Echo words in Tamil

Elinor Keane
Merton College, Oxford

Thesis submitted to the Committee for Comparative Philology and General Linguistics in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Trinity Term
2001

Echo words in Tamil

Elinor Keane
Merton College, Oxford
D.Phil. Thesis, Trinity Term, 2001

Echo words are characteristic of colloquial speech throughout the Indian subcontinent. They exhibit segment-changing reduplication: the introduction of fixed segments to a repeated string or reduplicant. A Tamil example is *maaṭu kiiṭu*, formed from the base *maaṭu* 'cow' and meaning 'cattle in general'. This thesis establishes the phonetic, phonological and morphological properties of such formations in Tamil, and investigates the consequences for theories of reduplication. Chapter 1 surveys developments in theoretical treatments of reduplication in general and segment-changing reduplication in particular. A new 'dual description' analysis is proposed, in which the reduplicant is defined by two descriptions, one the full description of the base, the other a partial description of the reduplicant, which can specify both prosodic structure and fixed segments.

Chapter 2 defines echo words in relation to other types of reduplicated expression, and reviews what is known about their properties in other Indian languages, including results of some original research on Hindi. Chapter 3 reports on the morphological constituency of Tamil echo words, from the responses of native speakers to a questionnaire. This was designed to explore which lexical categories can be echoed, the interaction of echoing with inflection and compounding, and the possibility of echo phrases. Acoustic analysis was used to investigate prosodic structure by considering stress placement and the distribution of different obstruent realizations. Chapter 4 provides the first experimental confirmation of differences in vowel quality and duration that can be associated with stress on an initial syllable, and concludes that the echo words bear a single stress.

The final chapter applies the dual description model to the Tamil data, and discusses how the phonological, morphological and syntactic aspects of echoing can be synthesized. In particular, the significance of phrasal echoing is discussed, not only for theories of reduplication but also the structure of the grammar.

Acknowledgements

Throughout my time as a graduate student I have been fortunate in the guidance and support I have received from my supervisor, John Coleman. He and other members of the Oxford University Phonetics Laboratory have contributed much to my thesis, and certainly to my enjoyment of it. Andrew Slater has provided technical support, with great patience and good humour, and Burton Rosner has supplied statistical advice on several occasions. Help of various kinds and encouragement have come from Paula West, Esther Grabe, Ian Watson and Celia Glyn. I have learnt much from lunch-time discussions in the lab about flying, gardening, cooking, and, occasionally, phonetics. In the wider linguistics community in Oxford I have appreciated advice and general support from various figures, notably Professors Anna Morpurgo Davies and Martin Maiden. The award of a senior scholarship from Merton College has allowed me to work on my thesis in comfort, and amidst beautiful surroundings.

I owe debts of gratitude to two Indian institutions, firstly the Lal Bahadur Shastri National Academy of Administration in Mussoorie. I spent a month there in April 1999, receiving intensive tuition in Tamil, and greatly appreciated the time and effort devoted to me by my teacher, Mr. Nallasamy. During that time I was privileged to stay with Pastor and Mrs. Cornelius, who welcomed me into their home as a surrogate daughter. The bulk of my fieldwork was conducted the following year during a spell as a visitor to the Pondicherry Institute of Linguistics and Culture. The director, Dr. Ramamoorthy, and his staff willingly helped me to organize recording sessions, and allowed me access to the institute's collection of Dravidian journals. I am very grateful to all those who acted as informants and cheerfully submitted to being questioned and recorded, both in Pondicherry and in Oxford.

On a personal level, I am thankful to many at Merton, and at St. Ebbe's church, for their friendship and encouragement. Special mention should perhaps be made of Samuel Kessler, whose involvement with my thesis landed him briefly in an Indian hospital. My fiancé (now husband), Chris, has contributed in all sorts of ways, providing typesetting suggestions, and keeping me company while I have worked, especially during the final stages. Finally, thanks go to my parents, sister and grandparents, for their unfailing love and support.

Contents

Abstract i

Acknowledgements ii

Contents iii

Chapter 1: Introduction 1

1.1. Introduction....1
1.2. Problems of interest....1
1.2.1. Structure of reduplicated expressions....1
1.2.2. Interaction with phonological rules or constraints....2
1.2.3. The problem of transfer....3
1.3. Development of the theory....4
1.3.1. Transformational rules....4
1.3.2. Affixation and parafixation....6
1.3.3. Prosodic templates....10
1.3.4. Full copy....15
1.3.5. Optimality theory....16
1.3.5.1. Generalized alignment....17
1.3.5.2. Generalized templates....18
1.3.5.3. Correspondence theory....19
1.3.5.4. Transfer....23
1.4. Dual description....25
1.5. Segment-changing reduplication....30
1.5.1. Segment-changing in the Indian languages....31
1.5.2. Review of proposals....33
1.5.2.1. Prespecification....33
1.5.2.2. Segmental insertion....34
1.5.2.3. Melodic overwriting....34
1.5.2.4. Default segmentism....35
1.5.2.5. Affixation....36
1.5.2.6. Dual description....43

Chapter 2: Echo expressions in Indian languages 47

2.1. Introduction....47
2.2. Survey of reduplicated forms....47
2.2.1. Points of difference....47
2.2.2. Onomatopoeic expressions....47
2.2.3. Expressives....48
2.2.4. Paired words....50
2.2.5. Complete reduplication....52
2.2.6. Syntactic reduplication....54
2.3. Echo expressions....55
2.3.1. Defining characteristics....55
2.3.2. Semantics....56
2.3.3. History....58
2.3.4. Geographical distribution....59
2.4. Constraints on echo expressions....59
2.4.1. Dravidian....60
2.4.1.1. Which lexical categories can be echoed?....60
2.4.1.2. Morphological constituency of echo bases....60
2.4.1.3. Compounds....62
2.4.1.4. Phrases....62

2.4.2. Hindi and Bengali 63
2.4.2.1. Which lexical categories can be echoed? 63
2.4.2.2. Morphological constituency of echo bases 64
2.4.2.3. Compounds and complex predicates 66
2.4.2.4. Phrases 68
2.4.3. Conclusion 69

Chapter 3: Investigation of constraints on Tamil echo expressions 71

3.1. Introduction 71
3.2. Data collection 71
3.2.1. Subjects 71
3.2.2. Design of the questionnaire 72
3.2.3. Issues addressed by the questionnaire 74
3.2.3.1. Context 74
3.2.3.2. Which lexical categories can be echoed? 75
3.2.3.2.1. Nouns 75
3.2.3.2.2. Pronouns 76
3.2.3.2.3. Verbs 76
3.2.3.2.4. Adjectives and adverbs 77
3.2.3.3. Morphological constituency of echo bases 78
3.2.3.3.1. Case markers 78
3.2.3.3.2. Postpositions 81
3.2.3.3.3. Verbal forms 83
3.2.3.4. Compounds 85
3.2.3.5. Phrases 87
3.2.3.6. Summary 89
3.3. Data analysis 90
3.3.1. Distribution of scores 90
3.3.2. Results 92
3.3.2.1. Context 92
3.3.2.2. Which lexical categories can be echoed? 93
3.3.2.3. Morphological constituency of echo bases 94
3.3.2.3.1. Case markers 94
3.3.2.3.2. Postpositions 96
3.3.2.3.3. Verbal forms 96
3.3.2.4. Compounds 98
3.3.2.5. Phrases 98
3.3.2.6. Elaborated echo expressions 100
3.3.2.6.1. Echo expressions and *ellaam* 100
3.3.2.6.2. Echo expressions and coordination 101
3.4. Conclusion 102

Chapter 4: Tamil stress 106

4.1. Introduction 106
4.2. Overview of the literature 106
4.3. Vowel reduction 110
4.3.1. Data collection 110
4.3.2. Description and analysis of data 112
4.3.2.1. Analysis of /a/ tokens 116
4.3.2.2. Analysis of /i/ tokens 119
4.3.2.3. Analysis of /u/ tokens 121
4.3.3. Conclusions 124
4.4. Is Tamil stress quantity-sensitive? 125
4.4.1. Nature of the system 125
4.4.2. Data analysis 126

4.4.3. Results 128
4.4.4. Stress-timing vs. syllable-timing 130
4.5. Monophthongization of /ai/ 133
4.6. Analysis of echo words 137
4.6.1. Lowering 137
4.6.1.1. Background 137
4.6.1.2. Results 138
4.6.2. Analysis of echo word structure 142
4.6.3. Comparison with compounds 144
4.6.3.1. Background 144
4.6.3.2. Results 146
4.6.3.2.1. Vowel reduction 146
4.6.3.2.2. Monophthongization of /ai/ 146
4.6.3.2.3. Glide deletion 147
4.6.4. Comparison with expressives 151

Chapter 5: Word-initial and word-internal /k/ and /kk/ **153**

5.1. Introduction 153
5.2. Overview of the literature 153
5.2.1. Phonology, phonetics and orthography of the Tamil obstruents 153
5.2.2. Gemination in Tamil 155
5.2.3. Correlates of gemination cross-linguistically 158
5.3. Investigation of word-initial obstruents 159
5.3.1. Data collection 159
5.3.1.1. Subject profile 159
5.3.1.2. Design and presentation of sentences 159
5.3.1.3. Parameters measured 160
5.3.2. Results 160
5.3.3. Conclusion 163
5.4. Investigation of word-internal /k/ and /kk/ 164
5.4.1. Design of data set 164
5.4.2. Results 165
5.5. Echo expressions 167
5.5.1. Results 167
5.5.2. Echo phrases 170
5.5.3. Comparison with expressives 171
5.6. Conclusion 172

Chapter 6: Discussion and conclusions **173**

6.1. Introduction 173
6.2. Discussion of findings 173
6.3. Theoretical implications 177
6.4. Conclusion 179

Appendices **182**

A. Tamil transliteration conventions 182
B. Hindi transliteration conventions 183
C. Glossing conventions 184
D. Questionnaire 185
E. Responses to the questionnaire 192
F. Stress sentences 194
G. Gemination sentences 200

References **205**

Chapter 1: Introduction

1.1. Introduction

Echo words exhibit complete repetition of a base element, typically a lexical item, and the introduction into the repeated string of a fixed segment or segments, a phenomenon known as segment-changing reduplication. Some examples found in English are formed on a pattern borrowed from Yiddish, and are used for pejorative effect, e.g. *party shmarty*. Such expressions are, however, characteristic of colloquial speech throughout the Indian subcontinent, belonging to a set of areal features shared between all four language families. A Tamil example is *paampu kiimpu*, formed from a base noun *paampu* 'snake', and having the generalized meaning 'snakes and other such creatures'. Theoretically they are of particular interest for two reasons: the technical difficulties posed by segment-changing reduplication, and the challenge of establishing the constituency of the base.

Although there is a body of research describing cross-linguistic variation in the choice of fixed segments, little attention has been paid to the morphological or prosodic structure of the echo words, or the possibility of phrasal reduplication. This thesis examines these issues for echo words in Tamil, employing two main methodologies to investigate their properties: evaluation of responses to a questionnaire, and acoustic analysis of recordings made in the field. The data, and their implications for the constituency of the Tamil echo words, are reported in chapters 3 to 5, which form the bulk of the thesis. The rest of the thesis is structured as follows: chapter 1 reviews developments in theories of reduplication, with particular attention to the problem of segment-changing reduplication. Chapter 2 provides some context by placing the echo words within a continuum of different reduplicated structures, and also surveying what is known about echo words in other languages. The final chapter draws together the empirical findings on different aspects of the echo words, and discusses their theoretical significance.

The first part of this chapter takes the form of a general review of theories of reduplication. It is approximately chronological in order, but is structured around three main issues, set out in section 1.2. The first two, the structure of reduplicated forms and their interaction with phonological rules or constraints, feature prominently in all the main theoretical proposals. The third, the issue of transfer, is particularly pertinent to the Tamil echo words: it has received some attention in the literature, but remains problematic. Section 1.3 traces the development of reduplicative theory, seeking to show not only the impact that changing models of phonology have had on theories of reduplication but also the way in which the study of reduplication has influenced those developments. Section 1.4 draws together the best aspects of different theories, and presents a new proposal, dual description. Finally, the theoretical questions posed by segment-changing reduplication are discussed in section 1.5, and the dual description analysis compared with other proposals in accounting for a range of data.

1.2. Problems of interest

1.2.1. Structure of reduplicated expressions

All reduplicated expression contain a string (not necessarily contiguous) that can be identified as repeated segments. This will be referred to as the 'reduplicant', and the remainder of the expression, which can commonly be identified with an unreduplicated counterpart, as the 'base'. In an echo expression such as *puli kili* 'tigers and such animals', for instance, *puli* is the base and *kili* the reduplicant. In this example the first syllable of *kili* is not copied from the base but made up of fixed segmental material, exemplifying segment-changing reduplication (see section 1.5). Such sequences are still classed as part of the reduplicant, even though their segmental content is not determined by the base. One of the main challenges facing a general theory of reduplication is characterizing the reduplicant, which may vary both in size and placement. This is illustrated in (1)–(5): (1) and (2) are

instances of total reduplication in Warlpiri (Nash 1980: 130), where the whole lexical item, regardless of its length, is repeated. (3) and (4) show prefixing partial reduplication in Yidiŋ (Dixon 1977: 156), and (5) suffixing partial reduplication in Dakota (Shaw 1980: 331).

(1)	*kurdu*	'child'	*kurdu-kurdu*	'children'
(2)	*mardukuja*	'female'	*mardukuja-mardukuja*	'females'
(3)	*mulari*	'initiated man'	*mula-mulari*	'initiated men'
(4)	*gindalba*	'(type of) lizard'	*gindal-gindalba*	'lizards'
(5)	*hâska*	'is tall'	*hâska-ska*	'are tall'

Much effort has been devoted to identifying reduplicants with constituents that are independently recognized as a theoretical entity, whether phonological or morphological. This is then employed as some kind of template in many theories, whether it is used to identify the relevant portion of the base to be copied, as a pattern to which copied material must conform, or as a constraint mediating the relationship between reduplicant and base. A related issue, but one that has received less direct attention, concerns the overall structure of the reduplicated expression, including the nature of the boundary between base and reduplicant. This is of particular interest for the Tamil echo expressions, and will be considered in depth in chapters 4 and 5.

1.2.2. Interaction with phonological rules or constraints

Accounting for all the different kinds of interaction between reduplication and phonological alternations poses a considerable challenge to theories of reduplication. A basic typology is to be found in Wilbur's thesis (Wilbur 1973): although her proposals have been largely disregarded and some of her examples since disputed, this remains an important statement of the problem. She sets out three main patterns of interaction, framing them in terms of rules.

A) 'normal' application, involving the conditioning environment for some phonological rule being met in either base or reduplicant, with the rule applying as expected.

B) 'exceptions' to the application of rules, commonly referred to as 'underapplication', where a phonological process fails to apply at all, despite its conditioning environment being met either in the base or the reduplicant. Examples from Madurese are given in (6) and (7), and involve the failure of regressive nasal assimilation to apply at the juncture between base and reduplicant (Wilbur 1973: 18). In cases of total reduplication like these it is generally not possible to establish the relative ordering of base and reduplicant.

(6)	*kun*	'order'	*kun-kun*	'orders'	**kuŋ-kun*
(7)	*baŋ*	'wing'	*baŋ-baŋ*	'wings'	**bam-baŋ*

C) 'overapplication', involving the environment for a rule being met in either base or reduplicant, and yet applying in both. In Javanese, for example, the deletion of an intervocalic /h/ is seen in both base and reduplicant in (8) and (9), although the conditioning environment is met only at the end of the second conjunct, before the *-e* suffix (data from Dudas 1976: 208–209).

(8)	*bədah*	'broken'	*bəda-bəda-e*
(9)	*dajɔh*	'guest'	*dajɔ-dajɔ-e*

Within each of these three categories further subdivision is needed to accommodate all the patterns described in the literature. One dimension of categorization concerns where the conditioning environment is met, whether in the base (B) or the reduplicant (R), at the juncture between them, or in the junctures with the material on either side. These five

possibilities are illustrated schematically in (10) (X and Y denote adjoining phonological material), and the number is doubled if the order of base and reduplicant is reversed.

(10) X + R – B + Y X + B – R + Y
1 2 3 4 5 6 7 8 9 10

Particular attention has focussed on the possibility that a phonological process applying in the reduplicant may be copied back to the base, a situation sometimes referred to as back copying.

Instances that potentially involve the morphological affiliation of particular segments changing have repeatedly attracted attention. An early recognition of this problem is found in Bloomfield (1933: 221–2), and involves nasal substitution in the Austronesian language Tagalog. In derivational terms, this involves a velar nasal, as in the prefix *paŋ-*, assimilating in point of articulation to a following voiceless stop, which is then deleted. Thus the prefixation of *paŋ-* to the noun *puːtul* 'a cut' regularly produces the form *pamuːtul* 'that used for cutting'. When the prefix is found with a reduplicated form, the /m/ is found not only in the preposed reduplicant, as expected, but also in the base, as in (11). (RED denotes the reduplicant, an abbreviation that will be employed at various points in the thesis.)

(11) *paŋ*-RED-*puːtul* *pa-mu-muːtul* 'a cutting in quantity'

Such examples, sometimes referred to as 'overcopy' or skeletal (as opposed to melodic) overapplication (Mester 1986: 181), have functioned as important test cases for different proposals. They pose considerable problems for theories of serial derivation, leading to so-called ordering paradoxes, and have therefore been in the forefront of discussion over where reduplication fits into the overall organization of the grammar, and whether that should involve derivational rules or the satisfaction of constraints.

1.2.3. The problem of transfer

Transfer refers to the preservation in the reduplicant of certain prosodic properties of the base, notably distinctive vowel length (quantitative transfer).[1] Cases of both transfer and non-transfer in reduplication are attested. A pair of simple examples of transfer involving total reduplication are given in (12) and (13): both are verbal adverbs in Hindi and in each the length of the root vowel is faithfully reflected in the reduplicant.

(12) *becte* 'selling' *becte becte* 'while selling'

(13) *kʰaate* 'eating' *kʰaate kʰaate* 'while eating'

Transfer of length is also seen in the partially reduplicated Mokilese progressive forms given in (14) and (15).[2]

(14) *sɔːrɔk* 'tear' *sɔr-sɔːrɔk* 'tearing'

(15) *pɔdok* 'plant' *pɔd-pɔdok* 'planting'

Cases of non-transfer show two kinds of discrepancy between base and reduplicant: a vowel may be long in the base but short in the reduplicant, or short in the base but long in the reduplicant. Examples of the first include the Sanskrit perfect forms given in (16) and (17), where the vowel of the reduplicant is short, whatever the length of the corresponding base vowel (Steriade 1988: 120).

---

[1] The term 'transfer' has also been applied to syllabic transfer (i.e. the tendency for segments to occupy the same syllabic position in base and reduplicant), e.g. by Clements (1985) and Steriade (1988). The explanation for this phenomenon, and for exceptions to it, is frequently bound up with the problem of defining the reduplicant, and will be discussed in relation to it. 'Transfer' as a term will therefore be reserved for quantitative transfer in this thesis.

[2] The Mokilese data have received considerable attention in the literature, including analyses by McCarthy and Prince (1986), Steriade (1988) and Blevins (1996).

(16) *pat* 'fly, fall' *pa-pát-a*

(17) *mna:* 'note' *ma-mná:-u*

The reverse situation is found in the so-called R2 reduplication of Tagalog, where a long vowel is consistently found in the second syllable of the reduplicant, regardless of the length of the corresponding vowel in the base (Carrier-Duncan 1984: 263).

(18) *baluktot* 'crooked' *balu:-baluktot* 'variously bent'

(19) *tahi:mik* 'quiet' *tahi:-tahi:mik* 'rather quiet'

Cases of total reduplication involving non-transfer are very difficult to find: Spring (1990) cites Umpila as such a case, describing the reduplicant as an identical copy of the segments but not the prosody of the base. However, the evidence in her source (Harris & O'Grady 1976) is not strong,[3] and the existence of this category is accordingly debatable.

According to standard autosegmental theory, length, both for long vowels and geminate consonants, is not a property of the segment itself but of its association to two positions, whether these be moras (e.g. Hyman 1985, Hayes 1989) or timing slots. Any theory of reduplication which involves the copying of the segmental melody alone therefore predicts that vowel length cannot be copied across to the reduplicant, since length is not represented on the segmental tier. Cases where vowel length does appear to transfer are accordingly problematic. Mutatis mutandis, theories that assume base and reduplicant match not only in segmental identity but also prosodic structure predict transfer, and the challenge is therefore to explain discrepancies in vowel length between base and reduplicant.

1.3. Development of the theory

1.3.1. Transformational rules

In early generative phonology reduplication was formalized as a rule applying to the base, marking out particular CV segments within it as subject to copying, and so producing the reduplicated output: this assumption is made in various accounts of reduplication from the 1970s (e.g. Munro & Benson 1973, Aronoff 1976). The most detailed formulation of this approach is found in Carrier-Duncan (1984), which explicitly defends the need for transformational rules in response to accounts such as Marantz (1982). She identifies relevant segments or groups of segments within the base by means of numbers, and then stipulates which should appear in the reduplicant, and in which order, according to standard transformational formalism. Her data are taken from Tagalog, for which she formulates three different reduplicative rules, including one labelled R2. This handles examples such as (18) and (19), and is given in (20), where [–seg] refers to a boundary (word or morpheme) and X to any further phonological material.

(20) [–seg] C V C_0 V (C +) X → 1 2 $3_{[+long]}$ 4 2 3 4 5

1 2 3 (4) 5

e.g. *bal u ktot* → *balu:-baluktot*

[3] Harris and O'Grady (1976: 183) quote several examples of progressive verb forms following the pattern:

ku:ngi + l + kungi + manha + ʔa

hide + conjugation marker + RED + present + they 'they are hiding'

However, they comment (p.192) that identification of the phonological rules applying to the progressives awaits work on the overall system of alternations. This certainly seems to leave open the possibility that the absence of vowel length in the reduplicant may have some cause independent of reduplication.

Note that this example allows both for transfer of the vowel length in the first syllable, and non-transfer in the second. The vowel of the reduplicant is specified as long, regardless of the length of the corresponding base vowel, although the mechanism used to achieve this result, the feature [+long], has since been superseded in mainstream phonology by a prosodic representation of length.

The possible interactions of such copying rules with phonological alternations were systematically addressed by Wilbur (1973), who surveyed several approaches to the problems of overapplication and underapplication. One possibility, adopted by Munro and Benson (1973) in their discussion of certain reduplicated adjectives in Luiseño, is the use of exceptional marking. This prevents the adjectives from participating in an alternation between [č] and [š], thereby giving the effect of underapplication. Munro and Benson argue that the [č] is underlying, deriving instances of [š] by means of the rule in (21).

(21) /č/ → [š] / _ {#, [–cont] (/l/ and /r/ considered non-continuant)[4]

Following the application of syncope to the second conjunct, the initial [č] is in the relevant conditioning environment, but fails to undergo the rule, as in examples (22) and (23).

(22) *čára-* 'to tear' *čaráčraš* 'torn' **čarášraš*

(23) *čóka-* 'to limp' *čukáčkaš* 'limping' **čukáškaš*

This was attributed to the introduction of reduplication features as part of the copying rule that applied to these particular adjectives. The phonological rule in (21) could then be assumed to apply after the copying, but would fail to produce the [č]~[š] alternation in these forms because of their exceptional marking.[5]

Another solution to both underapplication and overapplication within a derivational framework involves ordering the pertinent phonological rule prior to reduplication. Thus underapplication will result if the environment for the relevant phonological rule is met only after reduplication has occurred, by which stage it is too late for the rule to apply. This explanation can account for the underapplication of regressive nasal assimilation in the Madurese examples given in (6) and (7) above, and is also employed in Applegate's analysis of the underapplication of *l*-deletion in Chumash (Applegate 1976). Conversely, overapplication occurs if the conditioning environment is met in the unreduplicated form, and the effects of the phonological rule are subsequently copied. An example of this type cited by Wilbur (1973: 26) is the overapplication of palatalization in Dakota:[6] another analysis along similar lines is found in Hollenbach's discussion of the overapplication of tone sandhi in Copala Trique (Hollenbach 1974).

As Wilbur recognized, however, the ordering solution produces the correct results only under certain conditions: in cases of underapplication the conditioning environment must be met only in the reduplicated and not in the unreduplicated form, whereas in overapplication the reverse situation obtains, i.e. the environment must be met in the unreduplicated form. She hypothesized that such counterexamples could be found, and tentatively proposed Serrano and Chukchee as potential cases. Although these particular examples have since been discredited (see McCarthy & Prince 1995: 289), other cases have come to light, and will be discussed in section 1.3.5.3.

---

[4] Note, however, that the rule can be more stated more simply by an analysis in which [š] is underlying, with [č] appearing only before continuants. This is adopted in the analyses of Aronoff (1976: 76–77), Marantz (1982: 461–465), Mester (1986: 220) and McCarthy and Prince (1995: 356), who accordingly analyze Luiseño as an instance of overapplication.

[5] See Davis (1976) for further discussion of these data, arguing that the [č]~[š] alternation rule should also be marked not to apply to two further kinds of reduplicated form.

[6] The Dakota data have been the subject of several different analyses, including Shaw (1980), Sietsema (1988) and Patterson (1988).

Wilbur also discussed the unwelcome theoretical implications for a linear grammar comprising distinct components of ordering phonological rules prior to reduplication, which had been generally regarded as a morphological process (e.g. Sapir 1921: 63–64). One solution open to those maintaining such an ordering was to reclassify reduplication as a late phonological process, which was the approach taken by Applegate (1976) and Hollenbach (1974). Alternatively, a division could be made between the morphological aspects of reduplication and the phonological process of copying, thereby allowing other phonological rules to intervene.

Aronoff (1976: 64), for instance, analyzes reduplication as a WFR (word formation rule) characterized by an abstract marker [+redup]. This triggers copying at one of three points in the phonological derivation: either before or after the other phonological rules, or between the cyclic and word-level rules. Carrier-Duncan likewise adopts a two-stage account, assigning an abstract [+redup] feature in the word-building subcomponent of the lexicon with the copying rule being triggered at a later point, although she limits it to the terminal subcomponent of the lexicon, rather than the phonology proper. This allows affixation to apply between the two stages, and also a restricted class of allomorphy rules. Her account of the Tagalog cases noted by Bloomfield (see 11 above) therefore involves nasal substitution intervening between prefixation of *paŋ-* and the copying rule. The restriction that rules acting in this way must fulfil certain allomorphy-specific criteria represents a welcome attempt to constrain the theory, but not one that can be extended to other instances of overapplication.[7]

The object of containing reduplication within the lexicon is a response to the basic criticism (which Carrier-Duncan fails to resolve convincingly) that transformational rules are excessively powerful. The mechanisms they employ are inherently unconstrained, and so permit possibilities never encountered in natural language, such as the reversal of a sequence of more than two segments. Subsequent proposals have therefore sought to achieve descriptive adequacy by more constrained means.

1.3.2. Affixation and parafixation

The next group of theories extend to reduplication the autosegmental representations developed in studies of tone (Goldsmith 1979), and fruitfully applied to the nonconcatenative morphology of Semitic (e.g. McCarthy 1985). The first proposal of this kind is found in Marantz (1982), which employs CV sequences to define the reduplicant. Marantz's explicit aim is the integration of reduplication with other processes of affixation, the difference being only in the degree of specification of the affix. Unlike other affixes, which are phonologically fully specified, the reduplicant consists only of skeletal CV slots, with the remaining phonological content provided by the base. Marantz envisages this happening through the entire melody of the base being copied, and segments associating to the affixed skeletal slots in accordance with certain stated principles. A simple example is given in (24) to illustrate how these apply: it involves the prefixation of CVC to the Agta noun *takki*, meaning 'leg', to form the plural (Marantz 1982: 447).

(24)

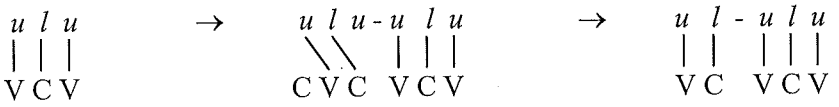
t	*a*	*k*	*k*	*i*	→	*t*	*a*	*k*	*k*	*i*	-	*t*	*a*	*k*	*k*	*i*	→	*t*	*a*	*k*	-	*t*	*a*	*k*	*k*	*i*
|	|	|	|	|		|	|	|				|	|	|	|	|			|	|	|		|	|	|	|	|
C	V	C	C	V		C	V	C				C	V	C	C	V		C	V	C		C	V	C	C	V

Association proceeds on the proviso that [–syllabic] segments can be linked only to C slots, and [+syllabic] segments only to V slots. Links are formed on a one-to-one basis, with any extra segments or CV slots being discarded at the end by stray erasure. The direction of association in this example is left-to-right, which Marantz regards as the unmarked case for prefixing reduplicants. Association is described as *'phoneme-driven'* (Marantz 1982: 447),

[7] There are no grounds, for instance, for believing that the intervocalic /h/ deletion seen in Javanese (see 8 and 9 above) applies only to an arbitrarily marked class of morphemes in the immediate environment of other arbitrarily marked morphemes, which are the defining characteristics of allomorphy rules in Carrier-Duncan's account.

meaning that it works from the copied segments to the skeletal slots, rather than vice versa. This allows for some variation in the composition of the reduplicant, and correctly derives VC- reduplicants for vowel-initial Agta plurals. The noun *ulu* 'head', for example, has the plural *ul-ulu*, not **lu-ulu*, which would result from a slot-driven derivation.

