgroups.

	total reading time (seconds)	number of intonation units	words per unit
NSs	40	19.7	6.3
NNS4	45	22.1	5.7
NNS3	52	20.7	6.1
NNS2	45	22.7	5.7
NNS1	55	22.8	5.5

TABLE 6.1 **Group averages for reading time and for total number of intonation units used**

Both longer speaking times and shorter intonation units have been noted before in nonnative speech (Jenner 1984; Hewings 1993), and neither of these findings is, perhaps, of great significance here; the NNS subgroups are small, and pause/hesitation is an intrusive factor in these data. However, I did note, in section 4.3, that shorter intonation units were also observed in the spontaneous speech of NNSs.

6.2 Pauses in the text reading

In English speech, pauses often vary significantly in length, with different-length pauses distinguishing intonational units of different status, functioning somewhat like breaks in written texts - commas, semicolons, full-stops, paragraph breaks etc. - to delineate levels of information structure and, hence, to facilitate clearer and more comprehensible speech.

There were eighteen points in the text where every speaker, including NNSs, paused. As noted above, both NSs and NNSs used slightly more than eighteen pauses in total. However, the eighteen points of pause common to all speakers were those which were in positions most obviously related to the structure of the text.

Taking the summed total of time spent pausing by each speaker in the production of these eighteen pauses, and expressing them as percentages of total speaking time, we see that group averages here are significantly different:

NS	64
NNS4	48
NUNS	42
NNS2	38
NNS1	44

TABLE 6.2 **Group averages for the percentage of total speaking-time occupied by the 18 pauses in the reading**

Pause, as I have previously noted, is one of the most reliable indicators of unit divisions. Indeed, Brown *et al.* (1980) find it to be 'the one reliable signal' (p. 47) in spontaneous speech, and others have also taken pause as a major device for creating unit divisions (Schuetze-Coburn, Shapley and Weber 1991). The position and length of pauses 'convey significant information about the speaker's discourse production process' (Du Bois *et al.* 1993: 61).

However, pause production has generally not been considered to represent a productive element in L2 speech, either in teaching or in language-assessment. Indeed, whereas I have already noted the generally sparse reference to prosodic parameters of speech in the guidelines of several major EFL schemes of assessment, reference to appropriate use of pause is entirely absent. This may be due to the particular properties of pause - its non-articulated nature - and the belief that pause is simply a naturally occurring facet of the speech process. Nevertheless, pause can reflect the structure of a

spoken text in systematic ways, and as such the non-appropriate use of pause may indeed contribute to the perception of nonnativelike production, even if this perception operates in a wholly intuitive way. Paradoxically, it may be the very inevitability of pause in human speech which allows it to be overlooked as a possible source of nonnativelike production in L2s.

Since pause is a marker in the prosodic organisation of speech, its nonnativelike use may contribute negatively to the perception of overall nativelikeness in an L2, through creating language which is not optimally comprehensible, or which lacks perceived cohesion. Also, pause is a curious prosodic phenomenon from the point of view of articulation; there can be no qualitative shortfall in its physical production, so we might say that it is the single component of speech which NNSs are (potentially) equipped to produce accurately in the L2.

If we assume that in some very broad sense all natural languages use pause as a marker of information divisions, then we might also assume that its correct acquisition in an L2 implies the fine-tuning of the system towards a nativelike sensitivity to particular kinds of information structure. However, making assumptions about interlingual development here is not particularly useful; we have data from only a single text reading, and no corresponding data on the L1 production of our NNSs. My observations here will therefore be suggestive only.

A group analysis of pause, even when the data are very similar across subjects, is not straightforward. Systematic use of different length pauses will be systematic only in relation to other pauses *for that speaker*, a point discussed by Edwards (1993) in a wider discussion of transcription conventions. The analysis of pause can be carried out using an approach based on broad perceived categories, and this is common in the literature. Du Bois *et al.*'s (1993) transcription conventions, for example, adopt a three-way categorisation of long, medium and short pauses, with typical durations of >70cs, 30-60 cs and <20 cs respectively (pp. 45-89). However, such categories are by their nature approximations. In the present study we want particularly to see whether pause is used differentially to help create paratonic structure in a text reading, and therefore we need to be able to quantify accurately the production of pause in some way.

In order to consider use of pause within the text readings, pause length was therefore compared to average word length in the speech of each speaker, and the length

of each pause was expressed using average word length as a measure.[1] Syllable length is perhaps a more conventional measure in this kind of analysis, but here there were problems with the NNS data; whereas total number of words was straightforward to establish, including repetitions and mis-readings, deviant segmental production, such as vowel epenthesis typical of Spanish-English ILs, made syllable counts unreliable.[2] The extra time needed to produce epenthetical and other deviant productions may have contributed to the overall longer reading times in the NNS data.

Average lengths of pause, as speaker-normalised measurements, were then averaged for NS and (all) NNS groups. Thus, the average length of pause used by native and nonnative speakers at each of the principal (eighteen) pause positions in the text were expressed.

This method of normalising pause readings is in some ways potentially very misleading. The percentage of a reading which is pause may well be affected by how long a reading has taken. In the NS group data here, there was only a low negative correlation (-0.20794) between total length of reading and percentage of reading time which was pause. On the other hand, this correlation was -0.51844 in the case of the whole NNS group. In other words, those NNSs whose text readings took longer also tended to spend proportionally *less* time pausing in that reading. This correlation has an effect on the calculation of normalised pause times. However, the alternative is to use bare readings of pause lengths. Given that these total reading times varied throughout the entire informant group by as much as one hundred percent, it would seem that bare readings constitute a far greater source of skewing.

The positions of the eighteen pauses are marked in the text in Figure 6.1, below. Average pause durations of the NSs are also shown, expressed as group averages of the normalised pause durations, described above:

---

[1] For example, a pause of 78 cs for a speaker whose reading is 123 words long and lasts 45 seconds is: 0.78/ (45/123). In this way, data on pause lengths becomes comparable across informant groups, being a kind of normalised measurement.

[2] For example, in the following excerpt from a NNS reading, within the first fourteen words of the text (two sentences) the speaker produced two extra syllables through vowel prothesis and one through vowel anaptyxis:

prothesis:	<study> - [estv:di]
	<students> - [estuden]
anaptyxis:	<programmes> - [programes]

These days <0.2> many students come to the United Kingdom to study, <0.81> often on exchange programmes. <2.83> If you want to study in the UK, <0.42> there are nearly one hundred universities. <2.16> And many towns and cities have two or more universities to choose from. <2.99>
Many institutions have a modular system, <0.93> in which you can study separate units <0.52> (called modules) <0.81> from different departments. <3.03> A modular system allows you to choose the things you want to study, <0.71> depending on your own particular interests. <3.07>
If you want to study in Hull, <0.98> there are plenty of modules to choose from. <2.63> Modern languages are a popular choice; <1.29> there are modules in French, <0.33> German, <0.19> and many other languages. <1.93> So, you should never get bored if you do come here to study.

FIGURE 6.1 Test reading with average NS pause lengths, expressed as a group average of each speaker's pause length divided by his/her average word length

In the NS readings, two of the three longest pauses occur at the end of the first and second (orthographic) paragraphs. The third of these long pauses follows *departments*; this may be explicable in terms of the very long sentence which precedes it, or as a result of speakers creating a distinction between information in this sentence, and the clarification and further explanation of that information in the following sentence.[3]

However, pause is in some ways the most difficult aspect of intonational form to make any direct form-to-function claims about; it is related to various aspects of processing as well as to intonational form, and specific explanations of longer or shorter pauses are unlikely to take account of all speaking conditions. Rather than make speculative assessments of nativelike production in this respect, we might rather consider these data in terms of the overall 'repertoire' of pauses which native speakers employ.

So, how does the production of pause differ in the NNS readings? A general tendency so far noted in this study is that clear, meaning-related variation in prosodic production is used less often, and to a lesser extent (less markedly), in nonnative than in native English speech. Is this true of pause here? If so, does nonnativelike speech contain shorter pauses or, on the contrary, do increased processing constraints result in longer pauses?

Figure 6.2 presents averages of normalised pause durations of the NS and the whole NNS group for the eighteen pauses. These averages are ranked for length, and do not relate sequentially to the structure of the text:

[3] Neither is this long average pause the result of the skewing effects of outlying data: readings for this pause in the NS group ranged from 2.44 to 4.18.

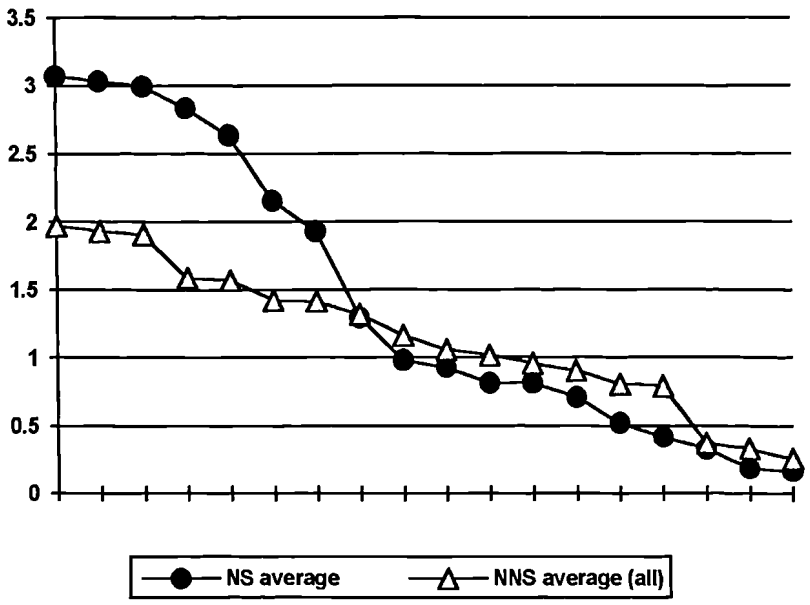


FIGURE 6.2 **Ranked pause lengths in NS and NNS, expressed as a group average of each speaker's normalised pause length**

The NNS average pause lengths do indeed show systematic differences from NS averages. As Figure 6.2 shows, the pause repertoire of the NNS group is significantly narrower in average length of pauses than in the case of the NS group. This relative bunching of pause lengths around a median in the NNS productions suggests a tendency to expand and contract the length of pauses to a far lesser degree than in the case of native English speakers in the same reading. Particularly striking is the fact that the NSs' relatively longer pauses are very much longer than the NNSs' longest pauses.

So, it appears that pausing, considered as a prosodic device, is exploited less by the nonnative speakers here. This narrower spread of pause lengths in the NNS group may also reflect a deficit in the systematic use of different length pauses as a means of marking discourse (paratonic) structure.

The consequences of this may be important in terms of our understanding of nonnative speech. If, as seems reasonable, one of the ways in which we instinctively respond to nonnative speech is in terms of its general comprehensibility, then pause may play a negative role in such responses. In the above graph it is not the length but the relative lack of variability in the length of NNSs' pauses which is crucial. These results are particularly significant in that the reading text here was designed to elicit a very clearly differentiated series of units of information.

6.3 Intonation units and pitch-(re)setting at onset

The role of pitch within macro-prosodic structure has been discussed in Chapter One (section 1.3). The text readings in the present study were analysed mechanically, and an F_0 reading taken on the central section of the onset syllable of each intonation unit.[4] Three onset positions were discounted because only a small number of subjects made unit divisions at these points. Nineteen common onset points were therefore analysed (corresponding to the nineteen units separated by the eighteen pauses discussed above).

The raw F_0 readings were converted into logarithmic expressions, and conversion to a semitone scale was performed.[5] Normalised (z) scores were then calculated for the semitone scores, so a group comparison could be made.[6] The resulting data expresses a speaker's F_0 readings in terms of a number of standard deviations from a mean of zero. Representation of pitch levels at all onsets expressed in this way illustrates the use of relative pitch height during pitch-setting. Since the scores are normalised, tendencies throughout whole speaker groups can be expressed as average z-scores.

6.3.1 NS readings

Firstly, I will discuss some of the interesting features of the native speakers' readings. Averages for the NSs' production of onsets are shown in Figure 6.3 as a progression of levels throughout the text:

[4] Measurements were generally taken without problem from within the stable middle part of a vowel. However, where clear movement was perceived auditorily, readings were taken from near the beginning of the movement on the vocalic section of the syllable, given that where an onset (for whatever reason) exhibits pitch movement, the pitch setting itself is established at the commencement of that movement, and not further along it.

[5] The formula was: $x=12.\log_2(f_1/f_2)$, where f_1 is the baseline for a speaker, and all subsequent (higher) readings $f_{2,\ n}$ are compared to that, so that a speaker's lowest F_0 becomes a zero semitone reading and others are expressed in terms of semitones above that zero level. This formula can be found in t'Hart, Collier and Cohen 1990, p.24.

[6] The formula was: = (m-s)/sd, where *m* is the mean of the semitone readings for a speaker, *s* is the reading to be normalised, and *sd* is the standard deviation for all readings for that speaker. This procedure is described in Schuetze-Coburn, Shapley and Weber 1991, p. 213.

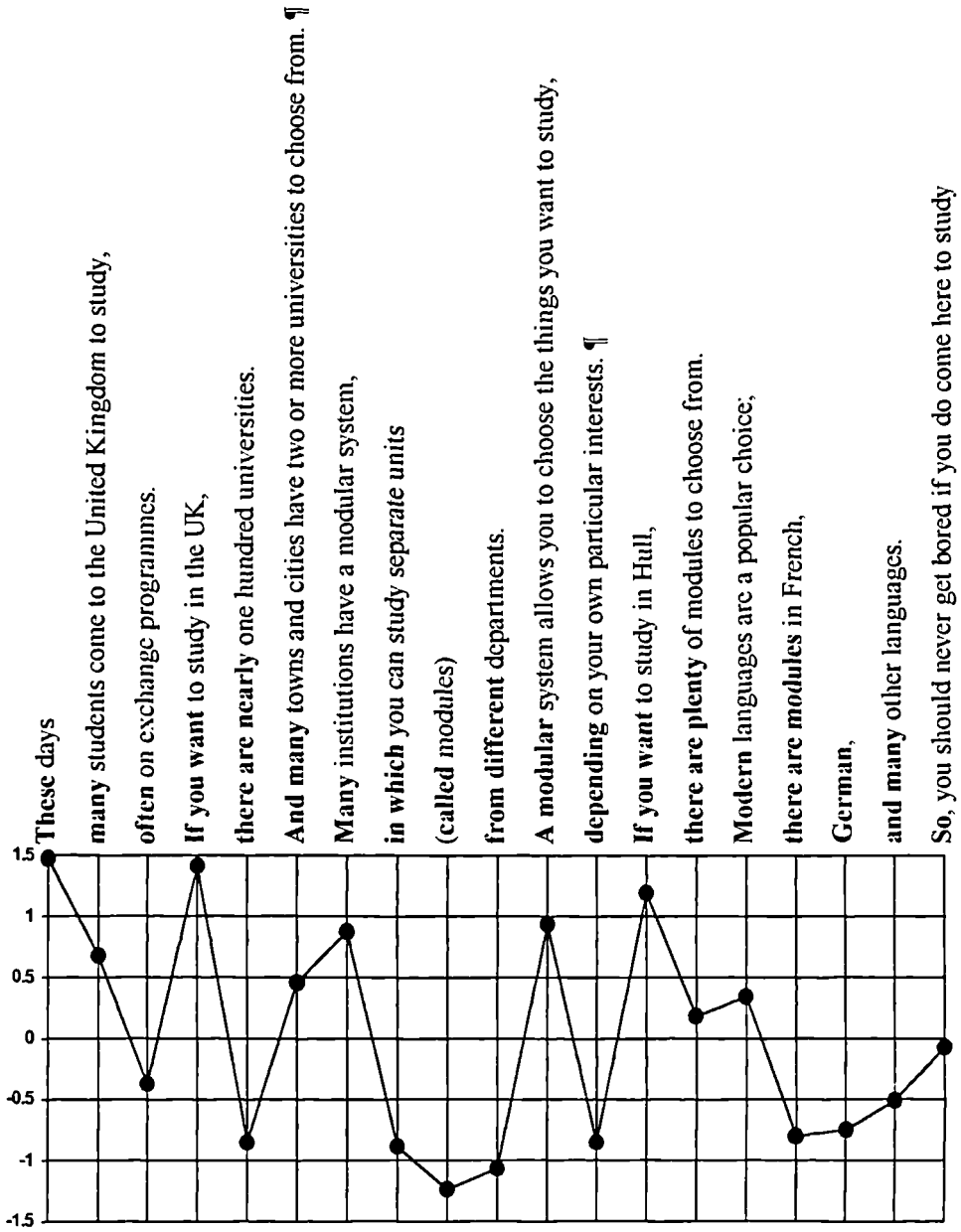


FIGURE 6.3 **NSs' onset levels of the text reading, as an average of all NSs' normalised semitone scores**

This, it must be remembered, is a line of *onsets* and not of nuclear prominences, which may indeed be significantly higher or lower than the onset within a unit. Figure 6.3 is intended to illustrate the progression of pitch-resettings which occurs throughout the reading. I will now discuss some of the features of these NS data.

There appear to be five jumps to a relatively high onset. All three orthographic paragraphs are so marked. However, there is an additional high onset within both the first and second of these 'paragraphs'. In all five cases of high onset, there follows one or more units with very markedly lower onsets. This strongly suggests that in this reading, native speakers instinctively recognised five points of informational transition, in which

particularly high onsets are used. Indeed, the presence of these five high onsets (as the highest five in the reading) is common to all NSs.

In Figure 6.3, the opening intonation unit of the reading has an extra-high onset, which conforms to models of paratonic structure described earlier. The next two intonation units have progressively lower onsets, and are both preceded by very short pauses (averages of 0.2 and 0.81, expressed in group-average word length); in terms of content they both contain text which forms part of the opening theme of the text, which in the literature has been referred to as the 'speaker topic'. More obviously, the three opening units correspond to the three opening orthographic sentences.Both the second and third unit, then, exhibit clear prosodic indicators of their dependent status relative to the opening unit, and can be seen as a simple example of how macro-prosodic structure of this kind evolves in production. The pattern of descending onset levels over the first three units of the reading can be seen equally clearly in the bare F_0 scores of the NS group (four men and four women):

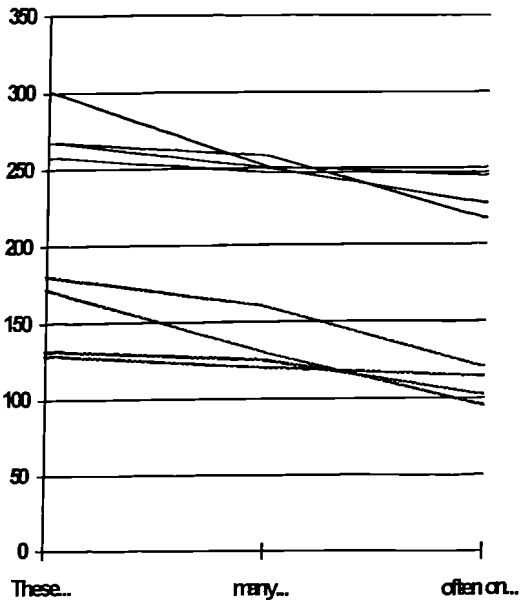


FIGURE 6.4 **Bare F_0 readings of NSs' first three onsets**

Taking the above NS averages as a whole, it might be argued that high onsets in this reading simply co-occur with new orthographic sentences. This is not the case. In the first orthographic paragraph there are three sentences, yet relatively high onsets fall at the beginning of only two of these, suggesting that indeed something like Couper-Khulen's information-based model of pitch re-setting may be in operation; the third

sentence, - *And many towns and cities have two or more universities to choose from.* - is perhaps more subsidiary in content than its predecessors, and has an onset nearer to mean level than the previous two. In other words, whereas *If you want...* constitutes a definite development of the topic, and as such is marked with a high onset, *And many towns...* is a subsidiary development, and receives only moderate onset.