(25)

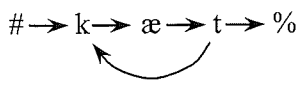


Recognizing that specification of CV slots is inadequate for cases of total reduplication, where the reduplicant is of varying length, Marantz also allows for syllabic and whole morpheme reduplication, although this extension of the theory is not worked out in much detail. He assumes a hierarchical constituent structure in which a morpheme node directly dominates syllables,[8] with single syllables or morphemes being copied, together with all their dependent structure. As Uhrbach (1987: 36) notes, this introduces something of an inconsistency, in that the structure of the base is respected when the reduplicative affix is defined as a constituent, but entirely disregarded when it is defined as a CV sequence.

Several subsequent proposals build on Marantz's basic insight that reduplication involves the association of segments from the base to some kind of skeletal affix, but have the reduplicant affixed in parallel with the base. They therefore order copying after association, rather than before it, and also avoid the need for stray erasure. Clements (1985), for instance, regards the reduplicant as a skeletal parafix introduced on a parallel tier to the base, with association occurring between the two CV tiers. This is followed by transfer of the melody of the base to the associated parafix, and finally sequencing of the reduplicant in relation to the base as either a prefix, infix[9] or suffix. Mester's Single Melody theory (Mester 1986) represents a slight simplification, associating the reduplicative parafix directly with the melody of the base, rather than the skeletal tier. Another proposal in the same vein is found in Uhrbach (1987), although she limits her affixes to prosodic templates (to be discussed in the next section), and so terms her theory 'prosodic parafixation'.

A further step towards representational simplicity is made by Raimy, whose precedence theory (Raimy 2000) is conceptually similar to parafixation theories in some respects, although chronologically much later. He dispenses with skeletal affixes altogether, encoding reduplication simply by an arrow representing a precedence relation. For instance, Raimy uses as his introductory example a hypothetical case of total reduplication 'catcat', represented in (26) (Raimy 2000: 1).

(26)



The hash and per cent symbols mark the beginning and end of the representation respectively. Precedence relations are explicitly encoded by the arrows between segments, and so reduplication is represented simply by the loop linking the last segment to the first. This obviates the need for any relationship between the base and reduplicant beyond that of *'self-identity'* (Raimy 2000: 9), whether association of segments or correspondence constraints. Adopting the framework of Distributed Morphology, which regards morphology and phonology as separate linearly-ordered modules (Halle & Marantz 1993), Raimy identifies reduplication as a morphological *'readjustment rule'* that supplies a complex phonological representation. As in parafixation theories, linearization applies at a later stage of the derivation: Raimy argues that this represents the early application of a process that is required at the interface with phonetics,[10] and is therefore not specific to reduplication.

---

[8] The relationship that Marantz assumes to hold between morphological and prosodic constituents is typical of analyses in the early and mid 1980s, being assumed also by Clements (1985) and Odden and Odden (1985).

[9] For extensive exemplification of infixing reduplication see Broselow and McCarthy (1983).

[10] Using the terminology of minimalist syntax (Chomsky 1995), Raimy sees linearization as the

The potential separation of affixation and association (or the addition of a precedence relation) from the final linearization of reduplicant and base can be exploited by ordering phonological and morphological rules between them to explain effects of over- and underapplication. The possibility of ordering the relevant rule before linearization is available for cases of underapplication where the relevant environment is not met in the synchronous representation. Moreover, a solution is suggested for the more challenging instances of underapplication, where the environment is met in the unreduplicated form: these prove intractable for a normal rule ordering approach. Mester (1986: 225–240) and Uhrbach (1987: 128) propose that the double association of the melody before linearization may block the application of certain rules. They appeal to the Geminate Constraint of Schein and Steriade (1986), arguing that the segments of the synchronous representation form *'non-linear geminates'* (Mester 1986: 225), and so fail to meet the structural description of structure-dependent rules.

Overapplication in theories of parafixation is attributed to the application of a rule to the synchronous representation, with its effect duplicated when linearization occurs. Raimy's approach is along similar lines: he orders rules that show over- or underapplication before linearization, but resorts to stipulation to determine which occurs. Specifically, he proposes a Uniformity Parameter (Raimy 2000: 20), which is given a setting for each phonological rule, according to whether or not its structural description needs to be met in all the environments in which it occurs or just one.

Most of the discussion of overapplication in the parafixation literature focusses on cases of overcopy, such as Tagalog nasal substitution. Clements (1985: 70), for instance, orders nasal substitution after affixation of both the reduplicant and the nasal prefix, but before the transfer of the melody. He assumes without comment that the nasal changes its morphological affiliation from the prefix to the base. A further example of overcopy that has been the subject of some discussion in the literature comes from the Bantu language Kíhehe, and was first presented by Odden and Odden (1985). They propose that the reduplicant coincides with a complex morphological constituent comprising the root, any number of optional suffixes and an affix known as the 'final vowel' (taking the form *-e, -aga, -ite* or *-a*), but no prefixes, as in (27) and (28) (Odden and Odden 1985: 500).

(27)	*kú-ceénga*	'to build'	*kú-ceenga-ceénga*	'to build a bit'
(28)	*kú-ceengéla*	'to build for'	*kú-ceengela-ceengéla*	'to build for a bit'

Overcopying is induced by three rules: the adjunction of a nasal consonant to the onset of a following syllable, glide formation, which applies to any high vowel before another vowel (accompanied by compensatory lengthening of the following vowel), and glide epenthesis. The effect of the latter two rules is illustrated in (29), which is said to derive from underlying /kú-i-eénda/. Glide formation applies to the underlying /u/ with compensatory lengthening of the /i/ vowel, and the epenthetic glide /y/ is inserted to avoid hiatus.

(29)	*kwíiyeénda*	'to love each other'	*kwíiyeendayeénda*	'to love each other a bit'

In the reduplicated form the epenthetic /y/ is repeated, even though it has no morphological affiliation to the reduplicant. Odden and Odden's analysis assumes a view of reduplication similar to Clements', with phonological rules applying after affixation of the morpheme node but before transfer of the dependent prosodic structure. Schlindwein (1986) takes issue with the assumption that morphemes directly dominate syllable structure, and proposes an alternative account in which morphological complexity is represented purely by tier structure. Each morpheme is assigned its own tier and these have to be aligned for a phonological rule to apply across a morphological boundary. Once this has occurred, the division between the relevant morphemes is destroyed and reduplicative copying applies to both together, producing the effect of overcopy. This does not, however, account for (29): for epenthesis of

---

satisfaction of a *'bare output condition'* requiring asymmetrical ordering of phonological representations (Raimy 2000: 15).

/y/ to have occurred the /i/ morpheme of (29) must have come into alignment with the root morpheme *-eenda*. Schlindwein's analysis would therefore predict that the whole sequence *-iiyeenda* should be repeated, yet it is only *-yeenda* that forms the reduplicant.[11]

The linearization process that aligns the reduplicant with the base forms a crucial part of the parafixation proposals, and is the subject of some controversy. Proponents of parafixation identify linearization with tier alignment or conflation, and so claim that it is independently motivated and not reduplication-specific, unlike a copying or transfer rule. This is challenged by McCarthy and Prince (1986: 78), on the grounds that the direction of linearization must be specified. Moreover, they argue that it cannot be independent of the direction of association between parafix and melody if contiguity of base and reduplicant is to be maintained. Information therefore needs to be transmitted from the association stage to the linearization stage, which may be separated by the application of other rules, and so necessitates some kind of global mechanism, according to McCarthy and Prince. Endowing the reduplicative affix with a subcategorization requirement would presumably meet this need, although it is not considered by either Mester or Uhrbach.

If linearization is to be identified with tier conflation and at the same time ordered after phonological processes that follow tier conflation, as Schlindwein argues, then the theory must accommodate multiple applications of tier conflation. This is achieved by integrating reduplication within the framework of Lexical Phonology (e.g. Kiparsky 1982), and identifying tier conflation with the bracket erasure which occurs at the end of each lexical stratum. This model permits the interleaving of phonological and morphological rules that proves problematic for theories in which morphology and phonology are distinct, linearly-ordered components of the grammar, but constrains their interaction by limiting the application of rules to particular lexical levels or strata.

Uhrbach (1987: 272) presents a detailed analysis of this kind for the interaction between prefixation and total reduplication in Indonesian, proposing a lexical component of three strata, with reduplication applying at levels 1 and 2. Unusually, she also devotes considerable attention to the overall structures of the reduplicated forms, comparing them with compounds. She concludes that the two have different structures on the basis of their interaction with stress: compounds form a single domain for stress assignment, whereas reduplicated forms in Indonesian bear two main stresses and are thus analyzed as comprising two (phonological) words, but a single morphological word (Uhrbach 1987: 270).[12] Uhrbach does, however, recognize crosslinguistic variation on this point, mentioning languages such as Bugis in which the application of phonological processes at the boundary between base and reduplicant points to them forming a single phonological word. This line of enquiry will be pursued in later chapters in relation to the Tamil echo expressions, with the aim of establishing their phonological constituency.

Finally, there is explicit recognition of the problem of transfer within theories of affixation and parafixation, and some attempt is made to address the issue. Indeed, Clements takes the problems posed by transfer (both quantitative and syllabic) as the motivation for his theory of parafixation, which is outlined in an article entitled 'The Problem of Transfer in Nonlinear Morphology' (Clements 1985). He assumes the standard autosegmental representation of length as association to two timing slots, and stipulates that copying of the

[11] A third analysis is offered by Aronoff et al. (1987): they apply the analyses of Odden and Odden (1985) and Schlindwein (1986) to a case of overcopying in the Indonesian language Makassarese, concluding that neither is adequate. They contend that reduplication targets prosodic structure but may be sensitive to morphological affiliation.

[12] Note that Uhrbach again assumes that morphological and prosodic constituents form part of a single hierarchy. For further discussion of the interaction between stress and reduplication in Indonesian see Cohn (1989). She presents a more detailed analysis than Uhrbach, noting that the stress pattern of reduplicated forms is affected by the ordering of reduplication relative to affixation (Cohn 1989: 184). An optimality theoretic analysis of Indonesian stress addressing some of the same data is found in Cohn and McCarthy (1998).

base melody to the affix skeleton (which he terms 'transfer') proceeds on the following condition:

> *'Let x be an element on the affix tier, and a(x) its associate on the base tier. If a(x) dominates elements $e_1 \ldots e_n$ on the melody tier of the base ($n \geq 0$), then x dominates a replica of $e_1 \ldots e_n$ on the melody tier of the affix, preserving order and structure.'* (Clements 1985: 51).

A similar approach is taken by Mester (1986), who refers to *'symmetry constraints'* on the association relations between the melody and skeletal tiers to ensure that distinctive vowel length is transferred from base to reduplicant. Uhrbach also discusses the issue, concluding that the transfer of vowel length in a theory that associates the reduplicative parafix directly to the melody of the base is *'still intractable'*. She suggests a stipulation that lexical association lines linking a segment to two skeletal slots in the base be mirrored in the reduplicant, but admits that it is unsatisfactory as a solution.

The opposite issue, that of non-transfer, is not formulated so explicitly, although some cases are considered. Both Marantz and Clements, for instance, comment on the Tagalog R2 pattern (e.g. *tali:-tali:noh*), and specifically the need to avoid a derivation in which the /o/ vowel associates to the final V slot of the affix, i.e. **talio-tali:noh*. Marantz (1982: 451–2, footnote 13) suggests that, exceptionally, doubly linked prosodic structure should be incorporated in the reduplicative affix. Clements uses his condition on transfer to derive the correct result in this case, but since this relies on the doubly linked prosodic structure being present in the base his principles of transfer are unable to derive examples in which the base vowel is short, e.g. *balu:-baluktot*. Uhrbach attributes all instances of non-transfer to the requirements of the reduplicative template, but her account therefore shares the problems associated with prosodic templates and transfer, to be discussed at the end of the next section.

1.3.3. Prosodic templates

The affixes and parafixes of the theories just discussed are variously defined: as a set sequence of CV slots, or as a node in a hierarchy that comprises both prosodic and morphological constituents. The affixes or 'templates' of McCarthy and Prince (1986), however, correspond to independently motivated constituents, and are exclusively prosodic. They argue that this is superior to an analysis involving unstructured sequences of CV segments, on the grounds that no other phonological process counts segments (McCarthy & Prince 1986: 1). Moreover, reduplication can then be classed with other kinds of templatic morphology that make crucial reference to prosodic constituency, and so falls within the purview of prosodic morphology. The fullest exposition of this theory is to be found in *Prosodic Morphology 1986*, a document finally revised in 1996, but circulating in manuscript form throughout the previous decade.

The Prosodic Morphology Hypothesis holds that:

> *'Templates are defined in terms of the authentic units of prosody: mora (μ), syllable (σ), foot (F), prosodic word (W), and so on.'* (McCarthy & Prince 1990: 209).

Reduplicants are thus expected to correspond to different units within the lexical part of the prosodic hierarchy (e.g. Selkirk 1980, Nespor & Vogel 1986). McCarthy and Prince adopt the moraic theory of syllable structure, and therefore distinguish between light, monomoraic (σ_μ) and heavy, bimoraic ($\sigma_{\mu\mu}$) syllables. They also recognize the need for the core syllable CV (σ_c) as a further template type, and admit a minimizing predicate on the grounds that a minimality requirement is always attached to the occurrence of the category 'word' in reduplicative and templatic systems (McCarthy & Prince 1986: 6).

In terms of its empirical coverage, prosodic template theory is well suited to cases where the base and reduplicant show discrepant syllabification since it involves *'prosodic*

reparsing' (McCarthy & Prince 1986: 9). Examples much cited in this connection are given in (30) and (31): they come from the Australian language Lardil and in each case show an /l/ that acts as a syllable onset in the base corresponding to a coda in the reduplicant.

(30)	*pareli*	'gather'	*parel-pareli*	'gather repeatedly'
(31)	*ŋaali*	'thirst'	*ŋaal-ŋaali*	'thirst repeatedly'

McCarthy and Prince analyze the reduplicative template as a minimal word or foot, which is satisfied by two light syllables in (30), and a single heavy syllable in (31). Principled variation in the segmental constituency of the reduplicant is thus predicted: in other theories two separate forms of the affix would be required to accommodate such cases. Further examples are given in (32)–(34), the first repeated from (15) above: these are Mokilese progressive forms in which the reduplicant consistently forms a bimoraic syllable. Since coda consonants are moraic in this language this is satisfied by either CVC or CVV, with lengthening of a short vowel if necessary to fill the template, as in (34).

(32)	*pɔdok*	'plant'	*pɔd-pɔdok*	'planting'
(33)	*kookɔ*	'grind coconut'	*koo-kookɔ*	'grinding coconut'
(34)	*pa*	'weave'	*paa-pa*	'weaving'

The copying and association mechanisms assumed bear more similarity to Marantz's theory than the parafixation proposals that followed it.[13] The template is affixed to the base, and full copy of the base melody is assumed, followed by mapping to the template. Unlike Marantz's theory, however, association is taken to be skeletally-driven, rather than phoneme-driven, and a principle of continuous association is adopted.[14] Like Schlindwein (1986), McCarthy and Prince assume that the reduplicant forms a tier of its own, arguing that this explains the absence of any phonological effects from the unassociated segments without the need for a special melodic erasure principle: rather, the independently motivated operation of tier conflation accounts for their disappearance.

In stipulating that the whole base melody be copied, neither Marantz nor McCarthy and Prince define precisely what constituent they mean by this, although the implication seems to be the morphological word. In some cases, however, a more restricted morphological constituent is involved: for example, the reduplicant in Kíhehe, which was discussed in the previous section, is said to be morphologically complex, containing suffixes but no prefixes. Examples from Hungarian showing reduplication of the verb prefix are cited by Moravcsik (1978: 306), and given in (35) and (36). Since the number of syllables and prosodic weight of the reduplicant clearly vary, this must presumably be analyzed as total reduplication of a specified morphological constituent.

(35)	*el-megy*	'away-goes'	*el-el-megy*	'away-away-goes'
(36)	*bele-néz*	'into-looks'	*bele-bele-néz*	'into-into-looks'

The influence of the morphological constituency of the base on the reduplicant is also noted by Mutaka and Hyman (1990) in their analysis of nominal and verbal reduplication in Kinande. They define the reduplicative template in each case as bisyllabic, but note that the base varies between the morphological word in nouns and the stem in verbs. They also note that in nouns there seems to be a condition (which they dub the Morpheme Integrity Constraint) that the whole of the base melody must be mapped to the template: if the base is

[13] Note that the last section of McCarthy and Prince (1986) explores an alternative, a line-crossing theory of mapping. This shares some properties with theories of parafixation, in proposing that the elements of the base melody are multiply-linked, to both the skeletal tier of the base and of the reduplicative affix, but the tiers are linearly ordered.

[14] Skipping is exceptionally permitted in cases where a complex onset is simplified in association to a core syllable (McCarthy & Prince 1986: 73), although this is recognized to be simply a stipulation.

polysyllabic, no corresponding reduplicated form exists. These different examples suggest that explicit morphological definition of the base may be required.

McCarthy and Prince do, however, discuss instances where the prosodic constituency of the base (rather than the reduplicative template) appears to be relevant. The initial motivation for this came from cases where a reduplicant is infixed: Broselow and McCarthy (1983) present a number of examples in terms of affixation to a prosodic constituent. For instance, the plural in Samoan, where a monomoraic reduplicant appears as either a prefix or infix, is analyzed as consistent prefixation to a (main-stress) foot (Broselow & McCarthy 1983: 30).

(37)	*taa*	'strike'	*ta-taa*
(38)	*nofo*	'sit'	*no-nofo*
(39)	*alofa*	'love'	*a-lo-lofa*

McCarthy and Prince (1990) develop this insight more fully, formalizing a circumscription process that factors the base into a prosodically defined kernel and residue. Moreover, they apply the notion of limiting the base not only to the linear positioning of the reduplicant but also its segmental content. One such example, which contrasts neatly with the reduplicated forms from Lardil in (30) and (31) above, is the pattern of reduplication in Yidiɲ: examples are given in (40) and (41), repeated from (3) and (4) above.

(40)	*mulari*	'initiated man'	*mula-mulari*	'initiated men'
(41)	*gindalba*	'(type of) lizard'	*gindal-gindalba*	'lizards'

Here, as in (30), the reduplicant comprises two syllables, but their make-up is crucially determined by the syllabification of the base, rather than conforming to an independent prosodic template. In this case, McCarthy and Prince (1986: 28) propose that the base is restricted to the first two syllables of the unreduplicated form, and that only these segments are copied and available for association. Lardil and Yidiɲ therefore differ crucially in the definition of their bases: the morphological stem in Lardil and prosodic minimal word in Yidiɲ. For the Lardil examples McCarthy and Prince need to assume some kind of principle of maximal association to explain why the /l/ consonants appear in the reduplicant.[15] According to Wilkinson (1988: 327), only vowels count as moras in Lardil, and so the prosodic template would be satisfied without these weightless coda consonants. Their inclusion would be the natural consequence of segment-driven association, but is difficult to explain on the assumption that association is skeletally-driven. Note also that the prosodic reparsing of the base involved in associating the copied base melody to the affix redundantly recapitulates the prosody of the base in cases where it is prosodically defined.

The possibility of either prosodic or morphological restrictions on the base is also recognized by Spring (1990). She differs from McCarthy and Prince (1986), however, in arguing that there is variation in whether or not the limitations on the base are accompanied by the imposition of a template on the reduplicant. For instance, in the case of Yidiɲ, and also Axininca Campa, the language on which her thesis focusses, she claims that the base is prosodically defined but that there is no independent prosodic specification of the reduplicant, and thus avoids the redundant reparsing of McCarthy and Prince's analysis. Indeed, in Axininca Campa the requirement that the base form a prosodic word is taken not only to limit the segments available for copying, but also to trigger processes of augmentation and epenthesis in the base, which are then reflected exactly in the copied reduplicant. In other cases, a template may be imposed on the reduplicant in addition to prosodic restrictions on the base. Her example for this is taken from Mayo, where the reduplicant is said to vary freely between a monomoraic and bimoraic syllable, as in (42) and (43).

[15] Interestingly, a maximization of association principle is included in McCarthy and Prince's line-crossing proposal (McCarthy & Prince 1986: 81).

(42) *nókwa* *nó-nokwa~nók-nokwa* 'known language'

(43) *nóka* *nó-noka~nón-noka* 'know language'

For the second alternant in each case Spring argues that the base is restricted to a stressed monosyllabic foot, i.e. *nók* in (42) and *nó* in (43). The reduplicant, however, must satisfy a bimoraic syllable template, and so a spreading rule from the onset is required to fill the second mora in (43).

Variation over whether or not a prosodic template is imposed on base or reduplicant is formalized in terms of two independent features: [±prosodic base] and [±affix] respectively. These combine to give a four-way typology, with [+prosodic base], [–affix] exemplified by Yidiɲ and Axininca Campa, and [–prosodic base], [+affix] by Lardil, which has a morphologically defined base. The evidence for the remaining permutations is rather sketchy, although Mayo seems to be an example of [+prosodic base], [+affix]. Instances of [–prosodic base], [–affix] are described as total morphological reduplication by Spring, and would involve perfect copying of the segments but not necessarily the prosody of the base. Since the same parsing principles applied to the same sequence of segments will produce the same results, it would appear that prosodic parsing is simply duplicated in this case. However, Spring does suggest one scenario where reduplicant and base may differ, when the base contains a distinctively long vowel. If length is represented prosodically, the reduplicant will have a corresponding short vowel, and thus non-transfer is predicted in a case of total reduplication, a possibility that McCarthy and Prince (1988: 6) explicitly exclude. Spring claims that this is exemplified by Umpila but, as reported in footnote 3 above, there are doubts about the data. Transfer is predicted in a case like Yidiɲ, where the base is prosodically defined and there is no reduplicative template, but not in cases like Lardil, where the prosody of the base is irrelevant. In examples of [+prosodic base], [+affix], like Mayo, the prosody of the base is copied, but may be subverted by the requirements of the reduplicative template, so either transfer or non-transfer may result. The predictive element of Spring's approach is certainly attractive, but currently lacks vindication by undisputed empirical data.

More generally, prosodic template theories deal well with cases of non-transfer, attributing differences between base and reduplicant to the prosodic shape imposed by the template. The Sanskrit perfect tense examples given in (16) and (17) above, and repeated here as (44) and (45), are thus explained by conformity to a core syllable template.

(44) *pat* 'fly, fall' *pa-pát-a*

(45) *mna:* 'note' *ma-mná:-u*

However, a problem still remains in accounting for cases where length is specified in the reduplicant, since there is no way to differentiate between CVV and CVC (in languages where both are bimoraic) by means of moraic prosodic templates. The same property noted as an advantage for accommodating the Mokilese forms in (32)–(34) above thus becomes a liability in this context. The only way to handle a form like the Tagalog example *balu:-baluktot* is therefore to specify doubly linked association lines from the two moras of the second syllable of the reduplicative template.

Transfer once again remains intractable on the twin assumptions of a prosodic representation of length and copying of the segmental melody alone. Recognizing this, McCarthy and Prince (1986: 79) propose that every aspect of the base's prosodic structure below the constituent specified by the template is copied along with the melody, making transfer the default case. This works well for examples of total reduplication, but problems arise when base and reduplicant show prosodic differences in determining what level of prosodic structure is preserved and under what circumstances. A slightly different approach to this issue is found in McCarthy and Prince (1988), which proposes that melody copying be replaced by the insertion of lexically specified structure. As a result:

> *'. . . all and only the lexically specified properties of the input are available for association (therefore 'transferred') to the output in reduplicative and templatic morphology'* (McCarthy & Prince 1988: 12).

The reduplicant is thus provided with any unpredictable mappings between the skeleton and the segments, such as an underlying mora associated in the lexical representation to a vowel that has distinctive length.[16] The stipulation that lexical specification be inserted rests on the proposal that reduplication is tautologous compounding, and so comprises two lexical entries. McCarthy and Prince (1988: 13) therefore predict that reduplicated forms will show the same phonological properties as any compounding that occurs at the same lexical level of the language. Moreover, they claim that tautologous compounding is responsible not only for total reduplication (the possibility investigated by Uhrbach), but also partial reduplication, since it supplies the segments for association, replacing the copying stage. Indeed, they argue explicitly against identifying partial reduplication with affixation, claiming that such cases may differ systematically from affixed forms, for example, in linear ordering. Despite the advantages for explaining transfer, however, this conception has not been developed in later work.

Analyses bearing on both the issue of transfer and also the structure of the reduplicated form are found in studies of tone, a prosodic property which, like vowel length, is represented on a separate tier from the base melody. At first glance the evidence seems rather mixed, with transfer occurring in some cases but not others, and default or assimilated tones being supplied in some instances. Various approaches have been taken, with the interaction of tone with reduplication being used by some as evidence for its location. Walsh (1992), for instance, attributes the transfer or non-transfer of tone to language-specific parametrization of tone as either a segmental or prosodic property. If tone is a segmental feature, transfer is taken to be obligatory, whereas if it is a property of a mora or a syllable, transfer is optional, depending upon how much prosodic structure is copied. However, by setting a single parameter for each language Walsh's account is unable to explain patterns of transfer and non-transfer coexisting within the same language.

The Bantu languages, which have formed the focus of discussions of tonal transfer, show systematic variability, with nominal reduplication tending to display tonal transfer, but not verbal reduplication. Mutaka and Hyman (1990: 103) associate this with the other differences observed in their comparison of nominal and verbal reduplication in Kinande. They attribute the transfer contrast to the stage at which association of tones occurs: for nouns prelinking of the tones ensures that the tonal melody is available for copying at the relevant stratum of word formation, whereas for verbs tone is apparently associated after reduplication.[17] This is linked with different structural properties: reduplicated verbs are analyzed as affixed, and are taken to form a single tonal domain, whilst nouns are analyzed as stem-stem compounds, with the reduplicant copying the tonal melody of the base.

Overall, therefore, the advances offered by prosodic templates lie primarily in defining the shape of the reduplicant in a constrained manner, and accounting for prosodic differences between base and reduplicant in a principled way. There is no real innovation in terms of representation or copying mechanism that is relevant to interaction with phonological or morphological processes. The literature consequently contains little discussion of this topic, and nothing of note.

---

[16] Note that in the prosodic system advocated by McCarthy and Prince (1988) long vowels are lexically specified with just one underlying mora, becoming associated with a second during syllabification: other approaches specify both underlyingly (e.g. Hyman 1985, Hayes 1989).

[17] Note, however, that Myers and Carleton (1996) claim that Chichewa is a counterexample to Mutaka and Hyman's hypothesis since tone transfers in verbs, even though it is predictable and hence unlikely to be underlying. Moreover, the extent of prelinking is controversial: Pulleyblank (1994), for instance, argues for consistent prelinking throughout Yoruba to account for the full range of tonal patterns observed.

1.3.4. Full copy

This section will discuss the theory of reduplication proposed by Steriade (1988), which shares some properties with theories that preceded it, and prefigures aspects of those that followed, but does not fit neatly into either group. Steriade refers to it as the 'full copy' theory, since she assumes copying of the base with all its attendant prosodic structure in every instance of reduplication, whether total or partial, and does not allow for any kind of restriction upon the base. Total reduplication with transfer of both distinctive vowel length and syllabic position is thus the default case. Any departures from that are attributed to the imposition of various constraints upon the base, including conditions on prosodic weight and syllabic markedness. The introduction of an abstract template, comprising the sum of different constraints, rather than a piece of prosodic structure, clearly anticipates more recent constraint-based analyses. Like them, it is open to the criticism that using disparate constraints to control prosodic shape may fail to capture certain generalizations. It does, however, permit a wide range of syllable types to be characterized by the syllabic markedness parameters. In particular, these may make reference to the onset, a position over which prosodic templates generally have no control.

The syllabic markedness parameters take the form of universal markedness relations, such as closed syllables being more marked than open ones. These are given a setting for each kind of reduplication, either as unmarked, i.e. the less marked option is found (in this case open syllables), or marked, indicating that the more marked option is permitted (in this case closed syllables). Interestingly, Steriade (1988: 80) makes the observation that templates are frequently unmarked in relation to the rest of the language, a point that has been developed extensively by some subsequent theories.[18] Her analysis of Sanskrit perfect tense reduplication (Steriade 1988: 121), which accounts for examples such as (16) and (17) above, illustrates how this works in practice.

(46) Weight parameters: unfootable domain (= light syllable)

Syllabic markedness parameters:

obligatory onset:	marked setting (= onset may be missing)
complex onset:	unmarked setting (= onset may not be complex)
vocoid nucleus:	unmarked setting (= nucleus must be a vocoid)

The weight condition effectively requires that the reduplicant be a light syllable, whilst the syllabic markedness parameters specify that it have a simple onset and a vocoid in the nucleus, criteria which converge on the core syllable CV.

Transformational rules operate to bring the reduplicant into conformity with these conditions, truncating or resyllabifying segments, expanding the base by vowel lengthening, and even inserting segments (see section 1.5). These clearly introduce considerable power into the theory, but Steriade claims that they all correspond to independently motivated processes in other areas of morphology. The formation of hypocoristics in French, for instance, is cited as a parallel to stem truncation and simplification, and ablaut in English strong verbs is analyzed as a case of feature insertion. Unusually, Steriade argues that segments copied from the base which are subsequently discarded by truncation may have phonological effects. She assumes this to be the case in her analysis of Nicobarese (Steriade 1988: 134), and also orders rules of syncope and vocalization between copying and truncation in her account of the Sanskrit perfects. One of her key examples is the perfect form *bu-bódh-a* from the root *baudh-* 'wake': syncope of the /a/ is proposed to explain why the reduplicant is *bu-*, rather than *ba-* (Steriade 1988: 122). Note, however, that the contrary position is held by Mester (1986: 169), and McCarthy and Prince:

> *'copied phonemic melody elements . . . are entirely silent in the phonological derivation'* (McCarthy & Prince 1986: 75).

[18] See references to the emergence of the unmarked in section 1.3.5.2 and the notion of default segmentism in section 1.5.2.4.

The empirical coverage of the full copy theory is broad, and it also accounts for transfer effects in a more principled way than any of the other theories considered thus far. Non-transfer is handled by the imposition of constraints, with vowel shortening resulting from weight restrictions. Cases where length is stipulated in the reduplicant, such as the troublesome *balu:-baluktot* example from Tagalog, can be attributed to a requirement that the second syllable be heavy, combined with a syllabic markedness parameter forbidding codas in the reduplicant. Note, however, that such markedness parameters apply to every syllable of the reduplicant, and are thus unable to handle cases where codas are permitted in the first syllable but not the second.[19] The theory also suffers from certain conceptual drawbacks inherent in derivational theory, notably the assumption of ill-formed representations at intermediate stages of the derivation.

1.3.5. Optimality theory

Constraint-based phonology purports to remedy the deficiencies of analyses involving derivational rules. Optimality theory is the example that has been most widely applied to reduplication, and assumes a two-level system comprising lexical input and output representations.[20] An optimal output candidate is selected from a potentially infinite set of candidates, according to how well it conforms to various well-formedness constraints (see, for example, Prince & Smolensky 1993). These are ranked in relation to one another, and may be either undominated (and thus universally obeyed) or minimally violated. This marks a significant departure from previous approaches, such as Steriade's, which also employed well-formedness conditions but required that they be surface true, i.e. satisfied by every output form.

Optimality theory has been extensively applied to reduplicative data, the first detailed example being McCarthy and Prince's re-evaluation of prosodic morphology in terms of constraint interaction (McCarthy & Prince 1993a). They assume that certain morphemes are marked as reduplicative (RED) in the lexicon and consequently lack phonological content: this is supplied from the base by various copying constraints. Three are given by McCarthy and Prince (1993a: 62),[21] chief among them being MAX(imization), which requires that the reduplicant be identical to the base (R = B). Total reduplication is therefore the default case, and any departures from it are attributable to constraints outranking MAX. These may be either prosodic or morphological, and may be relevant to the language as a whole or specific to reduplication. Moreover, variation in their rankings characterizes different types of reduplication, both cross-linguistically and within the same language.

The worked example provided by McCarthy and Prince (1993a) is reduplication in Axininca Campa. In addition to MAX, they propose three specific constraints on reduplication: DISYLL imposes a prosodic constraint on the reduplicant, requiring that it be disyllabic. It thus behaves like a prosodic template, although it does not in this instance refer to a unit of the prosodic hierarchy. R ≤ ROOT regulates the morphological content of the reduplicant, requiring that it contain only material from the lexical root, and R = SFX supplies the reduplicant with morphological status as a suffix. This last constraint can be dominated in Axininca Campa, according to McCarthy and Prince (1993a: 86), which recognizes both compounds and suffixed forms as possible reduplicated structures in the language. ONSET, a constraint on syllable structure active in the language generally, is also implicated in their analysis, crucially dominating the reduplicative constraints.

---

[19] Stemberger (1996) discusses the general difficulties posed by such cases for theories of reduplication.

[20] Note, however, that intermediate representations are creeping back into optimality theory in McCarthy's sympathy theory (McCarthy 1999).

[21] The other two are CONTIGUITY, which penalizes skipping of segments, and ANCHORING, which requires that base and reduplicant share an edge element, initial in prefixing reduplication and final in suffixing reduplication. This list of constraints is expanded in later versions of the theory.

At this stage of the theory, therefore, a wide range of different constraints on the reduplicant, both prosodic and morphological, are proposed. Indeed, even the earlier requirement that the reduplicant correspond to some member of the prosodic hierarchy has seemingly been abandoned.

1.3.5.1. Generalized alignment

Early versions of optimality theory provide no explicit morphological characterization of the base. It is defined simply as:

> *'the phonological material that immediately precedes the exponent of the suffix morpheme'* (McCarthy & Prince 1993a: 62).

They do, however, allow for prosodic restrictions on the base, and specifically Spring's insight that the base in Axininca Campa corresponds to the prosodic word. This is reflected in the constraint SFX-TO-PRWD (suffix to prosodic word), which specifies the prosodic subcategorization requirements of suffixes in general, including reduplicants. Prosodic circumscription, which controlled what was copied, i.e. the input to reduplication, is thus replaced by a constraint on the output form of the base.

Other data previously handled by prosodic circumscription, involving the variable positioning of certain affixes, are reanalyzed in terms of alignment constraints. One such case, much discussed in the literature, is *-um-* affixation in Tagalog, where *um-* appears as a prefix before a vowel-initial word, as in (47), but as an infix when the base is consonant-initial, as in (48) and (49).[22]

(47)	*aral*	*um-aral*	'teach'
(48)	*sulat*	*s-um-ulat*	'write'
(49)	*gradwet*	*gr-um-adwet*	'graduate'

Since prosodic circumscription isolates whole prosodic constituents, it is incapable of handling cases such as (48) and (49), in which the division between kernel and residue falls within the syllable. Moreover, since the shape of the affix and its placement are independent in the theory of circumscription, it is unable to capture the generalization that affix placement is determined by a constraint against closed syllables. This is reflected directly in the new analysis (McCarthy & Prince 1993b: 102ff.) by ranking prosodic constraints on syllable structure above a morphological constraint specifying prefixal alignment for *um-*, thereby forcing minimal violation (through infixation) when an ill-formed syllable would otherwise result. It thus exemplifies the general schema P >> M, i.e. prosody outranks morphology, which is taken to be the defining characteristic of prosodic morphology in optimality theory.

The theory of generalized alignment extends the notion of alignment, recognizing a whole family of constraints taking the following form:

> *'Align (Cat1, Edge1, Cat2, Edge2)* $=_{def}$ $\forall$*Cat1* $\exists$*Cat2 such that Edge1 of Cat1 and Edge2 of Cat2 coincide, where Cat1, Cat2* $\in$ *PCat* $\cup$ *GCat and Edge1, Edge2* $\in$ *{Right, Left}'* (McCarthy & Prince 1993b: 80).

Alignment may apply to corresponding or opposite edges, and PCat and GCat refer to prosodic and grammatical (i.e. morphological and syntactic) categories respectively, producing a four-way typology. Constraints specifying the alignment of a GCat with another GCat give straightforward morphological concatenation, and those aligning a PCat with another PCat produce effects of prosodic phonology, such as stress assignment. The alignment of PCats with GCats accounts for parsing, whilst aligning GCats with PCats produces prosodic morphology, as in the alignment constraint proposed for Axininca Campa: Align(Affix, L, PrWd, R). This range of possibilities, McCarthy and Prince (1993b: 141)

[22] The Samoan plural forms in (37)–(39) above exemplify a similar phenomenon involving the variable positioning of a reduplicant.

claim, allows a unified treatment of data previously handled by many different mechanisms, including extrametricality, prosodic circumscription, iterative foot-parsing and the phonological cycle.[23] They assume two distinct structural hierarchies: the prosodic hierarchy containing the familiar metrically motivated constituents (PrWd, Ft, σ), and also a morphological hierarchy defined by the rewrite rules in (50).

(50) MWd → Stem* Stem → Stem, Affix Stem → Root.

The distinction between stem and root is not made explicit, and the nature of morphological parsing (MPARSE) is left somewhat vague.[24] The central assumption of coexisting, potentially non-isomorphic prosodic and morphological structures owes much to work by Inkelas (1990), and represents an extension of edge-based theories of the syntax-phonology interface (e.g. Selkirk 1986).

1.3.5.2. Generalized templates

The theory of generalized templates (McCarthy & Prince 1994b), proposes the elimination of any direct prosodic restrictions on the reduplicant. Templates, which were in any case never required for total reduplication, are reduced to a simple identification of the morphological category of the reduplicant, so that:

> *'all that is left of the reduplicative 'template' is an irreducible minimum of morphological specification – no more than would be required for any morpheme – with the apparatus that is specific to reduplication or Prosodic Morphology essentially eliminated'* (McCarthy & Prince 1998: 302).

The shape of the reduplicant is then determined by various phonological properties associated with the relevant morphological category: these interact with the constraints defining the relationship between base and reduplicant. Some variation in prosodic structure is also accommodated by subcategorization constraints, which specify alignment and may, for instance, determine whether a reduplicated affix is integrated into the prosodic word or attached externally. A rather crude conception of morphological theory is assumed (as admitted by the authors – see McCarthy & Prince 1994b: A20), distinguishing only the categories Stem and Affix. These are associated with the prosodic categories PrWd and σ respectively (meaning that they form the most harmonic realization of the relevant morphological category), according to the constraints in (51).

(51) Stem = PrWd Affix ≤ σ

The matching of morphological and prosodic constituents is extended by Urbanczyk (1996: 89) to include the constraint ROOT = Ft, in the context of a systematic application of generalized template theory to reduplication in Lushootseed.

A sample analysis given by McCarthy and Prince (1994b: A6) concerns the reduplicant in Diyari, which is said to form a single prosodic word, bearing its own primary stress independently of the base (Austin 1981: 30). This is attributed to definition of the reduplicant as a stem, which entails that it is subject to the various phonological constraints relevant to the prosodic word. Moreover, since the reduplicant is related only to the base and has no lexical input of its own, it may display greater conformity to these constraints than is seen elsewhere in the language: this is referred to as the emergence of the unmarked (see McCarthy & Prince 1994a). It results from the relatively low ranking of MAX, which may be dominated by a phonological constraint that is itself outranked by faithfulness constraints penalizing deviations from the lexical input level, according to the following schema:

(52) Faithfulness >> Phonological Constraint >> MAX

---

[23] See Cole and Coleman (1992) for a demonstration that parallel constraints can account for cyclic effects.

[24] 'Affix' presumably does not form a level of its own in the hierarchy (cf. degenerate syllables in the prosodic hierarchy), and neither root nor affix can parse the morphological string exhaustively.

In the case of Diyari this means that the reduplicant always takes the form of a minimal prosodic word, comprising exactly two syllables, and will therefore show only partial reduplication of supraminimal bases, as in (53) (Austin 1981: 58).

(53) *kiṇṭala* 'dog' *kiṇṭakiṇṭala* 'doggy, puppy, little dog'

In instances where the reduplicant is defined as an affix, it provides some explanation for the structurally unmarked nature of the reduplicants in relation to lexical morphemes, a property that is shared with affixes in general.

By identifying the reduplicant as a stem or affix, McCarthy and Prince predict that the reduplicated form as a whole is analogous in its morphological structure to either an affixed form or a stem-stem compound. As Inkelas (1999: 179) points out, these reduplicated forms should therefore show the phonological, syntactic and semantic properties associated with affixed forms or compounds, a prediction that largely lacks empirical confirmation. She argues in favour of lexically specified prosodic templates for reduplication, noting that proponents of generalized template theory produce no principled arguments against them.

Whilst generalized template theory defines the reduplicant in exclusively morphological terms, the base, by contrast, is characterized phonologically, as the string adjacent to the reduplicant. This aspect of the theory is one that has been retained in more recent versions: Urbanczyk (1996: 272), for example, pays some attention to this issue, giving a detailed formulation in her Adjacent String Hypothesis. In common with other proposals, she gives no explicit limitation upon the length of this string, but seems to assume that it is bounded by the outermost prosodic word. Various cases raising problems for a purely phonological definition of the base have already been discussed in connection with other theories, e.g. Kíhehe, Kinande and Hungarian.[25] In response to these Urbanczyk admits some morphological influence in the constraint R ≤ Root, requiring that reduplicants be composed of material from the lexical root. This has the corollary that the base should never be restricted to affixes, and indeed Urbanczyk (1996: 275) comments that she knows of no languages exemplifying this possibility. The Hungarian verb forms given in (35) and (36) above, however, constitute a counterexample, suggesting that the theory needs to recognize the relevance of the base's morphological structure more systematically, rather than allowing it in by the back door.

1.3.5.3. Correspondence theory

Parallels between the base-reduplicant and input-output relationships were recognized at an early stage in optimality theory, and reflected by extending the PARSE/FILL constraints that mediated the input-output relationship to that between base and reduplicant. Correspondence theory (McCarthy & Prince 1995) builds on this, but crucially reverses the direction of influence, taking the base-reduplicant correspondence relationship as primary, and generalizing from that not only to the input-output relationship, but also, more recently, to various output-output relationships.[26] The consequence for reduplication is that the segments of the base participate in both a Base-Reduplicant (B-R) Identity constraint and also an Input-Output (I-O) Faithfulness constraint. The reduplicant has no underlying lexical representation of its own,[27] but it does have a direct relationship with the base's underlying representation, monitored by Input-Reduplicant (I-R) Faithfulness. This is distinct from the base input-output relationship labelled Input-Base (I-B) Faithfulness in McCarthy and Prince's Full Model (McCarthy & Prince 1995: 358), which is reproduced in (54).

---

[25] Futagi (1998) discusses the Kinande facts in an optimality theoretic framework, concluding that the reduplicant is subject to a correspondence relation with the root monitored by the constraint R-R Faith, in addition to its relationship with the phonologically defined base. Unlike Urbanczyk's R ≤ Root, this is violated gradiently rather than categorically.

[26] See, amongst others, Benua (1995), Itô, Kitagawa and Mester (1996) and Kenstowicz and Banksira (1999).

[27] With the possible exception of certain cases of fixed segmentism – see section 1.5.2.5.

(54) Input: /Af_{RED} + Stem/

I-R Faithfulness ↙↗ ↓↑ *I-B Faithfulness*

Output: R ⇄ B

B-R Identity

The justification for differentiating the two faithfulness constraints is found in instances where the reduplicant proves more faithful to the input to the base than the base itself (see 59 below). However, McCarthy and Prince provide extensive argumentation for I-B Faithfulness universally outranking I-R Faithfulness, on the grounds that affixes are unmarked relative to roots, regarding it as an instantiation of the universal metaconstraint given in (55).

(55) Root-Faith >> Affix-Faith (McCarthy & Prince 1995: 364).

The different correspondence relations are assumed to be mediated by related families of constraints,[28] the chief examples being the familiar maximization constraint MAX (every segment of string S_1 has a correspondent in S_2), also DEP (every segment of S_2 has a correspondent in S_1) and IDENT(F) (corresponding segments are identical in feature F). The constraints for each correspondence relation are independent (and therefore separably rankable), and are also specific to each reduplicative affix, allowing different patterns of reduplication within the same language.

Correspondence theory has been extensively applied to the full gamut of possible interactions between reduplication and other phonological constraints outlined in section 1.2.2. Indeed, its success in accounting for certain difficult cases has been taken as important proof of the superiority of parallel constraints over rule-based frameworks. The observation on which it builds is that the preservation of identity between base and reduplicant motivates departures from the 'normal application' of other phonological constraints. This is by no means original to McCarthy and Prince: Vennemann (1971: 125) had noted that in such cases differences between the two conjuncts in reduplicated forms are generally minimized, and Wilbur proposed the following Identity Constraint:

> *'There is a tendency to preserve the identity of R_o (i.e. the base) and R_r (i.e. the reduplicant) in reduplicated forms.'* (Wilbur 1973: 58).

She discussed and rejected various implementations before finally proposing a 'mate' relationship between base and reduplicant. Phonological rules were then defined for either overapplication (56) or failure to apply (57), according to the following schema (where X' signifies 'the mate of X').

(56) X (and X') → Y if AXB

(57) X (and X') → Y if X (and X')/A_B

Given cases like Chumash *l*-deletion, which overapplies in some circumstances and underapplies in others, making over- or underapplication part of the structural description of a rule is clearly problematic. More importantly, however, the proposal introduces a global relationship, one that persists throughout the derivation, and was on principle rejected by proponents of rule ordering. In a framework based on the parallel satisfaction of constraints, however, Wilbur's Identity Constraint is straightforwardly subsumed within B-R Identity, and varying application effects are achieved by different rankings relative to I-O Faithfulness and the relevant phonological constraints.

According to the summary typology in McCarthy and Prince (1995: 361), normal application involves domination of both I-R and I-B Faithfulness relations, and also B-R Identity by the phonological constraint (Phono-Constraint), as in (58). If the target of the phonological constraint is the base, then I-R Faithfulness is active, resulting in the reduplicant

[28] A full list is supplied in appendix A of McCarthy and Prince (1995).

being more faithful to the underlying representation of the base than the base is (i.e. the situation that motivates differentiation of the two faithfulness relations).

(58) Phono-Constraint >> I-B Faithfulness >> I-R Faithfulness >> B-R Identity

According to McCarthy and Prince (1995: 359), this is exemplified by prefixing partial reduplication in Klamath, which exhibits syncope or reduction of the first vowel of a prefix or stem, if preceded by at least one syllable. An example is given in (59): here the /o/ vowel of the lexical input surfaces in the reduplicant, but not the base, due to the normal application of syncope.

(59) /RED-*sm'oq'y-dk*/ *sm'o-smq'itk* 'having a mouthful (distributive)'

Both overapplication and underapplication involve the relevant phonological constraint and B-R Identity outranking the faithfulness relations, as in (60).

(60) Phono-Constraint, B-R Identity >> I-B Faithfulness >> I-R Faithfulness

Indeed, underapplication is, in a sense, parasitic upon overapplication, in that the distinction between them is due to the intervention of some additional constraint which specifically excludes the overapplicational candidate. For instance, McCarthy and Prince (1995: 345–348) attribute cases where Chumash *l*-deletion exceptionally underapplies to an undominated constraint on the prosodic weight of the reduplicant. *l*-deletion is described as a general process deleting /l/ before a coronal consonant, and the normal pattern of overapplication is illustrated in (61).

(61) /*s*-RED-*pil-tap*/ *s-pit-pi-tap* 'it is falling in'

Cases of underapplication include (62), where the /l/ of the reduplicant fails to delete, despite preceding a coronal consonant.

(62) /*s*-RED-*tal'ik*/ *š-tal-tal'ik'* 'his wives (i.e. of a chief)'

Examples of this type are attributed to a constraint R = $\sigma_{\mu\mu}$, requiring that the reduplicant form a heavy syllable: this outranks all the constraints given in (60), which retain their own positions relative to one another.

McCarthy and Prince also distinguish bidirectional from asymmetric overapplication: the first results from the constraint ranking given in (60), and applies regardless of where the environment for the phonological constraint is met. The second, given in (63), produces overapplication only if the phonological constraint targets the base, and normal application if the environment is met in the reduplicant.

(63) Phono-Constraint >> I-B Faithfulness >> B-R Identity >> I-R Faithfulness

Are all these predicted patterns actually encountered in natural language, or does the theory overgenerate? Examples of the ranking in (63) are difficult to find: McCarthy and Prince present only one case, the application of nasal substitution reported for Indonesian by Uhrbach (1987), and also discussed by Cohn and McCarthy (1998).[29] The triggering prefix */məN-/* may appear either preposed, as in (64), where the result is overapplication, or interposed between base and reduplicant, as in (65), where there is normal application.

(64) /*məN-potoŋ*-RED/ *məmotoŋ-motoŋ* 'keep cutting'

(65) /*pukul-məN*-RED/ *pukul-məmukul* 'hit each other'

Ranking (60) provides not only for instances where the phonological constraint applies to the base and then has its effects transmitted to the reduplicant by B-R Identity, but also for instances of back or reverse copying. There are two types of case: those in which the constraint is met in the reduplicant and transmitted back to the base, and those in which the reduplicant provides the context for a constraint that is met in the base, and transmitted to the

[29] Compare previous discussion of examples from Tagalog, another Austronesian language.

reduplicant. Uncertainty over the relative ordering of base and reduplicant means that it may not be possible to distinguish the two. One such instance discussed by McCarthy and Prince (1995: 289–294) is the interaction of nasal harmony with total reduplication in Malay (data from Onn 1976: 114). Nasal and oral vocoids are in complementary distribution in the language, with nasality spreading rightwards from a nasal consonant. Consequently the post-nasal vowel in the simple adjective *waŋĩ* 'fragrant' is nasalized, but not the first vowel. In the reduplicated form given in (66), however, all four vowels are nasalized, including the first, although this is not preceded by any nasal.

(66) *waŋĩ* 'fragrant' *wãŋĩ-wãŋĩ* 'fragant (intensified)'

If the first conjunct is the base, nasality spreads from base to reduplicant and the effect is then transmitted back to the base, but if the first conjunct is the reduplicant then nasality is triggered in the base by the reduplicant, and is then transmitted back to the reduplicant.

A further case of back copying recently brought to light involves dissimilation in the Semitic language Chaha (Kenstowicz & Banksira 1999). [x] and [k] are apparently in complementary distribution, with [k] appearing only when a following radical is drawn from the [+continuant] series, i.e. [f, s, z, ʔ]. In cases of total reduplication, however, [k] exceptionally appears before a non-continuant radical, as in (67).

(67) */xt/* 'crush' *kətkit*

The /x/ of the second conjunct is taken to provide the environment for dissimilation, although overapplication results in a [k] in the output in this position. Again the relative ordering of base and reduplicant are unknown, so the identity effect may be working from base to reduplicant or vice versa.

McCarthy and Prince also provide a couple of examples of back copy involving underapplication: the alternation between [w] and [ŋ^{w}] in Southern Paiute (data from Sapir 1930) and the reduction and syncope of vowels in Klamath. In all these cases, back copying is attributed to B-R Identity being ranked not only above the faithfulness constraints but also the relevant phonological constraint. Instances of overcopying, such as the Tagalog nasal substitution in (68) (repeated from 11 above), may also involve back copying, with information being transmitted from the reduplicant back to the base.

(68) *paŋ*-RED-*puːtul* *pa-mu-muːtul* 'a cutting in quantity'

McCarthy and Prince (1995: 310ff.) analyze this as the domination by B-R Identity of MORPHDIS, a constraint requiring that the exponents of distinct morphemes be disjoint. This produces '*a kind of fusion*' (McCarthy & Prince 1995: 309) between the prefix and reduplicant, which is then transmitted to the base. The technical ramifications (e.g. for IDENT(F)) are complex, and overall the theory represents no real advance over parafixation theories in handling these data.

The problems posed by back copy for rule ordering are also not insuperable. In particular, the cases involving underapplication can be straightforwardly handled by ordering the relevant phonological process before reduplication. Overapplication proves more difficult, requiring iterative applications of the copy rule. Moreover, it is possible to construct cases where the application of copying and the phonological rule produce infinite regress; for example, if the reduplicant is postposed in Malay (McCarthy & Prince 1995: 292), or preposed in Chaha (Kenstowicz & Banksira 1999: 580). McCarthy and Prince illustrate this with the oscillating derivation of the Malay example *wãŋĩ-wãŋĩ* given in (69). As noted above, however, there is no evidence that the reduplicant is postposed, and a workable serial derivation can be formulated if the opposite ordering of base and reduplicant is assumed, and the copy step applies repeatedly. The theoretical implications of persistent application are undesirable, but many of these could be avoided by setting the analysis within a lexical phonology framework.

(69)	Underlying form	/*waɲi*-RED/
	Spread Nasal	*waɲĩ*-RED
	Copy	*waɲĩ-waɲĩ*
	Spread Nasal	*waɲĩ-wãɲĩ*
	Copy	*waɲĩ-waɲĩ*
	Spread Nasal	*waɲĩ-wãɲĩ*
	Copy	*waɲĩ-waɲĩ*
	etc. . .	. . . . .

Back copying is specifically predicted by correspondence theory, since B-R Identity allows mutual influence between base and reduplicant. There is, however, an unwelcome consequence of this free bidirectional flow of information in correspondence theory, in that it potentially permits the prosodic structure of the base to be altered in line with that of the reduplicant.[30] Such prosodic back copying is unknown, and could only be ruled out by stipulating an undominated constraint against it. This would run counter to the basic tenet of optimality theory that any constraint is violable, and so presents a serious problem for the theory.

1.3.5.4. Transfer

Although they include little explicit discussion of quantitative transfer, McCarthy and Prince (1993a: 64) recognize the need for MAX to evaluate the prosodic affiliations of candidate reduplicants, and not just the segmental level. They suggest the generalization of the correspondence relation beyond the segment to:

> *'. . . higher-order units of prosodic structure such as moras, syllables, feet, heads of feet, as well as tones and even distinctive features or feature nodes, in support of theories of quantitative transfer, compensatory lengthening, and the effects of floating features.'* (McCarthy & Prince 1995: 14).

However, they do not elaborate further upon how this might be applied in practice, or give any examples involving transfer.

If the Mokilese progressive forms in (70) and (71) (repeated from 14 and 15 above) are analyzed in an optimality theoretic framework, the constraints need to be formulated and ranked in such a way as to mark *sɔ:-sɔ:rɔk* as optimal, and more harmonic than the competing candidate **sɔr-sɔ:rɔk*.

(70)	*sɔ:rɔk*	'tear'	*sɔ:-sɔ:rɔk*	'tearing'
(71)	*pɔdok*	'plant'	*pɔd-pɔdok*	'planting'

In the constraint tableau given in (72) both MAX and DEP constraints are employed, and specified for whether they apply to moras or segments: MAX constraints penalize candidates in which the reduplicant lacks a mora or segment found in the base, and DEP constraints penalize candidates in which the reduplicant contains an extra mora or segment not found in the base (see section 1.3.5.3). Considering firstly correspondence relations between segments without regard to their prosodic affiliations (i.e. constraints e and f in the constraint tableau in 72), the failed candidate would be favoured over the winning form, since it incurs one less violation of DEP(seg). The decision must therefore be made by some other, more highly ranked constraint. Both of the candidates are of equal prosodic weight, and so using MAX(μ) and DEP(μ) (i.e. constraints c and d) simply to count numbers of moras does not distinguish between the candidates. Rather, there must be constraints that do more than just note presence or absence of moraic structure: they must be sensitive to the segmental identity of the units filling the moraic positions. Constraints of this type are included as (a) and (b) in the tableau,

[30] This is sometimes referred to as the Kager-Hamilton problem, e.g. Raimy (2000: 173).

and labelled MAX(μ^) and DEP(μ^), to differentiate them from constraints (c) and (d), which are blind to segmental identity. If the candidates are evaluated using these segment-sensitive constraints, **sɔr-sɔːrɔk* incurs a MAX(μ^) violation, since the /r/ in the reduplicant has no corresponding moraic counterpart in the base, as onset consonants are non-moraic. Furthermore, **sɔr-sɔːrɔk* also incurs an extra DEP(μ^) violation, since the second mora of the long vowel in the base has no correspondent in the reduplicant. Given that both (a) and (b) are violated more by the failed candidate than the optimal one, there are no grounds for ranking one above the other, and so a dotted line is used to indicate that the ordering is not significant. It is, however, crucial that these constraints on moraic correspondence outrank the segmental correspondence constraints, and that they are sensitive to segmental identity if quantitative transfer is to be ensured. More generally, this points to a need for constraints to make reference to associations between the tiers of autosegmental representation, rather than just making pairwise comparisons between individual tiers without regard to their affiliations.

(72)

/RED-*sɔːrɔk*/	(a) MAX(μ^)	(b) DEP(μ^)	(c) MAX(μ)	(d) DEP(μ)	(e) MAX(seg)	(f) DEP(seg)
☞ *sɔr-sɔːrɔk*		**		**		*
sɔr-sɔːrɔk	*!	***!		**		

Recognition of the need for something more than the blind matching of autosegments is found in Urbanczyk's work on Lushootseed, which includes some discussion of the transfer of vowel length. She invokes two types of constraint upon moraic structure – the familiar MAX(μ) constraint, and also the Weight-Ident constraint given in (73).