A strict set of categories for jumps to high, medium and low onset, reflecting new, additional and subordinate developments of the active topic, may be too restrictive. Pitch is, above all, a relative system, and a certain flexibility in the use of categories of low, medium and high onsets needs to be employed here.

In the marking of parenthesis in this reading, a far lower onset than any other in the reading is jumped (down) to by all NSs. The unit which contains parenthetical information here (*called modules*) has the lowest average onset, and again this conforms to the general notion, reported widely in the literature, that subordinate information is often marked prosodically by low pitch in English:

> Parenthetic utterances, for example, seem to be characterised by pitch drop and diminished loudness more frequently than by anything else.
>
> (Crystal 1969: 174)

'Parenthesis' in intonational studies in fact usually refers to prosodic rather than orthographic parenthesis, 'where various grammatical structures which are parenthetical in nature' (see Cruttenden 1986: 78) are observed to be marked prosodically in characteristic ways.

I will therefore take the example of an orthographic parenthesis in the text reading as the clearest opportunity for the realisation of a *phonologically* parenthetical unit. In Figure 6.3 the phonologically parenthetical material in the NS readings in fact begins in the prior unit in the text. Here, the (NS) speakers appear to mark this prior unit (*in which you can...*) as carrying relatively unimportant or subordinate information in the reading, and the onset which marks the following (orthographically parenthetical) information unit (*called modules*) then drops even lower than this, to the lowest onset level in the reading. Moreover, the following unit also has a very low onset, so that these three units together have the three lowest onset levels in the averages for the NS readings. This seems to be a clear example of 'phonological parenthesis'.

In terms of the content of the text at this point, these three units appear to correspond to material which forms an extended explanation or definition of the

previously introduced concept 'a modular system', which Couper-Kuhlen's model would interpret as being particularly clear cases of subordinate information. Such specific use of pitch at onset here constitutes a useful example of how native speakers as a group seem to mark certain aspects of the information status of a text in very similar ways. Below I will compare the readings of the NNSs at this point in the text.

Wennerstrom (1994), in a similar analysis of a text reading, compared the F_0 on particular lexical items to illustrate their status within the discoursal structure of the text (see Chapter Two, section 2.4.5). She found significantly higher pitch and amplitude on words which were newly introduced into the discourse than when they were repeated. Similarly, in the present data, the first three utterances of the word *many* in the text should be distinguished in terms of pitch (all are onsets in our data): the first *many*, which is neither sentence-initial, nor initial within its own group of units, should have lower average onset than the second *many*, which falls on the first unit reflecting an orthographic sentence but a sentence which adds subsidiary information to the speaker topic. The third *many* opens new development of the topic (and here a new orthographic paragraph) and should be higher than either of the others:

> These days many students come to the United Kingdom to study,
> often on exchange programmes. If you want to study in the UK,
> there are nearly one hundred universities. And many towns and
> cities have two or more universities to choose from.
> Many institutions have a modular system,

FIGURE 6.5 **The first orthographic paragraph of the reading, showing its three occurrences of the word *many***

It has been noted above (Figure 6.3) that these predictions are borne out in the case of the second and third examples of *many*, but the first has a higher than predicted onset, perhaps because the establishment of the topic of the reading falls not only within the first unit of the reading (*These days,...*) but in fact continues into this second unit beginning, *many students....* We know, for example, that high onset is often maintained throughout the whole duration of the utterance of a topic when introduced for the first time. This seems to be the case here, where the first unit is an (initial) adverbial element, with the 'topic expression' proper following in the second unit.

6.3.2 NNS readings

I will now assess how the NNSs' readings compare to those of the NSs, firstly in terms of the general parameters of production, and then, in the following section, looking at some specific aspects of the text reading.

Figure 6.6 shows NS and NNS group averages for normalised onset levels throughout the reading:

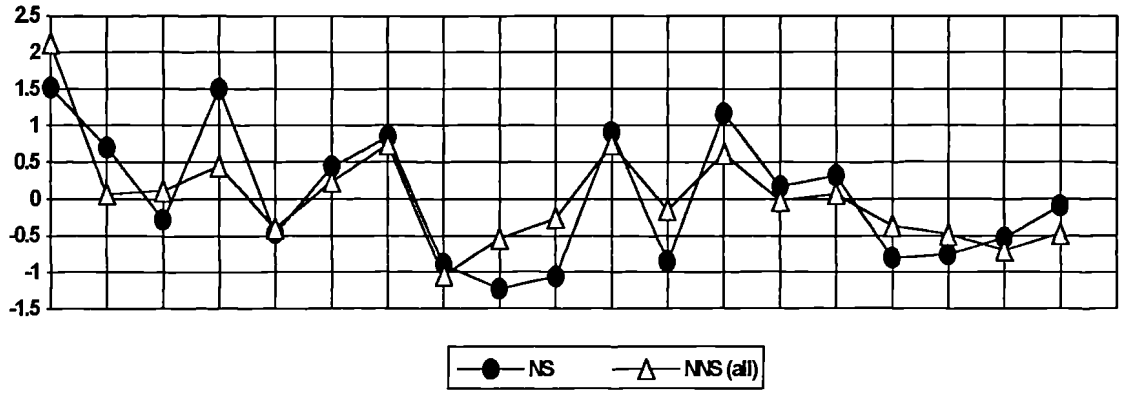


FIGURE 6.6 **Sequence of group averaged z-scores**

NNSs appear to follow the same general pattern of pitch re-setting as the NSs: an extra-high initial onset, some jump to relatively high onsets to mark points of informational transition as in the NS data, a series of low onsets in the final units as the reading comes to a close.

Average ranges of F_0 in the production of onsets by NS and NNS are significantly different:

	female	male
NS	80 (n4)	59.5 (n4)
NNS	57.9 (n15)	43.5 (n10)

TABLE 6.3 **Average F_0 ranges in the production of onsets**

I have already mentioned the claim that Spanish speakers, in their own language, use a relatively narrow pitch range. This is here not a simple physical inability, as shown by the average height of the opening onset of the reading produced by the NNSs, which

exceeds that of the NSs. If Spanish speakers do naturally use a less varied speaking range, we might ask whether this range expands as acquisition of an L2 with a typically broader pitch range progresses. Pitch range of onsets for each sub-group was expressed as average normalised pitch:

	average
NS	7.14
NNS4	5.64
NNS3	4.91
NNS2	5.64
NNS1	4.45

TABLE 6.4 **Subgroup averages of pitch range used in text reading, expressed as normalised semitones**

Quite clearly, NNSs do not significantly broaden the range at which onsets are produced as proficiency increases. The number of subjects here makes such an analysis merely suggestive. However, were these data to be reflected throughout the general population of Spanish speaking learners of L2 English, such a characteristic of production would constitute a significant area of fossilisation.

In the text reading data, the nineteen average onset levels of the NNSs bore a remarkable overall similarity to the NS group averages in terms of the pattern of pitch heights throughout the reading. However, the narrower overall pitch range in the NNS data is also clear; that is, the NNS onsets usually reflect the native-like model in terms of relative level, but within a narrower pitch range.

In Figure 6.7 the averaged scores for each of the nineteen onsets by NS and NNS groups were then ranked for pitch height, as in the ranking of pause averages:

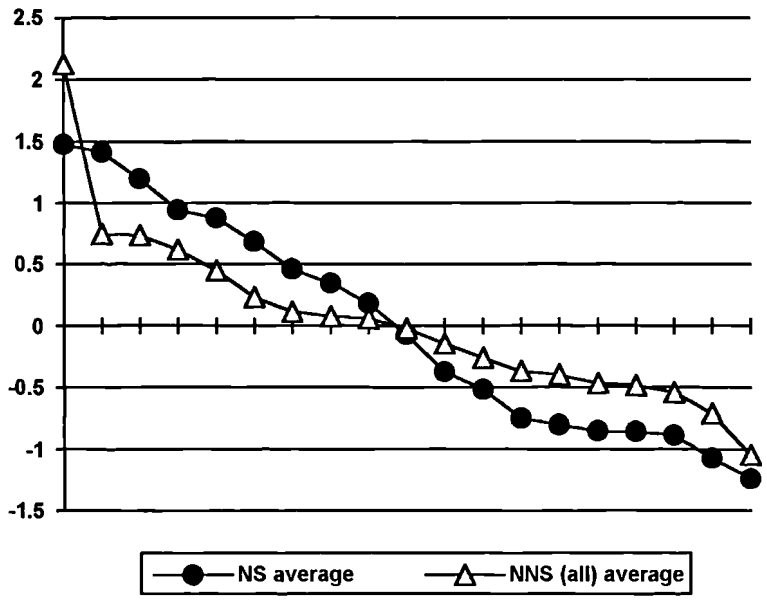


FIGURE 6.7 **The nineteen average onset levels for NS and NNSs, ranked for level**

In Figure 6.7 it can be seen that again production is narrower, in this case in terms of the breadth of the 'repertoire' of pitch onsets used. Again, NNS averages bunch more closely around the median, with the exception of one very high onset, which was in fact the first onset of the reading.

Less clearly differentiated pitch re-setting at onset may lead to less clearly comprehensible speech. Even if we were to assume that a differential system of levels is substantially in place in these NNS data, so that different information status is in fact marked with appropriate uses of different onset levels, it is nevertheless probable that the overall narrower pitch used by the NNSs will lead to less overtly marked distinctions, and, hence, to the possibility that distinctions may not be perceived in speech.

6.3.3 Details of NS and NNS production

Average onset pitch in the first orthographic paragraph (including the first onset of the next paragraph) was calculated for NNS sub-groups 1 + 2 and groups 3 + 4, yielding two larger NNS groups, representing the lower and higher half of the data in terms of proficiency. In this way, characteristic NNS production could be assessed for any development in acquisition:

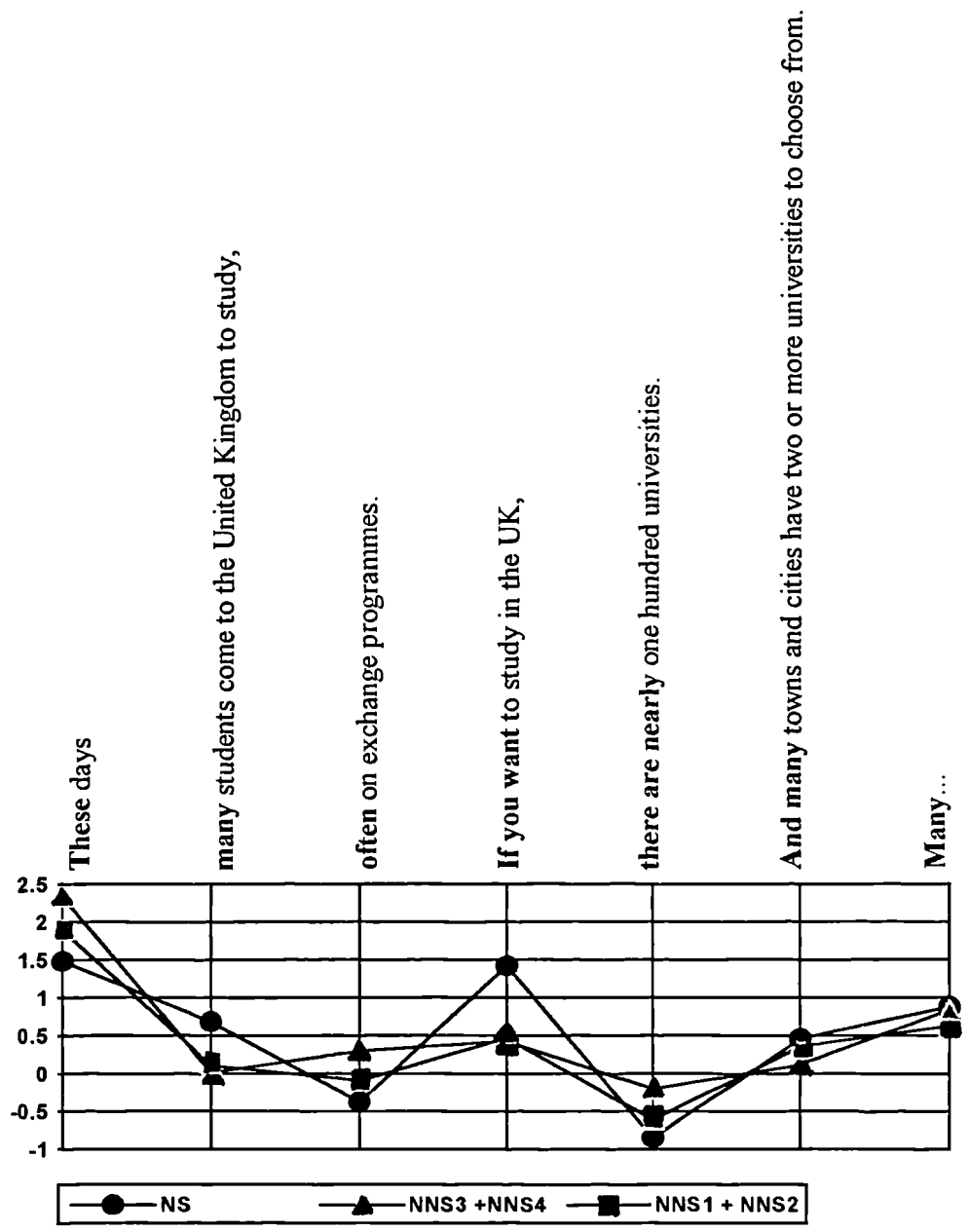


FIGURE 6.8 **Average onset levels of NSs, and two (larger) NNS groups in the opening paragraph of reading, including the first onset of the second paragraph**

Averages from both of the two larger NNS groups fail to show any maintenance of a reasonably high onset from the first onto the *second* unit, although we see here that the NSs do this. Earlier I suggested that this was probably because the initial exposition of the discourse topic is being established here over the first two units. Both NNS groups here show an average zero z-score on the second unit onset.

Both NNS groups also fail to create a second high onset on *If you...*; on average, then, NNSs simply fail to mark this with a sufficiently high onset. I earlier noted that it

was one of the NNSs' highest five onsets, but on closer inspection we see that in fact it is only marginally higher than the preceding onset.

Returning to onsets on the three occurrences of the word *many*, the NNSs produce onsets on the second and third occurrences of this word which conform closely to those of the NSs. In particular, we see the jump up to a medium pitch on the word *many* which introduces the final unit of the first paragraph (*And many towns and cities...*), which here includes an element of 'new' information, being more like a substantive development of the topic than merely a subordinate extension of it. The final utterance of *Many* in Figure 6.8 represents the first onset of the new orthographic paragraph, and averages for both NS and NNS groups show a jump to a relatively higher pitch.

In the discussion of parenthetical utterances above it was noted that orthographic parenthesis is an overt kind of subordinating marker, and that written parenthesis, when read out, will indeed share those prosodic properties which have usually been associated with parenthetical intonation generally. Against the NS averages for onsets within the example of phonological parenthesis which I noted above, we can compare group averages for the NNSs:

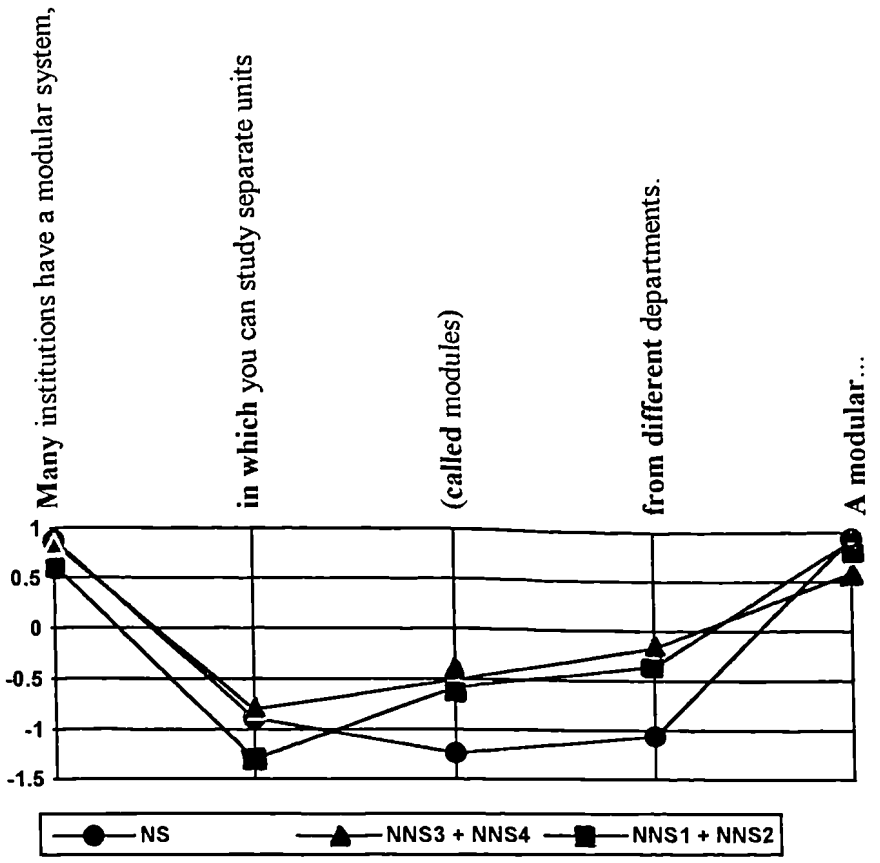


FIGURE 6.9 **Onsets during (phonological) parenthesis**

The section of the reading covered in the above chart, is:

Many institutions have a modular system in which you can study separate units (called modules) from different departments. A modular system...

The NS data provide a clear example of a parenthetical intonation, which is marked here with a characteristic drop in pitch not only on the bracketed material from the text, but also on the whole section *in which you can study separate units (called modules) from different departments.* The average onsets on these three units are the lowest in the NS data, with the lowest of these marking the bracketed material. The information conveyed here seems to be a kind of extended aside or a section of very clearly subordinate material. As I noted earlier, it is in this relatively loose, context-based sense that the nature of phonological parenthesis can be understood. The unit *called modules* is the most clearly parenthetical in this sense, being the most subordinate kind of information, a sort of added-on idea within the longer aside.

The NNS groups do mark this series with a lower onset on *in which...*, which I take as evidence that the subordinate complexion of this part of the text, in contrast to

the immediately prior unit, is recognised and marked. However, this fall in onset level is followed by two onsets which are both higher than in the NS data, and are indeed not the lowest onsets in the NNS data but rather climb back up towards average levels. Also, the NNSs' lowest onset does not fall on the unit containing the bracketed material, which, as we have said, is *the most subordinate* within a series of units marked as subordinate by the NSs. NNS productions here suggest that pitch re-setting to a very low level over several units is avoided, and that a drop to one low onset is followed by a return to more normal levels, in contrast to the NSs' maintenance of low onsets in the same context. NNSs thus appear to avoid the prolonged used of extreme variation in pitch, perhaps indicative of a general reticence to exploit the full breadth of the target system, or of an inability to do so in certain conditions.

Also, taking averaged scores for the new larger NNS subgroups, we see that there appears to be no development at all towards nativelike levels of pitch-setting; group 3 + 4 averages are often more distant from the NS score than the group 1 + 2 average. This appears to be very strong evidence that there is no development in this aspect of the production of differentiated onset levels as discourse markers here.