(73) Weight-Ident (S_1, S_2):
If α is mono(bi)-moraic and β = f(α), then β is mono(bi)-moraic (where α and β are segments belonging to strings S_1 and S_2 respectively). (Urbanczyk 1996: 214).

Moraic association is thus assessed in relation to the segment by this constraint, which is the optimality theoretic counterpart of Clements' transfer condition. Urbanczyk's motivation for maintaining MAX(μ) alongside Weight-Ident comes from an application of the constraint to the input-output relation, where underlying long vowels are realized as [V?], violating Weight-Ident but preserving the number of moras. She thus appreciates the need both for blind counting of moras and also sensitivity to segmental affiliation, commenting:

> '*Having two types of constraint monitor moraic structure has the advantage of treating vowel length like a feature (Weight-Ident) and like an autosegment (MAX-μ).*' (Urbanczyk 1996: 214–215).

Such a conclusion reinforces the points made in the discussion of Mokilese, demonstrating that both aspects of vowel length need to be monitored in an adequate constraint-based account of transfer.

If appropriate formulation and ranking of constraints make transfer the default case, non-transfer is due to domination of B-R Identity by constraints on prosodic structure. In the Sanskrit examples (see 16 and 17 above), for instance, the short vowel of the reduplicant can be attributed to constraints on syllable structure resulting in the emergence of the unmarked core syllable CV. The Tagalog case (see 18 and 19 above) once again proves to be less straightforward: *balu:-baluktot* and **baluk-baluktot* are indistinguishable in terms of number of moras, and the incorrect candidate would actually be more highly ranked by moraic correspondence constraints that are sensitive to segmental identity. Some other, more highly ranked constraint must therefore be responsible. NOCODA, a constraint penalizing the

presence of a coda, is one possibility, accompanied by a highly ranked MAX(μ) to account for the compensatory lengthening. This would, however, be insufficient for cases where the second syllable of the base is monomoraic: if such examples exist, a constraint stipulating double association of the second vowel would be needed instead. In conclusion, optimality theory certainly has the potential to accommodate both transfer and non-transfer, through constraints that monitor not only the features of autosegmental representation but also their geometry.

1.4. Dual description

This section will draw together the salient issues emerging from the various proposals surveyed, highlighting the best aspects of each and also suggesting a novel approach to prosodic constraints on the reduplicant. The focus throughout has been on how, given a certain base, the corresponding reduplicant is to be determined. The complete identity between base and reduplicant seen in total reduplication represents the simplest case, whether that is attributed to perfect copying, or unviolated correspondence constraints. The challenge is therefore to account in a principled manner for discrepancies between base and reduplicant, which may be of various types.

The difference in segmental substance between base and reduplicant seen in partial reduplication has been approached in various ways. One option is to limit the base by imposing a prosodic template, as in the prosodic circumscription theory of McCarthy and Prince (1986, 1990), and also Spring (1990). Alternatively, the base may be indirectly defined through a prosodic subcategorization requirement associated with the reduplicant. Little notice has been paid to the morphological make-up of the base, although the Hungarian pattern of reduplicated verbal prefixes suggests that morphological criteria may characterize the string that is repeated. I propose that the constituency of the base be defined explicitly, in either morphological or prosodic terms, and assume that the two types of structure are copresent, and may diverge (see, for example, Inkelas 1990).

An alternative to limiting the base in cases of partial reduplication is to attribute the difference in the length of the string to a requirement on the reduplicant. This may take the form of prosodic structure to which the melody copied from the base associates, or a constraint favouring candidate reduplicants that conform to the relevant prosodic specification. The absence from the reduplicant of segments in the base is then seen either as the result of stray erasure of unassociated segments, or as violations of the base-reduplicant identity constraint, due to domination by the pertinent prosodic constraint.

The copy-and-association approach is well suited to a second type of discrepancy between base and reduplicant, differences in prosodic structure, such as the syllabification of Lardil. These seem to be the exception, however, rather than the rule, and so the advantage of handling cases where base and reduplicant differ in their prosodic structure is outweighed by the difficulties faced by the copy-and-association theory in accounting for examples where there is no such difference. The transfer of vowel length in languages containing quantitative distinctions is an important subcase, and again seems to be the norm, with cases of non-transfer forming the exception. As has been seen, this is straightforwardly handled by theories such as Steriade's 'full copy' in which not only the melody but also the prosodic structure of the base is copied. The challenge for a theory of this type, however, is accounting for how this structure can be altered to conform with different prosodic specifications. As noted in section 1.3.4, the line pursued by Steriade is the imposition of constraints prompting transformational operations of deletion or resyllabification. Optimality theory, by contrast, invokes no such derivational operations, but uses constraint ranking to favour candidates that conform to the prosodic requirements on the reduplicant over those that correspond more closely to the base.

The root problem in each case is reconciling two different descriptions, which do not coincide exactly and may even conflict. One is the description of the base, which is fully specified for both prosodic structure and segmental identity; the other is a partial description

of the reduplicant. In optimality theory the conflict is modelled by competing constraints, on the nature of the reduplicant versus the relationship between base and reduplicant, and mediated by relative ranking of the relevant constraints. In other constraint-based frameworks, which distinguish between description and object (see, for instance, Bird 1995: 3), the two descriptions can themselves be regarded as constraints on a single object, the reduplicant. Crucially, the description of the reduplicant has priority over that of the base, and will win out where the two conflict, thus acting as a kind of filter.

A formal implementation of this approach, which will be termed dual description, can be given using logical notation to describe prosodic structure. The details of what follows are loosely based on the proposals of Bird (1995), although the results are not dependent on any specific formulation. The representations given here assume mora theory (Hayes 1989), but the general approach is also compatible with onset-rhyme theory. Hierarchical prosodic structure is described in terms of labelled nodes associated with one another by dominance and precedence relations. So, for instance, (74) shows the Sanskrit root *pat* 'fly, fall', comprising a single bimoraic syllable. Each node is referred to by a variable with a unique subscript, and a labelling function can be used to assign labels indicating prosodic constituency to non-terminal nodes, i.e. $L(x_1) = \sigma$, $L(x_2) = \mu$, $L(x_3) = \mu$, and segmental content to terminal nodes, i.e. $L(x_4) = p$, $L(x_5) = a$, $L(x_6) = t$.

(74)

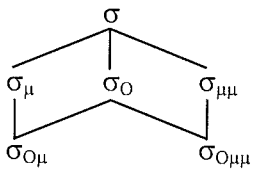
If δ denotes immediate dominance, and $<$ linear precedence then the associations between nodes will be expressed by the following set of relations: $x_1\delta x_2$, $x_1\delta x_3$, $x_1\delta x_4$, $x_2\delta x_5$, $x_3\delta x_6$, $x_2<x_3$, $x_4<x_5$, $x_5<x_6$. Various general constraints will be needed to ensure the well-formedness of the structure, such as autosegmental licensing, and Bird's appropriateness constraint (Bird 1995: 58), which constrains which type of nodes can immediately dominate other nodes, in accordance with the prosodic hierarchy.

The challenge posed by reduplication lies in reconciling the two descriptions of the same object, and formalizing the priority of one description over the other. One approach is to use a default inheritance sytem, in which the description of the reduplicant overrides the default description of the base. Such systems have been extensively applied to natural language lexicons in computational approaches to linguistics (see Gazdar & Daelemans 1992). In the context of reduplication, the priority of the reduplicant description could be implemented in various ways, such as by defining a special logical operator for qualified conjunction. Reiter (1980: 82), for example, proposes an operator $M(\phi)$ 'it is consistent to assume ϕ' to handle defaults. Using this approach, the set of formulae describing the reduplicant would be conjoined with $M(\phi)$, where ϕ is the description of the base.

More detailed conventions are also needed to define exactly how conflicts in the descriptions are resolved. Clearly those nodes occurring in the description of the base which also occur in the reduplicant are licit. Material that is not dominated by those nodes will not appear in the reduplicant, but there is variation over whether or not material that is dominated is included in the reduplicant. Terminal nodes specified in the description of the reduplicant always inherit segmental content from the base where there is any available. Non-terminal nodes, however, may or may not be inherited: in the Yidiŋ examples *mula-mulari* and *gindal-gindalba* (see 3 and 4 above), for instance, the whole of the first foot of the base appears in the reduplicant, regardless of its syllabic structure or segmental content. In Sanskrit, by contrast, the reduplicant is of the form (C)V, and so may contain a discontinuous subpart of the base, as in *ma-mna:-*. In this example, neither the second element of the complex onset nor the distinctive length of the vowel appear in the reduplicant, even though they are dominated by the same syllable node as the material that is inherited.

This distinction can be captured by recognizing a range of different possibilities for each prosodic constituent, and specifying the relevant type or types in the description of the reduplicant. Syllables, for instance, can be classified according to the simplified type hierarchy[31] in (75), in which the unadorned syllable label σ represents the supertype, the union of all possible syllable types. The subscripts represent division according to two criteria, moraic structure and the presence of an onset. Inclusion of the subscript indicates that the syllable is positively specified for that property, so σ_O is required to have an onset, whereas absence of the subscript, as in σ, σ_μ and $\sigma_{\mu\mu}$, covers both onsetless syllables and syllables with onsets. Note that σ_μ and $\sigma_{\mu\mu}$ are mutually exclusive, so σ_μ means that the syllable must contain exactly one mora, not at least one mora.

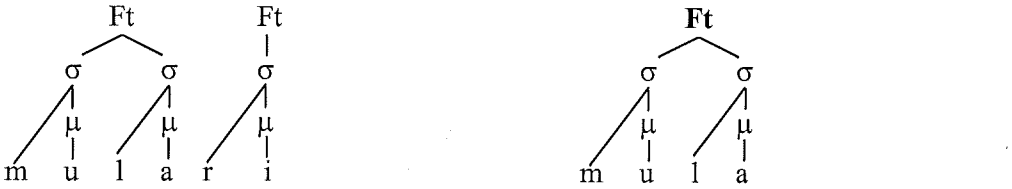
(75)



Further dimensions may be included, such as whether the onset is complex or simple, and whether the syllable node dominates a weightless coda. Similar hierarchies are also to be assumed for other prosodic constituents: the foot, for instance, can be specified as monosyllabic or disyllabic, or in terms of the numbers of moras it dominates, if its syllabicity can vary.

In the case of Yidiɲ the description of the reduplicant consists simply of the unadorned foot node (Ft), i.e. ∃x such that L(x) = Foot. (76) and (77) show how combining this description with that of the base works for the examples *mula-mulari* and *gindal-gindalba*, with the base on the left and the reduplicant on the right. Node labels and any association lines that appear in the descriptions of both base and reduplicant are emboldened.

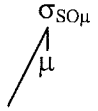
(76)



(77)

These examples, in which everything that is dominated by the first foot node in the base also appears in the reduplicant, contrast with the Sanskrit examples in (79) and (80). Here the reduplicant is defined as a monomoraic syllable with simple onset. This can be specified as $\sigma_{SO\mu}$, corresponding to the representation in (78).

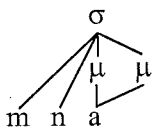
(78)

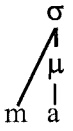


(79)

[31] For other examples of type hierarchies see, for instance, Mastroianni and Carpenter (1994: 16).

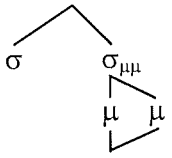
(80)





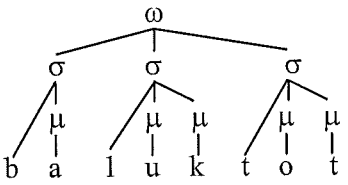
The Sanskrit examples thus illustrate how dual description handles non-transfer in cases where a long vowel in the base corresponds to a short vowel in the reduplicant. The opposite scenario, where a short vowel in the base corresponds to a long vowel in the reduplicant, is exemplified in Tagalog R2 reduplication. As discussed in section 1.2.3, the reduplicant is consistently disyllabic in these examples, with the second vowel specified as long. The description of the reduplicant therefore specifies two syllable nodes, the second dominating two moras: both of these dominate the same terminal node, so that the vowel is doubly linked.

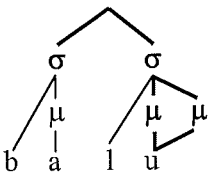
(81)



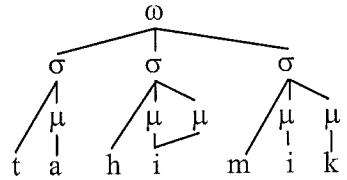
Examples (82) and (83) show the results of combining this description with the description of the base in the cases of *balu:-baluktot* and *tahi:-tahi:mik*, again with the base[32] on the left hand side and the reduplicant on the right. The autosegmental licensing constraint ensures that any segment associated to a second mora in the base, such as the /k/ of *baluktot*, does not appear in the reduplicant.

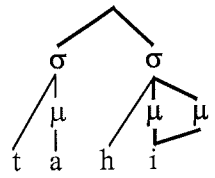
(82)





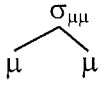
(83)





Cases of non-transfer of distinctive vowel length thus result from the specification of particular moras or terminal nodes in the description of the reduplicant. The principle that these always inherit segmental material whenever it is available is demonstrated in cases where base and reduplicant show disparities in prosodic structure. For instance, the Mokilese progressive example given in (15) involves an onset in the base corresponding to a moraic element in the reduplicant. The description of the reduplicant specifies a single bimoraic syllable $\sigma_{\mu\mu}$, represented in (84).

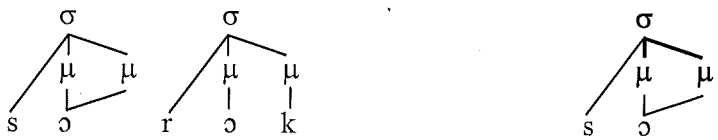
(84)



In contrast to the Tagalog example, where the relations between mora and segment are specified, there is no such specification in this case, allowing variation between a long vowel, as in *sɔ:-sərɔk* in (85), and a vowel-consonant sequence, as in *pɔd-pɔdok* in (86).

[32] A relatively flat structure has been used in the interests of simplification, omitting the foot level: this is not intended to indicate a theoretical commitment to ternary branching.

(85)

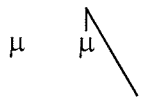


(86)

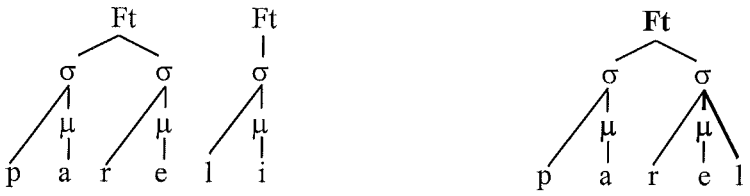
In this last example the /d/ of the reduplicant is recruited from the second syllable of the base, with a new relation of dominance being formed, marked with a dotted line. This shows firstly that segmental material is inherited whenever possible, even if a new relation has to be formed, and also that the whole description of the base is available for inheritance, not just a prosodically delimited subsection. In certain situations there may be no segmental material available for inheritance, and then a terminal node will end up being empty, as in examples of the Sanskrit perfect where the base is vowel-initial, e.g. *u-aúc-a* (> *uvóca* 'please') (Steriade 1988: 122), and so the onset position in the reduplicant cannot be filled.

The Lardil examples *parel-pareli* and *ŋaal-ŋaali* (see 30 and 31 above), which contrast minimally in their make-up with the Yidiŋ cases, also show discrepancies between the syllabification of base and reduplicant, with syllable-final consonants in the base corresponding to onsets in the reduplicant. Moreover, the reduplicants, although consistently bimoraic, vary between one and two syllables in length. The description of the reduplicant therefore contains a foot node, which is specified as bimoraic, i.e. $Ft_{\mu\mu}$, but no reference to nodes at the syllable level. This allows all the material that is dominated by the foot in the base to be inherited, whether it is monosyllabic or disyllabic. The second mora also has to be followed by the specification of a coda consonant, i.e. $Ft_{\mu\mu c}$, as in (87), to account for the /l/ consonants recruited from the next foot of the base in *parel* and *ŋaal*, represented in (88) and (89). The analysis follows Wilkinson (1988: 327), who claims that coda consonants are non-moraic in Lardil: they are therefore represented as immediately dominated by the syllable node, like onsets.

(87) $Ft_{\mu\mu c}$



(88)



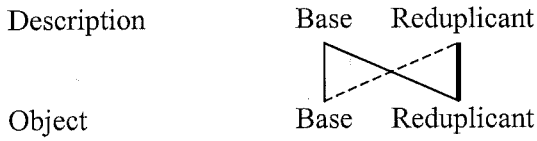
(89)

This brief demonstration suggests that a constraint-based analysis in which descriptions of the base and reduplicant mutually constrain the form of the reduplicant is capable of handling a wide range of data with minimal theoretical apparatus, and overcomes several of the drawbacks inherent in other theories. It offers a natural account of transfer,

since the reduplicant inherits a full description of the base, subject to compatibility with the description of the reduplicant. In the simplest case, therefore, in which the description of the reduplicant contains only the topmost labelled node of the base, total reduplication is the result, preserving all the prosodic structure of the base. Examples of non-transfer result when the description of the reduplicant specifies the necessary configuration. Note that, unlike the theory of prosodic templates, where possible descriptions of the reduplicant are very limited, there is great flexibility in the descriptions permitted. Indeed, since the descriptions are only partial, they need not even conform to the general constraints on well-formedness mentioned above, as long as the object of description, the reduplicant, does so. Nodes from different levels of the prosodic hierarchy may therefore feature in the description, with or without specification of the relations between them. Moreover, relations between moras and the segments that they dominate can be specified, such as the doubly linked structure for distinctively long vowels.

The main respect in which dual description differs from optimality theory is that the constraints involved are the descriptions themselves, and are therefore language-specific. In optimality theory all constraints are universal, and all rankings are potentially instantiated in some language, making wide-ranging predictions that frequently lack empirical validation. As discussed in section 1.3.5.3, this means that optimality theory predicts cases of back copying, where the base is altered in line with the reduplicant. Proponents of optimality theory make much of this, although, as discussed earlier, there are associated problems and cases of back copying are rare. They could be accommodated by the dual description analysis by allowing not only the reduplicant but also the base to have a dual description. (90) illustrates the proposal schematically: descriptions and objects are distinguished, appearing on different lines, but linked where appropriate. The description of the base is therefore linked with the base as an object, and likewise the description of the reduplicant with the reduplicant. The reduplicant as an object is also linked with the description of the base, although the link with its own description has priority, indicated by a thicker line. In cases of back copying a further link is required, from the reduplicant description to the base, and this is shown as a dotted line.

(90)



The analysis could therefore potentially be elaborated to allow the reduplicant to affect the base, whilst maintaining a language-specific explanation that avoids the universal claims made by optimality theory. A further advantage of dual description is that it offers a straightforward account of cases of segment-changing reduplication, the topic of the next section.

1.5. Segment-changing reduplication

Accounting for segment-changing is the main problem posed to phonological theory by echo words. As noted in the introduction, they are characterized phonologically by the appearance of certain fixed segments in the reduplicant, although the nature of these segments varies across the languages of the Indian subcontinent, the source of all the data to be considered here (see section 2.3.4 for a description of the geographical distribution of echo forms). The main patterns of segment-changing will be surveyed in the first subsection, and some of the theoretical questions that they raise introduced. The productivity of different patterns of echo formation varies (see section 2.3.1), but the emphasis in this section will be on the highly productive type investigated in the remainder of the thesis. Section 1.5.2 will then review several proposals for handling this phenomenon, leading into a demonstration of how the dual description analysis of reduplication can be easily extended to accommodate segment-changing data.

1.5.1. Segment-changing in the Indian languages

A good starting point for a review of echo patterns is Trivedi's list in the 1990 report of the All India Linguistic Traits Survey. This project studied a number of shared traits across 160 languages or dialects within the subcontinent, including the occurrence of echo formations. In analyzing these data, Trivedi proposes that the vast majority of cases fall within the following eight basic patterns:

A) C_1X-C_2X where C_2 is a consonant other than C_1, and X represents the rest of the phonological string. This is the most productive type, with a whole range of options listed for C_2: /v/, /b/, /p/, /p^h/, /t/, /ʈ/, /s/, /m/, /h/, /g/, /k/.

B) VX-CVX where C may be any of the segments listed under (A).

C) C_1V_1X-C_2V_2X where V_1 and V_2 are of the same length. Possible C_2V_2 sequences are listed as: /mu/, /muu/, /bi/, /bii/, /gi/, /gii/, /ki/, /kii/.

D) V_1X-CV_2X where CV_2 may be any of the sequences listed under (C).

E) CV_1X-CV_2X where V_1 and V_2 are of the same length.[33] There appear to be two possibilities, found in complementary distribution: V_2 = /u/ or /uu/, or V_2 = /a/ or /aa/.

F) V_1X-V_2X where V_2 = /u/ or /uu/, unless V_1 is /u/ or /uu/; then V_2 = /a/ or /aa/.

G) CV_1X-V_2X where V_2 = /u/ or /uu/, unless V_1 is /u/ or /uu/; then pattern (H) is followed.

H) CVX-VX

Note that, for the most part, these fall into pairs differentiated only by the presence or absence of an initial onset, so this list can be further reduced to five basic patterns.

Patterns (A) and (B) involve a single consonant attaching to the left periphery of the reduplicant, according to the generalized schema $(C_1)X$-C_2X. If the initial onset position is filled, as in the Hindi example in (91), it is replaced by the fixed segment, /v/ in this case.

(91) *k^{h}aanaa* 'food' *k^{h}aanaa vaanaa* 'food and so forth'

The second element of a complex onset, however, may be left intact, as in (92) and (93), further Hindi examples.

(92) *d^{h}yaan* 'attention' *d^{h}yaan vyaan* 'attention etc.'

(93) *traas* 'grief' *traas vraas* 'grief etc.'

According to Ohala (1983: 39), the native vocabulary imposes strict sequential constraints, allowing only glides (i.e. /v/ and /y/) to be the second element of a complex onset (hence 93 must be a borrowing). Where the onset is complex, /v/ simply replaces the first segment, as in (92), and the same pattern is followed in (93), since /vr/ is also admissible. In some other loanwords, however, simply replacing the first segment with /v/ would result in a phonotactically ill-formed cluster, and in these cases the whole complex onset is replaced.

(94) *pleʈ* 'plate' *pleʈ veʈ* 'plates and so forth'

(95) *klaas* 'class' *klaas vaas* 'class etc.'

In cases such as (96) the cluster produced by attaching the fixed segment before an existing onset is acceptable (see 92), but does not appear in the echo form.

(96) *yaad* 'memory' *yaad vaad* 'memory and so forth'

This suggests that the fixed segment is in some way parasitic, preferring to take over an existing prosodic position. If there is no initial consonant, however, the fixed segment will appear in the previously empty onset position, as in (97).

[33] Note, however, exceptions to this generalization about vowel length reported in Singh (1969) and discussed below in section 1.5.2.3.

(97) *aam* 'mangoes' *aam vaam* 'mangoes etc.'

Patterns (C) and (D) form a parallel pair to (A) and (B), differentiated only by the presence or absence of an initial consonant, and can therefore be represented by the generalized schema $(C_1)V_1X\text{-}C_2V_2X$. The *ki(i)/gi(i)-* pattern is characteristic of the Dravidian languages, and its occurrence in Tamil will be investigated at length. As in the cases of transfer of distinctive vowel length discussed in section 1.2.3, the length of the vowel in the reduplicant is dependent on the vowel in the base. This is illustrated in examples (98) and (99), taken from the Dravidian language Kolami (Emeneau 1955: 101).

(98) *puvul* 'flowers' *puvul givul* 'flowers etc.'

(99) *maasur* 'men' *maasur giisur* 'men etc.'

The doubly-linked prosodic structure of a long vowel thus seems to be maintained, whilst its segmental identity is changed. Furthermore, it does not appear to be the case that the fixed vowel simply takes over two moras wherever they are available, since consonants that form a second mora are preserved in the reduplicant. Amongst the limited set of segments that occur in this position in Tamil are nasals which are homorganic to the following onset consonant, as in the following example given by Asher (1985: 207).

(100) *campaḷam kimpaḷam* 'salary and the like (e.g. perks, bribes)'

The evidence on whether or not the fixed segments replace the second element of a diphthong in Tamil is rather inconclusive. The orthographic system contains two characters which may be transliterated as <ai> and <au>. The second of these is very infrequent, being largely restricted to a small number of borrowings from Sanskrit, but the first has a wide distribution, and is represented by a single symbol: <ஐ> word-initially and <ை> elsewhere. It is sometimes described as a sequence of vowel and glide, rather than a diphthong,[34] and this is reflected by an alternative transliteration as <ay>. In echo forms, /ai/ is replaced by a long vowel, as in (101).

(101) *pai* 'bag' *pai kii* 'bag etc.'

This supports a bimoraic representation of /ai/:[35] were it to be a monomoraic diphthong, *pai ki* would be the expected echo form. This contrasts with the Munda language Gtaʔ, to be discussed at greater length in section 1.5.2.5, where diphthongs are consistently replaced by a short vowel (Mahapatra 1976: 822). This points to a monomoraic representation, and ties in with independent arguments for the diphthongs being short (see Zide 1976). In Tamil it is unclear whether both moras are replaced by fixed segments, or just the /a/, with the original final /i/ being preserved. The data are certainly consistent with the fixed segments replacing whatever is in the first moraic position, and so appearing as a long vowel only if that is already doubly linked.

Patterns (E) and (F) again form a pair, differentiated only by the presence or absence of an initial consonant, and can be represented as $(C)V_1X\text{-}(C)V_2X$. Since the fixed segment is simply a vowel, with its length again determined by the base, it is internal to the reduplicant in pattern (E), where there is an initial consonant. This is exemplified by the Hindi examples (102)–(105), taken from Singh (1969: 192). The distribution of /aa/ and /uu/ is determined by the identity of the vowel in the first syllable of the base, /uu/ appearing when the first vowel is /aa/, and /aa/ elsewhere, in accordance with the common pattern of suppletive alternation (see

[34] See, for instance, Rajaram (1980), Christdas (1988) and Vasanthakumari (1989), none of whom include diphthongs in their vowel inventories. For further discussion and acoustic analysis of /ai/ in initial and non-initial syllables, see section 4.5.

[35] Languages appear to differ over whether their diphthongs pattern with long or short vowels. See Lehiste (1985), for example, for arguments based on evidence from a word game that the diphthongs of Estonian behave like long vowels.

section 1.5.2.5 below). Note that in (103) not only the length of the vowel but also the presence of nasality is determined by the base.

(102)	*saamaan*	'luggage'	*saamaan suumaan*	'luggage etc.'
(103)	*kʰr̃ĩcnaa*	'to pull'	*kʰr̃ĩcnaa kʰããcnaa*	'to pull and so forth'
(104)	*caaʈnaa*	'to lick'	*caaʈnaa cuuʈnaa*	'to lick and so forth'
(105)	*dʰuum*	'pomp'	*dʰuum dʰaam*	'pomp and such'

Finally, patterns (G) and (H) share the apparent deletion of an initial consonant, differing in whether or not this is accompanied by the introduction of a fixed vowel. In (106) and (107) below, the reduplicant is suffixing, the normal pattern for Hindi, but (108)–(110), also from Hindi, all have a prefixed reduplicant.

(106)	*somvaar*	'Monday'	*somvaar omvaar*	'Monday etc.'
(107)	*vaayu*	'air'	*vaayu aayu*	'wind, climate etc.'
(108)	*paas*	'near'	*aas paas*	'near about'
(109)	*gird*	'around'	*ird gird*	'all about, around'
(110)	*saamne*	'in front of'	*aamne saamne*	'face to face'

This pattern is not particularly common in Hindi, or productive synchronically, although the semantic connections between (108)–(110) suggest that it may have been productive at one time within a restricted section of the lexicon. Furthermore, small numbers of examples, again frequently with the reduplicant preceding the base, are cited for a number of other Indian languages. Gnanasundaram (1972: 254), for instance, gives the examples in (111) and (112) for Tamil.[36]

(111)	*pakkam*	'side'	*akkam pakkam*	'this way and that way'
(112)	*picaku*	'error'	*icaku picaku*	'awkward predicament'

Taken cumulatively, such examples suggest that a pattern of this kind may be productive elsewhere, and that some account of subtractive reduplication is needed.

1.5.2. Review of proposals

This section surveys, in roughly chronological order, the various theoretical proposals that have been made for handling the kind of data just described. These are inevitably bound up with general theories of reduplication and have therefore followed the development from skeletal slots to prosodic templates, and from rule-based derivations to constraint-based frameworks.

1.5.2.1. Prespecification

Marantz (1982) regards segment-changing reduplication as the result of prespecification of distinctive features, which are attached to the skeletal slots that form the reduplicative affix. These take precedence over the features of the base melody, either overriding them completely or leaving the segment free to attach to the next slot (Marantz does not adjudicate between these two possibilities). Moreover, Marantz allows for variability in the number of features specified, thereby covering not only cases where a whole segment or sequence of segments is fixed, but also instances where a single feature seems to be prespecified: he gives the example of [+high] to explain the forms in (113) and (114) from Akan.

(113) *seʔ* 'say' *siseʔ* (multiple activity form of the verb)

[36] Arunachalam (1977: 4) provides another pair of examples for Tamil, and Hettiaratchi (1959: 50) presents a small group of similar cases from Sinhalese.

(114) *sɔʔ* 'light' *susɔʔ*

Another instance where a single feature is regarded as prespecified is found in Myers and Carleton's analysis of reduplicated nouns in Chichewa. In some instances a high tone appears on the reduplicant even when there is none on the base, and this is described as:

> *'a kind of prespecification, analogous to prespecification of the template for invariant C/V features.'* (Myers & Carleton 1996: 58).

Marantz's examples are limited to cases where the reduplicant corresponds to a fixed number of CV slots: he does not explain how the analysis could be extended to cases such as the echo expressions where his theory would require whole morpheme reduplication. However, the basic strategy of specifying distinctive features as part of the reduplicant will be picked up and developed in the dual description analysis of section 1.5.2.6.

1.5.2.2. Segmental insertion

The analysis proposed by Steriade (1988) proceeds by copying the whole word and then altering it as necessary (see section 1.3.4), thereby avoiding the disjunction between total and partial reduplication. Fixed segments are introduced by segmental insertion, one of the operations that modifies the reduplicant. The locus is taken to be the existing syllable, with elimination of segments whenever a phonotactically ill-formed cluster is produced. Apart from the undesirability of having an ill-formed representation at an intermediate stage of the derivation, this approach fails to account satisfactorily for the Hindi example *yaad vaad* 'memory and so forth', given in (96) above. If phonotactic ill-formedness is what motivates deletion, why does the /y/ not appear in the reduplicant of (96), given that /vy/ is clearly a well-formed initial cluster, as demonstrated by (92)? Steriade's account needs to allow deletion as a result of substitution: the fixed segment replaces the /v/ rather than simply being inserted alongside it.

A further disadvantage of segmental insertion, noted by Yip in her extensive comparison of insertion and overwriting (Yip 1992: 470), is that the theory does not constrain the combinations of segments that can be inserted. Therefore it is theoretically no more costly to insert a consonant into the onset and a consonant into the rhyme than it is to insert a consonant into the onset and a vowel into the rhyme, even though the former situation is not attested, whereas the second is relatively common.

1.5.2.3. Melodic overwriting

Melodic overwriting was first devised by McCarthy and Prince (1990) as a response to the problems encountered by prespecification in cases involving total reduplication. The fixed segment(s) are taken to be purely melodic and to occupy their own autosegmental plane (an explicit parallel is drawn with the component morphemes of root-and-pattern morphology). Association with the reduplicant then proceeds:

> *'in a "feature-changing" manner, overwriting the original melodic material of the base'* (McCarthy & Prince 1990: 245).

McCarthy and Prince state explicitly that no prosodic structure can be contained in the fixed material, and that it is the base that provides the prosodic positions to which the melody anchors. Transfer of vowel length is therefore predicted, since a vocalic melody is expected to take over the moraic association of the segment being overwritten, be that double or single. Cases where distinctive vowel length is not transferred are explicitly ruled out, on the grounds that templates, not simple melodic sequences, are needed to impose invariant prosody. Indeed, McCarthy and Prince state:

> *'we predict the non-existence of an echo-word system that takes arbitrarily long input and that specifies both the quality and the quantity of some segment in the output (e.g. an echo-word system with kota → kota-giita and koota → koota-giita or one with kota → kota-gita and koota → koota-gita).'* (McCarthy & Prince 1990: 245).

This follows logically from their assumptions about overwriting, but is it supported by the empirical evidence? One pattern of vowel-changing reported for Hindi appears to provide a counterexample. The pattern is described by Singh (1969: 192) as *'very common'*, and is given schematically as CVX CV_1X, where V_1 is either /uu/ or /aa/. Examples containing long vowels in the base were given in (102–105) above, and these show straightforward transfer of vowel length, with the overwriting vowel adopting the double moraic association of the vowel it replaces. In the examples in (115)–(117), however, the first vowel of the base is short, and yet the overwriting vowel is still long.

(115)	*moɽnaa*	'to fold'	*moɽnaa maaɽnaa*	'to fold and so forth'
(116)	*p^huʈ*	'lonely'	*p^huʈ p^haaʈ*	'lonely and so forth'
(117)	*cup*	'quiet'	*cup caap*	'quiet etc.'

This is precisely the situation predicted not to occur, and poses considerable, if not insuperable, problems for any analysis in which fixed segments are assumed to be purely melodic, and vowel length is represented autosegmentally. In such examples it would appear that the overwriting vowel is prespecified as long, suggesting that the overwriting string may need to contain prosodic structure. As with the subtractive examples given in (106)–(112), a caveat should be made concerning the scarcity of examples, and their probable unproductivity. Again, however, what appears as a marginal and possibly fossilized pattern in Hindi may be productive elsewhere, and the serious challenge it poses to the theory cannot be disregarded.

Another type of overwriting string explicitly predicted not to occur is the specification of a moraic coda consonant (McCarthy & Prince 1990: 245). Counterexamples to this are to be found in Yip's extended comparison of Steriade's insertion analysis with McCarthy and Prince's overwriting model (Yip 1992). She regards the replacement of a coda with a fixed segment as a relatively rare phenomenon. Nevertheless, if such data is to be comprehended by the overwriting model at all, it will require some modification, so that a single overwriting consonant can form either an onset or a coda.

1.5.2.4. Default segmentism

A rather different, but not necessarily competing approach, has grown out of the realization that fixed segments often act as defaults in the language as a whole, notably by featuring in epenthesis. In Yoruba, for instance, one of the languages discussed by Marantz, gerundives are formed by partial reduplication, with the insertion of a fixed /i/, as in (118) and (119).

(118)	*lo*	'go'	*lí-lọ̀*	'going'
(119)	*dùn*	'to be tasty, sweet'	*dí-dùn*	'being sweet'

Pulleyblank (1988) presents a whole series of arguments for the default status of /i/ in this language, arguing that it is underlyingly totally unspecified for place features. His analysis was challenged by Clements and Sonaiya (1990), and has since been altered by Pulleyblank himself (Pulleyblank 1998).[37] More generally, an analysis in which fixed segments are underspecified poses difficulties for prespecification and overwriting approaches to segment-changing reduplication. In both cases the default segment would be underspecified and yet somehow resist being filled by the specific features of the base melody, eventually receiving the default features by rule.

[37] Clements and Sonaiya (1990: 98) object to Pulleyblank's radical underspecification on the grounds that certain lexical rules in Yoruba refer to /i/, and cannot refer explicitly to the absence of a feature. Broe (1993: 229–231) discusses the Yoruba data in terms of structured specification, proposing that rules vary according to whether they hold of all features or only marked ones. Pulleyblank (1998) argues that the Yoruba data are best analyzed by an optimality theoretic constraint set governing faithfulness to vocalic specifications and based on considerations of sonority.

A rather different approach is taken in optimality theoretic analyses of similar phenomena, which regard the appearance of a default segment as an instance of the emergence of the unmarked (McCarthy & Prince 1994a, 1994b, 1995 – see section 1.3.5.2). In terms of constraint ranking, the key domination relation is as in (120), where Phono-Constraint refers to an unspecified phonological constraint favouring unmarked segmental structure (McCarthy & Prince 1995: 329).

(120) I-O Faithfulness >> Phono-Constraint >> B-R Faithfulness.

The fact that I-O Faithfulness outranks the constraint determining segmental markedness means that faithfulness to the underlying lexical representation will be maintained, even at the expense of incurring markedness violations. Hence marked segments and structures will appear in the phonology of the language, as long as they are present in the lexical input. However, in two situations I-O Faithfulness is irrelevant, allowing Phono-Constraint to become active. One is reduplication, since I-O Faithfulness holds only between the base and its input and hence the reduplicant is free of its demands.[38] It will therefore conform to the markedness constraint, even if this means that identity between base and reduplicant is sacrificed. The other situation is epenthesis, when the segment has no correspondent in the underlying lexical representation and so has its featural identity determined by the language's segmental markedness constraints.

An example analyzed in some detail by Urbanczyk (1996) is diminutive reduplication in Lushootseed. The data encountered here present considerable difficulties for either prespecification or segmental insertion, since the 'fixed' segment, an /i/ vowel, as in Yoruba, does not appear consistently in all reduplicated forms; in some, a segment identical to its correspondent in the base appears instead, as in examples (121)–(125) (Urbanczyk 1996: 206).

(121)	*təláw-il*	'run'	*títəlaw'-il*	'jog'
(122)	*s-du:k^w*	'knife'	*s-dídu:k^w*	'small knife'
(123)	*č'λ'á?*	'rock'	*číč'λ'á?*	'little rock'
(124)	*čáləs*	'hand'	*čáčaləs*	'little hand'
(125)	*hiw-il*	'go ahead'	*hihiwil*	'go ahead a bit'

The generalization made by Urbanczyk is that the fixed segment appears only when the root contains a schwa, as in (121), a long vowel, as in (122), or begins with a consonant cluster, as in (123). The common denominator uniting these cases is that they all contain phonologically marked structure[39] and so incur violations of the constraints militating against markedness in the language. Identity between base and reduplicant thus appears to be subordinated to considerations of markedness, as in the constraint ranking given in (120).

1.5.2.5. Affixation

Whilst the approach described above accounts for cases where the fixed segment coincides with a default, it is plainly inadequate for cases like English *table-schmable* (following the pattern borrowed from Yiddish), where the sequence *schm-* is highly marked in the language as a whole. In such cases overwriting provides a far more satisfactory analysis. Consequently Alderete and his fellow researchers (Alderete et al. 1997) adopt both approaches in their survey of fixed segmentism, and attempt to demarcate their respective domains. In so doing, they appeal to the phonology-morphology distinction, characterizing overwriting as morphological and default segmentism as phonological. They regard the

[38] See section 1.5.2.5, however, for a suggestion that overwriting strings and even tones are subject to an input-output correspondence relation of their own.

[39] The schwa is not of itself a marked segment, but in the reduplicant it would be stressed, and a stressed schwa is marked.

overwriting string or segment as an affixal morpheme, but realized simultaneously with the base. In the framework of correspondence theory, this means that the reduplicant is related both to the base by the B-R correspondence constraints, and to the underlying lexical input of the fixed string, as illustrated in (126). This is a significant elaboration of McCarthy and Prince's Basic Model (McCarthy & Prince 1995: 327),[40] set out in (127) as a comparison, in that it introduces a new correspondence relation, that between the fixed affix and the reduplicant. This is a distinct input-output relation from that holding between the stem and the base, and subscripts are accordingly used in (126) to differentiate them.

(126) Input: /Af_{FIXED} / /Af_{RED} + Stem/

I_R-O_R Faithfulness ↘ ↕ *I_B-O_B Faithfulness*

Output: R ⇄ B

B-R Identity

(127) Input: /Af_{RED} + Stem/

↕ *I-O Faithfulness*

Output: R ⇄ B

B-R Identity

In all cases where overwriting occurs and the fixed affix surfaces, the I_R-O_R Faithfulness constraints outrank B-R Identity. The tonal data from Chichewa mentioned in section 1.5.2.1 can be used as an example of how this works out in practice. When a high tone surfaces without a correspondent in the base, B-R Identity is violated, and this is attributed to its being outranked by a constraint requiring faithfulness between the reduplicant and its lexical input, the high tone. This is referred to as MAX(I-O) (Myers & Carleton 1996: 59), and would constitute an I_R-O_R Faithfulness constraint in the model illustrated in (126). In the situation where the base already has a high tone, Myers and Carleton argue that the lexically specified tone is able to satisfy both this constraint and that requiring identity between base and reduplicant simultaneously.

Since all the correspondence relations in (126) are two-way, this model potentially allows for the fixed segments to be introduced via the reduplicant into the base. This represents an extreme case of back copying, and one that does not seem to have been reported for any language. In order for this to be avoided, I_B-O_B Faithfulness would need to be ranked above B-R Identity in all cases of fixed segmentism. Since it is a tenet of optimality theory that all constraint rankings are possible in principle, such a stipulation runs counter to the spirit of the theory. Moreover, it makes the prediction that cases of fixed segmentism never show over- or underapplication effects, since these crucially involve B-R Identity outranking faithfulness relations (see section 1.3.5.3). It is not clear whether this prediction is borne out, although there are no obvious counterexamples among the Indian echo formations.

How well does the affixation analysis fare in its coverage of the data surveyed in section 1.5.1? As noted in the discussion of patterns (A) and (B), i.e. (C_1)X-C_2X, a single fixed consonant appears to take over an existing prosodic position, rather than simply attaching to the left of the reduplicant. The *yaad vaad* example that proved problematic for Steriade's segmental insertion analysis cannot be explained simply by phonotactic constraints discriminating against illicit clusters. Since **vyaad* would incur fewer violations of B-R constraints on the featural identity of segments than *vaad*, these constraints must be outranked either by a constraint favouring parallel prosodic structure, or a constraint aligning the left

[40] McCarthy and Prince (1995) also have a Full Model, given in (54) above, showing a further faithfulness relation holding between the base's lexical input and the reduplicant, which is omitted here in the interests of clarity. Incorporating this dimension into the model in (126) would result in the reduplicant participating in three different correspondence relations simultaneously, but the prediction would be that I_R-O_R Faithfulness would always outrank both I-R Faithfulness and B-R Identity.

edge of the fixed affix with the left edge of the reduplicant. The transfer of distinctive vowel length seen in patterns (C) and (D), i.e. $(C_1)V_1X\text{-}C_2V_2X$, can be handled by correspondence constraints that monitor not only phonological entities but also their prosodic affiliations, as discussed in section 1.3.5.4. Moreover, cases of non-transfer, such as examples (115)–(117) above, could be accommodated by making doubly linked prosodic structure part of the fixed affix input.

Alderete et al. (1997, footnote 48) admit the inadequacy of an affixation approach for instances of subtractive reduplication, in patterns (G) and (H). They comment vaguely that a more general *'processual'* view of morphology may be required for them, and downplay their relevance. However, there are analyses within optimality theory of similar cases from other languages, where the reduplicant is identical to the base with the exception of a peripheral segment. For instance, Benua (1995), in her discussion of truncation, cites the pair of examples given in (128) and (129), the first from Axininca Campa, the second from Balangao.

(128) *osampi*-RED *osampi-sampi* 'ask'

(129) *maŋ*-RED-RED-*tagtag* *ma-nagta-tagta-tagtag* 'running everywhere'

She attributes the first to domination of the B-R Identity constraint by a constraint ONSET, which favours syllables with onsets. The second is explained by the high ranking relative to B-R Identity of NOCODA, a constraint discriminating against the presence of codas (Benua 1995: 80). Although both are plausible constraints favouring the core CV syllable, a further constraint (CONTIGUITY) is required to limit their effect to one edge of the reduplicant (see Stemberger 1996 for extensive discussion of this problem).

The difficulty with extending a similar analysis to the Hindi and Tamil examples given in (106)–(112) is that a constraint favouring *omvaar* as the reduplicant of *somvaar* over **somvaar* or **vomvaar* would have to discriminate against onsets, i.e. precisely the opposite of Benua's ONSET. Since optimality theory constraints are said to be universal, the postulation of two exactly contradictory constraints is problematic, nor does there seem to be any independent motivation for such a constraint in either Hindi or Tamil. Since it does not appear possible to derive this pattern by general constraints on syllable structure, the only alternative seems to be specifying a zero fixed affix, which would overwrite the initial segment of the base. Overall, therefore, the affixation analysis is able to handle a wide selection of examples, although there is no entirely satisfactory explanation of the subtractive cases.

The distinguishing aspect of the affixation analysis is the explicit identification of the overwriting segments as an affix, with an underlying lexical representation that is independent of the phonologically empty RED affix. In support of their analysis they claim that the segments involved display various properties characteristic of affixal morphemes. These include a propensity for peripheral alignment, but with the possibility of infixation under compulsion from higher ranked syllable structure constraints. Another property that they expect to hold of affixal morphemes is contiguity, and so the prediction is that overwriting strings will be uninterrupted, even if they could be interleaved with material from the base without violating phonotactic constraints. The identification of the overwriting string as an affix also engenders an expectation that it will conform to the properties of the other affixes of the language, e.g. in the subset of the segment inventory upon which it draws. Finally, they point to the possibility of phonologically conditioned suppletive alternation, i.e. the situation where phonological environment determines a choice between allomorphs that cannot be related to a single underlying representation. The remainder of this section will consider how far each of these claims is met in the segment-changing reduplicated forms found in the Indian languages.

Alignment

The patterns described in section 1.5.1 do seem to bear out the claims about the alignment of fixed affixes. They differ, however, from most other affixes in taking up prosodic positions within existing syllables, even if they have to replace segmental material to do so. For the most part they are peripheral, appearing at the left edge of the reduplicant, but in some cases, including those conforming to pattern (E), the overwriting affix appears infixed in the reduplicant. This is attributed in optimality theory to syllable structure constraints outranking constraints on peripheral alignment (see, for instance, McCarthy & Prince 1993a).

The fixed affixes are somewhat exceptional in the direction of their alignment. If the Dravidian cases are considered, the echo segments differ from most, if not all, other affixes in being realized at the left, rather than the right, periphery of the reduplicant (note that the reduplicant itself usually appears to the right of the base). Cross-linguistically suffixation is far more frequent than prefixation (Greenberg 1963: 73, Hawkins & Gilligan 1988), and the question of whether Dravidian possesses any prefixes at all is disputed. Masica, for instance, (following Bloch 1934 and Emeneau 1956) lists absence of prefixation in his survey of the areal features of the Indian languages (Masica 1976: 188), and Zvelebil also states that affixation in Dravidian languages is confined to suffixation (Zvelebil 1990: 16). Christdas (1988: 390), however, argues that the vowels /a, i, e/, which appear in initial position in demonstrative pronouns and time and manner adverbials, form an exception. These denote remote, proximal and interrogative respectively, as in the third person pronouns *avan* 'that man', *ivan* 'this man', *evan* 'who'. The existence of some kind of morpheme division between the vowels and what follows is not disputed: Zvelebil, however, regards the vowels as roots, whilst Christdas analyzes them as prefixes. Whatever the analysis of these particular forms, they are the only, rather isolated, parallel for the prefixal alignment of the echo segments. Hindi too is predominantly suffixing, but various prefixed forms of Sanskrit origin provide a parallel for the fixed affixes.

Contiguity

The claim that fixed affixes are contiguous is illustrated, according to Alderete et al., by the Kannada example *b^{h}rame-gime* 'illusion and such like' (Sridhar 1990: 312), where the echo sequence /gi/ is uninterrupted. It is preferred to **b^{h}rame-grime*, although this is phonotactically acceptable and more faithful to the segments of the base. However, the strict constraints that hold on syllable structure in most of the Dravidian languages, ruling out any initial cluster in the native vocabulary, make this question rather difficult to judge. The Kannada example is a Sanskrit loanword, and the handful of other examples from the literature are also loans, generally from Sanskrit or English:

(130)	*skuulu*	'school'	*skuulu giilu*	'school and the like'
(131)	*sp^{h}uurti*	'inspiration'	*sp^{h}uurti giirti*	'inspiration and the like'
(132)	*klaasu*	'class'	*klaasu giisu*	'class etc.'
(133)	*preema*	'love'	*preema giima*	'love etc.'

(130) and (131) are both Kannada examples given by Sridhar (1990: 285, 268) but prove little since *gk-* and *gp^h-* clusters are highly implausible. (132) and (133) are from Telugu, cited by Bhaskararao (1977: 7) and Abbi (1992a: 21) respectively, and show complex onsets being replaced in their entirety, rather than **gliisu* and **griima*, thereby respecting the integrity of the overwriting string. However, the paucity of examples and their status as loanwords make the evidence less than conclusive.

Alderete et al. appear to assume that the overwriting material itself is always contiguous, a view expressed explicitly by Raimy:

> *'Prespecification only appears to occur in a single region or span in a reduplicative construction.'* (Raimy 2000: 172).

Examples of non-contiguous overwriting strings are certainly not common, but one such case is found in the echo words of Gta?, a South Munda language spoken in Orissa.[41] The data on these formations are given by Mahapatra (1976), with further comments in Zide (1976). The cases of particular interest are those in which the base comprises two or more syllables, and more than one of these shows changed vocalism in the reduplicant. There appears to be considerable flexibility in the choice of vowels, with a potential tally of seven patterns for a disyllabic base.

(134)

(a)	V_1 = /a/, V_2 unchanged	(e)	V_2 = /a/, V_1 unchanged
(b)	V_1 = /i/, V_2 unchanged	(f)	V_2 = /i/, V_1 unchanged
(c)	V_1 = /a/, V_2 = /a/	(g)	V_1 = /i/, V_2 = /i/
(d)	V_1 = /u/, V_2 = /a/		

Mahapatra attributes different semantics to some of the patterns, suggesting that they do form distinct types of reduplication. For example, he associates different forms of the reduplicant of the base *kesu* (a word denoting a thick cloth used to protect the body from the cold) with the meanings given in (135)–(140) (Mahapatra 1976: 824–5).

(135) *kesu kasa* 'any kind of cloth equivalent in size and texture'

(136) *kesu kisi* 'small or thin piece of cloth'[42]

(137) *kesu kesa* 'large piece of thick cloth, torn and worn out, serving as a *kesu*'

(138) *kesu kasu* (as for 137)

(139) *kesu kesi* 'small piece of thin cloth, torn and worn out, serving as a *kesu*'

(140) *kesu kusa* 'any other material usable against the cold'

Three of these examples (135, 136 and 140) involve two fixed segments, which replace non-adjacent vowels in the base. This poses a challenge to existing theories: a default segmentism account is clearly inappropriate, given all the different vowel qualities involved, and the overwriting approach as it stands would require some expansion. An analysis by McCarthy (1982) uses these data to motivate the representation of consonants and vowels on separate tiers. The vowels of the base and the fixed segments thus share a tier, but the principles determining which surface in the reduplicant, and in which order are unclear. Odden (1987a) takes a slightly different approach, proposing that the two overwriting vowels constitute a morpheme in their own right, and so occupy a tier of their own, associating to adjacent V slots, with a prohibition against many-to-one mapping between melody and skeleton. This last restriction could be dispensed with on the assumption that overwriting segments take over existing prosodic structure rather than projecting their own. The affixal status of the two vowels is questionable: they certainly do not conform to the properties characteristic of other Indian affixes, bearing more similarity to the non-concatenative morphology of the Semitic languages, as McCarthy and Prince comment (1990: 245).

---

[41] Neighbouring Munda languages apparently show comparable patterns, though with a more restricted range of possibilities: Mahapatra gives examples from Remo and Gorum, and also the Desia dialect of Oriya. In addition, there are instances of echo words involving fixed vowel melodies, including sometimes sequences of non-adjacent vowels, in the languages of south-east Asia (see section 2.3.4).

[42] Note the parallel to the pattern of sound symbolism in English, whereby the high vowel [i] tends to represent smallness, e.g. *teeny-weeny*, and the low vowel [a] largeness. Diffloth (1994), however, cautions against claiming universality for this pattern: he cites examples from Bahnar, a Mon-Khmer language of Vietnam, in which precisely the opposite relation holds, i.e. the iconic value of a high vowel is 'big' and of a low vowel 'small'.

Segmental content

In terms of their segmental make-up, the Indian echo segments are unexceptional in comparison with other affixes. In Tamil, as in many other languages, affixal segments are drawn from a subset of the total segment inventory. Christdas (1988: 45) lists the following segments as being excluded from this subset: /t̪,[43] ʈ, s, ʂ, c, j, ɳ, ñ, r̪/, and indeed none of these appear as fixed segments. Furthermore, all the echo sequences attested for the Dravidian languages take the form CV(:), and thus conform straightforwardly to the strict constraints upon syllable structure. The status of single echo consonants as affixes in Hindi, however, is more dubious: other affixes seem to contain at least one vowel. Note also that the fixed segments of the *schm-* echo words in English are highly marked phonotactically. Their exceptionality is attributable to their Yiddish origins, and is arguably exploited for effect.

Suppletive alternation

The possibility of suppletive alternation, with the choice of allomorph phonologically conditioned, is claimed to characterize morphological, as opposed to phonological, fixed segmentism (Alderete et al. 1999: 356). This seems to be borne out cross-linguistically in echo formations, which tend to employ alternative fixed segments in cases where the base begins with the usual echo segment. The Yiddish pattern, for instance, either fails to reduplicate words beginning with the sequence /ʃm/ entirely, or else uses /ʃp/ as the echo segments, as in (141) and (142).

(141) *schmata schpata* 'rag, ratty clothes'

(142) *Schmerling schperling* (name)

Such cases are problematic for a simple prespecification approach, in that there is a choice of echo segment(s), determined by the phonological nature of the base. Any reduplicant, of course, is characterized by phonological dependence on the base, but this involves a rather different kind of dependency, and one that certainly cannot be explained in terms of copying. In a derivational framework it is treated either in terms of a rule that replaces the expected echo segment with the alternative, or one that pre-empts the rule inserting the usual echo segment by virtue of being more specific (i.e. in accordance with the Elsewhere Condition). McCarthy and Prince (1986: 68), for instance, view echo formations as the product of a dissimilatory rule, restricted in its application to reduplicated configurations.

More recently, Yip has addressed the issue, proposing that a constraint militating against exact identity between the two halves of the echo formation outrank the reduplicative correspondence constraints that favour identity (Yip 1998a). She also seeks to integrate it with other instances of 'identity avoidance', cases where sequences of identical phonological material are avoided under certain morphological conditions, by means of various strategies. One is the occurrence of just one of the sequences, where two might be expected. Examples include the English plural possessives and classical Greek determiners, both discussed by Stemberger (1981) as instances of morphological haplology.[44] Yip also gives the example of the so-called 'Doubl-ing Filter'[45] in English, where the offending configuration (i.e. contiguous *-ing* affixed words) is blocked and simply avoided altogether. This solution is

[43] Christdas distinguishes between dental /t/, alveolar /t̪/ and retroflex /ʈ/, noting that the three-way contrast of written Tamil is found in the Kanniyakumari dialect on which her research is based. Many other dialects have only a two-way distinction, and this is reflected in the transliteration conventions adopted in the thesis (see appendix A).

[44] See also Golston (1995) for an analysis covering some of the same data from an optimality theoretic perspective, in terms of a constraint against phonological homophony outranking morphological exponence.

[45] A Doubl-ing Constraint was proposed by Ross (1972: 78), prompting a response from Milsark (1972), and further discussion in Milsark (1988).

paralleled by cases where lexical items beginning with the characteristic echo segment(s) simply cannot form echo words at all.

Other strategies include movement in order to separate homophonous morphemes and the use of alternative syntactic constructions. The situation with the echo forms is slightly different, in that the solution involves the choice of an alternative allomorph: the closest parallel for this that Yip provides is the avoidance in Hindi of two arguments taking the accusative/dative case marker *ko* in the same clause. Various options are used, but for the pronouns there is a choice between two alternative forms, e.g. *hamko* or *hamẽ* 'to us', and the second is chosen to avoid two arguments in adjacent phonological words being marked with *ko*. However, this is still slightly different from the echo word situation, in that the variant forms of the pronouns exist independently, and the pertinent phonological strings are not directly adjacent, as they are in echo words.

Yip's proposal to unify these phenomena involves an extension of the OCP constraint, which was first proposed by Leben (1973). Yip suggests that this can be instantiated with different arguments, not only phonological but also morphological, e.g. OCP(Stem) is suggested as a constraint applying to the habitual-repetitive reduplicated forms of Javanese to avoid exact identity between the two halves. The issues of morphological and phonological identity are arguably somewhat confused here: if the form is reduplicated, one half is defined morphologically as a reduplicant, rather than a straightforward stem. It is thus not entirely analogous to a stem-stem compound, as Yip claims, and the two halves are morphologically different entities, although their phonological exponence may be the same. It is exactly this, i.e. identical phonological exponence of adjacent morphemes that is apparently disfavoured, and the constraint requires some elaboration to make this clear.

A functional explanation can be proposed for many of the Indian languages that avoid exact identity, since they also have cases of total lexical reduplication (see section 2.2.5). The use of an alternative echo segment thus prevents ambiguity between echo formations and cases of complete non-segment changing reduplication, which have different semantics. This is certainly the view of Trivedi, who claims that:

> *'languages maintain a conscious difference between* [complete] *reduplication and echo formation'* (Trivedi 1990: 78).