6.4 Summary

In the text reading:

- reading speed in NNSs increased with ability
- NNSs use on average more units than the NSs

In terms of pause production:

- NNSs use on average shorter pauses as expressed in terms of speaker's speech rate
- NNSs use a narrower range of pause durations

Analysis of pitch re-setting reveals that:

- NNSs follow the general pattern of nativelike onset levels, including typical features of paratone marking such as extra-high initial onset, jump to relatively high onsets to mark points of informational transition, low onsets in the final units as the reading comes to a close
- NNSs' range of onset levels is on average narrower than those of the NSs
- this narrower range does not increase with ability
- those specific uses of onset levels observed in the NS data to mark different status information are not always seen in the NNS data
- there seems to be no evidence that these deficits reduce as ability increases.

Chapter 7 - The role of L1 transfer in L2 intonation

7.0 Introduction

So far in this study I have assessed the production of prominence, tone and paratonic structure by a group of nonnative speakers. There has been little reference to their L1; instead, the basis of my approach in Chapters Four, Five and Six has been of a comparison between the prosodic production of the NNS group and comparable data from native speakers of English.

However, in order to interpret more satisfactorily these findings we also need to know which phenomena identified are plausibly explicable in terms of L1 transfer. The use of one large, single-L1 informant group here makes this possible. Such an interpretation also indicates which aspects of L2 production appear *not* to be L1-related, and this may tell us something about more general patterns in the prosodic production of L2s.

In the present chapter, then, I will discuss the main findings reported above in relation to Spanish phonology. I will also consider briefly data which were collected from two smaller informant groups of Italian and Chinese (Mandarin) native speakers. There were a total of eight speakers of Italian and five of Chinese.

7.1 Prominence

7.1.1 Prominence in Spanish and English

Contreras (1980) argues that accentuation patterns in Spanish are most convincingly explained in terms of an informational rather than a syntactic account:

> If we assume the following decreasing scale of "prominence",
> Initial element within main stress (foregrounded rheme)
> Final element with main stress (normal rheme)
> ...
>
> (p. 50)

Bolinger also claims that English and Spanish share 'the same basic accentual and intonational shapes' (1989: 48).

However, it has been noted that Spanish relies on a variety of different linguistic devices to create kinds of prominence which might typically be marked prosodically in English (Kelm 1987; Cid-Uribe 1989). For example, changes in word order are commonly used to create narrow focus:

La casa es bonita → ¡Es *bonita* la casa!

Alternately, an extra word provides more emphasis:

La casa es bonita → ¡La casa sí es bonita!

Thirdly, prosodic marking can be used.

As a consequence of this relative profusion of devices, speakers of Spanish in our data may be less disposed to mark prominence prosodically in their L2 English, simply because they do it less consistently in their L1. However, in order to assert this, we also need to know whether prosodic marking is indeed qualitatively comparable in Spanish and English.

'Acento' in Spanish is an under-researched area. Accounts sometimes blur the distinction between word and sentence accent, and, in terms of intonational form, prominence/accent has rarely been described in any great phonetic detail.

The prosodic marking of Spanish *acento* has often been seen as realised through increases in amplitude; terms such as *acento de energia* and *acento de intensidad* are common (see Quilis 1981: 312). Navarro-Tomás (1961) asserts that 'the Spanish ear is more sensitive to modification of accent through intensity than through other phonetic elements' (p. 183).[1] Fernandez-Ramirez (1951), though, claims that both intensity and pitch contribute, with neither predominating. However, the Spanish Royal Academy in its *Gramática* of 1962 sees pitch as the principle element in the accentual system, although adding that intensity frequently co-occurs on accented syllables (Academia Real Española 1962: 459-60). Contreras' (1963) acoustic experiments also indicate that pitch is the primary marker of accent, perceived either in terms of its obtrusion or its kinetic movement, and that both intensity and duration may reinforce this perception, and indeed may mark accent where pitch marking is absent. Interestingly, Contreras' work also provides evidence that duration, and not intensity, is the second most common accent marker in Spanish. This finding was subsequently replicated in the experiments of Quilis

[1]All quotations from Spanish texts are my own translations.

(1981), for whom '...intensity performs a practically negligible role in the function of Spanish accent' (p.330).

Canellada and Madsen (1987) refer to the *núcleo informativo* (p. 132), but their extensive treatment of Spanish intonation is concerned largely with the structure of pitch contours within the unit as a whole, and they say little about the nucleus itself, despite the instrumental nature of their account.[2] Canellada and Madsen observe that all three prosodic features may maximise on the nucleus, although they argue that the tonal line has its own structure and that, given the need to emphasise a particular syllable within this line, accent and pitch maxima may not coincide. They also observe that intensity may in fact reduce on accented syllables, especially where a pitch rise occurs after the lengthened, accented syllable, in these cases both pitch and intensity coinciding on a kinetic movement outside the immediate scope of perceived accent.

Canellada and Madsen also discuss Spanish narrow focalisation, which conforms in many respects to the phenomenon of contrastivity in English (p. 89). Crucially, they note that extra intensity is the major marker of narrow focus. Other accounts of Spanish concur on this point.

7.1.2 Spanish-English transfer in the prosodic marking of prominence

In the current findings, it was seen that the overall prosodic weight of 'normal' prominences in the NNS data was not greatly different from that found in the NS group, although in the four NNS sub-groups this overall weight gradually increased from the under-marking which characterised the production of less proficient speakers. This suggests that transfer does not generally impede the extent to which such prominences are produced here. The fact that in Spanish prosodic means are employed to create prominence less *frequently* than in English does not appear to influence the *quality* of production of normal prominences in L2 English to any great extent.

Indeed, an alternative explanation of the initially weaker prosodic marking in the less proficient subgroups is that speakers exhibit a general reticence to exploit fully the prosodic features during speech. We know that the production of prosodic features

[2] Indeed, Cid-Uribe and Roach (1990) note that Spanish intonation 'has always been described in a global rather than an atomistic way, in terms of intonation contours where the end of the contour is the factor that bears the linguistic significance' (p. 1).

relates directly to the physical effort of speech and to the degree of physical involvement with which a speaker communicates. Perhaps there is a tendency for L2 speakers to restrict the extent of their involvement with what they are saying, or for the heavy constraints of processing or anxiety to lead to speech which is 'held back' in terms of its gestural dynamism.

Such a reticence would have no relation to L1 at all, but would lead to flatter and less varied prosodic production. Certainly, speaking in an L2 involves profound psychological adjustments to the ego of the speaker, which may result in this general kind of reticence. For example, Guiroa *et al.* (1972) report increased accuracy in the pronunciation of L2 Thai after informants were given a dose of alcohol, which was interpreted as the result of the increased permeability of ego boundaries which alcohol is known to induce.[3] In terms of language ego and L2s, others have gone further:

> To seek to change someone's pronunciation - whether of the L1 or of an L2 - is to tamper with their self-image, and is thus unethical - morally wrong.
>
> (Porter and Garvin 1989: 8)

7.1.3 Contrastivity

A puzzling finding in the Spanish-speaking subjects is the over-use of maximum amplitude to mark both normal and contrastive prominences. It was noted that this over-use may represent a flooding strategy, in which both normal and contrastive prominences at first receive amplitude marking, and only at higher levels of proficiency is the distinction between normal and contrastive seen to emerge through the gradual *reduction* of amplitude marking on normal prominences.

We have seen that amplitude is a more reliable marker of contrastive than normal prominence in Spanish. So, Spanish-speaking learners here may initially rely on their own L1 phonological model for contrastive (i.e. strongest, most forceful) prominence, and mark *all* prominences with what they intuitively feel to be the most salient form of production, in order to maximise the communicativeness of their L2 speech. That is, their L2 production strategy is to rely on an over-generalised model of the production of prominences. Such exaggerated use of prosodic features is known to exist in other

[3] Unfortunately, the positive effects of alcohol become negative when doses exceed two ounces.

speaking conditions where communication needs to be maximised overtly (caretaker talk, foreigner talk). Perhaps in SLA too such a communication strategy exists.

The failure of NNSs to mark contrastive prominences sufficiently strongly is a very significant finding in our data. We saw how striking this failure was, both in medial and final positions, and how in general contrastive marking was no heavier than on normal prominences, whereas comparable NS productions showed a substantial increase in marking. Yet we observe that Spanish speakers have a similar system of contrastive marking. Of course, we have no directly comparable data on the weight of marking in Spanish, but we know that narrow focus is explicitly marked, largely with increased amplitude. Sensitivity to the prosodic requirements of contrastive marking may simply be a very late developing feature in the case of Spanish-English ILs, or indeed in the SLA of English generally. This may have to do with the generally high processing requirements in the production of an L2, which lead to an upper limit on what can be handled in real-time production. Contrastivity in this sense might be seen as a dispensable aspect of production, a relatively complex system relating to the fine-tuning of meaning, which is incorporated into L2 speech only at very high levels of proficiency.

Interestingly, I have also noted that NNSs take longer than NSs to shift prosodic weight from normal to contrastive elements in successive contextualised utterances. I suggested that there might be generally more severe constraints on how much meaning an L2 speaker can map onto the prosodic layer during speech. This would support the notion of a late development of models of contrastivity generally.

Such findings suggests that limitations in the production of prosodic form in our NNSs manifest themselves in very specific ways, and at very predictable points in the developing L2 phonology. The difference between the marking of normal and contrastive prominence is clear; whereas the general reticence in production discussed above may also be present during the production of contrastive prominences, the NNS-NS differences here are far greater than in the case of normal prominences. Indeed, there is no appreciable increase in the ability to differentiate normal and contrastive prominences prosodically as proficiency increases. So, the data here on contrastive prominence seem very clear evidence of a late development of a phonological system of contrastivity.

Two other aspects of contrastive marking were seen to be absent or inadequately used in the data: the marking of two elements as prominent where double prominence marking within a unit is necessary, and the contrastive marking of final (i.e. default 'normal' position) prominences where a narrow focus is required. Both these observed

deficits appear to correspond to my suggestion that it is the more complex demarcative (non-default) functions of prosody, such as contrastivity, which are absent, or which develop particularly late, in L2 speech. If this is so, such a notion may have important consequences for our understanding of how 'foreignness' is perceived in L2 speech, as well as for our approach to teaching L2 speech skills.

7.1.4 Other L1s and the marking of normal and contrastive prominence

Comparable data on the production of normal and contrastive prominences were also elicited from the L2 English speech of Chinese and Italian speakers:

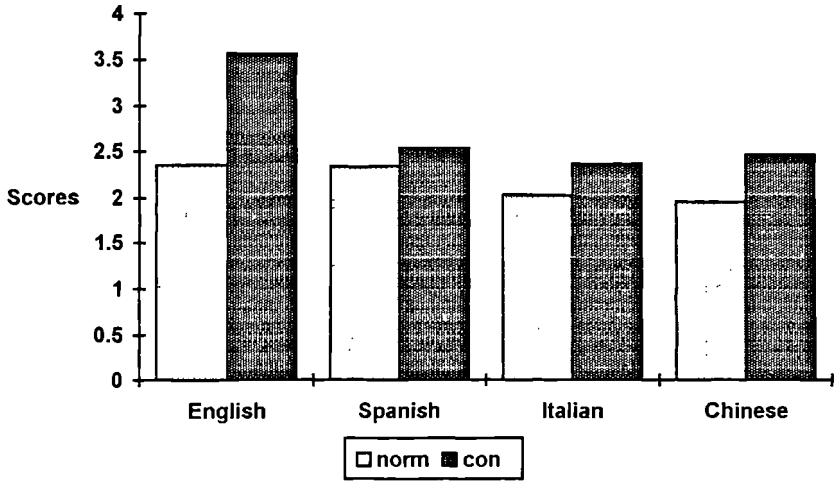


FIGURE 7.1 Prosodic marking on productions of normal and contrastive prominences

Scores for prosodic marking in both Chinese and Italian speaking groups show that, on average, normal prominences carry less prosodic weight than in the case of the NS group. This appears to give support to a general claim that speakers of L2 English have a tendency to under-utilise prosodic means here relative to target norms. Also, all L1-defined groups here show a considerable deficit in the extent to which contrastive prominence is marked prosodically, both relative to their own marking of normal prominences, and also compared to NS levels of marking. Thus, my previous comments on the acquisition of contrastive marking may be generalisable to the second language acquisition of English generally. The data here reflects the L2 speaking norms of speakers with very different (L1) prosodic intuitions, and it is perhaps most surprising how closely their relative scores for normal and contrastive prominences exhibit very similar tendencies here.

7.2 Intonation units

Shorter than NS-average intonation units were observed in both my (Spanish) NNS data from screen-question tests and also in the text reading. Across all single units in the screen-questions, the NS group produced units with an average 6.47 words per unit, and the NNSs an average of 4.83.

In Spanish, a unit comparable to the English intonation unit is generally recognised. Canellada and Madsen (1987) talk of a *phonic group* (pause defined intonation unit) as containing 'at least one accentual group or clause' (p. 103). Elsewhere, Spanish *phonic groups* have usually been characterised in similar ways as have English intonation units: they are said to coincide most often with the grammatical clause (Navarro-Tomás 1966: 37-8; Quilis and Fernandez 1990: 164); they are delimited by pause or potential for pause (Quilis and Fernandez 1990: 163-4); and are referred to as *breath groups* ('grupos espiratorios') (Quilis 1981: 350, 363). So, some attempt can be made to compare average length of units in both languages, although we cannot hope to quantify these exactly, since transcription conventions, the precise definition of a unit, and the register and genre of different spoken data under analysis may all distort comparisons. However, given my own data, it will be of interest to see whether L1 Spanish can be cited as a possible factor in the NNSs' production of shorter units than those of NSs informants under the same conditions.

There are three small corpus studies available: Navarro-Tomás (1966), and two corpora in Canellada and Madsen (1987), one of which is based on literary readings and one of a more 'popular' register. The data from these three corpora are expressed in syllables per unit, and have been reanalysed here and expressed in terms of their average words per unit count, using a rough syllable-per-word ratio:[4]

[4] To arrive at a reasonable syllable-to-word ratio of modern Spanish, an analysis of a small corpus was made, containing 573 words from *El Pais* (an article on family finances) and further 728 from *Hola!* magazine (an interview). Proper names were removed and all figures, dates etc. were counted as if having been read aloud. The syllable per word counts were 2.179 and 1.973 for *El Pais* and *Hola!* respectively; the count for the whole corpus was 2.067.

	Navarro-Tomás (1966)[5]	Canellada and Madsen (1987) - literary readings[6]	Canellada and Madsen (1987) - popular register[7]
syllables per unit / equivalent words per unit	% of all units	% of all units	% of all units
< 5 s.p.u / < 2.4 w.p.u	13	15.2	29.4
5 - 10 s.p.u / 2.4 - 4.8 w.p.u	68	49.5	58.8
> 10 s.p.u / > 4.8 w.p.u	20	34.4	10.2

TABLE 7.1 Percentage occurrence of units of different length in Spanish

The most common unit length is found to be between 2.4 and 4.8 words in all three studies.[8] Cid-Uribe (1989), in a study of an extensive corpus of spoken Spanish, finds an average of 3 words per 'minor tone unit'. She claims that Spanish and English operate with similar length units.

Studies of English provide comparable findings. Hewings (1993) finds an NS average of between 2.9 and 3.5 words (pp. 145-50) in read-out dialogues and 2.8 (p. 266) in free speech. The Lancaster/IBM Spoken English Corpus (SEC) contains an average of 4.82, and Gussenhoven (1986) cites 4.5 words per unit (p. 82).

What evidence there is, then, does not seem to suggest that the two languages are substantially different here. It appears that transfer does not offer a simple explanation of the shorter units produced by the NNSs.

Hewings (1993) suggests that shorter units in NNS speech may reflect speech production in which word-selection predominates at the expense of prosody. In his data, it appears that NNSs 'target-in' on too many lexical items, marking them as (inappropriately) prominent, which leads to a tendency towards unmotivated unit divisions. As a result of this focus on *word selection*, he argues, the usual prosodic planning based on *sense selection* of whole units suffers (p. 158). A word-based processing strategy, he claims, is also evidence of a reduced capacity for forward planning. If this is so, then in my data the free speech of the screen-questions would

[5] Navarro-Tomás 1966: 84 ff. Orginal statistics (%): <5 syllables 12.55; 5-10 syll. 67.60; 7-8 syll. 26.32; >10 syll.19.54. These data do not appear in earlier editions of the book.

[6] Canellada and Madsen: (p. 103) literary readings, original statistics (%) (which total only 99.1 in their data): <5 syllables 15.3; 5-10 syll. 49.5; 7-8 syll. 19.5; >10 syll. 34.4

[7] Canellada and Madsen (p. 104) from reading of a more popular register, original statistics (%) (which total only 98.4 in their data): <5 syllables 29.4; 5-10 syll. 58.8; 7-8 syll. 21.0; >10 syll.10.2

[8] There are also data for units of between 7 and 8 syllables (3.4-3.7 words): Navarro-Tomás (26%), Canellada and Madsen (literary reading 19.5%, popular register 21%).

indeed require substantial forward planning, and the relatively shorter units in the free speech of NNSs relative to that of NSs may reflect this.

7.3 Intonation and information

7.3.1 Listing intonation patterns in Spanish and English

Transfer can be positive and potentially facilitative in SLA. Canellada and Madsen's (1987) descriptions of Spanish listing patterns (pp. 144-5) are substantially the same as those which I have previously described as typical English patterns. Their description of Spanish listing patterns includes a division into *tensive* and *distensive* parts of a whole contour (represented by '//' in their transcription):

A//a-a-a C[9] = (broad) rise // (narrow) fall-rise - fall-rise - fall-rise - (broad) fall

which, using Canellada and Madsen's diagrammatic system, we can represent as:

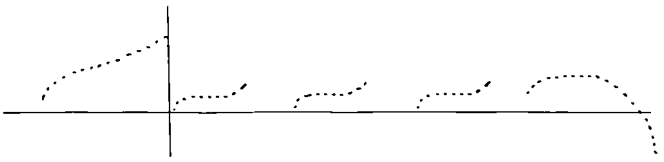


FIGURE 7.2 Listing pattern in Spanish, according to Canellada and Madsen

This pattern, at least in the last two units, is broadly comparable to Quilis and Fernandez's description (1990: 173):

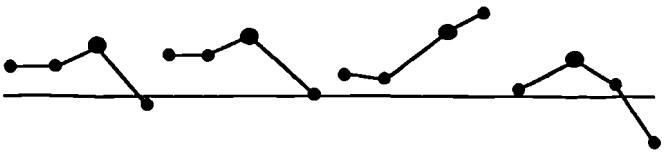


FIGURE 7.3 Listing pattern in Spanish, according to Quilis and Fernandez

[9] A ('anticadencia') // a ('semianticadencia') - a - a - C ('cadencia')

Both describe a rise-to-fall pattern in the final two tones; the tone sequences of the whole contours reflect the two most common patterns in English (RRRF and FFRF). It is no surprise, then, that my NNSs managed to produce nativelike listing patterns; such patterns are common to both languages.

7.3.2 Marking of a 'new' element within a list

The NS group in my data disturbed the predictable open tone in penultimate position in a list, when that tone marked a newly-introduced item relative to a preceding list. However, the NNSs did not do so to any significant extent.

Little work has been done on 'informational' aspects of intonation in Spanish. However, the following can (somewhat circuitously) be said: finality is generally marked in Spanish with a falling tone (Quilis 1981: 416); the nucleus and any following material in a unit are recognised as the rheme (Canellada and Madsen 1987: 132); generalising for unit structure, an ascending-descending contour typifies Spanish intonation units (Canellada and Madsen p. 126 ff). Given that informational order in Spanish is (default) theme-rheme (Contreras 1980: 50-3; Candellada and Madsen p. 132), we can infer that rhematic, or most probably non-given, information, is more likely to receive a falling than a rising tone in Spanish.

Yet the Spanish-speaking NNSs do not produce appropriate falling tones during the utterance of (L2) lists. Perhaps they are unable to alter listing patterns in this way. In other words, their L2 speech may be typified by a tendency towards a less flexible prosodic organisation, where the production of a particular conventionalised pattern, here a listing pattern, somehow blocks more elaborate, non-default manipulations of that pattern, making it difficult for more complex prosodic configurations to be incorporated into speech.

In Chapter One I discussed Cruttenden's suggestion that there may be a hierarchy in the various competing uses of fall and rise in English, with core, linguistic uses giving way to more discoursal ones (for example, 'new' information marking, or pragmatic uses of tone) where necessary. In this way perhaps we can understand these NNS data as reflecting the availability of more basic uses of intonation (the 'default' ones), but a shortfall in more complex uses of tone. Such complex uses perhaps require too great a

cognitive effort during L2 speech, and thus default intonation patterns become relied upon, not only contrary to target-like models, but even perhaps contrary to L1 models.

Hewings (1993) finds a similar phenomenon. Using Brazil's proclaiming (fall) and referring (fall-rise) opposition, he states that in his data NNSs are demonstrably 'in command of the uses of the p/r opposition in English to mark matter as either contextually recoverable or part of the common experiential knowledge of the participants' (p.216). He finds, however, that NNSs fail to use tones in other, discoursally appropriate ways, such as for 'socially integrative purposes' (his data here is from dialogue readings, where conversational/interactional use of tones is necessary). This bears some similarities to my own findings; NNSs use the open/closed opposition to structure well-formed listing patterns, but other, 'superimposed' uses of tones are largely absent. As I have noted in Chapter One, intonation is by its nature polysystemic, and its realisation includes decisions as to how competing systems are reflected within the prosodic layer. This may be an area where speakers of an L2 are capable of processing decisions of less depth, which may result in contextually inappropriate production.

The Italian and Chinese data here are far too small to extract percentage uses of different sequences, but the general tendency at least suggests that the same occurs here. There were no significant deviations in typical tone selection on penultimate list items when these items constituted new or given information.

7.4 Tone repertoires

It was noted that in the NNS data there was a very significant tendency to over-use open tones (including rises, fall-rises and level tones not close to baseline) relative to target-like usage. However, the proportion of open to closed tones became more nativelike as proficiency increased. To assess the role of L1 transfer here, the tone repertoire of Spanish will be compared to that of English.

7.4.1 Tone repertoires in Spanish and English

Stockwell and Bowen (1965) claim that, like English, Spanish has a three-tone 'terminal juncture' set of fall, level and rise (p. 25). Quilis also specifies the same three (1981: 416). Quilis and Fernandez (1990: 165 ff) describe how, in realisation, a total of five distinct tones operate, with both fall and rise having two allotones. These five are also used by Candellada and Madsen (p. 16).

Cid-Uribe (1989) is the only large-scale quantitative study available of Spanish tone repertoires. She uses an enlarged tone set, which is designed to allow for cross-linguistic comparisons with the SEC (English) corpus. These data are presented below along with data from the SEC corpus, and also those corpus studies of English cited previously, which use smaller tone sets:

	Spanish		English					
Study/ tone	Cid-Uribe		SEC	Crystal (1969)	Quirk *et al.*(1964)	Davy (1968) - reading	Davy (1968) - conver-sation	smaller tone set
high fall	11.2		37.3	51.2	52.5	50.2	58.7	fall
low fall	13.2		17.4					
high rise	0.8		3.2	20.8	24.7	24.6	24.6	rise
low rise	9.9		12.1					
rise-fall	13.4		0.2	6.9	4.5	2.7	4.6	rise-fall[10]
fall-rise	10.6		14.3	16.2	16.2	16.6	12.5	fall-rise
level	40.9		15.6	4.9	-	5.5	8.0	level

TABLE 7.2 **Different corpus studies of English and Spanish, showing percentage occurrence of tones**

From Cid-Uribe's data, we see that high rise and high fall, especially the latter, are more common in English than in Spanish; in Spanish, tones which describe a relatively narrow kinetic movement are favoured. Also, there is here a very high percentage use of level tones in Spanish. Both these suggest that Spanish speakers make less extensive use of variation in the pitch line. Indeed, Cid-Uribe notes that a level in Spanish may be an allophone of either a rise or a fall, and this adds to the general impression that there is a tendency in Spanish to favour flatter pitch lines (p. 69). If we look again at the data for tone use in the multi-unit utterances of the NNSs in the present study, where speech was free and did not include lists, we see how close to Cid-Uribe's

[10] Percentages for rise-fall and rise + fall are summed, as are those for fall-rise and fall + rise.

data on the production of level tones in Spanish (40.9%) speakers of sub-group NNS1 (the least proficient) come:

	levels
NSs	5.2
NNS (all)	38.6
NNS 4	27.3
NNS 3	38.3
NNS 2	46.5
NNS 1	43.8

TABLE 7.3 Percentage use of level tones in multi-unit utterances

It seems likely, then, that transfer does influence this aspect of the data, and that the use of levels here may be evidence of two related kinds of transfer: the predisposition to use the level tone, and also the general tendency to produce flatter pitch lines. If, as Cid-Uribe suggests, levels often function as allotones of rises or falls in Spanish, this particular aspect of L1 may very well transfer. The significant diminution of use of levels in the highest two proficiency sub-groups suggests that such a feature undergoes alteration through acquisition, although remaining far above target-like norms even at late stages in acquisition.

However, not all evidence of tone selection is so easily attributed to L1 transfer. For example, in all the listing intonations, NNSs significantly over-use rising tones compared to the NS group data. Nevertheless, we cannot claim from these data that the NNS lists are intonationally *nonnativelike*, nor assert with confidence that L1 transfer is in operation; as I have noted, the use of a series of rising tones in lists (e.g. R R R F) constitutes a well-formed listing pattern in English, and Spanish speaking subjects may simply have used this correctly-realised L2 pattern.

If we consider overall tone selection in all listing patterns in the Chinese and Italian data, however, we see that indeed, the preference for rises and for levels is far stronger in the Spanish data:

	fall	rise-fall	rise	fall-rise	level
NS n=90	63.3	5.6	8.9	7.8	14.4
Spanish (all) n=398	28.1	0.0025	32.2	1.5	56.7
Italian n=87	52.9	0	20.7	6.9	17.2
Chinese n=54	55.6	0	25.9	11.1	7.4

TABLE 7.4 Total use of tones in all four listing utterances throughout L1-defined subject groups

This does suggest that L1 transfer may account for the unexpectedly high occurrence of rising tones here.

7.4.2 The mystery of over-used rising tones

My data from the multi-unit utterances represent a numerically substantial body of data, and also reflect something closer to normal, spontaneous production than the data on listing. In these multi-unit utterances, Spanish NNSs used higher than target-like percentages of rises. However, the over-use was statistically rather small, and reduced towards nativelike levels throughout the four sub-groups. Taking fall-rises and rises together, the percentage used by NNSs in fact remains very stable at around 20%:

	fall and rise-fall	rise and fall-rise	levels
NSs	70.8	24.0	5.2
NNS(all)	41.4	20.0	38.6
NNS 4	53.3	19.5	27.3
NNS 3	40.3	22.4	38.3
NNS 2	33.8	19.7	46.5
NNS 1	37.5	18.8	43.8

TABLE 7.5 **Summed totals of tones involving rising movement, and tones involving falling movement in the multi-unit utterances**[11]

This distribution of all rising-type tones in the NNS data is very close to the summed percentages for low rise, high rise and fall-rise in Cid-Uribe's data on Spanish speech (22.3%). Crucially, it is far *lower* than comparable summed percentages for all rising-type tones in English corpus studies: SEC 29.59%; Crystal (1969) 36.2%; Quirk *et al.* (1964) 40.9%; Davy (1968) 39.2% (reading), 37.1% (conversation), and also higher (though less markedly) than my own NS data here. This appears to indicate that L1 preferences here are highly resistant to acquisition, remaining almost unchanged across the sub-groups. It also suggests that if there is any validity in the claim that my NNSs use a higher proportion of 'open' tones, it is not rising tones which constitute this higher distribution.

[11] The compound tones fall + rise and rise + fall do not appear in my data.

In the earlier discussion of these data, it was noted that 'open' tones were significantly over-used in the NNS data, and that their use diminished significantly as proficiency increased. Level tones were taken as open when in high or high-medium range, and as falls when close to baseline or, for non-final levels, when medium-low or low in range. A reduction in over-use of open tones was observed over time, so that data from the most proficient subgroup matched approximate open/closed ratios construed from major corpus studies of English (at around 40:60 for open:closed). I also noted, though, that within the comparable NS data, this ratio favoured closed tones more greatly than in the corpus studies, and this was explained in terms of the overtly factual, declarative nature of speech elicited in my tests, leading to a preponderance of falls. Also, I suggested that NNSs here did in fact acquire a native-like open/closed system at general nativelike levels of distribution, but were perhaps not able to vary it according to context in the ways that the NSs did.

However, we now see that these open/closed data hide the finding on rising tones, which show little development in their distribution. Given that the rise-fall was used only rarely in the NNS data, the reduction in level tones over acquisition must be interpreted with reference to the only other significant development in tone distribution: an increase in the use of the falling tone.

By further analysing all the tones classified as levels, we see that the percentage of levels which were interpreted as 'closed' does not reduce significantly as proficiency increases, leading us to conclude that the NNSs here retain their characteristic use of flattened, fall-like low levels. The use of 'open' levels, on the other hand, drops significantly as proficiency increases:

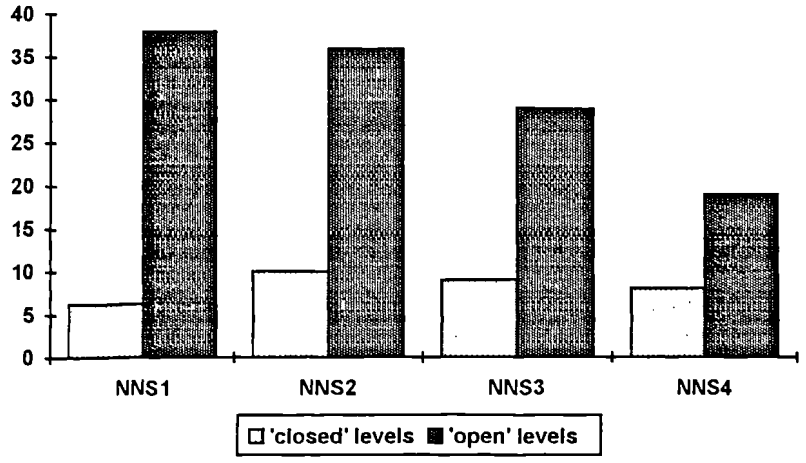


FIGURE 7.4 Distribution (%) of 'closed' and 'open' level tones over four ability subgroups

The number of tokens per subgroup here is small. By grouping the four NNS groups into two larger subgroups, we get a clearer picture of how these changes in use of 'open' and 'closed' levels appear to follow increases in overall ability:

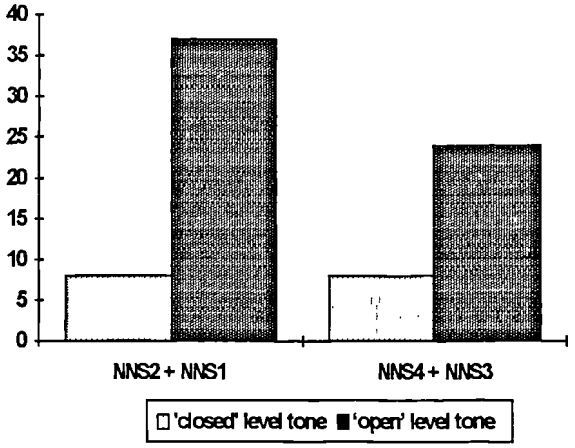


FIGURE 7.5 Distribution (%) of 'closed' and 'open' level tones over two larger ability sub-groups

The principal development is that 'open' levels decrease in use, whereas 'closed' level tones remain at similar distribution. As we already know, the only other significant parallel change in tone repertoires as these open levels decrease is that falling tones increase.

In a discussion of the possible reasons for high distribution of open tones, above, it was suggested that the typical senses of doubt, hesitancy etc. associated with rising tones may explain the preference for open tones. This cannot now be claimed to be the only possible factor; preference for open tones is seen to be the result of a preference for open levels and not for rises. There is very strong evidence that part of this preference is the result of transfer; distribution of all level tones in lowest-proficiency NNSs is at L1 percentages. The diminution in the use of open levels also follows increased proficiency, and can thus be seen as a clear developmental pattern in the acquisition of the L2 tone repertoire. Of course, open level tones may also indicate hesitancy and anxiety.

7.5 Text reading and the use of pause

I have suggested that the use of pause as a device for the delineation of the structure of a spoken text is consistently under-used in L2 English speech, and that this may have a

detrimental effect on comprehensibility. This was true not only of the Spanish NNSs, but also of the Italian and Chinese speaking subjects. Figure 7.6 illustrates the clear similarities between different L1 groups. Again, pause-lengths at each point in the text were normalised for each speaker, then averaged for the whole group, and finally these group averages were ranked for size, to express a 'pause repertoire' for the groups:

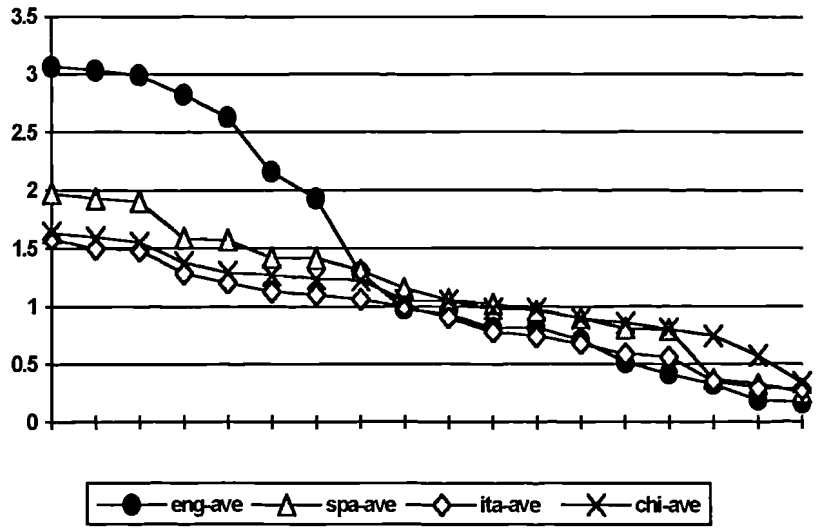


FIGURE 7.6 **Ranked spread of pauses used in the text reading**

7.6 Pitch range in L1 and L2

7.6.1 Free speech

Does speaking in an L2 effect the production of pitch? In order to assess this, pitch readings were taken on both the L1 and L2 speech of all informants. For all the nonnative speakers, readings from responses to the screen-questions 3.4-6 (see Appendix A) were used for L2 data; for L1 data, informants were asked to describe in their own language a series of screens which had no accompanying text questions, but which were similar in design to the other screen-question tests. All data, then, represent something approaching the conditions of free speech. Readings were converted to semitones and averaged for informant groups:

	L1	L2
NSs n=8	5.7	
Spanish (all) n=25	7.6	6.6
Italian n=8	8.6	7.4
Chinese n=5	10.3	9.0

TABLE 7.6 **Average pitch band in semitones for NNSs and NSs speaking their respective L1s and L2 English**

Even though the tendency towards flatter tone lines in L1 Spanish has been noted, pitch ranges themselves are not narrower but in fact wider than the native English data here. However, there is an average reduction in pitch range of one semitone in the Spanish speakers' L2 English speech. A comparable degree of narrowing in L2 speech also occurs in the other L1 groups.

We can see from subgroup data on the Spanish NNSs that in fact this narrowing of pitch range does not appear to alter as proficiency increases:

	L1 Spanish	L2 English
NNS4	8.4	6.3
NNS3	7.5	7.5
NNS2	8.0	5.8
NNS1	6.5	6.9

TABLE 7.7 **Average pitch range in semitones for NNS subgroups, for both L2 and L1 speech in similar conditions**

The extent of the narrowing varies, and once again the size of the sub-groups may interfere here. It seems, though, that narrowing does not diminish as proficiency increases. By dividing the NNS data into two larger proficiency-determined sub-groups, the following averages are obtained:

	L1 Spanish	L2 English
NNS4 + NNS3	7.9	6.8
NNS2 + NNS1	7.2	6.4

TABLE 7.8 **Average pitch range in semitones for two larger (Spanish) NNS subgroups, for both L2 and L1 speech in similar conditions**

There does, then, appear to be a narrowing of just less than one semitone when the NNSs speak English. What is perhaps of more interest is that this does not decrease

as proficiency rises; indeed, of the four subgroups above, it is the most proficient which exhibits the greatest degree of narrowing.

7.6.2 Text reading and onset pitch

The ranking of average onset levels presented in Chapter Six is repeated here, with measurements of Chinese and Italian subject groups included. In terms of pitch onset, the Chinese and Italian data describe less clearly narrower pitch ranges than the Spanish NNSs do. The NSs' onsets describe the broadest pitch range, but significant trends in the Italian and Chinese groups are hard to find, and are unreliable due to the small number of informants:

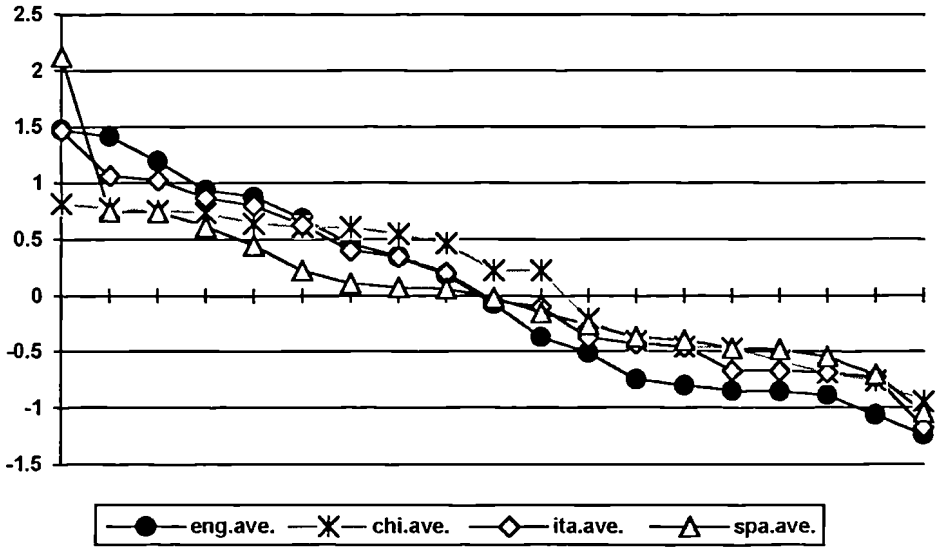


FIGURE 7.7 **The nineteen average onset levels for NS, and for Spanish, Italian and Chinese speaking groups**

Below is a comparison of the production of pitch range, expressed as group averages in semitones, which illustrates how the NS and NNS groups differ in their pitch production during text reading and free speech:

	text reading	free speech
Nss n=8	7.1	5.7
Spanish (all) n=25	5.2	6.6
Italian n=8	6.3	7.4
Chinese n=5	5.8	9.0

TABLE 7.9 Group averages of pitch ranges (semitones) during English speech under two conditions: text reading, and free speech

The English NSs use a broader pitch range for the text reading than for free speech, whereas all NNS groups show narrower ranges for the reading here. Why should this be?