It would be a mistake, however, to regard this as a strictly rule-governed phenomenon, unlike some of the cases that Yip discusses. There appears to be some variation in whether identity avoidance occurs at all: Trivedi, for instance, reports that /pʰ/ is used instead of /ʈ/ in Bengali as the echo segment for words with an initial retroflex /ʈ/ (Trivedi 1990: 56), as in (143).

(143) *ʈak* 'bald' *ʈak pʰak* 'bald etc.'

Fitzpatrick-Cole (1994: 131), by contrast, claims that Bengali, unlike Hindi, permits total identity in such cases, giving the example in (144).

(144) *ʧebil* 'table' *ʧebil ʧebil* 'table and such'

If an alternative segment is used, its identity forms yet another dimension of variation. Indeed, even when identity avoidance is not an issue, the choice of echo segment used may vary between lexical items. In Hindi, for instance, /v/ is overwhelmingly the most popular echo segment, but there are cases where /s/, /m/ or /dʰ/ are used instead. /dʰ/, for example, is reportedly restricted to the pair of lexical items given in (145) and (146) (Trivedi 1990: 77), whilst no parallel is to be found in the literature for the use of /ʃ/ in (147).

(145) *ronaa* 'to cry' *ronaa dʰonaa* 'to weep and wail'

(146) *puraanaa* 'old' *puraanaa dʰuraanaa* 'old and so forth'

(147) *gap* 'gossip' *gap ʃap* 'tittle-tattle'

Clearly lexical storage of individual echo forms is to be assumed for such cases. The existence of such minor alternative patterns seems to be exploited by speakers when identity avoidance is an issue. Thus a Hindi informant, when asked to produce an echo expression based on *vyaapaar* 'trade', offered the form in (148), possibly on the analogy of *gap ʃap*.

(148)	*vyaapaar*	'trade'	*vyaapaar ʃyaapaar*	'trade and so forth'
(149)	*viinaa*	'Indian lute'	*viinaa giinaa*	'lute and such like'

Another informant produced the echo form in (149), which happens to follow the characteristic Dravidian pattern. Given the prevailing multilingualism in India, it is not unreasonable to speculate that echo segments used in other Indian languages known to the speaker may influence their choice of segment when the usual segment is dispreferred. There are some claims in the literature that particular semantic connotations may be associated with certain segments. For instance, Trivedi reports for the Awadhi dialect of Hindi that the echo segment /w/ has the usual meaning of generality, whilst the notion of *'contemptibility'* (Trivedi 1990: 58) is reserved for the alternative, /s/.

In conclusion, the claim made by Alderete et al. (1999: 356) that suppletive alternation may be a characteristic of morphological fixed segmentism is amply borne out in the Indian echo expressions. Comparison of conflicting claims in the literature, and variation in the forms produced by informants suggest precise categorizations may not be applicable. There certainly seems to be a general tendency to avoid identity, but the strategies employed show considerable variation. The results of testing the other predictions entailed by Alderete et al.'s characterization of overwriting strings as affixes against data from the Indian languages have proved variable. The propensity for peripheral alignment is borne out, although in the languages considered the fixed affixes are exceptional in the direction of their alignment. The claims about contiguity, however, are challenged by the non-contiguous overwriting vowels of Gtaʔ. More generally, the notion that fixed segmentism is to be explained by introducing yet another set of constraints on the reduplicant seems unnecessarily complicated. In the model proposed by Alderete et al. and given in (126) above, the output of the reduplicant is potentially subject to constraints on its relationship with the output of the base, and indirectly with the lexical input to the base, constraints on its relationship to the lexical input of the fixed affix, as well as constraints specific to the reduplicant, e.g. defining its prosodic structure. A simpler approach, and one that does not divorce the constraints on the reduplicant imposed by fixed segmentism from other constraints on its morphological or prosodic structure, is offered by the dual description analysis outlined in section 1.4.

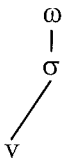
1.5.2.6. Dual description

Segment-changing reduplication is straightforwardly explained by incorporating the fixed segments within the description of the reduplicant. As before, the reduplicant object is defined by reconciling just two descriptions, that of the base and that of the reduplicant. Since the reduplicant is given priority, segments that are specified as part of its description will overwrite the segments of the base. This approach is also easily able to accommodate cases where the segment is not fully specified, such as the Akan examples given in (113) and (114), and the cases of tonal overwriting from Chichewa (see sections 1.5.2.1 and 1.5.2.5). In such examples only a single feature or subset of features is specified in the description of the reduplicant, and combine with the features inherited from the base according to the principle that the reduplicant description receives priority.

In the Indian data considered thus far all of the base segments appear in the reduplicant, modulo the overwriting effect of the fixed segments. The topmost prosodic node of the base is therefore included in the description of the reduplicant, indicating that all subordinate structure is inherited from the base. This is assumed to be the prosodic word in all the examples that follow. For patterns (A) and (B) from section 1.5.1, represented schematically as $(C_1)X\text{-}C_2X$, the description of the reduplicant will comprise the prosodic word node, and also a specification of the fixed consonant in onset position. The syllable node

specified in the description of the reduplicant is assumed to be the leftmost syllable node dominated by the prosodic word. This can be represented as in (150), which shows the description of the reduplicant for the Hindi *v-* echo words.

(150)



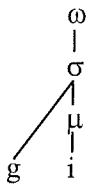
In an example like *k^{h}aanaa vaanaa*, shown in (151), the /v/ will overwrite the existing onset, whereas in *aam vaam* in (152), the onset position is empty in the base and so there is no conflict between the two descriptions. As in section 1.4, the parts of the reduplicant specified in the reduplicant description are emboldened.

(151)

(152)

Distinctive vowel length is inherited from the base unless the description of the reduplicant specifies otherwise, as demonstrated in section 1.4. This is true even when the vowel is specified in the description of the reduplicant, as in patterns (C) and (D), since it simply takes over the moraic associations of the vowel it is overriding. Examples (154)–(155) illustrate this for *puvul givul* and *maasur giisur*, the Kolami echo words introduced in (98) and (99): these are precisely parallel to the Tamil echo formations which will form the focus of this thesis.

(153)



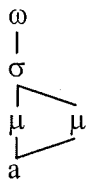
(154)

(155)

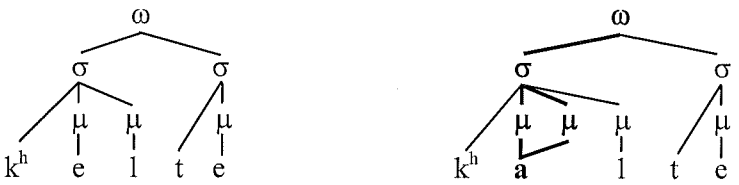
Whilst transfer of distinctive vowel length represents the default case, it is also possible to specify the vowels of the reduplicant as long or short, independently of the base. This allows for the Hindi pattern reported by Singh (see examples 102–105 and 115–117), which was of the type specifically predicted not to exist by McCarthy and Prince (1990: 245). Example

(156) represents the description of the reduplicant, and (157) shows how it applies in the case of *kʰelte kʰaalte*.

(156)

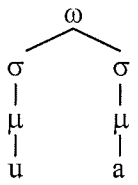


(157)



The overwriting vowels of patterns (E) and (F), which may be non-peripheral, are straightforwardly handled. In contrast with the predictions of Alderete et al., even non-contiguous examples are unproblematic for the dual description analysis, as shown for the Gta? echo word *kesu kusa* in (158) and (159).

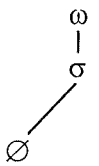
(158)



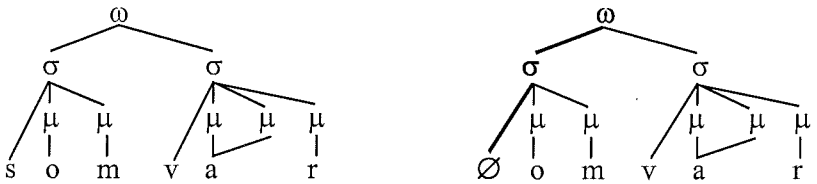
(159)

Finally, the instances of subtractive reduplication in patterns (G) and (H) of section 1.5.1 can be explained as the specification of an overwriting zero, as in (160). When this description is combined with that of the Hindi base *somvaar*, the resulting reduplicant is *omvaar*, as shown in (161).

(160)



(161)



Dual description analysis is therefore able to handle the full range of segment-changing reduplication patterns, including cases of non-contiguous fixed segments, non-transfer of distinctive vowel length and subtractive examples, all of which proved problematic for other accounts. Moreover, unlike other theories which require special apparatus for such cases, e.g. the extra source of constraints in the account of Alderete et al., segment-changing reduplication involves a natural extension of the mechanisms used to implement other

constraints on the reduplicant. In addition to the other advantages outlined in section 1.4, this suggests that dual description analysis is the most promising account of reduplication.

All the theories surveyed in this first chapter share the assumption that reduplication as a general phenomenon requires a general explanation. Analyses therefore attempt to capture all possible parameters of variation, using language-specific examples as evidence for the properties of a unified theory. The remaining chapters of the thesis will take a slightly different approach, presenting an in-depth study of one specific type of reduplication in one language, i.e. echo words in Tamil. The next chapter sets out the context in which these echo expressions are found, presenting a general survey of reduplication in the Indian subcontinent and reviewing descriptions of echo expressions in languages other than Tamil.

Chapter 2: Echo expressions in Indian languages

2.1. Introduction

Whilst all languages seem to employ reduplicated forms in varying degrees and for different functions, extensive use of reduplication is particularly characteristic of the Indian subcontinent. Indeed, it has long been recognized to be one of a number of features shared between all four of the Indian language families. Bloch, for instance, included *'les mots doubles et à écho'* (Bloch 1934: 328) amongst various morphological and syntactic traits shared by different Indian language families.[1] Emeneau has discussed at length the distribution of various reduplicative expressions in articles arguing that India constitutes a linguistic area (Emeneau 1956, 1969, 1980). Finally, Abbi has devoted a book to exploring what she terms the *'linguistic feature area'* defined by the occurrence of reduplicated structures in India (Abbi 1992a).

Section 2.2 presents an overview of the different kinds of reduplicative expressions found in the subcontinent, with particular attention to Tamil. This sets the context for the echo formations, which will be introduced in detail in section 2.3. The final section will review the literature on echo formations in other Dravidian languages. Data on echo expressions in two Indo-Aryan languages, Bengali and Hindi, will also be presented, as background to the detailed investigation of similar issues in Tamil.

2.2. Survey of reduplicated forms

2.2.1. Points of difference

The categorization outlined in the coming sections represents an attempt to divide up what seems to be a continuum. This encompasses not only the categories given below but also other reduplicative phenomena, such as language games and rhyming conventions in verse (cf. Yip 1998b: 13), and even baby talk. The boundaries between different categories cannot always be clearly drawn, as is reflected by the considerable terminological variation in the literature. Various criteria have been taken into account in making the classification, notably the semantic structure of the expressions. Onomatopoeic expressions and expressives share the general property that neither of the two halves of the expression are independently meaningful, whereas the base of an echo expression forms a lexical item in its own right, and this is true of both elements of paired words. The semantics associated with reduplication are relevant too, including the question of iconicity, which has attracted some attention in the literature.[2] Lexicality and productivity are also pertinent issues, and will receive comment, although they are not always easily determined. Finally, the syntactic roles played by the reduplicated forms show some variation, but a tendency towards adverbial function will emerge.

2.2.2. Onomatopoeic expressions

Indian languages in general display a *'predilection for onomatopoeia'* (Masica 1992: 44), and Tamil is no exception. Onomatopoeic expressions may or may not be reduplicated in structure, and the repetition may be exact, as in (1), or there may be segmental differences between the two halves, as in (2). Example (3) is a non-reduplicated case.

(1) *kiiccukiiccu-ŋŋu* 'chirping'

(2) *ʈaakkuʈikk-enal* 'tick-tack sound produced by slippers'

[1] For discussion of other areal features see Kuiper (1974), Southworth (1974), Masica (1976) and the 1990 All India Linguistic Traits Survey (Krishan 1990).

[2] See Moravcsik (1978: 316–325) and Regier (1998). For a general review of the literature on iconism see Anderson (1998).

(3) *kup-ŋŋu* 'sound of sucking'

As these examples show, onomatopoeic expressions are mostly followed by a suffix. Typically this is *-ŋŋu*, the colloquial form of the past participle of the verb *en* 'say', which also appears regularly after reported speech in Tamil. As Schiffman comments:

> *'The structure of these emulates a 'quotative' construction, i.e. it is as if there is a sound of some sort being quoted'* (Schiffman 1999: 181).

Syntactically, these forms are classed as manner adverbials (Lehmann 1989: 376): nouns may also be formed by the addition of other suffixes, as in (2) and (4), and onomatopoeic expressions can function as verbal roots, as in the verb form in (5).

(4) *kilukilu-ppu* 'tinkling'

(5) *kiɭukiɭukkiraan* 'he laughs out loud'

Phonotactic constraints observed elsewhere in Tamil may be violated,[3] which seems to be a characteristic of onomatopoeic expressions cross-linguistically (see, for example, Hinton et al. 1994). In particular, there are many examples of onomatopoeic forms beginning with retroflex consonants, as in (6) and (7), or even voiced obstruents, as in (8),[4] although these do not otherwise appear word-initially in Tamil (see section 5.2.1). Further examples include the non-homorganic consonant cluster in (3), and the word-final position of the velar nasal in (7).

(6) *ʈukku-ŋŋu* 'knock, knock'

(7) *ŋaŋ* 'clang'

(8) *guʈu guʈu* 'rumbling'

The meaning of such forms is partly determined by sound symbolism, although there are obviously language-specific conventions at work. As in other languages, the lexical status of such forms is unclear: some of the examples given are included in standard dictionaries,[5] but the class of items does not appear to be closed.

The main features exemplified here for Tamil also hold true of the other Dravidian languages. Emeneau (1969: 279), for instance, mentions various other languages (Toda, Kota, Kannada, Telugu, Konḍa, Pengo and Kui) in which the construction with the quotative verb is found, and Kuiper (1974: 149) argues that it has been imitated by some of the Munda languages, in a discussion of various instances of structural convergence. Parallels between the onomatopoeic systems of Dravidian and Indo-Aryan are also noted (e.g. Zvelebil 1990: 73), and a number of areal etymologies proposed. Emeneau regards the Dravidian languages as the source in this instance, with borrowing into Indo-Aryan followed by independent proliferation on the same pattern (Emeneau 1969: 287).

2.2.3. Expressives

This next category shares many similarities with the onomatopoeic expressions, but involves terms that are not sound symbolic in the strict sense. Example (9), the reduplicated

[3] For comments in the literature see Gnanasundaram (1972: 247), Emeneau (1969: 284), Steever (1987: 743), Annamalai and Steever (1998: 103), Schiffman (1999: 182), and Wiltshire (1999: 129).

[4] Emeneau and Hart (1993) consider the occurrence of initial voiced stops in reduplicated expressives (in which they include both onomatopoeics and the forms to be considered in the next section). They present numerous examples from modern colloquial Tamil, and argue that they can be reconstructed for Proto-Dravidian (or subgroups thereof) on the basis of etyma appearing in related languages. It also appears that a sound change causing palatalization of Proto-Dravidian *k before front vowels was blocked in expressives, the class of words to be introduced in section 2.2.3 (Annamalai 1968). For a parallel case of resistance to sound change see Morin (1972: 98) on onomatopoeics in early French and Latin.

[5] Cataacivam (1966) is a Tamil dictionary devoted to onomatopoeic expressions.

counterpart of (6), is such a case, showing the development of a more abstract meaning, beyond the realm of sound.

(9) *ʈukkuʈukku-ŋŋu* 'regularly, like clockwork'

Writers vary in how they classify such expressions: some include them within a broad definition of onomatopoeia (e.g. Emeneau 1969, Steever 1987), others regard onomatopoeics as a subset of a more general expressive category (e.g. Abbi 1992a, Wiltshire 1999).[6] The latter approach seems to have been adopted by the classical grammarians: the *Tolkaappiyam* distinguishes between two kinds of reduplication, defining *iraʈʈaikkiɭavi* 'doubled words' as those in which neither part has any meaning on its own and gives examples involving both sound and other qualities. The lack of independent semantics for either half differentiates these forms from another class of expressions, in which one part corresponds to a lexical item, but the other is a copy altered in some way. These include the *ki(i)-* echo expressions investigated in detail in later chapters but also a range of other echo forms to be considered in section 2.3. They bear many similarities to the expressive category, and are often discussed in connection with them: Malten (1989), for instance, in his study of reduplicated verb roots in Tamil, considers both expressive examples and some echo cases, classifying them as primary and secondary respectively.

Morphologically, there are clear parallels with the onomatopoeic expressions, notably the suffixation of *-ŋŋu* and other affixes, and the incorporation of expressive forms into the verbal system. Malten (1989: 47) notes that such cases tend to follow a four syllable pattern, containing only short vowels and single consonants, as in (10) and (11).

(10) *catacata* 'to be damp or wet'

(11) *narunaru* 'to feel grit in the mouth along with food'

Malten lists 229 examples altogether (including both primary and secondary forms), but comments that that number could be multiplied more than five times if social and regional variants were taken into consideration.[7] Other writers also comment on the large numbers of such forms: Steever remarks that they could fill an entire dictionary in themselves (Steever 1987: 743), and a considerable proportion of Winslow's entries can be classified as onomatopoeics or expressives (Winslow 1862). Malten's examples all involve reduplication but, as with the onomatopoeics, there are also non-reduplicated instances, as in (12).

(12) *caʈak-ŋŋu* 'spontaneously'

In cases where reduplication is not exact, certain regularities have been observed.[8] For instance, Wiltshire (1999: 130–131) remarks that changes primarily affect the first consonant and vowel of the base, with coda consonants in the first syllable remaining unchanged and the second syllable usually being copied exactly, or at the most undergoing a vowel change. Furthermore, it is only initial consonants /k/ and /c/ that are likely to be replaced, and the substituted consonant is always articulated further forward in the vocal tract than the base consonant, so /k/ can be followed by /c/, /p/ or /m/, as in (13), but /c/ only by /p/ or /m/, as in (14).

(13) *kasamusa-ŋŋu* 'dirtiness, disorderliness'

---

[6] Asher (1985: 242) refers to both onomatopoeic expressions and expressives as *'ideophones'*, whilst Gnanasundaram (1972) groups onomatopoeics, expressives and echo expressions together under the general heading of '*echo words*'.

[7] Note that Malten is concerned only with written forms, and thus includes no information on the voicing of initial obstruents, since the orthographic system cannot distinguish between voiced and voiceless obstruents in initial position (see section 5.2.1).

[8] Emeneau and Hart (1993) provide information on combinations of consonants in four-syllable reduplicated examples. They claim that there is a systematic constraint against dental consonants preceding /l/ (Emeneau & Hart 1993: 81).

(14) *caṭṭuputṭu-ṇṇu* 'fast, quickly'

As the examples given thus far illustrate, the semantics of the expressive forms vary considerably, and extend well beyond the sphere of sound to which the onomatopoeics are restricted. They appear to be loosely sense-based, but under a broad definition of sense that includes not only taste, touch, feel etc. but also motion.[9] Emeneau comments on the issue:

> *'Perhaps it would be more just to say that the class denotes varied types of sensation, the impingement of the material world, outside or within the person, upon the senses'* (Emeneau 1969: 284).

In some cases it is possible to trace a metaphorical association of meaning from an onomatopoeic source: Steever, for instance, notes that the meaning of 'bitching, complaining' develops naturally from (15), and the notion of 'guilt' can be derived from (16).

(15) *toṇutoṇu* 'sound of beating drums'

(16) *kurukuru* 'scratching, throbbing'

In other cases the sound-meaning connection is more opaque, and the expressives thus seem to span the gap between sound symbolic and 'normal' vocabulary, which displays Saussure's arbitrariness of the sign (Saussure 1987: 100). However, the repetition in reduplicated expressives is arguably of iconic significance, representing the continuing or repeated nature of an event.[10] According to Wiltshire, onomatopoeics and expressives are largely limited to positive utterances: the only negative contexts in which they can appear are imperatives and metalinguistic comments. Her explanation is that:

> *'expressive language transmits a sensation which cannot be negated, once spoken'* (Wiltshire 1999: 133).[11]

Interestingly, this is the converse distribution to the echo expressions, which appear predominantly in negative contexts (see sections 3.2.3.1 and 3.3.2.1): the distinction therefore supports the separate classification adopted here.

The remarks made above concerning the parallels between the Dravidian and Indo-Aryan onomatopoeic systems also hold for the expressives. Some representative examples of expressives from Hindi are given in (17) and (18), and from Bengali in (19) and (20).

(17) *chī chī* 'dirty, filthy'

(18) *cip cip* 'sticky'

(19) *thik thik* 'sense of teeming with maggots'

(20) *pil pil* 'sense of being overcrowded'

2.2.4. Paired words

These examples involve the juxtaposition of two lexical items and are variously referred to in the literature as *'synonymic compounds'* (Dongare 1975), *'synonymic repetition'* (Politzer 1961), *'semantic reduplication'* (Vacek 1994) and *'redundant compounds'* (Singh 1995).[12] Both parts belong to the same semantic field, and may stand in several possible relations to one another. Vacek (1994: 146) proposes a three-fold typology for Tamil, according to which the constituents may be (near) synonyms, as in (21), may have similar or logically connected meanings, as in (22), or may be antonyms, as in (23).

---

[9] Note that a similarly broad range of meanings are covered by parallel forms in the Chinese lexicon: see the discussion of phonaesthetic features in Po-Ching (2000).

[10] Continuity and repetition are amongst the set of meanings associated with reduplication cross-linguistically (see, for example, Moravcsik 1978, Regier 1998).

[11] This is a general property of expressives, according to Diffloth (1994: 108).

[12] Abbi classifies these expressions as a subtype of lexical reduplication and refers to them simply as *'compounds'* (Abbi 1992a: 24).

(21) *tikku ticai* 'point of direction direction' i.e. 'direction'

(22) *kuṭi makkaḷ* 'subjects children' i.e. 'citizens'

(23) *taay takappan* 'mother father' i.e. 'parents'

Parallel examples of each type can be given for Hindi, as in (24)–(26).

(24) *dʰan daulat* 'wealth wealth' i.e. 'wealth, riches'

(25) *saanii paanii* 'cattle-cake water' i.e. 'cattle'

(26) *raat din* or *din raat* 'night day' i.e. 'continually'

The meaning of the whole expression is fixed and often transparently compositional. (25) is something of an exception, since the meaning of the whole bears a less direct relation to the meanings of the constituents. The rhyming relationship between the two parts may have facilitated their juxtaposition: phonetic similarity seems to be favoured, which is a point of contact with the other categories described. In some instances there may be a generalizing effect, similar to that associated with the echo expressions. This is quaintly illustrated in (27), an example from the Awadhi dialect of Hindi taken from an early grammar (Saksena 1937: 324).

(27) *haatʰii gʰoṛaa* 'elephant horse' i.e. 'elephants and horses and also other high class conveyances and paraphernalia of a king'

Paired synonyms frequently involve words taken from different registers, dialects or even languages. This has been remarked upon by several writers, who propose that language contact particularly favours the development of such expressions. Singh (1995), for example, discusses Hindi cases such as (24), in which a native Hindi word is followed by a Persian synonym. He argues that the prevalence of reduplication in the language provides the necessary structural resources, and contact with Persian the right climate for such expressions to be formed. Vacek (1994: 145), in his study of the phenomenon in Tamil, also notes that one member of the pair may be a borrowing, listing Hindi, Urdu, Marathi, Bengali, Sanskrit and Mongolian as potential sources. Finally, in a rather different context, Politzer's study of the increase of synonymic repetition in late Latin and Romance concluded that Germanic-Romance and Latin-Romance bilingualism or quasi-bilingualism at the time was a contributory factor (Politzer 1961: 487).

Like the other categories considered thus far, paired words are amongst the shared features characterizing the Indian linguistic area. Their geographic distribution also extends beyond the subcontinent: Abbi (1992a: 24–25) refers to south east Asian languages in general being rich in such formations. She gives a couple of antonymous examples, reproduced in (28) and (29), the first from Malay, the second from Thai.

(28) *tuwa muda* 'old young' i.e. 'all, everybody'

(29) *sukh thukh* 'happiness sufferings' i.e. 'ups and downs in life'

Languages differ in the syntactic categories that make up such forms: in Tamil it is predominantly nouns and verbs, although Vacek (1994: 146) claims that any lexical item except pronouns and conjunctions can be involved. Hindi paired words are apparently limited to nominal, verbal, adjectival and adverbial combinations, whilst Marathi also permits prepositions to be paired, according to Dongare (1975: 251). In the same way, there are language-specific restrictions on categories that can be echoed (see section 2.3.1). Whilst paired words differ from the other types of expression in showing only semantic, and not phonological, reduplication, there are several parallels with the echo formations, and their pervasive presence throughout India suggest that they belong to the same continuum of reduplicated structures.

2.2.5. Complete reduplication

This next category involves the complete, unaltered reduplication of lexical items, and belongs to the *Tolkaappiyam*'s second reduplicative type, the *aṭukkuccol* (literally 'row words'). The associated semantics vary according to the lexical category of the base, but all belong to the nexus of meanings established for reduplication cross-linguistically. As with the onomatopoeics and expressives, instances of complete reduplication frequently function as adverbs, and may be followed by *-aa(ka)* in Tamil. Apart from this tendency, reduplication is not exploited as a category-changing device in either the Dravidian or Indo-Aryan languages, unlike the Munda family, where it plays a crucial part in building up the lexicon (Abbi 1992a: 114). Abbi, whose book focusses predominantly on complete reduplication and its distribution throughout India, stresses its status as a separate class, remarking that:

> '*word reduplication in particular can be unequivocally posited as a grammatical/semantic category distinct from Echo-formations, Expressives and Compounds*'[13] (Abbi 1992a: 28).

These examples are primarily differentiated from the other categories by their productivity, although their semantics also set them apart, and will be described for each lexical class in turn.

Nouns

A distributive meaning is frequently associated with the repetition of nouns in Tamil, and this is also true of all the other Indian languages investigated by Abbi, with the sole exception of Malayalam. A representative Tamil example is given in (30), with the exact Kannada equivalent in (31), and a comparable Hindi case in (32).

(30) *viiti viiti* 'each, every street'

(31) *biidi biidi* 'every street'

(32) *g^{h}aṇṭe g^{h}aṇṭe* 'every hour'

Intensity and emphasis can also be effected by nominal repetition, as in the Tamil example in (33).

(33) *ratta rattamaa(ka)* 'only blood'

This particular case also demonstrates the addition of an adverbial suffix at the end, and the loss of the final *-m* of the base, a possibility noted by Malten (1989: 39). He also mentions a further variant, in which the absence of *-m* is accompanied by compensatory lengthening of the preceding vowel, giving the example reproduced in (34).

(34) *vaaraa vaaram* 'week by week'

There seems to be some variation in the interaction of complete reduplication with inflectional endings.[14] In some cases it appears that the case marker may be added once, to the end of the reduplicated nominal stem, as in (35).

(35) *kaṭai kaṭaikku* 'to every shop'

The same pattern is found in Kannada, as in (36), an example given by both Murthy (1975: 159) and Sridhar (1990: 253).

(36) *mane maneguu* 'to every house'

In other cases, a single occurrence of the dative case marker is found between the two halves, as in the Tamil example in (37).

---

[13] Abbi's *'expressives'* comprise both onomatopoeic expressions and expressives, as they are defined here, and she uses *'compounds'* to refer to the category termed 'paired words'.

[14] This topic will be considered in detail for the echo formations in sections 3.2.3.3.1 and 3.3.2.3.1.

(37) *teruvukku teru* 'in every street'

Again, it has a counterpart in Kannada (Sridhar 1990: 257), although this also involves the addition of the emphatic *-ee* clitic, and seems to have an intensive rather than distributive meaning.

(38) *uurige uuree* 'the whole town'

Pronouns

The distributive sense associated with nouns extends to pronouns and also interrogative and demonstrative expressions, as illustrated for Tamil in (39)–(41).

(39) *avan avan* 'he he' i.e. 'each and every man'

(40) *yaar yaar* 'who who' i.e. 'which different persons'

(41) *aŋkaŋkee* 'there there' i.e. 'everywhere'

Parallel examples can be cited for other Indian languages, such as the reduplicated interrogative from Hindi in (42).

(42) *kyaa kyaa* 'what different things'

Numerals

The complete reduplication of cardinal numerals is reported for all the languages investigated by Abbi, with the exceptions of Assamese and Oriya. Abbi (1992a: 79) lists numerous examples of reduplicated numerals with distributive sense in different languages, including Tamil. Lehmann (1989: 114) cites the forms in (43)–(45), which show only partial reduplication.

(43) *ovv-onru* 'each one, one by one'

(44) *ivv-iraɳʈu* 'each two, two by two'

(45) *mu-muunru* 'each three'

Adjectives

The semantic result of reduplicating a Tamil adjective seems to be primarily a matter of intensification, as in (46) and (47). This is the main semantic effect associated with the repetition of adjectives in Abbi's pan-Indic survey (Abbi 1992a: 55ff.).