Increased anxiety during reading aloud may inhibit prosodic production in an L2. If this is so, these data appear to suggest two aspects of production detract from L2 speech. The first of these is that increased stress leads to generally narrower pitch ranges, and this occurs most overtly in a text reading, despite the fact that here less linguistic forward planning is necessary than in free speech. Reading from a printed text in experimental conditions no doubt promotes a rather overt kind of monitoring of speech production. What is interesting here is that such a reading task, which we might suppose promotes highly monitored production style, appears to result in less pitch variation than in free speech. Such a (nonnativelike) reduction in pitch range seems to constitute exactly the contrary to what usually occurs during 'monitored' L2 production. It was noted in Chapter Three that high levels of monitoring usually result in increased nativelikeness in L2 production, even in the production of speech sounds. If the reverse occurs in the production of pitch, our approach to the whole issue of L2 teaching in this area may need revision.

Secondly, native speakers of English have broader pitch ranges in the text reading than in free speech. It is as if the reading promotes a very clear use of pitch variation, perhaps to delineate meaning most explicitly where the structured text is clearly laid out on the page and requires an explicit, structured rendition. On the other hand, the free speech (which involved responding to screens with simple pictures on them) may have been deemed by the NSs as too simplistic to require any such use of particularly clear pitch structure. The NNSs appear not to respond to the two conditions in this way at all.

From the point of view of research into intonation, these data confirm what Bolinger has long asserted regarding the elicitation of data on intonation: that read-out productions are the least appropriate samples of production of intonational form.

Chapter Eight - Summary

8.1 Intonational form and prosodic production

8.1.1 Difficulties in the approach to research

I noted in the introduction to this thesis that very little is known about intonation in second languages. The principal aim of this study, then, has been to provide a detailed account of some aspects of the acquisition of L2 English intonation. I have also sought to identify characteristics of prosodic production here which are not evidently related to the L1 of learners, and hence might point to more general tendencies in this area of L2 production.

The standard approaches to the study of SLA, particularly those based on the identification of types of errors in L2 production, are in many ways unsatisfactory in the study of intonation. We can make various quantitative evaluations; measuring pitch range, or noting tone distribution, for example. However, such quantification effectively describes the nature of the prosodic parameters of production, and not in any convincing sense how erroneous this production is. I have noted various tendencies in production here, most significantly, the generally narrower, under-exploited nature of production in L2s.

Apart from the bare facts on prosodic production, we can also look at the linguistic and discoursal uses of intonational forms in speech, and at whether the ability to use the appropriate intonational devices in a given context increases as overall spoken ability does. In other words, research in this area needs to consider simultaneously two distinct aspects of intonation: elements of bare prosodic production, *and* the relationship between phonological decisions here and the meaning of speech. I have in this study attempted to consider both these aspects of intonation.

Many explanatory accounts of how SLA takes place are also unsatisfactory in the case of L2 intonation. Although I have made reference to such processes as 'transfer of training', in general the cognitive nature of many proposed processes of SLA do not accommodate easily the broad gestural and physiological nature of prosodic production. *Intonational* form is perhaps more closely related to linguistic form than other aspects of prosodic production (such as voice quality). Even so, intonational systems are not rule-

governed in any strict sense, and hence the identification of acquisition 'processes' is not straightforward, since evidence of 'errors' is hard to find and justify.

I have previously noted Sweet's distinction between 'analytic' and 'synthetic' systems of language. Another difficulty with the relatively fluid nature of the 'synthetic' systems of intonation is that their very identification in an L2 is troublesome. There is no clear point at which a learner begins to use target-like intonation, since the prosodic layer develops target-like characteristics (if at all) over time, building directly on the base of L1 production. The production of L2 prosody , then, is hardly about right and wrong at all, but about degrees. And as a consequence of this, the measurement of production in terms of meaningful, linguistic use is usually about degrees of appropriateness. Appropriateness can be framed within a study of *tendencies* of production in specific contexts, which is one reason why I have used highly controlled contexts in eliciting speech data here.

In much current work on the phonology of SLA the distinction between 'new' and 'L1-similar' L2 sounds is used to discuss developmental acquisition of L2 sounds. Again, such an approach is not immediately applicable to the study of intonation, since what exactly might constitute a 'new sound' in the case of L2 intonation is not clear. Consequently, notions such as 'equivalence classification' cannot be employed easily.

Yet another problem for us here is the intrusive nature of paralinguistic effects in L2 speech; levels of anxiety, timidity and frustration may affect the bare prosodic production in ways which impinge on the phonological character of intonation. That is, psychological constraints rather than any shortfall in intonational competence may explain deviant production. Second language speech may also contain high levels of acquiescence or doubtfulness, which might lead to a communication strategy involving the proliferation of concessive, non-volitive and conciliatory rising tones. Indeed, the use of 'open' tones has been an interesting features of the data here.

8.1.2 Typifying NNS production

For all those reasons noted above, research in this area is potentially very difficult to frame in a way which allows for interesting findings. My aim in this study has been to take a small number of elements of intonational form which are reasonably easy to identify, and also have a clear function within the discourse of speech. Elements such as

normal and contrastive prominence, listing intonation, the marking of 'new' information and paratonic structure are intended to provide a window on production, where tendencies in NNS production can be compared to those of native-speaking informants.

Several writers have made similar attempts to prescribe a model of intonational form. Wennerstrom (1994), for example, found 'overwhelming consistency' in a group of native speakers of English in the realisation of such elements as high tone boundaries and paratonic structure. In this way a model of target-like tendencies in production was established, against which the production of NNSs could be compared. Such an approach is useful in that it considers speech within a framework of contextualised speech production.

To elicit speech data I have used something similar to The Bilingual Syntax Measure, in an attempt to get even closer to the notion of the obligatory suppliance of forms in context, an approach used in much SLA research. However, the specific nature of elicitation tests here also displays similarities to methods normally used in research into discourse and intonation in L1s. In this sense, I have tried to incorporate approaches to my experiments which are the most appropriate for precisely the kind of data I want to elicit and assess. Consequently, there are no studies of an exactly comparable nature in the literature; my findings, then, are not always directly relatable to the small amount of existing work in this area.

8.1.3 The bare facts and physical properties of production

- **Marking of prominences**

The 'scoring' of prominences used here was a means of expressing numerically the notion of prosodic weight. It is clearly not a *measure* of prosodic strength, but an approximation, incorporating elements of both perceptual and mechanical analysis.

In general, all nonnative speakers in the experiments made less use of prosodic features as markers of normal prominence in L2 English than did native speakers of English. This relative lack of prosodic weight was seen also in contrastive utterances, where despite correct placement of these narrowly focused points of prominence within the sentence, contrastivity here was not marked with the far heavier use of prosodic features which native English speakers used. In the case of 'double' prominences, where

two prominent points are obligatorily marked in the sentence, the same relative lack of prosodic weight was observed in NNS production.

So, speakers of various L1s under-utilise prosodic means of creating points of prominence in L2 English. In particular, they fail to increase the extent of marking to create clear contrastive marking, and also do not appear sensitive to the availability of extra-strong marking as a way of creating emphatic prominences, such as in the case of final-position contrastive prominence. In general there is not only a reduction in the weight of marking, but also insufficient differentiation between normal and contrastive/emphatic levels of marking.

- **Individual prosodic features**

Some increase in the use of pitch as a marker of prominence was observed as proficiency increased. However, NNSs all tended to use a generally narrower pitch range in L2 speech: in text readings there was a narrower than target-like range of onset pitches, and in free speech a significant reduction in total pitch range occurred relative to that in the free speech in informants' L1 (in a comparable task).

The length of pauses in the text reading spanned a greater range of durations in NS than NNS readings, again suggesting a 'narrowing' of target-like production here, where (NNS) speakers do not exploit pause as a means of delineating discourse structure as effectively as NS speakers do.

Conversely, loudness is (in the Spanish-speaking informants) significantly over-used as a marker of all prominences, and I noted that this may be due to a simplified model of prominence marking, in which loudness is used as the principal means of creating prosodically strong points in all cases.

- **Tones**

The Spanish-speaking informants used tone to structure a listing intonation pattern correctly. However, unlike the NS speakers, they were not able to disturb this pattern when 'new' information marking was required.

The distribution of tones in the Spanish group was also interesting, with an overall preference for 'open' tones in excess of target-like norms. Indeed, I suggested that a high percentage of open tones may be a feature of second language production

generally. I noted some specific evidence of transfer in the use of open level tones (section 7.4.2), but also that more general features of production, such as anxiety, may also affect the distribution of tones, as noted elsewhere in the literature. Spanish speaking NNSs are able to use intonational means to bind two units together where these two units are likely to be the result of processing limits forcing an otherwise unwarranted unit division to be made.

NNSs here tended to speak in shorter intonation units, which reflects other findings in the literature. However, some pauses in NNS speech seemed to be the result of simple breakdowns in fluency, and hence were not seen as intonationally significant in terms of unit divisions.

- **General**

The data here, elicited from highly controlled contexts, indicate that speakers have a broad ability to mark discourse prosodically. For example, they locate prominences at the appropriate point in the sentence. However, they fail to shift prosodic marking from right-most to medial position as quickly as NSs do. This suggests that, whereas a basic, core system of (default) prominence placement is employed correctly, deviations from it, in this case contrastive rather than default, normal prominence, are less easily or successfully realised. This may reflect a general reticence or inability to manipulate prosodic features during production away from default configurations. We have seen something of this in the failure to 'disturb' default listing patterns when new information needed to be marked.

Hence, production tends towards basic, core patterns, and some of the more marked uses of intonational form (contrastivity, the disturbing of default patterns) are not realised consistently.

Both these features of NNS production are seen in the marking of 'phonological parenthesis' in the text reading. Parenthesis is indeed marked using pitch by NNSs, but there is a tendency not to *maintain* low pitch throughout the entire informationally subordinate section of the text; this might be seen as the result of their tendency to favour 'normal' patterns of production (i.e. away from the 'disturbance' of the pitch line which parenthetical intonation requires), or, more directly, as a move away from very low (i.e. extreme) pitch levels themselves.

8.1.4 Developmental aspects of production

The small amount of available work on development of L2 intonation was reviewed in Chapter Two.

A variety of findings from my analysis of data from NNS (Spanish-speaking) informants are suggestive of developmental patterns in acquisition. There are important problems in the interpretation of data from cross-sectional studies of this kind, not least with the criteria for the construction of ability-based subgroups. Nevertheless, some potentially significant tendencies here were noted.

- **Prosodic features**

The generally weaker prosodic marking on normal prominences in NNSs increases slightly as proficiency does. This may be due to specific developments in the production of the three prosodic features here. Use of pitch marking increases; analysis of durational phenomena was inconclusive; loudness is initially over-used on normal prominences, but is used less as ability increases.

However, in contrastive prominences, a different picture emerges. NNSs do *not* significantly increase the extent of overall prosodic marking as proficiency increases, with pitch showing no increased use. Loudness, however, is employed more in higher-ability speakers, and is the only identifiable developmental feature in the marking of contrastivity here.

The ability to create two points of prominences within a single unit in NNSs is neither consistent nor does it increase in more proficient speakers.

In multi-unit utterances, increasingly appropriate tone choices by NNSs appear to reflect a developing ability to use sequences of tones to create complex informational structure, incorporating tonal dependency.

In the text reading, both the relatively narrow range of pitch used at onset, and also the narrow range of pause durations used at unit boundaries, remain consistently narrower than target-like norms, showing no tendency to expand as proficiency increases.

- **Tone repertoires in Spanish and English**

In the tone repertoires of the Spanish-speaking NNSs, there were several significant developments over time. The lower-than-target distribution of falling tones increased significantly with proficiency, and the substantial over-use of level tones fell very significantly. Also, the L1-related tendency to use very flat tones, resulting in high distribution of both high and low level tones, changed in an interesting way: the distribution of 'low' level tones (which I interpret as closed-type tones) remains steady, whereas the high distribution of non-low levels (open-type tones) reduces significantly. Paradoxically, the distribution of rising and falling-rising tones remained very stable throughout all groups, reflecting L1 levels of distribution. So, the principal development is that 'open' levels decrease in use, whereas 'closed' level tones remain at similar distribution.

- **Pitch range in L1 and L2**

In free speech, the pitch range used by Spanish-speaking NNSs is on average one semitone narrower in L2 English than in comparable tasks in their L1 speech. This relatively narrow pitch in the L2 shows no tendency to expand in the more proficient speaker groups.

8.2 Processes in SLA

8.2.1 Universals features

Discussions of the universality of intonation are at best rather general. There are no well-attested universal orders, such as the sonority orders or vowel hierarchies of segmental phonology. I have already discussed the very broad kind of universality often attributed to intonation, particularly to pitch tones. We have perhaps seen some evidence of a 'universal' use of loudness, in the flooding of prominences with maximum amplitude. As I suggested earlier, perhaps this is related to the very direct physical nature of how we most easily create prominences, through increased loudness. We saw that this overuse

declined over time, perhaps as a better sensitivity to the requirements of target-like production was acquired.

Certain aspects of tone use, for example listing patterns and terminality marking, are shared by the informants' L1 (Spanish) here, and the positive transfer of such uses was observed. More complex uses, such as the marking of discourse structure in multi-unit utterances, are seen to develop along with proficiency; this suggests that the very general meanings which can be attached to tones, specifically to the open/closed opposition, are often employed in contextually appropriate ways by NNSs, and that this appropriate use becomes more nativelike over time. However, we have also seen that certain tonal sequences, specifically the listing patterns, were not altered to incorporate necessary 'new' information marking, despite the fact that the marking of new information with a fall seems to be a feature of Spanish. This was attributed to an inability to subvert default intonation patterns in contextually appropriate ways.

So, through processes of (positive) L1 transfer, and also universal intuitions as to the meaning of tones, speakers gradually approximate to the target model. However, certain uses of tone, in particular the more complex ones, are not evident in L2 production. Perhaps, then, some very general productive constraints operate in L2 intonation at the level of context (rather than pure 'competence'). I will discuss this below.

8.2.2 Cognitive load and context

I have noted that there may be evidence that general cognitive constraints on production, rather than deficits in the speaker's intonational competence, are significant aspects of the nonnativelikeness of L2 production here. One example of this is that whereas both NSs and NNSs 'shift' prominence marking from default right-most position onto a (medial-position) contrastive element in the sentence in response to context, the NNSs appear to 'shift' less of the available prosodic means in the same context, and over several utterances take longer to achieve fully this shift. Other aspects of contrastive marking were absent or inadequately used in the NNS data, include the marking of double prominence in a unit, and the contrastive marking of final (i.e. default 'normal' position) prominences where a narrow focus is required.

There may be gaps in NNSs' L2 intonational competence here. However, these deficits also share the characteristic of being relatively complex, demarcative uses of intonation, in which a particularly overt relation between prosodic form and meaning is created, requiring a relatively high processing load during the encoding of speech. Certain forms may not emerge in L2 speech simply because the speaker is unable to accommodate them during production, despite a sensitivity to the relevant aspects of meaning in his speech.

Those very few data we have on the production of Italian and Chinese speakers here support the general assertion: they also failed to mark new information with a falling tone when this entailed disturbing a listing pattern.

8.2.3 Speaking styles and affective states

In the discussion on speaking styles and monitoring in Chapter Two, I noted that more careful styles will tend to lead to more accurate L2 production generally. Since all my data were elicited in a rather unnatural setting, where the informants were aware of being under observation, one might expect that their speaking style would be formal and, as a consequence, relatively form-focused.

There are several accounts in the literature of how formal styles tend to produce more accurate segmental production. However, prosodic production is a largely spontaneous feature of speech, and little is known about the effects of speaking styles, or indeed of monitoring. I found that whilst the native speakers used broader pitch ranges for the text reading than for free speech, all NNS groups show narrower ranges for the reading here.

In the previous chapter I suggested that the increased stress associated with reading aloud leads to generally narrower pitch ranges, despite the fact that here less linguistic forward planning is necessary than in free speech. Even though such a task presumably promotes very self-conscious, and thus careful production, this leads to a reduced pitch range, where native speakers actually expand their range. The production of pitch, then, suffers contraction as a result of either anxiety, or 'careful' speech, or both.

8.2.4 Transfer

Wennerstrom's (1994) findings suggest that L1 influenced the degree of nativelikeness of L2 English intonation.

I have already mentioned some examples of probable (positive) transfer in the case of Spanish-English in my data. However, I have also noted that in the use of tone, direct L1-L2 statements are difficult to support, and that a more complex picture emerges here than a simple contrastive account of both languages would suggest.

Considering also the (few) data from the Italian and Chinese speaking informants, no significant differences in the degree of prominence marking was found. Indeed, the similarity between the data from the three language groups here were surprisingly high. Pitch and pause production in paratones is also similar.

8.2.5 Input

Elements in speech such as contrastive prominence, double contrastive prominence, unit-final contrastive prominence, closed tone marking on a penultimate item in a list, are all likely to occur in normal speech less often than their 'unmarked' counterparts (which are observed to be acquired more successfully by the NNSs). So, perhaps the relative absence of these marked elements in production is simply the result of learners having received insufficient input in these areas.

My informants all lived and studied in England. Hence, we might expect that the majority of their current input was naturalistic, and did *not* exhibit that very clear and exaggerated intonation associated with 'foreigner-talk'.

Indeed, the whole issue of L2 intonation and input is inadequately understood. White (1987) has suggested that SLA is substantially 'input-free'; input itself will never lead to a reduction in an SLA feature such as overgeneralisation, and explicit corrective feedback is needed. In the case of those intonational forms under investigation here, corrective feedback from native-speaking interlocutors may not occur; whereas corrective feedback of grammatical, lexical and even segmental errors are all common in and out of the classroom, the correction of an inappropriately strong or weak prominence, or of an errant rising tone at the end of a sentence, would be very hard to achieve successfully. Indeed, we generally seek to correct only a few of the most overt

errors which a NNS makes during a conversation. Given that intonation is in some ways 'at the edges of language', we probably stay clear of pointing out problems in the production of intonation, lest the NNS feel that we are being pedantic or superficial. Moreover, when an NNS uses intonational devices inappropriately, meaning will usually be extracted by the (NS) listener from context, and many NSs may indeed not be *aware* of the 'error' at all. The same may be said of 'inferred negative feedback', that is, the lack of comprehension in an interlocutor which indicates to the NNS that his language contains errors.