(46) *cikappu cikappaa(ka)* 'red all over'

(47) *nalla nalla* 'very good'

Verbs

Reduplicated infinitives and participles in Tamil express repetition and continuity or intensity, all fairly standard semantic properties of reduplication cross-linguistically and with a clear iconic basis. Repeated infinitives function as the predicate of subordinate clauses, as in (48) and (49), examples taken from Malten (1989: 43) and Lehmann (1989: 264) respectively.

(48) *viiʈu neruŋka neruŋka*
house draw-close.infin draw-close.infin
'as he came closer to the house'

(49) *neeram aaka aaka*
time become.infin become.infin
'when/because it became later and later'

The repetition of participles produces an adverbial construction, as in (50) and (51), taken from Steever (1998: 124) and Steever (1987: 744) respectively.

(50) *aɻutu aɻutu* 'crying crying' i.e. 'sobbing'

(51) *vantu vantu* 'coming coming' i.e. 'coming time and again'

Such reduplicated verbal adverbs are found throughout the subcontinent: Abbi (1992b: 134), indeed, claims that their occurrence uniquely identifies south Asia as a linguistic area. She gives a detailed survey of the associated aspectual meanings in different languages, and also considers their historical origin, concluding that there is limited evidence for such structures in Proto-Dravidian. She hypothesizes that reduplicated structures in general may have originated in the Austroasiatic languages, and been adopted by the other language families through borrowing and areal diffusion. She proposes, however, that the reduplicated verbal adverbs were a subsequent innovation of the Indo-Aryan languages, where they are most prominent, and spread thence into Dravidian. According to Abbi (1992a: 49), neither the Indo-Aryan nor Dravidian languages reduplicate main verbs. One counterexample, given in (52), is reported for Tamil by Raghunathan (1995: 57), but he remarks that it is a *'marked'* phenomenon in the modern language.

(52) *kataiyai keeʈʈu avan cirittaan cirittaan*
story.acc hear.vbp he laugh.past.3sm laugh.past.3sm
'Hearing the story he laughed and laughed.'

2.2.6. Syntactic reduplication

In addition to cases where an infinitive or participle is repeated in its entirety, Tamil possesses a number of complex syntactic constructions involving different grammatical forms derived from the same verbal root. One such example is an emphatic construction comprising an infinitive, to which the clitic *-ee* is added, followed by a finite verb form. In (53) it is the main verb that is reduplicated, whereas in (54) it is the auxiliary.[15]

(53) *kumaar varavee vara maaʈʈaan*
Kumar come.infin.emph come.infin will-not.3sm
'Kumar just won't come.'

(54) *vaaŋga maaʈʈavee maaʈʈeen*
come.imper.pl will-not.infin.emph will-not.1s
'I won't come at all.'

Kannada seems to have a parallel construction: Schiffman's grammar of the spoken language contains the examples in (55) and (56), in which the stem form of the verb is followed by the emphatic marker *-ee* and then a finite verb (Schiffman 1983: 124).

(55) *avɭ hoogee hoogtaaɭe* 'she will definitely go'

(56) *avn maaɖee maaɖtaane* 'he will definitely do it'

Tamil also has various constructions expressing different degrees of possibility and probability that involve juxtaposed forms with the same verbal root. The following pair of examples are given by Schiffman (1999: 169–170) to illustrate less than chance probability in (57) and greater than chance probability in (58).[16]

(57) *maɻe vantaalum varalaam*
rain come.past.conc come.infin-may
'Rain might just possibly come.'

[15] Further examples of the construction are to be found in Lehmann (1989: 86, 157) and Schiffman (1999: 172).
[16] See also Lehmann (1989: 378) for further examples.

(58) *iruntaalum iruppaan*
be.past.conc be.fut.3sm
'He will probably be (there).'

Repeated verb forms are also used in expressions of simultaneity, where past participles, as in (59), or verbal nouns, as in (60), appear first in the positive and then the negative, with the *-um* conjunction linking the two (Schiffman 1999: 171).

(59) *caappiṭṭum caappiṭaame*
'before eating, before even getting a chance to eat, as soon as (someone) ate'

(60) *kuḷiccatum kuḷikkaatatumaa*
'before bathing was finished, before doing anything else, i.e. early in the morning'

Such examples tend to behave as adverbial phrases, and may be followed by the *-aa(y)* suffix, as in (60). This provides a point of contact with the other reduplicative categories considered: adverbial function has been seen to be a recurring characteristic of such forms. Finally, Wiltshire (1999: 122) gives the example in (61), which bears obvious similarity to the onomatopoeic and expressive categories. This follows the phonological pattern of four core syllables observed by Malten in the reduplicated verb forms, but uses an independent lexical root which then recurs in the following finite verb.

(61) *paṭi paṭi enru*[17] *paṭikkiraan*
study study say.vbp study.pres.3sm
'He is studying intensely.'

2.3. Echo expressions

2.3.1. Defining characteristics

The echo expressions, variously known as 'echo words', 'echo formations', or occasionally 'echo-compounds' (e.g. Steever 1987: 744), form yet another category of reduplicated forms. In terms of their semantic structure, they fit between the onomatopoeics and expressives, in which neither part is independently meaningful, and the paired words and examples of complete and syntactic reduplication, where both parts can be assigned a meaning of their own (although the semantics of their sum may not be straightforwardly compositional). The base of an echo formation, which in the vast majority of cases appears first,[18] is always a lexical item in its own right, but the reduplicated part has no independent lexical meaning. A consistent cluster of meanings tend to characterize echo expressions cross-linguistically, setting them apart from other sections of the continuum. These will be outlined in some detail in the next section.

Varying degrees of productivity are associated with the other categories considered so far. The last two sets (complete and syntactic reduplication) are entirely open-ended, whereas there seems to be less freedom to form new onomatopoeics and expressives, and certainly new paired words. Within the echo expressions there is also considerable variation, which has implications for the question of lexical storage. As discussed in section 1.5.1, different languages favour different fixed segments, with *ki(i)-/gi(i)-* being characteristic of the Dravidian languages.[19] Such formations are highly productive: almost any Tamil word, or even loanwords,[20] can form the base for an echo expression in *ki(i)-*, and the same is true of

[17] Note that *enru* is the formal equivalent of colloquial *-ŋŋu*, the participle of *en* 'say'.
[18] Some exceptions have been noted in section 1.5.1.
[19] There is alternation between *ki(i)-* and *gi(i)-* amongst different Dravidian languages: Tamil, Toda and Malayalam (where echo words are a marginal phenomenon) use the former, whilst Telugu, Kannada, Kolami and Tulu are among those using the latter.
[20] Arunachalam (1977: 9) presents various examples, including *table kiible* 'table and other furniture' (English), *puurnaa kiirnaa* 'full and so forth' (Sanskrit) and *aadmi kiidmi* 'men etc.' (Hindi).

Hindi *v-*, Bengali *t-* etc. One systematic set of exceptions involves words which begin with the same segments as the echoed portions and are therefore indistinguishable from cases of total reduplication: these have been discussed in section 1.5.2.5 under the heading of suppletive alternation. Given the otherwise high level of productivity, it seems reasonable to assume that speakers create such expressions spontaneously by performing some kind of reduplicative operation on the base, rather than storing each example independently in the lexicon. The principles and constraints responsible for producing such forms are therefore an important topic for investigation, and will be considered in depth.

At the other end of the spectrum are expressions with entirely idiosyncratic echo patterns. An example of this kind in Tamil is given in (62), based on the word meaning 'little', but with the reduplicated part altered in a way that is not paralleled in other Tamil expressions.

(62) *koñcam nañcam* 'itsy bitsy'

According to Arunachalam (1977: 4), there are hundreds of such idiomatic expressions in Tamil, where a lexical base word is followed by a meaningless repetition phonologically altered in some unpredictable way. The changes are, however, generally restricted to the initial consonant and vowel, as seen in (62). Semantically, the effect seems to be predominantly one of intensity or emphasis, although the echoed form may develop its own meaning, as in (63), an example given by Arunachalam based on the noun meaning 'salt'.

(63) *uppu cappu* 'taste'

One semantically homogeneous set of such expressions in Tamil involves colour terms, as in (64) and (65).

(64) *paccai paceer enru* 'intensely green'

(65) *kannaŋ kareen enru* 'intensely black'

These seem to form an intermediate category: they bear obvious affinities to the expressives, in terms of the addition of *enru*, and their adverbial function. The lexicality of the first half, however, aligns them with the echo expressions, although the phonological alteration unusually applies to the second, rather than the first, syllable. Echo expressions with idiosyncratic phonology, and possibly semantics, seem to be a common feature of the Indian subcontinent. The proportion of productive and non-productive types seems to vary, however, between different Dravidian languages: Steever (1987: 745) observes a trend towards higher levels of unpredictable examples in the North Dravidian group. Echo expressions with unpredictable phonology must presumably be stored lexically. Between these examples and the fully productive echo forms, there are various patterns that recur in a restricted number of words: some of these have been considered in section 1.5. They may either represent a synchronic process of limited productivity, or the lexicalized residue of a past process. Finally, whatever their degree of productivity, the syntactic role played by the echo words corresponds consistently to that of their base: they do not display the tendency towards adverbial function noted for several of the other categories.

2.3.2. Semantics

The meaning associated with echo expressions varies, according to context and personal usage. Trivedi (1990: 77), for instance, commenting on the use of such forms throughout India, notes that there are *'many idiolectal, stylistic and idiosyncratic effects'* attached to them. Several key semantic aspects recur, however, and particularly the notion of generality. The object or event referred to in the base word is taken as an example of a wider set, to which the whole expression refers. If the object is a member of a clearly defined superordinate category, then the other members of the set may be specified, as in (66)–(68), where the superordinate categories are musical instruments, drinks and reptiles respectively.

(66) *koṭṭu kiṭṭu* 'drums and other musical instruments'

(67) *kaappi kiippi* 'coffee and other beverages'

(68) *paampu kiimpu* 'snakes and other reptiles/pests'

In some cases several different semantic fields will be potential candidates, as in (69), an example given by Annamalai and Steever (1998: 125), which nicely illustrates context-dependence. According to them, this could refer either to a group of domestic animals, part of a dowry or traffic obstacles.

(69) *maaʈu kiiʈu* 'cattle and the like'

Frequently the reference is even more vague, as reflected in translations like 'and related stuff' (Lidz 2000: 147), 'and other things of the same sort' (Asher 1985: 207), 'and related activities', 'and so forth', 'and such', even 'etc.'.[21]

Usually a multiplicity of objects or events is referred to, but in some instances the effect is not so much generalizing as indefinite, and again this will usually be determined by the context. Thus in (70) and (71), examples taken from Arunachalam (1977: 8), a single object or locality is being referred to, and so the translation is disjunctive rather than conjunctive.

(70) *peenaa kiinaa* 'pen or any similar writing instrument, e.g. pencil'

(71) *viiʈʈukku kiiʈʈukku pookaatee*
house.dat echo go.neg.imp
'Don't go home or to any similar place such as a hotel or hostel room.'

In terms of semantic structure, the second part is clearly parasitic on the first, and any meaning assigned to it is perhaps best seen as a function of the whole expression, rather than the specific phonological material that makes up the second part. There are, however, a few isolated examples where the echoed part seems to take on an independent status, reflected not only by (at least partially) autonomous semantics but also a non-adjacent position in the word order. Bhaskararao, for instance, comments in a footnote (Bhaskararao 1977: 10) on a tendency amongst younger speakers of Telugu to produce sentences such as (72).

(72) *peḷḷi ayindi kaani giḷḷi avaleedu*
'The marriage took place but not *giḷḷi* (the consummation).'

Another Dravidian example, this time from Malayalam, is given by Malten (1989: 44), although here the usage is partly metalinguistic.

(73) *ummaan koʈuttaal ammaavan alleŋkil kimmaavan*
'As long as he feeds me, he's called "Uncle", but otherwise "Kincle".'

Finally, Fitzpatrick-Cole (1994: 132) provides a Turkish example, reproduced in (74).

(74) *işte onlar sanatçı üst tarafı manatçı*
'It is precisely those who are artists; the remainder are etceteras!'

This is certainly a marginal phenomenon, but represents a potential development from the echo expressions, with the two halves beginning to diverge. Interestingly, a folk-tale motif common to several of the Dravidian languages relies on precisely this kind of dissociation between the two parts of an echo word. The expression in question takes the form *puli kili* 'tiger and such like animal' in Tamil. According to Emeneau (1938: 116), the story involves a tiger hearing this expression, and identifying itself with the first part, but mistakenly assuming

[21] Similar generalizing translations are given for echo expressions in numerous other Indian languages and dialects, including Telugu (Bhaskararao 1977: 6), Kannada (Sridhar 1990: 268), Toda (Emeneau 1938: 113) and Kolami (Emeneau 1955: 101–2) among the Dravidian languages, and Marathi (Apte 1968: 22), Kamrupi (Goswami 1957: 164), Awadhi (Saksena 1937: 323), Sinhalese (Hettiaratchi 1959: 50) and Hindi (Abbi 1992a: 20).

that the second part refers to some other creature, which it then imagines to be active in subsequent events.

The uncertainty or vagueness inherent in the echo expressions may imply a lack of care on the part of the speaker, who does not bother to specify exactly. This connects straightforwardly with the pejorative connotations sometimes associated with echo expressions. Zvelebil (1990: 73), for instance, gives the example in (75) to illustrate the possibility of ridicule.

(75) *miicai kiicai* 'a (funny/silly) sort of moustache'

A clearly dismissive attitude towards the *gawḍa*, or headman, underlies the example from Kannada in (76), taken from Murthy (1975: 164).

(76) *gawḍa giwḍage naanu hedaroolla*
'I am not afraid of the *gawḍa* (who is like any other ordinary person).'

Scornful or sarcastic connotations are widely reported for echo words, and, indeed, form one of the meanings associated with reduplication cross-linguistically (see Regier 1998 and references therein). One further alternative is that the echo expression merely intensifies or emphasizes the meaning of the base word: this is reported to be possible by both Zvelebil (1990: 73) and Ferro-Luzzi (1992: 21). More generally, the echo words are said by Steever (1987: 745) to *'encode the speaker's affective state'*. Malten (1989: 11) expands at some length upon the kinds of emotions that may be involved, directed either towards the addressee or to what is described in the base word, listing playfulness, hesitation, ridicule and emphatic negation.

Throughout the Indian languages, echo expressions are predominantly restricted to colloquial speech (see, for example, Malten 1989: 46, Trivedi 1990: 51 and Abbi 1992a: 20). This is particularly clear in Tamil, where the colloquial and formal varieties diverge so systematically that the language is said to be diglossic (see section 3.2.2). Since the colloquial variety has no standard written form, the occurrence of echo expressions in texts is therefore minimal. They are, however, in widespread use in the spoken language, at all levels of society (Arunachalam 1977: 6). Malten (1989: 46) comments, however, that the associated connotations of familiarity mean that they would not be used in speaking to a social superior.

2.3.3. History

The restriction of echo expressions to colloquial speech and their consequent absence from written texts poses obvious difficulties for tracing their historical development. Moreover, the tendency to regard such forms as sub-standard means that they are generally ignored in traditional grammars. As noted above, the *Tolkaappiyam* does refer to reduplicated forms, including onomatopoeics, but makes no specific mention of echo expressions. Steever (1998: 28) dates their first attestation in Dravidian to the mediaeval period, commenting that they appear in the Viiraśaiva literature of Middle Kannada and Middle Telugu. Arunachalam (1977: 9) claims to have identified an echo word in *nampi pimpi*, an expression implying a name of no significance, which appears in an eighth century Tamil text, the *Periyaalvaar*. There seem to be no equivalent early examples of *ki(i)-* forms, but Emeneau (1938: 116) reconstructs **ki(i)-/gi(i)-* for Proto-Dravidian on the basis of its distribution throughout the language family. Zvelebil (1990: 73) also notes that there is good evidence of echo formations at an early stage of Sora, which belongs to the Munda language family. There are, however, no indications of echo formations in Sanskrit or any of the related Indo-European languages, so it seems likely that they were borrowed into Middle Indo-Aryan from one of the neighbouring language families, possibly Dravidian.

2.3.4. Geographical distribution

The distribution of echo formations throughout the Indic languages is widely recognized,[22] but their occurrence is not restricted to the Indian subcontinent. The languages of Asia in general make heavy use of reduplicated structures, for a variety of syntactic and semantic purposes (Abbi 1992a: 135). These include instances variously referred to in the literature as 'imitative reduplication', 'chiming' or 'chameleon affixation'. Williams (1991), discussing this phenomenon, argues that they should be recognized as the same type of formation as the Indian echo words. He gives various examples, including an apparently regular pattern in Thai involving vowel change (Williams 1991: 108), illustrated in (77) and (78).

(77) *mɔ̂ɔ mɑ̂ə* 'pot echo' i.e. 'pots and pans'

(78) *hŭu hăə* 'ear echo' i.e. 'ears and stuff'

He also gives examples from Vietnamese, Khmer and Indonesian languages, most of which appear to belong to the phonologically unpredictable class of echo expressions. Yip (1982, 1998b) notes the occurrence of similar forms in Chinese, citing data from the Chaoyang and Tengxian dialects. Finally, Noyer (1998: 83) presents data on a productive echo pattern from northern Vietnamese that is semantically identical to the Indian examples, and involves fixed segments in the rhyme, as in (79) and (80).

(79) *học hiệc* 'study and stuff'

(80) *bạn biệc* 'friends and stuff'

A specific phonological parallel to the productive echo formations considered in this thesis is found to the west, in Turkish. The favoured echo segment is /m/, although cases with /p/ are also reported, and the associated semantics are precisely comparable to the Indian examples. The instances in (81) and (82) are taken from Marchand (1952: 65).

(81) *ayna mayna* 'mirrors etc.'

(82) *kitap mitap* 'books etc.'

According to Marchand, Brinzeu claims that these forms spread over the whole area covered by the early Ottoman empire, being adopted at that time into non-Turkic languages such as Greek and Armenian. This may well be the source of the *m*-echo words reported by Bruening (1997) for Abkhaz, a Caucasian language spoken in Georgia and Turkey. Finally, there is an equivalent Yiddish formation, involving the sequence /ʃm/, which has been borrowed into English, particularly American English. The connotation of contempt, noted in section 2.3.2 above, is especially strong in these examples.

2.4. Constraints on echo expressions

This section will examine some of the constraints on the formation of echo expressions, looking at issues such as which lexical categories may form the base. The morphological make-up of the base will also be investigated, particularly the interaction of echoing with inflectional and derivational morphology. The possibility that bases for echoing may be not only morphologically but syntactically complex is also considered: this is an issue that has not been addressed by mainstream theories of reduplication. All of these questions will be investigated in detail for Tamil in chapter 3: this section considers them firstly in relation to other Dravidian languages, and then the Indo-Aryan languages Hindi and Bengali.

[22] See, for example, Bloch (1934: 328), Emeneau (1956: 10), who describes them as a *'pan-Indic trait'*, Trivedi (1990: 51) and Abbi (1992a: 20).

2.4.1. Dravidian

Apart from brief comments in grammars, information is most readily available for two literary Dravidian languages: Telugu, in a monograph on reduplication and onomatopoeia in the language (Bhaskararao 1977), and Kannada, in a recent article devoted to echo reduplication (Lidz 2000). These data should be treated with some degree of caution: sources of information are generally not made explicit,[23] and claims may be based on the intuitions of only a single speaker. There is also a paper by Emeneau on echo words in the tribal language, Toda (Emeneau 1938), and reference to the same phenomenon in his grammar of Kolami (Emeneau 1955).

2.4.1.1. Which lexical categories can be echoed?

Echo formations in Telugu and Kannada seem to involve a fairly wide range of lexical categories: Abbi (1992a: 21–22) gives examples of echoed nouns, pronouns, verbs, adjectives and adverbs in Telugu. According to Sridhar's grammar (Sridhar 1990: 312), echo formation in Kannada applies to all classes of words except interrogative pronouns and demonstrative adjectives, and even includes some clitics. Murthy (1975: 161) mentions tighter restrictions, excluding pronouns, adjectives, interjections and numerals, but he too affirms the general productivity of echoing in Kannada. For both Toda and Kolami Emeneau only gives examples of echo formations involving nouns and verbs. These two categories appear to be the prototypical bases for echoing in Dravidian generally, although other lexical categories may also be involved.

2.4.1.2. Morphological constituency of echo bases

Emeneau's analysis of Toda includes some comment on the morphological constituency of the bases for echoing. For instance, he provides examples of case-marked nouns being echoed, with the case markers included in the base. (83) shows a dative case-marked noun corresponding to the nominative form *mox* 'child, boy, son', and (84) a locative corresponding to the nominative *meˈɳ* 'tree' (Emeneau 1938: 113). Square brackets are used from this point onwards to mark the base and echo.

(83) *moxk* 'to the child' [*mox(k)*][*kixk*]

(84) *meˈɳt̲s* 'in the tree' [*meˈɳt̲s*][*xiˈɳt̲s*]

In contrast with the nominal morphology, Emeneau states that verb complexes in Toda cannot be echoed in their entirety, but that echoing is restricted to secondary stems (Emeneau 1938: 114).

The situation in Kannada is fairly well-documented; both Murthy (1975: 162) and Lidz agree that nouns can be echoed with or without their case markers. Thus Lidz (2000: 148), for instance, presents the examples in (85), involving an accusative case marker, as equally acceptable variants.

(85a) [*baagilannu*] [*giigilannu*] *muchide*
door.acc echo close.past.1s
'I closed the door and related things.'

(b) [*baagil*] [*giigil*]*annu muchide*
door echo.acc close.past.1s

He also claims that the inclusion of inflectional morphology within the base for echoing is optional in verbs. Hence markers of tense, and also person and number endings, may be echoed with the verb root, as in (86a), but can alternatively follow the echo expression, as in (86b).

[23] The third footnote of Lidz's article does mention three informants.

(86a) *baagilannu* [*muchide*] [*gichide*] *anta heeɭabeeḍa*
door.acc close.past.1s echo that say.prohib
'Don't say that I closed the door or did related activities.'

(b) *baagilannu* [*much*] [*gich*]*ide anta heeɭabeeḍa*
door.acc close echo.past.1s that say.prohib

Murthy (1975: 162) also notes the possibility of echoing finite verb forms as a whole, and comments that it is rarely the primary verb of the sentence that is echoed in this way. Sridhar (1990: 313) provides some cases of verbal complexes being echoed, with an infinitive, as in (87), or participle, as in (88) and (89), serving as the base, and further verbal markers following the echo expression: he does not comment on whether or not the complex can be echoed in its entirety.

(87) [*ooḍa*] [*giiḍa*]*beeḍa*
run.infin echo.prohib
'Don't run or do such things.'

(88) [*bandu*] [*gind*][24]*iiye*
come.vbp echo.cont
'Don't you dare come (back) or something.'

(89) [*eddu*] [*giddu*]*biṯṯare kaṣṭa*
get-up.vbp echo perf.cond trouble
'(We'll be in) trouble if (he) gets up or something.'

Lidz also considers derivational markers, concluding that they generally interact with echoing in the same way as inflectional morphology. He gives examples of both valency changing and category changing morphology, such as the reflexive markers in (90), and the verbalizing use of the causative *-is(u)-* in (91).

(90a) *rashmi tannannu* [*hogaɭikoṇḍaɭu*] [*gigaɭikoṇḍaɭu*] *anta heeɭabeeḍa*
Rashmi self.acc praise.refl.past.3sf echo that say.prohib
'Don't say that Rashmi praised herself and did related activities.'

(b) *rashmi tannannu* [*hogaɭi*] [*gigaɭi*]*koṇḍaɭu anta heeɭabeeḍa*
Rashmi self.acc praise echo.refl.past.3sf that say.prohib

(91a) *rashmi vijayige* [*patris*] [*gitris*]*idaɭu anta heeɭabeeḍa*
Rashmi Vijay.dat letter.caus echo.past.3sf that say.prohib
'Don't say that Rashmi wrote Vijay a letter and did related activities.'

(b) *rashmi vijayige* [*patra*][*gitr*]*isidaɭu anta heeɭabeeḍa*
Rashmi Vijay.dat letter echo.caus.past.3sf that say.prohib

In the first pair the reflexive marker and tense and personal endings are kept together, but in the second the causative marker is divided from them by the echoing process. Either of these options is possible in either case, according to Lidz, who claims that echo reduplication:

> *'can apparently decide to break the word at any of its morpheme boundaries'* (Lidz 2000: 158).

Certain affixes, however, are exceptions to Lidz's generalization for Kannada, and prove impervious to echoing. These include the nominalizing *-ike-* suffix, which may attach at either the stem-level or the word-level, according to the analysis of Aronoff and Sridhar (1983: 12), and can in neither case be excluded from the base for echoing. This difference in behaviour amongst the affixes is explained by Lidz in terms of whether or not they are syntactically represented: more discussion of his interpretation of the data will follow in section 6.3.

[24] The final *-u* of the base has been elided before the following vowel.

2.4.1.3. Compounds

None of the references in the literature on Dravidian make specific mention of how compounds interact with echoing. There does, however, seem to be one example given by Lidz of an exocentric or coordinate compound type, where two semantically related verbs are juxtaposed in a pairing of the type discussed in section 2.2.4. Kannada appears to allow repetition of the whole compound, as in (92a), but not of the first element on its own, as in (92b) (Lidz 2000: 152).

(92a) [*ooḍ-aaṭa*] [*giiḍ-aaṭa*]
run.play echo
'running around and related activities'

(b) *[*ooḍ*] [*giiḍ*] *aaṭa*
run echo play

There are also a couple of examples of phrasal idioms that Lidz includes to make the point that echo reduplication is not restricted to semantically compositional elements. These are comparable to the Tamil compound verbs (see section 3.2.3.3.3), which are also frequently non-compositional in their semantics (Annamalai & Steever 1998: 124). Lidz (2000: 154) claims that either the noun+verb combination or just the noun can be echoed, as in (93a) and (93b) respectively.

(93a) *hari* [*kannu muchida*] [*ginnu muchida*]
Hari eye close.past.3sm echo
'Hari died (lit. closed his eyes).'

(b) *hari* [*kannu*] [*ginnu*] *muchida*
Hari eye echo close.past.3sm

2.4.1.4. Phrases

As Lidz's article argues, the possibility of phrasal echoing is of particular theoretical interest. He claims that echo reduplication in Kannada:

> '*applies equally to words, subparts of words and entire syntactic phrases*' (Lidz 2000: 146).

Only two examples of phrasal echoing are given to substantiate this claim (Lidz 2000: 149), so the real extent of the phenomenon in Kannada is hard to gauge. One example, given in (94), involves a postposition following a genitive case-marked noun.

(94) *pustavannu* [*meejina meele*] [*giijina meele*] *nooḍide*
book.acc table.gen on echo see.past.1s
'I saw the book on the table and in related places.'

The other example, in (95), involves a verb phrase containing an accusative case-marked object and past tense verb.

(95) *nannu* [*baagilannu muchide*] [*giigilannu muchide*] *anta heeḷabeeḍa*
I.nom door.acc close.past.1s echo that say.prohib
'Don't say that I closed the door or did related activities.'

There are passing references to echo phrases including more than one word for three other Dravidian languages, and in each case the example involves the negative marker. In Telugu, for example, Bhaskararao (1977: 9–10) gives the pair of sentences in (96), in which the negative marker *leedu* is optionally repeated.

(96a) *vaaḍiki* [*illuu leedu*] [*gilluu leedu*]
he.dat house not echo
'He has no house etc.'

(b) *vaaḍiki* [*illuu*] [*gilluu*] *leedu*
he.dat house echo not

An example from Malayalam following the same pattern is given in (97) (Malten 1989: 44).

(97) [*vaḍa illa*] [*kiḍa illa*]
vada not echo
'There are no vadas or other edible things here.'

Emeneau appears to present a parallel pair of sentences from Kolami (Emeneau 1955: 102).

(98a) [*me'kel to'tev*] [*gi'kel to'tev*]
goat not echo
'There are no goats at all.'

(b) [*me'kel*] [*gi'kel*] *ta'na to'tev*
goat echo emph not

There are indications, therefore, that the possibility of phrasal echoing is a general feature of the Dravidian languages, but the available evidence suggests that it is a marginal phenomenon restricted to certain negative contexts.

2.4.2. Hindi and Bengali

As noted in section 2.3, echo expressions are characteristic of the whole Indian subcontinent, occurring in all four of India's language families. Cross-linguistic phonological variation of the echo segment is well documented and has already been discussed: information about restrictions on the base for echoing is more scarce. Data are available, however, on two Indo-Aryan languages: there is a description of the phenomenon in Bengali in the doctoral thesis of Fitzpatrick-Cole (1994),[25] and Reynolds (1998) describes research on Hindi.