So, learners may never become sensitive to the precise nature of their nonnativelike intonation. White (1987) also argues that at least for some L2 structures, it is *incomprehension of input* which may induce a greater effort on the part of the learner to attempt to construe rules from the (confusing) linguistic evidence before them. This is also the position of Krashen, whose Input Hypothesis (see Chapter Two) proposes that input is useful when it is (linguistically) slightly beyond the reach of the learner's present competence ($i + 1$) but still broadly comprehensible within context. Krashen's model relies largely on data from the acquisition of grammatical morphology. Intonation, however, can often be relatively marginalised in comprehension, especially when a learner is attending to meaning-retrieval. Intonation, as I have noted, is in some ways 'at the edges' of language, in the sense that it is usually not strictly *necessary* in order that meaning be communicated. There are usually alternative ways of constructing meaning within the context of utterance where intonation and general prosodic form are either not fully understood or appear to be misleading. Thus, a shortfall in comprehension resulting from incomplete knowledge of intonational form may not render input incomprehensible. Certainly, the learner is unlikely to identify intonation as the *source* of incomprehension. In an SLA context, then, deficits in intonational competence may rarely act as a stimulus to a learner's acquisition in this way. One can easily imagine how comprehension in many contexts can function quite adequately without maximum understanding of prosodic systems, and indeed L2 intonational competence has been seen as underdeveloped in terms of perceptive skills (Cruz-Ferreira 1987)

The likelihood that L2 listeners' attention is not heavily orientated towards prosodic form might also mean that 'indirect negative evidence' of intonational form does not form part of typical input in SLA. Such negative input, where the *absence* of a form in input provides the learner with (indirect) evidence that it is not a well-formed target language form, may be relatively inactive in the SLA of prosodic form. Where

indirect negative evidence operates at all, it is likely to be in the realms of linguistic structure, where, even on an inferred level, input may lead to the perception that certain discrete forms never arise in the target language. It is unlikely that such inferences could be made with regard to the relatively loose and abstract systems of prosodic structure.

8.3 Research into L2 intonation and pedagogy

There is a widespread belief that L2 intonation is highly resistant to acquisition (see Grosser 1993), and that intonation is 'a difficult subject' (see Wooland 1993).

My findings suggest that, in terms of the phonetics of intonation, there is a definite development in the production of prosodic features towards native norms in the realisation of certain intonational forms. I have also noted, though, that in my data, L2 production tends to be prosodically flatter and narrower than target norms. This general narrowness of production doesn't seem to disappear naturally as acquisition progresses. Phonologically, I have also noted significant developmental trends in the appropriate use of pitch and tone in various contexts.

In terms of the linguistic and discoursal uses to which specific intonational forms are put, it appears that whereas the more core, basic forms are acquired - due to the transfer of L1 or to the availability of universals - other, more complex uses, especially contrastivity, are very slow to emerge. In discussing intonation, we need to distinguish the phonetic character of forms from the phonological form which they carry. In the case of contrastivity it is largely the phonetic aspects in the present data which fall short of target norms: prosodic marking is not sufficiently heavy, whereas the *need* to make narrow focus appears to be a fully functioning part of speakers' intonational competence.

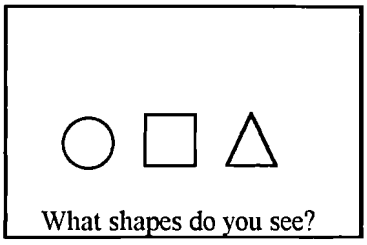
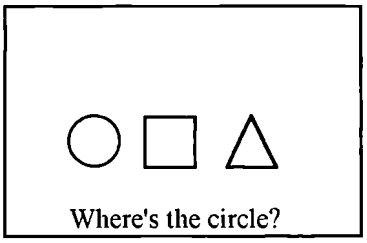
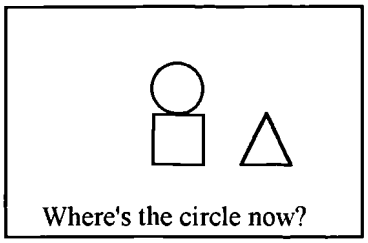
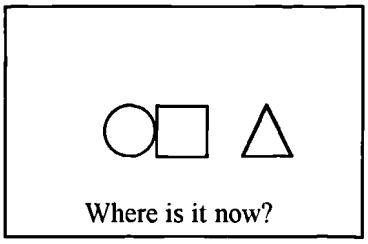
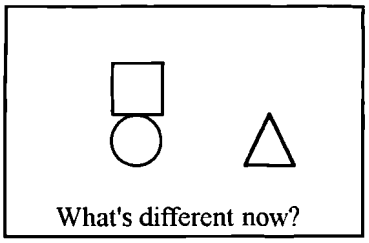
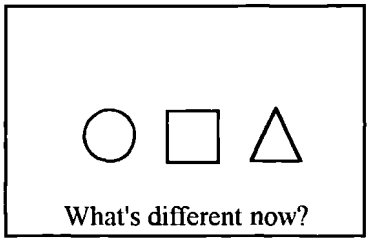
Many of the forms I have examined in this thesis are almost entirely absent from English pedagogical texts. In particular, I have noted that descriptive accounts of intonation are seen by many as complex and inaccessible and that we still lack a commonly recognised description of intonational form. I have suggested that the lack of teaching in this area may have a negative effect on what learners manage to acquire. However, this is speculative. The very nature of intonation - both its close relation to physical gesture and affect, and also its universal aspects - may make it impossible to teach adequately. There is no available work in this area.

What I have tried to do in this study is to provide a description of what is acquired naturalistically by students. Two questions arise: whether those areas which to some extent develop naturally would do so *more* as a result of instruction; whether those areas which do not develop at all naturally *would* do so under instruction.

These issues, then, need to be addressed.

Appendix A

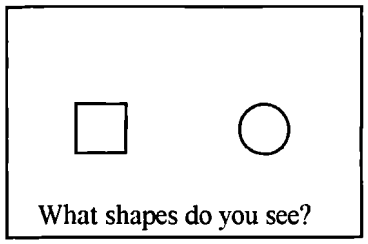
Screen-question tests

Test One (six screens[1]):	Example response[2] and gloss of responses:
1.1 	(Not used)
1.2 	*It's to the left of the square* (normal) Default, last-lexical-item prominence placement.
1.3 	*The circle's on the square* (contrastive) Prominence contrasting with previous response *...to the left of the square.*
1.4 	*Now it's touching the square* (contrastive) Another contrastive element, in the same slot as the previous response.
1.5 	*Now the square's on top of the circle* (double contrastive) The two items from previous response have reversed functional roles, and are thus marked contrastively.
1.6 	*It's like it was at the beginning* (normal) A variety of different responses, most of which carried normal prominence, making no reference to the previous screen.

[1] Each shape had its own colour, which it maintained throughout the battery of tests.
[2] The responses cited are generally the most frequent ones in the data.

Test Two (seven screens)

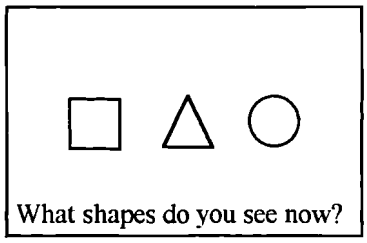
2.1



A square and a circle (list)

Both items in list are new, although not new in the context of the sequence of tests.

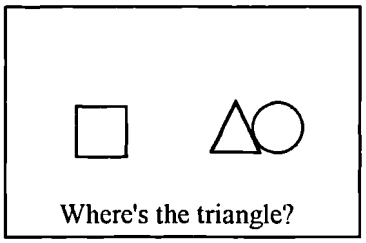
2.2



A square, a triangle and a circle (list)

As above but with a new item introduced, which is predicted to carry 'new' marking in relation to the list items repeated from the previous response.

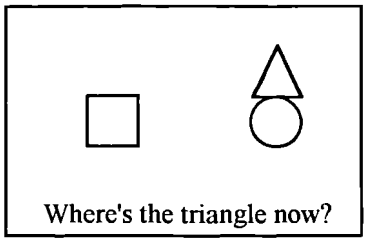
2.3



The triangle's next to the circle (normal)

Default pattern.

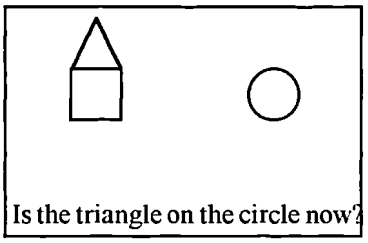
2.4



Now it's on top of the circle (contrastive)

Shift of focus onto new preposition within the repeated PP.

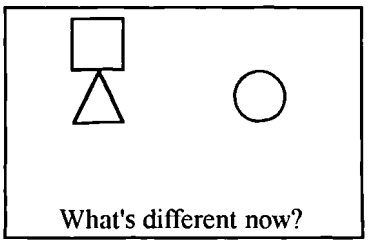
2.5



No, it's on top of the square (contrastive/emphatic)

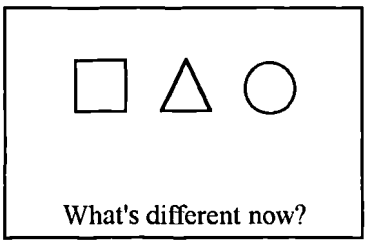
Contrastivity here also corrects the question.

2.6



Now the square is on top of the triangle (double contrastive)
Reversal of functional roles of two items active in previous screen.

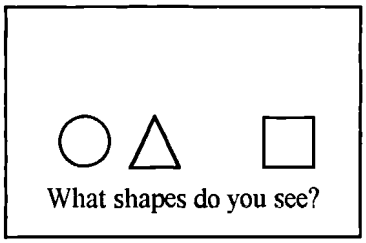
2.7



It's like it was in the beginning (normal)

Default pattern.

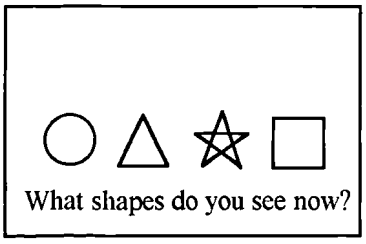
Test Three (six screens)



3.1

A circle, a triangle and a square (list)

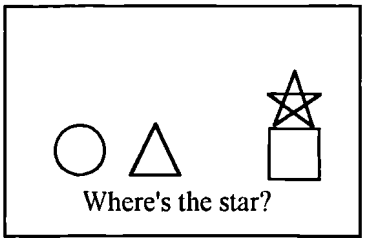
All new items in the test, although familiar from previous tests.



3.2

A circle, a triangle, a star and a square (list)

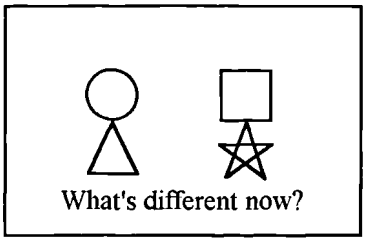
The same list, but with the addition of a totally new item, which is predicted to be marked as such relative to the other list items.



3.3

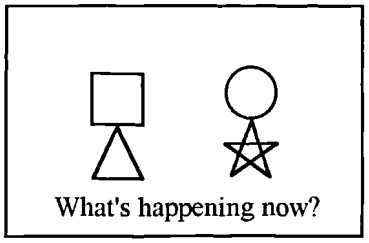
It's on top of the square (normal)

Default pattern.



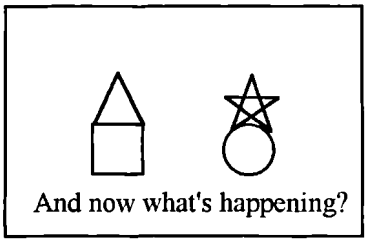
3.4

The circle is on the triangle, and the square is on the star (two-unit response containing four focused items).
In this and the following two questions, the sequence of tones (one was observed on each shape-name) were analysed.



3.5

The square is on the triangle, and the circle is on the square



3.6

The triangle is on the square, and the star is on the circle

Appendix B

Paratone text-reading

The paratone text-reading used in the experiments (see section 3.2.2) was designed to contain a specific informational structure, in accordance with the discussion of paratonic form in Chapter One, part three.

The text is printed here in full, in the form it was presented to informants, and then, below, in sections in which the 'idea' groups are divided by the symbol ↔:

> These days many students come to the United Kingdom to study, often on exchange programmes. If you want to study in the UK, there are nearly one hundred universities. And many towns and cities have two or more universities to choose from.
> Many institutions have a modular system, in which you can study separate units (called modules) from different departments.
> A modular system allows you to choose the things you want to study, depending on your own particular interests.
> If you want to study in Hull, there are plenty of modules to choose from. Modern languages are a popular choice; there are modules in French, German, and many other languages. So, you should never get bored if you do come here to study.

section one (first sub-topic):

> [1]These days many students come to the United Kingdom to study, often on exchange programmes. ↔ [2]If you want to study in the UK, there are nearly one hundred universities. ↔ [3]And many towns and cities have two or more universities to choose from.

section two (second sub-topic):

> [4]Many institutions have a modular system, in which you can study separate units [5](called modules) from different departments. ↔ [6]A modular system allows you to choose the things you want to study, depending on your own particular interests.

section three (third sub-topic):

> [7]If you want to study in Hull, there are plenty of modules to choose from. ↔ [8]Modern languages are a popular choice; there are modules in French, German, and many other languages. ↔ [9]So, you should never get bored if you do come here to study.

The following notes describe some of the features which, given the design of the text, we might expect to observe during reading. Here each of the three sections is referred to as a major paratone, but with the extra-high onset only in the first of these:

1. Extra-high onset to reflect paratone-initial status.
2. New minor paratone with medium onset in first unit. Contains additions to the speaker topic.
3. New minor paratone with low onset. Contains subordinate information.
4. New major paratone, with high onset.
5. Parenthetical (subordinate) information, with low (or extra-low) onset.
6. New minor paratone, with medium onset. Contains additions to current sub-topic.
7. New major paratone, with high onset.
8. New minor paratone, with medium onset. Contains additions to current sub-topic.
9. New minor paratone, with medium or high onset. Reintroduces speaker topic.

Appendix C

Instructions given to judges

On the tape there is a series of short excerpts (between fifteen and thirty seconds each) of a text reading, made by thirty-three different speakers. Some speakers are nonnative speakers of English, whilst others are native speakers. Each excerpt is identified by a number before it begins.

Listen to each speaker and assess which category his/her speech belongs to. In each case make an assessment of the overall standard of the speaker's English. Try and answer the general question, How well does this person speak English? *Do not take into consideration the content of what is being said, since each speaker is reading from the same text.*

There are no absolute beginners amongst the nonnative speakers on the tape.

The categories are:

1. Poor
2. Moderate
3. Good
4. Advanced
5. Native-speaker of English

References

Abecassis, J. and Tardy, M. 1978. 'Intonation et communication'. *Bulletin de Psychologie* 31: 572-80. (Referred to in Marcos 1987)

Abdul-Ghani, C. 1978. *Accent in Discourse: An analysis of function*. PhD dissertation, University of South California.

Academia Real Española. 1962. *Gramática de la Lengua Española*. Madrid: Espasa-Calpe.

Allen, G. D. and Hawkins, S. 1978. 'The development of phonological rhythm'. In Bell, A. and Hooper, J. B. (eds.) *Syllables and Segments*. Amsterdam: North Holland.

Alvarez-Caacamo, C. 1996. 'Building alliances in political discourse: Language, institutional authority, and resistance'. *Folia Linguistica* XXX, 3/4: 245-270.

American Council on the Teaching of Foreign Languages. 1986. *ACTFL Proficiency Guidelines*. Hastings-on-Hudson: ACTFL.

Armstrong, L. E. and Ward, I. C. 1931. *A Handbook of English Intonation* (2nd edition). Cambridge: Heffer.

Backman, N. 1979. 'Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English'. *Interlanguage Studies Bulletin* 4: 239-65.

Bahms, J., Burmeister, H. and Vogel, T. 1986. 'The pragmatics of formulas in L2 learner speech: use and development'. *Journal of Pragmatics* 10: 693-723.

Bannert, R. 1980. 'Svarighteter med svenskt uttal: inventering och prioritering'. *Praktisk Lingvistik* 5. University of Lund.

Bates, E. 1979. *The Emergence of Symbols*. NY: Academic Press.

Bates, E., Camaioni, L. and Volterra, V. 1975. 'The acquisition of performatives prior to speech'. *Merril-Palmer Quarterly* 21: 205-26. (Referred to in Marcos 1987)

Berman, A. and Szamosi, M. 1972. 'Observations on sentential stress'. *Language* 48: 304-325.

Bever, T. G., Fodor, J. A. and Weksel, W. 1965. 'On the acquisition of syntax: a critique of contextual generalisation. ' *Psychological Review* 72: 467-82.

Bierwisch, M. 1968. 'Two critical problems in accent rules'. *Journal of Linguistics* 4: 173-178.

Bohn, O.-S. and Flege, J. E. 1992. 'The production of new and familiar vowels by adult German learners of English'. *Studies in Second Language Acquisition* 14: 131-58.

Bohn, O.-S. and Flege, J. E. 1997. 'Perception and production of a new vowel category by adult second language learners'. In James and Leather (eds.) 1997.

Bolinger, D. 1958a. 'A theory of pitch accent in English'. *Word* 14: 109-149.

Bolinger, D. 1958b. 'Stress and intonation'. *American Speech* 33: 3-20.

Bolinger, D. 1958c. 'Intonation and grammer'. *Language Learning* 8: 31-37.

Bolinger, D. (ed.) 1972a. *Intonation*. Harmondsworth: Penguin.

Bolinger, D. 1972b. 'Accent is predictable (if you're a mind-reader)'. *Language* 48: 633-644.

Bolinger, D. 1976. 'Meaning and memory'. *Forum Linguisticum* 1: 1-14.

Bolinger, D. 1978. 'Intonation across languages'. In Greenberg (ed.) 1978a.

Bolinger, D. 1983 'Intonation and gesture'. *American Speech* 58: 156-74.

Bolinger, D. 1986. *Intonation and Its Parts*. London: Edward Arnold.

Bolinger, D. 1989. *Intonation and Its Uses*. Stanford, CA: Stanford University Press.

Bowler, B. and Cunningham, S. 1990. *Headway Pronunciation - Intermediate*. Oxford: Oxford University Press.

Bowler, B. and Cunningham, S. 1991. *Headway Pronunciation - Upper-Intermediate*. Oxford: Oxford University Press.

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

Bradford, B. 1997. 'Setting the scene'. In Vaughan-Rees (ed.) 1997. p.6.

Brazil, D. 1975. *Discourse Intonation*. Discourse Analysis Monographs No. 1, Birmingham: English Language Research, University of Birmingham. (Later published as: Brazil, D. 1985. *The Communicative Value of Intonation in English*. Birmingham: Bleak House)

Brazil, D. 1986. 'Investigating the intonation of language learners'. In Cling, M. and Humbley, J. (eds.) *Hommage à A. C. Gimson*. Villetaneuse: CELDA.

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

Broen, P. 1972. 'The verbal environment of the language learning child'. *American Speech and Hearing Association Monographs 17*.

Brown, G. 1977. *Listening to Spoken English*. London: Longman.

Brown, G., Currie, K. L. and Kenworthy, J. 1980. *Questions of Intonation*. London: Croom Helm.

Brown, G. and Yule, G. 1983. *Discourse analysis*. Cambridge: Cambridge University Press.

Brown, R. 1973. *A First Language*. London: Allen and Unwin.

Bruner, J., Roy, C. and Ratner, N. 1982. 'The beginnings of request'. In Nelson, K. E. (ed.) *Children's Language*, Vol. 3. NY: Garden Press.

Bryan, K.L. 1989. 'Language prosody and the right hemisphere'. *Aphasiology* 3: 385-399.

Burmeister, H. and Ufert, D. 1980. 'Strategy switching'. In Felix, S. (ed.) *Second Language Development*. Gunter Narr: Tubingen.

Burt, M., Dulay, H. and Hernandez-Chavez, H. 1975. *Bilingual Syntax Measure*. NY: Harcourt Brace Jovanovich.

Canellada, M. J. and Madsen, J. K. 1987. *Pronunciación del Español*. Madrid: Castalia.

Chafe, W. L. 1974. 'Language and consciousness'. *Language* 50: 111-133.

Chafe, W. L. 1976. 'Givenness, contrastiveness, definiteness, subjects, topics, and point of view'. In Li, C. N. (ed.) *Subject and Topic*. NY: Academic Press.

Chafe, W. L. 1979. 'The flow of thought and the flow of language'. In Givón (ed.) 1979.

Chapman, S. R. 1994. *Accent and the Sentence-utterance Distinction*. Unpublished PhD thesis: University of Newcastle upon Tyne.

Chomsky, N. 1971. 'Deep structure, surface structure, and semantic interpretation'. In Steinberg, D. D. and Jakobovitz, L. A. (eds.) *Semantics*. Cambridge: Cambridge University Press.

Chomsky, N. and Halle, M. and Lukoff, F. 1956. 'On accent and juncture in English'. In Halle, M., Lunt, H. and McLean, H. (eds.) *For Roman Jakobson*. The Hague: Mouton.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. NY: Harper and Row.

Chun, D. M. 1988. 'The neglected role of intonation in communicative competence and proficiency' *The Modern Language Journal* 72 (3): 295-303.

Cid-Uribe, M. E. 1989. *Contrastive Analysis of English and Spanish Intonation Using Computer Corpora - A Preliminary Study*. Unpublished PhD Thesis, University of Leeds.

Cid-Uribe, M. E. and Roach, P. 1990. 'Spanish intonation: design and implementation of a machine-readable corpus'. *Journal of the International Phonetic Association* 20: 1-8.

Clark, H. H. 1992. *Arenas of Language Use*. Chicago, IL: University of Chicago Press.

Clark, J. and Yallop, C. 1995. *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.

Clumeck, 1980. 'The acquisition of tone'. In Yeni-Komshian, G. H., Kavanagh, J. F. and Ferguson, C.A. (eds.) *Child Phonology*, Vol. 1. NY: Academic Press.

Contreras, H. 1963. 'Sobre el acento en español'. *Boletín de Filología* XV: 223-37.

Contreras, H. 1980. 'Sentential stress, word order, and the notion of subject in Spanish.' Waugh, L. R. and Schooneveld, C. H. *The Melody of Language*. Baltimore: University Park Press.

Corder, P. 1967. 'The significance of learner's errors'. *IRAL* 5: 161-70.

Corder, 1974. 'Error Analysis'. In Allen, J. P. B. and Corder, S. P. (eds.) *The Edinburgh Course in Applied Linguistics*, Vol. 3. Oxford: Oxford University Press.

Corder, S. P. 1981. *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Couper-Kuhlen, E. 1986. *An Introduction to English Prosody*. London: Arnold.

Couper-Kuhlen, E. and Selting, M. (eds.) 1996. *Prosody in Conversation Interactional Studies*. Cambridge: Cambridge University Press.

Cowin, E., Mann, V., Schoenheimer, J. and Berman, J. 1984. 'Age-related changes in the syntactic interpretation of prosodic cues'. Paper presented to the 108th Meeting of the Acoustical Society of America. (Referred to in Cutler and Swinney)

Cruttenden, A. 1979. *Language in Infancy and Childhood*. Manchester: Manchester University.

Cruttenden, A. 1981. *The intonation of English sentences with special reference to sentence adverbials*. Unpublished PhD thesis, University of Manchester.

Cruttenden, A. 1986. *Intonation*. Cambridge: Cambridge University Press.

Cruttenden, A. (ed.) 1994. *Gimson's Pronunciation of English* (5th edition). London: Edward Arnold.

Cruz-Ferreira, M. 1983. *Non-native Comprehension Patterns in Portuguese and in English*. Unpublished PhD thesis, University of Manchester.

Cruz-Ferreira, M. 1987. 'Non-native interpretive strategies for intonational meaning: an experimental study. In James and Leather (eds.) 1987a.

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

Crystal, D. 1975. *The English Tone of Voice.* London: Arnold.

Crystal, D. 1979. 'Prosodic development'. In Fletcher, P. and Garman, M. (eds.) *Language Acquisition.* Cambridge: Cambridge University Press.

Crystal, D. and Davy, D. 1975. *Advanced Conversational English.* London: Longman.

Cunningham-Andersson, U. 1997. 'Native speaker reactions to non-native speech'. In Leather and James (eds.) 1997.

Cutler, A. and Swinney, A. A. 1987. 'Prosody and the development of comprehension'. *Journal of Child Language* 14: 145-67.

Danesi, M. 1988. 'Neurological bimodality and theories of language teaching'. *Studies in Second Language Acquisition* 10: 13-31.

Davy, D. 1968. 'A study of intonation and analogous features as exponents of stylistic variation, with special reference to a comparison of conversation with written English read aloud'. University of London, M.A. Thesis.

de Bot, K. 1986. 'The transfer of intonation and the missing data base'. In Kellerman, E. and Smith, M. S. (eds.) *Crosslinguistic Influence in Second Language Acquisition.* NY: Pergamon.

de Villiers, J. and de Villiers, P. 1973. 'A cross-sectional study of grammatical morphemes in child speech'. *Journal of Psycholinguistic Research* 2: 267-78.

Dechert, H., Mohle, D. and Raupach, M. (eds.) 1984. *Second Language Production.* Tubingen: Gunter Narr.

Dickerson, W. 1987. 'Explicit rules and the developing interlanguage phonology'. In James and Leather (eds.) 1987a.

Diller, K. C. (ed.) 1981. *Individual Differences and Universals in Language Learning Aptitude.* Rowley, Mass.: Newbury House.

Dinnsen, D. and Eckman, F. 1978. 'Some substantive universals in atomic phonology'. *Lingua* 45: 1-14.

Donegan, P. J. and Stampe, D. 1979. 'The study of natural phonology'. In Dinnsen, D. A. (ed.), *Current Approaches to Phonological Theory.* Bloomington: Indiana University Press.

Dore, J. 1974. 'A pragmatic description of early language development'. *Journal of Psycholinguistic Research* 4: 343-50.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S. and Paolino, D. 1993. 'Outline of discourse transcription'. In Edwards and Lampert (eds.) 1993.

Dulay, H. and Burt, M. 1974. 'Errors and strategies in child second language acquisition'. *TESOL Quarterly* 8: 129-36.

Duwe. 1534. *An Introductorie for to learne to rede, to prononce, and to speke French trewly.* (Referred to in Mallinson, V. 1953. *Teaching a Modern Language*. London: Heinemann)

Dziubalska-Kolaczyk, K. 1987. 'Phonological rule typology and second language acquisition'. In James and Leather (eds.) 1987a.

Eckman, F.R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315-30

Eckman, F. 1987. 'Word-final consonant clusters in interlanguage'. In James and Leather (eds.) 1987a.

Edwards, J. A. 1993. 'Principles and contrasting systems of discourse transcription'. In Edwards and Lampert (eds.) 1993.

Edwards, J. A. and Lampert, M. D. (eds.) 1993. *Talking Data.* Hillside, NJ: Lawrence Erlbaum Associates.

Edwards, J. R. 1982. 'Language attitudes and their implications among English speakers'. In Ryan, E. B. and Giles, H. (eds.) *Attitudes Towards Language Variation*. London: Arnold.

Eichler, A. (ed.) 1910. *Charles Butler's Grammar*. Halle: Niemeyer.

Ekstrand, L. H. 1976. 'Age and length of residence as variables related to the adjustment of migrant children, with special reference to second language learning'. In Nickel, G. (ed.) *Proceedings of the Fourth International Congress of Applied Linguistics* 3. Hochschulverlag, Stuttgart.

Ellis, R, D. 1985. *Understanding Second Language Acquisition.* Oxford: OUP

Ellis, R. D. 1994. *The Study of Second Language Acquisition.* Oxford: OUP

Ervin-Tripp, S. 1974. 'Is second language learning like the first?' *TESOL Quarterly* 8: 111-27.

Esling, J. H. and Wong, R. F. 1983. 'Voice quality settings and the teaching of pronunciation'. *TESOL Quarterly* 17: 98-95.

Færch, C. K., Haastrup, K. and Phillipson, R. 1984. *Learner Language and Language Learning*. Clevedon: Multilingual Matters.

Fathman, A. 1975. 'Language background, age, and the order of acquisition of English

structures'. In Burt, M. and Dulay, H. (eds.) *On TESOL '75: New Directions in Second Language Learning, teaching and Bilingual Education.* Washington D.C.: TESOL.

Fathman, A. 1982. 'The relationship between age and second language productive ability'. In Krashen, Scarcella and Long (eds.) 1982.

Felix, S. 1980. 'Interference, interlanguage and related issues'. In Felix, S. (ed.) *Second Language Development.* Tübingen: Gunther Narr.

Fergusen, C. A. 1976. 'The structure and use of politeness formulas'. *Language in Society* 5: 137-151.

Fernald, A. 1982. *Acoustic determinants of infant preference for 'motherese'.* Unpublished doctoral dissertation, University of Oregon. (Referred to in Kelmert Nelson *et al.* 1989)

Fernald, A. 1984. 'The perceptual and affective salience of mothers' speech to infants'. In Feagans, L., Garvey, C. and Golinkoff, R. (eds.) *The Origins and Growth of Communication.* Norwood, NJ: Ablex.

Fernald, A. and Simon, T. 1984. 'Expanded intonation contours in mothers' speech to newborns'. *Developmental Psychology* 20: 104-13.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies B and Fukui, I. 1989. 'A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants'. *Journal of Child Language* 16(3): 477-501

Fernandez-Ramirez, S. 1951. *Gramática Española.* Madrid: Revista de Occidente.

Fillmore, L. W. 1976. *The Second Time Around: cognitive and social strategies in second language acquisition.* Unpublished dissertation, Stanford University. Quoted in Vihman, M. M. 1982. 'Formulas in first and second language acquisition'. In Menn, L. and Obler, L. (eds.) *Exceptional Language and Linguistics.* NY: New York Academic Press.

Firbas, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication.* Cambridge: Cambridge University Press.

Flege, J. E. 1981. 'The phonological basis of foreign accent: A hypothesis'. *TESOL Quarterly* 15: 443-455.

Flege, J. E. 1985. 'The production and perception of foreign language speech sounds'. In Winitz, H. (ed.), *Human Communication and Its Disorders,* Vol. 1. Norwood: Ablex.

Flege, J. E. 1987a. 'Effects of equivalence classification'. In James and Leather (eds.) 1987a.

Flege, J. E. 1987b. 'A critical period for learning to pronounce foreign languages'. *Applied Linguistics* 8: 162-77.

Flege, J. E. 1997. 'English vowel production by Dutch talkers: more evidence for the 'similar' vs. 'new' distinction'. In James and Leather (eds.) 1997.

Flege, J. E. and Port, R. 1981. 'Cross-language phonetic interference: Arabic and English'. *Language and Speech* 24: 125-146.

Flege, J. E. and Hillenbrand, R. 1984. 'Limits on production accuracy in foreign language speech production'. *Journal of the Acoustical Society of America* 76: 708-21.

Flege, J. E., Bohn, O-S, and Jang, S. 1997. 'Effects of Experience on Non-Native Speakers' Production and Perception of English Vowels'. *Journal of Phonetics* 25: 437-70.

Fox, A. 1973. 'Tone sequences in English'. *Archivum Linguisticum* (New Series) 4: 17-26.

French, P. and Local, J. 1983. 'Turn-competitive incomings'. *Journal of Pragmatics* 7: 17-38.

Fudge, E, C. 1984. *English Word-stress*. London Allen & Unwin.

Fudge, E. C. 1994. 'Natural Phonology'. In Asher, R.E. (ed.) *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon.

Gass, S. 1984. 'A review of interlanguage syntax: language transfer and language universals'. *Language Learning* 34: 115-32.

Gass, S. and Madden, C. (eds.) 1985. *Input in Second Language Learning*. Rowley, Mass.: Newbury House.

Garnica, O. 1977. 'Some prosodic and paralinguistic features of speech to young children'. In Snow and Fergusen (eds.) 1997.

Gatbonton, E. 1978. 'Patterned phonetic variability in second language speech: a gradual diffusion model. *Canadian Modern Language Review* 34: 335-47.

Gaziel, T., Obler, L., Benton, S. and Albert, M. 1977. 'The dynamics of lateralisation in second language learning: sex and proficiency effects'. Paper presented at Boston University. (Referred to in Obler 1981)

Gilbert, J. B. 1984. *Clear Speech*. Cambridge: Cambridge University Press.

Gilbert J. B. 1996. Visual and kinaesthetic aids for teaching pronunciation. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group*. IATEFL.

Giles, H., Coupland, N. and Coupland, J. 1991. 'Accommodation theory: communication, context and consequence'. In Giles, H., Coupland, N. and

Coupland, J. *Contexts of Accommodation. Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Gimson, A. C. 1975. *A Practical Course of English Pronunciation. A Perceptual Approach*. London: Arnold.

Gimson, A. C. 1980 (1962). *An Introduction to the Pronunciation of English* (3rd edition). London: Arnold.

Givón, T. (ed.) 1979. *Syntax and Semantics Volume 12: Discourse and Syntax*. NY: Academic Press.

Gleitman, L. R. and Wanner, E. 1982. 'Language acquisition; the state of the art'. In Gleitman, L. R. and Wanner, E. (eds.) *Language acquisition: the state of the art*. Cambridge: Cambridge University Press.

Goldsmith, J. 1976. *Autosegmental Phonology*. Dissertation, MIT. Bloomington: Indiana University Linguistics Club.

Greenberg, J. H. (ed.) 1978a. *Universals of Human Language* (4 Vols.). Stanford: Stanford University Press.

Greenberg, J. H. 1978b. 'Some generalizations concerning initial and final consonant clusters'. In Greenberg (ed.) 1978a.

Gregg, K. 1984. 'Krashen's Monitor and Ocam's razor'. *Applied Linguistics* 5: 79-100.

Grimes, J. E. 1975. *The Thread of Discourse*. The Hague: Mouton.

Grosser, W. 1993. 'Aspects of intonational L2 acquisition'. In Kettemann and Wieden (eds.) 1993.

Guiora, A., Beit-Hallahmi, B. Brannon, R., Dull, C. and Scovel, T. 1972. 'The effects of experimentally induced changes in ego states on pronunciation ability in a second language: an exploratory study'. *Comprehensive Psychiatry* 13: 421-8.

Gunter, R. 1966. 'On the placement of accent in dialogue'. *Journal of Linguistics* 2: 159-179.

Gussenhoven, C. 1983. Focus, mode and the nucleus'. *Journal of Linguistics* 19: 377-417.

Gussenhoven, C. 1983. 'A semantic analysis of the nuclear tones of English'. In Gussenhoven, C. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, C. 1986. 'The intonation of 'George and Mildred': post-nuclear generalisations'. In Johns-Lewis (ed.) 1986.

Halliday, M. A. K. 1967a. 'Notes on transitivity and theme in English, Part 2. ' *Journal of Linguistics* 3: 199-244.

Halliday, M. A. K. 1967b. *Intonation and Grammar of British English*. The Hague: Mouton.

Halliday, M. A. K. 1970. *A Course in Spoken English*. The Hague: Mouton.

Halliday, M. A. K. 1975. *Learning How to Mean: Explorations in the Development of Language*. London: Arnold.

Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London: Longman.

t'Hart, J., Collier, R. and Cohen, A. 1990. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.

Hatch, E. 1978. 'Discourse analysis and second language acquisition'. In Hatch, E. (ed.) *Second Language Acquisition: a book of readings*. Rowley, Mass.: Newbury House.

Hatch, E. 1979. 'Interaction, input and communication strategies'. Paper presented at the First Nordic Symposium on Interlanguage, Helsinki. (Referred to in Tarone 1980)

Hatch, E., Shapira, R. and Wagner-Gough, J. 1978. 'Foreigner talk discourse'. *ITL Review of Applied Linguistics* 38/40: 39-60.

Henderson Eugenie J. A. (ed.) 1971. *The Indispensable Foundation: A Selection of the Writings of Henry Sweet*. London: Oxford University Press.

Hewings, M. J. 1993. *The English Intonation of Non-native Speakers*. PhD Thesis, University of Birmingham.

Hickey, T. 1993. 'Identifying formulas in first language acquisition'. *Journal of Child Language* 20: 27-41.

Hinofotis, F. A. B. 1976. 'An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language'. Unpublished PhD thesis, University of South Illinois. (Referred to in Oller 1979)

Hirschberg, J. and Ward, G. 1995. 'The interpretation of the high-rise question contour in English'. *Journal of Pragmatics* 24: 407-412.

Hockett, C. F. 1958. *A Course in Modern Linguistics*. NY: Macmillan.

Horowitz, E., Horowitz, M. and Cope, J. 1986. 'Foreign Language Classroom Anxiety'. *Modern Language Journal* 70: 125-32.

Hultzén, L. S. 1956. 'The poet burns again'. *American Speech* 31: 195-201.

Ioup, G. 1984. 'Is there a structure foreign accent? A comparison on syntactic and phonological errors in second language acquisition'. *Language Learning* 34: 1-17.

Jakobson, R. 1968 [1941]. *Child Language, Aphasia and Phonological Universals.* Mouton: The Hague. Translated from the German by A. R. Keiler.

James, A. and Leather, J. (eds.) 1987a. *Sound Patterns in Second Language Acquisition.* Dordrecht: Foris.

James, A. and Leather, J. 1987b. 'Introduction'. In James and Leather 1987a.

James, A. and Leather, J. (eds.) 1997. *Second Language Speech: Structure and Process.* NY: Mouton de Gruyter.

Jamieson, J. 1995. 'The cognitive styles of reflection/impulsivity and field independence/dependence and ESL success'. In Brown, D. H. and Gonzo, S. *Readings on Second Language Acquisition.* Englewood Cliffs, NJ: Prentice Hall Regents.

Jenkins, J. 1995. *Variation in Phonological Error in Interlanguage Talk.* Unpublished doctoral dissertation, University of London.

Jenkins, J. 1997. 'Testing pronunciation in communicative exams'. In Vaughan-Rees (ed.) 1997.

Jenner, B. 1984. *Dutch English: the Phonological and Prosodic Systems of a Non-Native Variety of English.* Unpublished PhD Thesis, University of Lancaster.

Johansson, S. 1978. 'Studies of error gravity'. *Collection Gothenburg Studies on English* 44: 78-118. (Referred to in Lepetit 1989)

Johns-Lewis, C. (ed.) 1986. *Intonation in Discourse.* London: Croom-Helm.

Johnson, J. and Newport, E. 1989. 'Critical effects in second language learning: the influence of maturational state on the acquisition of ESL'. *Cognitive Psychology* 21: 60-99.

Jones, D. 1909. *Intonational Curves.* Berlin: Teubner.

Jones, D. 1956 (1909). *The Pronunciation of English* (4th edition). Cambridge: Cambridge University Press.

Jones, R. H. and Evans, S. 1995. 'Teaching intonation through voice quality'. *ELT Journal* 49 (3): 244-51.

Kellerman, E. 1992. 'Another look at an old classic: Schacter's avoidance'. Lecture notes. Tokyo: Temple University. (Referred to in Ellis 1994)

Kelm, O. R. 1987. 'An acoustic study on the differences of contrastive emphasis between native and non-native Spanish speakers'. *Hispania* 70: 627-33.

Kenworthy, J. 1987. *Teaching Pronunciation*. London: Longman.

Kettemann, B. and Wieden, W. (eds.) 1993. *Current Issues in European Second Language Acquisition Research*. Tubingen: Narr.

Kiriloff, C. 1969. 'On the auditory discrimination of tones in Mandarain'. *Phonetica* 20: 63-67.

Kirk, L. 1973. 'An analysis of speech imitations by Ga children'. *Anthropological Linguistics* 15: 267-75.

Krashen, S. 1973. 'Lateralization, language learning and the critical period: some new evidence'. *Language Learning* 23: 63-74.

Krashen, S. 1977. 'Some issues relating to the Monitor Model'. In Brown, H. Yorio, C. and Crymes, R, (eds.) *On TESOL '77*. Washington, D.C.: TESOL.

Krashen, S. D. 1981a. 'Effective second language acquisition: insights from research'. In Alatis, J., Altman, H. and Alatis, P. (eds.), *The Second Language Classroom: Directions for the 1980's*. New York: Oxford University Press.

Krashen, S. D. 1981b. *Second Language Acquisition and Second Language Learning*. Oxford: Permagon Press.

Krashen, S. 1985. *The Input Hypothesis: Issues and Implications*. NY: Longman.

Krashen, S. D. and Scarcella, R. 1978. 'On routines and patterns in language acquisition and performance'. *Language Learning* 28: 283-300.

Krashen, S. D., Long, M. and Scarcella, R. 1979. 'Age, rate, and eventual attainment in second language acquisition'. *TESOL Quarterly* 13: 573-82.

Krashen, S. D., Scarcella, R. and Long, M. 1982. (eds.) *Child-adult differences in Second Language Acquisition*. Rowley, Mass.: Newbury House.

Krashen, S. and Terrell, T. 1983. *The Natural Approach*. NY: Pergamon.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia, P.A.: University of Pennsylvania Press.

Ladd, R.D. 1980. *The Structure of Intonational Meaning: evidence from English*. Bloomington and London: Indiana University Press.

Ladd, R. D. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lado, R. 1964. *Language Teaching: A Scientific Approach*. NY: MacGraw-Hill.

Lakoff, R. T. 1975. *Language and Woman's Place*. NY : Harper & Row.

Larsen-Freeman, D. and M. Long. 1991. *An Introduction to Second Language Acquisition Research*. London: Longman.

Leather, J. 1987. 'F0 pattern inference in the perceptual acquisition of second language tone.' In James and Leather (eds.) 1987a.

Leather, J. 1997. 'Interaction of perceptual and productive learning in the initial acquisition of second-language tone'. In James and Leather (eds.) 1997.

Leather, J. and James, A. (eds.) *New Sounds '97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*. University of Klagenfurt.

Leben, W. 1976. 'Tones in English Intonation'. *Linguistic Analysis* 2: 69-107.

Lehiste, I. 1975. 'The phonetic structure of paragraphs'. In Cohen, A. and Nooteboom, S. G. (eds.) *Structure and Process in Speech Perception*. Berlin: Springer-Verlag.

Lehiste, I. 1979. 'Perception of Sentence and Paragraph Boundaries'. In Lindblom, B. and Ohman, S. (eds.) *Frontiers of Speech Research*. London: Academic Press.

Lehiste, I. 1980. *Phonetic Characteristics of Discourse*. Tokyo Acoustical Society of Japan.

Lehiste, I. and Fox, R. A. 1992. 'Perception of prominence by Estonian and English listeners'. *Language and Speech* 35: 419-434.

Lehman, C. 1977. 'A re-analysis of givenness: *Stress in discourse*'. Papers from the 13th Regional Meeting, Chicago Linguistics Society, pp. 316-324.

Lenneberg, E. 1967. *Biological Foundations of Language*. NY: Wiley and Sons.

Lepetit, D. 1989. 'Cross-linguistic influence in intonation: French/Japanese and French/English'. *Language Learning* 39: 397-413.

Liberman, M. 1978. *The Intonational System of English*. Dissertation, MIT. Bloomington: Indiana University Linguistics Club.

Li, C. N. and Thompson, S. A. 1977. 'The acquisition of tone in Mandarin-speaking children'. *Journal of Child Language* 4: 185-99.

Lieberman, P. 1967. *Intonation, Perception, and Language*. Cambridge, Mass.: MIT Press.

Lightbown, P. and Spada, N. 1993. *How Languages are Learned*. Oxford: Oxford University Press.

Long, M. 1983. 'Native speaker/non-native speaker conversation and the negotiation of comprehensible input'. *Applied Linguistics* 4: 126-41.

Longacre, R. E. 1979. 'The paragraph as a grammatical unit'. In Givón (ed.) 1979.

Longman. 1995. *Dictionary of Contemporary English* (3rd edition). Harlow, Essex: Longman.

Loveday, L. 1981. 'Pitch, politeness and sexual role: an explanatory investigation into the pitch correlates of English and Japanese politeness formulae'. *Language and Speech* 24: 71-89.

Luthy, M. J. 1983. 'Nonnative speakers' perceptions of English "nonlexical" intonation signals'. *Language Learning* 33: 19-36.

McLaughlin, B. 1978. 'The Monitor Model: some methodological considerations'. *Language Learning* 28: 309-32.

MacIntyre, P. and Gardner, R. 1991. 'Methods and results in the study of foreign language anxiety: a review of the literature'. *Language Learning* 41:25-57.

MacWhinney, B. and Bates, E. 1978. 'Sentential devices for conveying givenness and newness: a cross-cultural developmental study'. *Journal of Verbal Learning and Verbal Behaviour* 17: 539-58.

MacWhinney, B., Pleh, C. and Bates, E. L. 1985. 'The development of sentence interpretation in Hungarian'. *Cognitive Psychology* 17: 178-209.

Major, R. C. 1987. 'The natural phonology of second language acquisition'. In James and Leather 1987a.

Major, R. C. 1997. 'L2 acquisition, L1 loss, and the critical period hypothesis'. In James and Leather 1997.

Marcos, H. 1987. 'Communicative functions of pitch range and pitch direction in infants'. *Journal of Child Language* 14: 255-268.

Martin, Ph. 1981. 'A phonological theory for intonation models of English and French'. In Goyvaerts, D. L. (ed.) *Phonology in the 1980s*. Ghent, Belgium: J. Story-Scientia.

Marton, W. 1981. 'Contrastive analysis in the classroom'. In Fisiak, J. (ed.) *Contrastive Linguistics and the Language Teacher*. Oxford: Permagon.

Mechthild, M. and Hwang, Shu-Fen C. 1991. 'Tone and intonation in Mandarin babytalk to presyllabic infants: comparison with registers of adult conversation and foreign language instruction'. *Applied Psycholinguistics* 12(4): 481-504.

Menyul, P. and Bernholtz, N. 1969. 'Prosodic features and children's language production'. *Quarterly Progress Report No. 93*. MIT Research Laboratory of Electronics. (Referred to in Cruttenden 1979)

Morgan, J. L. 1986. *From Simple Input to Complex Grammar*. Cambridge: MIT Press.

Morley, J. 1979. *Improving Spoken English*. University of Michigan: Ann Arbor.

Morris-Wilson, I. 1997. 'Attitudes of British and Finnish listeners to specific segmental features of mispronunciation in Finnish-accented English'. In Leather and James (eds.) 1997.

Morse, P. A. 1974. 'Infant speech perception: a preliminary model and review of the literature'. In Schiefelbusch, R. L. and Lloyd, L. L. (eds.) *Language Perpectives - Acquisition, Retardation, and Intervention*. London: Macmillan.

Nakazima, S. 1962. 'A comparative study of the speech developments of Japanese and American English in childhood'. *Studia Phonoloqica* 2: 27-39.

Nash, R. 1971. 'Phonemic and prosodic interference and intelligibility'. In Rigault, A. and Charbonneau, R. (eds.) *Proceedings of the 7th International Congress of Phonetic Sciences*. The Hague: Mouton, pp. 570-74.

Nattinger, J. R. and DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Navarro-Tomás, T. 1966 (1948). *Manual de Entonación Española* (6th edition). Colección Málaga, SA: Mexico.

Navarro-Tomás, T. 1961 (1944). *Manual de Pronunciación Española* (10th edition) Madrid: Consejo Superior de Investigaciónes Científicas (CSIC).

Neufeld, G. 1977. 'Language learning ability in adults: a study on the acquisition of prosodic and articulatory features'. *Working papers in Bilingualism* 12: 45-60.

Neufeld, G. 1978. 'On the acquisition of prosodic and articulatory features in adult language learning'. *Canadian Modern language Review* 34: 163-74.

Neufeld, G. 1979. 'Towards a theory of language learning ability'. *Language Learning* 29: 227-41.

Obler, L. 1981. 'Right hemisphere participation in second language acquisition'. In Diller (ed.) 1981.

O'Connor, J. D. 1967. *Better English Pronunciation*. Cambridge: Cambridge University Press.

O'Connor, J. D. and Arnold, G. F. 1973. *Intonation of Colloquial English* (2nd edition). London: Longman.

Oyama, S. 1976. 'A sensitive period in the acquisition of a non-native phonological system'. *Journal of Psycholinguistic Research* 5: 261-85.

Palmer, H. E. 1922a. *Everyday Sentences in Spoken English, with Phonetic Transcriptions and Intonation Marks (for the use of foreign students)*. Cambridge: Heffer.

Palmer, H.E. 1922b. *English Intonation, with Systematic Exercises.* Cambridge: Heffer.

Patowski, M. 1980. 'The sensitive period for the acquisition of syntax in a second language. *Language Learning* 30: 449-72.

Penfield, W. and Roberts, L. 1959. *Speech and Brain Mechanisms.* NY: Atheneum Press.

Pennington, M. C. 1989. 'Teaching pronunciation from the top down'. *RELC Journal* 20: 20-38.

Peters, A. 1977. 'Language learning strategies: does the whole equal the sum of the parts'. *Language* 53: 560-573.

Pica, T. 1988. 'Interlanguage adjustments as an outcome of NS-NNS negotiated interaction'. *Language Learning* 38: 45-73.

Pierrehumbert, J. B. 1980. *The Phonetics and Phonology of Intonation.* PhD Thesis, M.I.T., Published Indiana University Linguistics Club, 1987.

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

Porter, D. and Garvin, S. 1989. 'Attitudes to pronunciation in EFL'. *Speak Out!* 5: 8-15.

Porter, D. and Weir, C. 1997. 'Valid criteria for the assessment of proficiency in pronunciation'. In Vaughan-Rees (ed.) 1997.

Purcell, E. T. and Suter, R. W. 1980. 'Predictors of pronunciation accuracy: a re-examination'. *Language Learning* 30: 271-88.

Quilis, A. 1981. *Fonética Acústica de le Lengua Española.* Madrid (CSIC).

Quilis, A. and Fernandez, J. A. 1990. *Curso de Fonética y Fonología Españolas.* Madrid: CSIC.

Quirk, R., Duckworth, A. P., Svartvik, J., Rusiecki, J. P. L. and Colin, A. J. T. 1964. 'Studies in the correspondence of prosodic to grammatical features in English'. *Proceedings of the Ninth International Congress if Linguistics, Boston, 1962.* The Hague: Mouton.

Quirk, R. and Greenbaum, S. 1973. *A University Grammar of English.* London: Longman.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. and Crystal, D. (Index). 1985. *A Comprehensive Grammar of the English Language.* London Longman.

Raupach, M. 1984. 'Formulae in second language speech production'. In Dechert, H., Möhle, D. and Raupach, M. (eds.) 1984. *Second Language Production.* Tübinger: Gunter Narr.

Reichardt, C. and Cook, T. 1979. 'Beyond qualitative versus qualitative methods'. In Cook, T. and Reichardt, C. (eds.) *Qualitative and Quantitative Methods in Education Research*. Beverly Hills: Sage Publications.

Richards, J. C. 1974. 'A non-contrastive approach to error analysis'. In Richards (ed.) 1974.

Richards, J. C. (ed.) 1974. *Error analysis: Perspectives on Second Language Acquisition*. London: Longman.

Ryan, E. B., Carranza, M. and Moffie, R. 1977. 'Reactions towards varying degree of accentedness in the speech of Spanish-English bilinguals'. *Language and Speech* 20: 267-73.

Sachs, J. 1977. 'The adaptive significance of linguistic input in prelinguistic infants'. In Snow and Fergusen (eds.) 1977.

Sacks, H. and Schegloff, E. A. 1979. 'Two preferences in the organization of references to persons in conversation and their interaction'. In Psathas, G. (ed.) *Everyday Language: Studies in Ethonomethodology*. NY: Irvington.

Sato, C. 1985. 'Task variation in interlanguage phonology'. In Gass and Madden (eds.) 1985.

Schachter, J. 1974. 'An error in error analysis'. *Language Learning* 27: 205-14.

Schmerling, M. 1974. 'A re-examination of 'normal stress''. *Language* 50: 66-73.

Schmerling, S. F. 1976. *Aspects of English Sentence Stress*. Austin: University of Texas Press.

Schubiger, M. 1958. *English Intonation: its Form and Function*. Tubingen: Niemeyer.

Schuetze-Coburn, S., Shapley, M. and Weber, E. G. 1991. 'Units of intonation in discourse: a comparison of acoustic and auditory analysis'. *Language and Speech* 34: 207-34.

Scovel, T. 1988. *A Time to Speak: A Psycholinguistic Inquiry Into the Critical Period for Human Speech*. NY: Newbury House/Harper and Row.

Scuffil, M. 1981. 'Differential interpretation of intonation patterns by learners and native speaker'. In Kühlwein and Raasch, A. (eds.) *Sprache; Lehren-Lernen*. Band 1. Tübingen.

Selinker, L. 1972. 'Interlanguage'. *IRAL* 10: 209-31.

Sharwood-Smith, M. 1981. 'Conscious-raising and the second language learner'. *Applied Linguistics* 2: 159-69.

Slobin, D. I. 1985. 'Cross-linguistic evidence for the language-making capacity'. In Slobin, D. I. (ed.) *The Cross-linguistics Study of Language Acquisition*. Vol. 2. Hillsdale, NJ: Erlnaum.

Snow, C. and Fergusen, C. (eds.) 1977. *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press.

Snow, C. E. and Hoefnagel-Hohle, M. 1977. 'Age differences and the pronunciation of foreign sounds'. *Language and Speech* 20: 357-65.

Solan, L. 1980. 'Contrastive stress and children's interpretation of pronouns'. *Journal of Speech and Hearing Research* 23: 688-98.

Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stampe, D. 1979. *A Dissertation on Natural Phonology*. NY: Garland.

Stockwell, R. P. 1972. 'The role of intonation: Reconsiderations and other considerations'. In Bolinger (ed.) 1972a.

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

Svartvik, J. and Quirk, R. (eds.) 1980. *A Corpus of English Conversation*. Lund: CWK Gleerup.

Sweet, H.1901. *A Primer of English Speech*. Oxford: Oxford University Press.

Swerts, M. and Geluykens, R. 1994. 'Prosody as a marker of information flow in spoken discourse'. *Language and Speech* 37: 21-43.

Tarone, E. 1979. 'Interlanguage as chameleon'. *Language Learning* 29: 181-91.

Tarone, E. 1980. 'Communication strategies, foreigner talk, and repair in interlanguage'. *Language Learning* 30: 417-31.

Tarone, E. 1982. 'Systematicity and attention in interlanguage'. *Language Learning* 32: 69-87.

Tarone, E. 1983. 'On the variability of interlanguage systems'. *Applied Linguistics* 34: 142-63.

Tarone, E. 1985. 'Variability in interlanguage use: a study of style-shifting in morphology and syntax'. *Language Learning* 35: 373-404.

Taylor, G. 1986. 'Errors and explanations'. *Applied Linguistics* 7: 144-166.

Taylor, D. S. 1993. 'Intonation and accent in English: what teachers need to know'. *International Review of Applied Linguistics in Language Teaching (IRAL)* 31: 1-21.

Tench, P. 1990. *The Roles of Intonation in English Discourse*. Frankfurt: Peter Lang.

Tench, P. 1996. *The Intonation Systems of English*. London: Cassell.

Tench, P. 1997. 'Towards a design of a pronunciation test'. In Vaughan-Rees (ed.) 1997.

Thompson, S. 1995. 'Teaching intonation on questions'. *ELT Journal* 49: 235-43.

Tonkova-Yampol'skaya, R. V. 1973. 'Development of speech intonation in infants during the first two years of life'. In Ferguson, C. A. and Slobin, D. I. (eds.) *Studies of Child Language Development*. NY: Holt, Rinehart and Winston.

Trager, G. L. and Smith, L. Jnr. 1951. *Outline of English Structure*. Norman, Okla.: Battenburg Press. (Revised edition, Trager and Smith 1957).

Trager, G. L. and Smith, L. Jnr. 1957. *An Outline of English Structure*. Studies in Linguistics: Occasional Papers 3, American Council of Learned Societies, Washington, D.C.

Trim, J. L. M. 1959. 'Major and minor tone-groups in English'. *Le Maître Phonétique* 112: 26-9.

Tropf, H. 1987. 'Sonority as a variability factor in second language phonology'. In James and Leather (eds.) 1987a.

Trubetzkoy, N. 1939 (1969). *Grundzüge der Phonologie*. Travaux du Cercle Linguistique de Prague 7. Translated and published in English by Baltaxe, C. 1969. Berkeley, CA: University of California Press.

Vanderslice, R. and Ladefoged, P. 1972. 'Binary suprasegmental features and transformational word-accentuation rules'. *Language* 48: 819-38.

van Dijk, T. A. 1977. *Text and Context*. London: Longman.

Vaughan-Rees, M. 1994. *Rhythms and Sounds*. London: Macmillan.

Vaughan-Rees, M. (ed.) 1997. *Testing Pronunciation. Speak Out!* Newsletter of the IATEFL Pronunciation Special Interest Group.

Walsh, T. M. and Diller, K. C. 1981. 'Neurolinguistic considerations on the optimal age for second language learning'. In Diller (ed.) 1981.

Weeks, T. E. 1980. 'Intonation as an early marker of meaning'. In Key, M. E. (eds.) *The Relationship of Verbal and Non-verbal Communication*. Mouton: The Hague.

Weinert, R. 1995. 'The role of formulaic language in second language acquisition: a review'. *Applied Linguistics* 16: 180-205.

Weinreich, U. 1953. *Language in Contact*. NY: Linguistic Circle of New York.

Weir, R. H. 1962. *Language in the Crib*. Mouton: The Hague.

Wells, W. H. G. 1986. 'An experimental approach to the interpretation of focus in spoken English'. In Johns-Lewis (ed.) 1986.

Wennerstrom, A. 1994. 'Intonational meaning in English discourse: a study of non-native speakers'. *Applied Linguistics* 15: 399-420.

White, L. 1987. 'Against comprehensible input: the input hypothesis and the development of second language competence'. *Applied Linguistics* 8: 95-110.

Wieden, W. 1993. 'Aspects of acquisitional stages'. In Kettemann and Wieden (eds.) 1993.

Wieden, W. and Nemser. 1991. *The Pronunciation of English in Austria: A Developmental and Regional Study*. Tübingen, Narr.

Wieman, L. 1976. 'Stress patterns in early child language'. *Journal of child Language* 3: 283-6.

Wilkins, D. A. 1974. *Second Language Learning and Teaching*. London: Arnold.

Wode, M. 1981. *A Definition of Idiom*. Manchester, UK: Centre for Computational Linguistics.

Wolff, H. 1950. 'Partial comparison of the sound systems of English and Puerto-Rican Spanish'. *Language Learning* 3: 38-41.

Woolard, G. 1993. 'Intonation matters'. *Modern English Teacher* 2: 23-43.

Yule, G. 1980a. 'Speakers' topics and major paratones'. *Lingua* 52: 33-47.

Yule, G. 1980b. 'Intonation and givenness in spoken discourse'. *Studies in Language* 4: 271-86.

Zawadski, H. 1994. *In Tempo: An English Pronunciation Course*. National Centre for English Language Teaching and Research, Macquarie University.