The Hindi data were collected from a small number of native speakers resident in Oxford, with the bulk of the material being provided by two female graduate students in their mid twenties. In each case the speaker had learnt Hindi in the home but their education had been predominantly in English. Accordingly, their use of Hindi was mainly confined to informal settings: they commented, for instance, that they would not feel comfortable about producing a formal piece of academic writing in Hindi. The interviews involved some direct questions about their use of echo words, but for the most part informants were asked to comment upon the acceptability of various echo forms in the context of simple sentences. These were presented orally or in Devanagari script: unlike Tamil (see section 3.2.2), there is no significant divergence between spoken and written Hindi. The next sections summarize the results of these interviews, comparing them with the data on Bengali reported by Fitzpatrick-Cole.

2.4.2.1. Which lexical categories can be echoed?

As reported in the literature for Hindi (Abbi 1992a: 23), echo forms based on nouns, adjectives and various verbal forms all proved acceptable to the informants. Echoing of pronouns, however, is said to be unacceptable in Hindi, and this was confirmed by the subjects' rejection of sentence (99).

(99) *[*ham*] [*vam*] *hindii siik*^{h} *rahe* *haĩ.*
we echo Hindi learn prog.mpl be.pres.3pl
'We are learning Hindi.'

Bengali allows nouns, adjectives and verbs to serve as echo bases, but differs from Hindi in permitting examples involving pronouns, such as (100) and (101) (Fitzpatrick-Cole (1994: 128).

(100) [*tomra*] [*tomra*]
you.pl echo

[25] Fitzpatrick-Cole (1996) presents much of the same data: where possible, reference will be made to this article as it is more widely available.

(101) [*o*] [*ʈo*]
he, she, it echo

Various prepositional and adverbial echo forms were also acceptable in Hindi, such as *niice viice* 'below and so forth' and *jaldii valdii* 'soon etc.'. As with the Dravidian languages, therefore, nouns and verbs are the prototypical cases, but various other lexical categories may be involved on a language-specific basis.

2.4.2.2. Morphological constituency of echo bases

Inflectional endings in Hindi, whether these appear on nouns, verbs or adjectives, are consistently included in the base for echoing. The nominal morphology differentiates three types of stem (commonly referred to as direct, oblique and vocative), each inflected for number and gender (singular or plural, masculine or feminine). For example, the *-e* in (102) signifies the masculine plural direct stem form of the noun *laʈkaa* 'boy', and has to be included in the base of the echo formation.

(102a) [*laʈke*] [*vaʈke*]
boy.mpl.dir echo

(b) *[*laʈk*] [*vaʈk*]*e*
boy echo.mpl.dir

Adjectives in Hindi also inflect to agree with the noun that they qualify in number and gender, and according to whether it is direct or oblique. Again subjects reject examples where these endings are excluded from the base for echoing.

(103a) [*piilii*] [*viilii*]
yellow.fs.dir echo

(b) *[*piil*] [*viil*]*ii*
yellow echo.fs.dir

Further case distinctions for the nouns are marked by the so-called 'case clitics' or case markers. These appear after oblique stems in a position which can otherwise only be occupied by postpositions (with the sole exception of locative destinations, which are followed by neither[26]). The distinction between case markers and simplex postpositions[27] is somewhat fuzzy, but both appear to be routinely excluded from the echo base. This is seen in (104) for the case marker *ko,* and in (105) for *tak* 'up to, until', which is generally classed as a postposition.

(104a) *[*laʈkõ ko*] [*vaʈkõ ko*] *kursiyãã do*
boy.mpl.obl dat[28] echo chair.fpl.obl give.imp
'Give the boys the chairs.'

(b) [*laʈkõ*] [*vaʈkõ*] *ko kursiyãã do*
boy.mpl.obl echo dat chair.fpl.obl give.imp

(105a) *[*mezõ tak*] [*vezõ tak*]
table.fpl.obl up-to echo
'up to the tables and so forth'

(b) [*mezõ*] [*vezõ*] *tak*
table.fpl.obl echo up-to

There are other indications that the association between oblique stem and case marker in Hindi is not particularly close: an orthographic gap often precedes the Hindi case markers, and they can be separated from the preceding word by a pause (Mohanan 1994: 60), or even other phonological material, such as the emphatic particle *hii*.

---

[26] Note that the oblique form of the second nominal in the example below is followed by neither case clitic nor postposition.

baccaa ɖaakk͟haane gayaa
child.ms.dir post-office.ms.obl go.perf.ms
'The child went to the post-office.'

[27] The majority of postpositions in Hindi are complex, e.g. *ke liye* 'for'.

[28] The glosses follow Mohanan (1994) in regarding *ko* as a marker of either accusative or dative case, depending upon context. In (105) there is no *ko* after the object *kursiyãã* due to a constraint against the double appearance of *ko* in any clause (see the discussion of identity avoidance in section 1.5.2.5).

The situation in Bengali is rather different, however: like Kannada, inflectional material and clitics may either be included in the echo base or added to the end of the echo formation, with the latter option apparently preferred (Fitzpatrick-Cole 1994: 154). (106b), for example, which contains only one occurrence of the plural marker, is acceptable, whilst (106a), in which it is repeated, is of dubious grammaticality, as indicated by the question mark.

(106a) ?[*bari-r*] [*ʈari-r*]
house.pl echo

(b) [*bari*] [*ʈari*]-*r*
house echo.pl

There appears to be some flexibility in the echoing of verbal forms: Fitzpatrick-Cole (1996: 316) gives the examples in (107) as acceptable alternatives in Bengali, the first taking the fully inflected word as the base, the second the verbal stem.

(107a) [*mer-e-č^h-i*] [*ʈer-e-č^h-i*]
beat.perf.pres.1pl echo

(b) [*mer-e*] [*ʈer-e*]-*č^h-i*
beat.perf echo.pres.1pl

Hindi again proves more restrictive: verbal inflections must be included in the base, as in (108).

(108a) [*bolaa*] [*volaa*]
speak.perf.ms echo

(b) *[*bol*] [*vol*]*aa*
speak echo.perf.ms

The same is true of the infinitival *-naa* ending, which is added to the verb root, and can itself be inflected.

(109a) *muj^{h}e* [*lik^{h}naa*] [*vik^{h}naa*] *pasand hai*
I.obl read.infin echo pleasing be.pres.3s
'I like reading.'

(b) **muj^{h}e* [*lik^h*] [*vik^h*]*naa pasand hai*
I.obl read echo.infin pleasing be.pres.3s

Auxiliary verbs are excluded from echo forms, and this is true even of the conjunctive particle *ke*, which, unlike the other auxiliary verbs, is orthographically joined to the preceding verbal root.

(110a) **mez ko saaf* [*karke*] [*varke*]
table.obl acc clean make.conj echo
'After cleaning the table, . . .'

(b) *mez ko saaf* [*kar*] [*var*]*ke*
table.obl acc clean make echo.conj

Hindi has a rich derivational morphology, which provides a further source of word-internal complexity. Unlike Kannada, echoing in Hindi is generally unable to break the boundary between a stem and its derivational affixes, as the following sets of examples demonstrate. The first two both contain suffixes: *-aa* is a transitivizing suffix that attaches to the verbal root *jal* 'burn (intrans.)', and *-pan* follows the nominal stem *laʈak-* 'boy' to create an abstract noun meaning 'boyhood'.

(111a) [*jalaa*] [*valaa*]

(b) *[*jal*] [*val*]*aa*

(112a) [*laʈakpan*] [*vaʈakpan*]

(b) *[*laʈak*] [*vaʈak*]*pan*

The same pattern is observed when the affix precedes, as in (113), where the prefix *paraa-* has the sense 'opposite to' and combines with the noun *jay* 'victory' to give the meaning 'defeat'. Again the nominal element cannot be echoed independently of the affix, nor, indeed, can the affix be echoed on its own.

(113a) [*paraajay*] [*varaajay*]

(b) **paraa*[*jay*] [*vay*]

One exception to the pattern thus far established involves the noun *pʰuuldaan* 'vase', given in (114), which contains a suffix *-daan*, indicating some kind of container, and the nominal stem *pʰuul* 'flower'.

(114a) ?[*pʰuuldaan*] [*vuuldaan*] (b) [*pʰuul*] [*vuul*]*daan*

(c) **pʰuul*[*daan*] [*vaan*]

The expected echo form would be *pʰuuldaan vuuldaan*, as in (114a), and this is indeed acceptable, although marginal: one informant commented that it is something of a tongue-twister, and gave *pʰuul vuuldaan*, in (114b), as the preferred form. There is apparently no semantic distinction involved, since both simply mean 'vases and such items'. As with all the affixed forms discussed, using the suffix on its own as the base for echoing, as in (114c), was decisively rejected.

The common *-vaalaa* suffix differs from the other affixes considered thus far in being able to attach not only to words but also phrases. It inflects for number, gender and the direct/oblique distinction, and is adjectival in function, although it is also frequently used to form substantives. When it attaches to a single element, this is echoed on its own, following the pattern seen in *pʰuul vuuldaan*. This can be an adverb, as in (115), where it follows *uupar* 'up, above', and gives the meaning 'upstairs', or a noun, such as *śahar* 'town' in (116). This combination forms an adjective meaning 'of the town', which can also be used substantivally to mean 'town-dweller'. In (117) it is in agentive function, attaching to the noun *miʈʰaaii* 'sweet' to mean 'sweet-seller'. An oblique infinitive can also form the base for *-vaalaa* in its agentive use, as in (118), where it attaches to *dauɽne*, the oblique infinitival form of the verb meaning 'run' to give 'one who runs, a runner'. In each case, interaction with echoing follows a consistent pattern.

(115) *[*uupar-vaalaa*] [*vuupar-vaalaa*] (b) [*uupar*] [*vuupar*]*vaalaa*

(116) *[*śahar-valaa*] [*vahar-vaalaa*] (b) [*śahar*] [*vahar*]*vaalaa*

(117) *[*miʈʰaaii-vaalaa*] [*viʈʰaaii-vaalaa*] (b) [*miʈʰaaii*] [*viʈʰaaii*]*vaalaa*

(118) *[*dauɽne-vaalaa*] [*vauɽne-vaalaa*] (b) [*dauɽne*] [*vauɽne*]*vaalaa*

The situation in Bengali appears to be slightly different from Hindi, where the properties of individual derivational affixes appear to have some effect on interaction with echoing. According to Fitzpatrick-Cole, any derivational morphology must be included in the base for echoing, although she gives only one example, *dudʰ-ɔla* 'milk-man' (Fitzpatrick-Cole 1994: 155). This is the direct equivalent of Hindi *duudʰ-vaalaa*, which echoes only the first element, without the agentive suffix: in Bengali, however, both components must be echoed, as demonstrated in (119).

(119a) [*dudʰ-ɔla*] [*vudʰ-ɔla*] (b) *[*dudʰ*] [*vudʰ*]*ɔla*

2.4.2.3. Compounds and complex predicates

In general, Hindi seems to favour echoing both elements of a compound together. For example, the endocentric compound noun *rasoii-gʰar* meaning 'kitchen' (literally 'meal-house') can be echoed as in (120a), but not (120b) or (120c).

(120a) [*rasoii gʰar*] [*vasoii gʰar*] (b) *[*rasoii*] [*vasoii*]*gʰar*

(c) *[*rasoii gʰar*] [*var*]

An exocentric example combining the antonyms *raat* 'night' and *din* 'day' to mean 'day and night, continually' also follows the same pattern, as shown in (121).

(121a) [*raat din*] [*vaat din*] (b) *[*raat*] [*vaat*] *din*

(c) **raat* [*din*] [*vin*]

There appears, however, to be further flexibility in some cases: all three possibilities are admitted for echoing *aaraam kursiyãã* 'arm-chairs', a noun+noun compound made up of *aaraam*, meaning 'rest, comfort' and the plural of *kursii*, meaning 'chair'.

(122a) [*aaraam kursiyãã*] [*vaaraam kursiyãã*] (b) [*aaraam*] [*vaaraam*] *kursiyãã*

(c) *aaraam* [*kursiyãã*] [*vursiyãã*]

As discussed in section 2.2.4, Hindi, like many other Indian languages, is rich in exocentric compounds comprising two synonyms, or near synonyms. Singh (1995) claims that these do not participate in echo formations, and this was confirmed by my subjects. Examples involving nouns (e.g. *baal bacce* 'family'), adjectives (e.g. *dublaa patlaa* 'thin and scrawny') and verbs (e.g. *gaate bajaate* 'festivities' lit. 'singing, dancing') were all tested, but echoing of either part, or both together, was unanimously rejected. Singh's explanation is that these 'redundant compounds', as he terms them, are the outputs of a rule of reduplication operating on a purely semantic (and not a phonological) level, and are therefore prohibited from undergoing any further reduplication process. It may simply be, however, that it would be superfluous to echo such pairs, since they already contain the semantic notion of generality.

Endocentric compounds in Bengali, according to Fitzpatrick-Cole (1996: 316), echo either the entire compound, or just the second stem. Her example is the noun+noun compound *gari g^{h}õra* meaning literally 'car-horse', which interacts with echoing as shown in (123), with either (a) or (c) acceptable, but not (b).

(123a) [*gari g^{h}õra*] [*ʧari g^{h}õra*] (b) *[*gari*] [*ʧari*] *g^{h}õra*

(c) *gari* [*g^{h}õra*] [*ʧõra*]

When an inflectional ending appears on the compound, such as the plural *-gulo*, this is preferentially attached after the echoed portion (Fitzpatrick-Cole 1994: 158). This follows the general pattern seen in Bengali (shown in 106 above), but again contrasts with Hindi, where the plural marker in (122) cannot be separated from the second stem by echoing.

The occurrence of complex predicates formed by the combination of verbs, adjectives or nouns with finite verbs is characteristic of Indo-Aryan languages in general, and considerable attention has been devoted to their syntactic properties.[29] In Hindi, both parts can be echoed individually, but not together, whilst in Bengali all three possibilities are apparently permitted. Hindi examples are given below, with a noun+verb sequence in (124), and a verb+verb predicate in (125).

(124a) **raam ko kahaanii* [*yaad aayii*] [*vaad aayii*]
Ram dat story memory come.perf.fs echo
'Ram remembered the story and so forth.'

(b) *raam ko kahaanii* [*yaad*] [*vaad*] *aayii*
Ram dat story memory echo come.perf.fs

(c) *raam ko kahaanii yaad* [*aayii*] [*vaayii*]
Ram dat story memory come.perf.fs echo

(125a) **pramod relgaaɽii se* [*utar aayaa*] [*vutar aayaa*]
Pramod train.obl from descend come.perf.ms echo
'Pramod came down out of the train etc.'

(b) *pramod relgaaɽii se* [*utar*] [*vutar*] *aayaa*
Pramod train.obl from descend echo come.perf.ms

[29] See Mohanan (1989) and Butt (1995) for analyses of complex predicates in Hindi and Urdu respectively, both in the framework of Lexical Functional Grammar.

(c) *pramod relgaaɽii se utar* [*aayaa*] [*vaayaa*]
Pramod train.obl from descend come.perf.ms echo

The Bengali example given by Fitzpatrick-Cole (1996: 316) involves the complex predicate meaning 'bathe', containing *čan* 'bath' combined with *kor*, meaning 'do'.

(126a) [*čan kore*] [*ʈan kore*] (b) [*čan*] [*ʈan*] *kore*

(c) *čan* [*kore*] [*ʈore*]

2.4.2.4. Phrases

Thus far, Bengali has proved generally more permissive than Hindi in the restrictions it places on echoing, allowing stems, for instance, to be echoed independently of their inflectional endings. Subparts of words can therefore be echoed more readily than in Hindi, and this is also true of sequences of words: Bengali apparently permits some instances of phrasal echoing, whilst the Hindi subjects were not prepared to echo sequences comprising more than one lexical item. Echoed adjectival phrases were rejected, even when they fell together within the scope of the *-vaalaa* suffix, which functions as a phrasal clitic in (127).

(127) *niilii ããkʰõ -vaalii laɽkiyãã*
blue.fpl.obl eye.pl.obl suff.fpl.dir girl.fpl.dir
'girls with blue eyes'

When asked to echo example (127), subjects offered both (128a) and (128b). Interestingly the meaning given in (128a) holds for (128b) also, and any other translation, such as 'girls with blue eyes and other blue features' was strongly rejected. This suggests that the semantics is tied to the phrase as a whole, although only part of it can be echoed.

(128a) [*niilii*] [*viilii*] *ããkʰõ -vaalii laɽkiyãã*
blue.fpl.obl echo eye.pl.obl suff.fpl.dir girl.fpl.dir
'girls with blue and other light-coloured eyes'

(b) *niilii* [*ããkʰo*] [*vããkʰo*]*-vaalii laɽkiyãã*
blue.fpl.obl eye.pl.obl echo suff.fpl.dir girl.fpl.dir

This contrasts with the situation in Bengali, where adjective+noun sequences are involved in all of Fitzpatrick-Cole's examples of phrasal reduplication. She describes the possibility of echoing either the noun or the adjective or the two together for the phrase *kalo makorša* 'black spider(s)', with the translations given in (129–131) (Fitzpatrick-Cole 1996: 329).

(129) [*kalo makorša*] [*ʈalo makorša*]
'black and other coloured spiders'

(130) [*kalo*] [*ʈalo*] *makorša*
'black and other coloured spiders'

(131) *kalo* [*makorša*] [*ʈakorša*]
'black spiders and other (not necessarily black) beasties'

Fitzpatrick-Cole (1996: 344) makes much of the fact that echoing noun and adjective together has the same semantics as echoing just the adjective, rather than a composite meaning such as 'black and other coloured spiders and other beasties'. In this case the second element does not seem to contribute to the semantics: in the Hindi example given in (128b), the semantics seems to be determined by the previous word, rather than the one that is echoed. Both suggest that the semantics of the echo formation is not rigidly tied to the part of the adjectival phrase that is actually repeated.

The pattern already seen in simple nouns and compounds of preferentially excluding inflectional elements from echo formations is apparently true also of adjective+noun phrases

in Bengali. Fitzpatrick-Cole (1996: 345) gives the following example, in which the genitive *-r* marker appears only at the end of the reduplicated form.

(132) [*notun gɔʏna*] [*ʈotun gɔʏna*]*r*
'of the new, shiny, attractive jewellery'

Fitzpatrick-Cole's other examples concern the slightly more complex situation of having either a further adjective or an adverb precede the adjective+noun sequence. In the former case, the second adjective and noun are echoed together if the first adjective modifies both elements; if the first adjective modifies only the second, then these can be echoed together to the exclusion of the noun. For example, for the sequence *halka lal šari* 'light red sari', (133) and (134) are both possible echo forms, with the translations as given (Fitzpatrick-Cole 1996: 340).

(133) *halka* [*lal šari*] [*ʈal šari*]
'light sari which is red and so forth'

(134) [*halka lal*] [*ʈalka lal*] *šari*
'sari which is light red and so forth'

Fitzpatrick-Cole also discusses the echoing of verb phrases consisting of a nominal object and verb: these, she claims, cannot be echoed, unless they form a complex predicate, and gives the example in (135) (Fitzpatrick-Cole 1996: 331).

(135) *[*gɔrommɔšla čiboto*] [*ʈɔrommɔšla čiboto*]
whole spice chew.past.habit.2 echo

As in Kannada, therefore, the available evidence suggests that phrasal reduplication is a somewhat marginal phenomenon, apparently confined to adjectival and postpositional phrases in Bengali.

2.4.3. Conclusion

Drawing all this evidence together, is it possible to give a consistent characterization of the base for echoing in either Bengali or Hindi? The Bengali data show considerable variation in the make-up of the base, ranging from subparts of words to phrases. Fitzpatrick-Cole accordingly does not offer a precise characterization of the base, but suggests instead bounds on the formation of echo expressions. The base is said to be minimally a phonological stem and the whole expression maximally a phonological phrase (Fitzpatrick-Cole 1996: 330) (see sections 3.4 and 6.2 for further discussion of her proposals).

The Hindi bases, by contrast, show much less variation in their constituency, corresponding in each case to some word-like unit. Which of the various competing definitions of the word fits the data most closely? There is some evidence that a morphological characterization of the base is inadequate from examples involving the echoing of compounds, such as *aaraam kursiyãã vursiyãã* (example 122c). In such a case prosodic and morphological structure diverge:[30] the plural marker is morphologically attached to the compound as a whole, although its phonological shape is determined solely by the second element. The two structures are represented in (136), with the morphological structure on the left and prosodic on the right.

(136) [*aaraam kursi*]*-yãã* [*aaraam*] [*kursiyãã*]

When the second element, i.e. *kursiyãã*, is taken as the base for echoing it cuts across the morphological structure, not corresponding to any single morphological constituent. Instead, it seems to be the prosodic structure that is relevant, with the base corresponding in each case

[30] The notion that morphological and prosodic structure coexist, and are not necessarily isomorphic, is developed in Inkelas (1990).

to a whole prosodic constituent. The prosodic word seems the most likely candidate for the base: this is certainly consistent with the inclusion of inflectional and derivational morphology within the base, and the exclusion of case clitics and auxiliary verbs. One apparent exception to the inclusion of derivational affixes is *p^{h}uul vuuldaan*, the echo form of *p^{h}uuldaan* 'vase', where the suffix seems to be preferentially excluded from the base. The generalization that the base always corresponds to a prosodic word can, however, be maintained if a nested prosodic structure is permitted (see section 6.2 for discussion of this point). This is shown in (137), with the suffix attaching to a prosodic word (represented by the subscript omega). Since the suffixed form as a whole is also a prosodic word, this too can be echoed, although informants indicate that *p^{h}uul vuuldaan* is preferred.

(137) [[*p^{h}uul*]$_\omega$-*daan*]]$_\omega$

In contrast, the *-vaalaa* suffix is consistently excluded from the base, which fits with evidence suggesting that it is essentially a phrasal, rather than word-level, clitic. Those compounds where either individual elements or the compound as a whole can form the base for echoing may be assumed to have the prosodic structure illustrated in (138).

(138) [[*aaraam*]$_\omega$ [*kursiyãã*]$_\omega$]$_\omega$

It appears, therefore, that the base for echoing in Hindi corresponds consistently to a single phonological constituent, which can be identified as the prosodic word (for further exemplification and argumentation see Reynolds 1998).

The next chapter will investigate whether the base for echoing in Tamil is amenable to a similar characterization. It will consider the same kinds of issues that have been raised in this section in relation to other languages, presenting data gathered from native speakers of Tamil in response to a questionnaire, and analyzing the results.

Chapter 3: Investigation of constraints on Tamil echo expressions

3.1. Introduction

The existence in Tamil of reduplicative constructions in general and echo expressions in particular has attracted some comment in the literature. Several short articles have been devoted to the topic (Gnanasundaram 1972,[1] Arunachalam 1977 and Kothandaraman 1983), and there are brief descriptions of the phenomenon in most modern grammars. A couple of books on reduplication also touch on the issue: Abbi (1992a) discusses reduplication in Indian languages in general, and Malten (1989) focusses on Tamil, but looks predominantly at the formal variety. Despite this attention, however, there has been no systematic attempt to establish in any detail what restrictions may apply to the formation of echo expressions. Data were therefore collected in an attempt to resolve this issue, by means of a questionnaire presented to native speaker informants. The questions were designed to elicit data on several aspects of the echo formations, including the categories of word that could be echoed and their morphological make-up. The method of data collection is described in section 3.2, looking firstly at the subjects who were interviewed and then at the design of the materials used. Section 3.3 presents the results from the questionnaire and section 3.4 discusses what they indicate about the morphological and syntactic constituency of the Tamil echo expressions.

3.2. Data collection

3.2.1. Subjects

The data were collected through a series of interviews conducted with twelve speakers of Tamil. In each case, the informant was presented with a questionnaire containing examples of echo formations and asked to comment upon their acceptability. Subjects wore unobtrusive lavalier microphones (Audio-technica AT803b), allowing the interviews to be recorded for subsequent analysis. A couple of pilot runs were held in Oxford using a slightly different set of questions, but the bulk of the material was gathered in Pondicherry, a Tamil-speaking town south of Madras, during a field trip.

Each subject was asked a short series of questions so that a basic profile could be compiled, and some of the information is summarized in table 1. Languages are listed in descending order of competence, with a key to the abbreviations given beneath the table. As this shows, ages ranged between 26 and 45, and there was a mixture of male and female subjects, although the former predominated by a ratio of 2:1. Educational backgrounds varied, with a large proportion of the subjects engaged in doctoral research (anthropological or literary), but there were others with much more basic schooling: all were literate, however. The two described as sanitary assistants, a mother and daughter, were exceptional in not having Tamil as the language spoken in their home. They both claimed Kannada, a Dravidian language spoken in the neighbouring state of Karnataka, as their first language. The mother, however, had lived in Pondicherry since the age of fifteen, and the daughter for the whole of her life, and both were fluent Tamil speakers. As discussed in section 3.3.1, their responses proved not to be sufficiently different from those of the rest of the subjects to warrant exclusion.

The first ten subjects were all residents of Pondicherry, and most had lived there, or very nearby, for the whole of their lives. Moreover, the first of the subjects interviewed in Oxford (subject 11) had also grown up in Pondicherry and lived there until the age of 18. The set of subjects was therefore intentionally homogeneous from a regional perspective, to try to eliminate the potentially confounding factor of dialectal variations associated with different regions. The only exception was subject 12, who had lived throughout his childhood in the

[1] Note, however, that Gnanasundaram's definition of echo words is much broader than that employed here: he includes within its scope onomatopoeic expressions of all sorts.

Tanjore district of Tamilnadu. His responses, however, did not prove markedly different from those of the other subjects, and were therefore included in the analysis.

Table 1. Subject profiles

Subject	Gender	Age	Occupation	Languages
1	M	40	Academic	Ta, En, Fr
2	M	36	Librarian	Ta, En
3	M	45	Clerical worker	Ta, En
4	F	45	Sanitary assistant	Ka, Ta, En
5	M	28	Maintenance worker	Ta, En, Ma, Fr
6	F	26	Ph.D. student	Ta, En, Hi, Sa, Fr
7	M	27	Ph.D. student	Ta, En, Te, Fr
8	F	27	Sanitary assistant	Ka, Ta, En
9	M	36	Shopkeeper	Ta, En
10	F	26	Ph.D. student	Ta, En, Te
11	M	29	Ph.D. student	Ta, En, Fr
12	M	26	Ph.D. student	Ta, En, Hi, Ka

Language abbreviations: Ta = Tamil, En = English, Fr = French, Ka = Kannada, Ma = Malayalam, Hi = Hindi, Sa = Sanskrit, Te = Telugu

3.2.2. Design of the questionnaire

The questionnaire comprised a total of 105 sentences, grouped into 49 sets. The full set of sentences is reproduced in appendix D, with glosses[2] and translations. The sets contained between one and five sentences, and in most cases there is a single translation (given under the first gloss in the appendix). The sentences were differentiated by minor variations in the form of the echo expressions that were not expected to have any significant effect on the semantics. Subjects were asked to consider each sentence and indicate which option or options they might use and would regard as acceptable. In the minority of cases in which only a single sentence appeared under each number the informant was asked simply to comment on its acceptability. The informants' command of English varied considerably, so some of the interviews were conducted partially in Tamil, and most subjects were unwilling to make detailed comments on the appropriateness of the English translations provided.

The sentences were presented to the subjects in written form, in Tamil orthography accompanied by an English translation. As mentioned in section 2.3.2, Tamil has both a colloquial, spoken variety and a formal, literary variety used for writing, and the two are sufficiently divergent for the language to be classed as diglossic (see, for example, Britto 1986). Echo words are acceptable only in the former, which is largely unwritten, so the status of the written sentences in the questionnaire requires explanation.

The written or literary form of the language still largely conforms to standards set in the thirteenth century by the Tamil grammarian Pavanandi. Some changes have occurred since, such as the relaxation of sandhi rules, so that it is possible to distinguish, as Britto does, between modern Literary Tamil and Classical (or Pandit) Tamil (Britto 1986: 130), but the differences are minor. Literary Tamil (sometimes referred to as *centamiḻ*) is used for spoken communication in various formal settings, including radio news broadcasts, and in almost all

[2] A key to the abbreviations used in glosses is given in appendix C.

written media. Spoken colloquial Tamil (*koṭuntamiḻ*) is used for all other purposes, and shows considerable regional and caste-based variation. There is, however, a growing consensus that some kind of standard variety is emerging,[3] based on, though not identical to, the eastern dialect. Zvelebil, for example, commenting on the gradual divergence of colloquial Tamil from the standardized classical variety, says:

> *'During the period of the actual development of common spoken Tamil as a system, the centres of prestige are . . . in Tanjore, in Tiruchirapalli, in Puduchcheri* (i.e. Pondicherry), *in short, in the so-called Carnatic, and finally in Madras, that is in the regions where the Eastern and Northern dialects are spoken.'* (Zvelebil 1964: 249).

According to Schiffman, a key context for its development has been college hostels housing students who speak very different, even mutually unintelligible, regional dialects, whilst 'social' films have helped in its dissemination. There is, however, no formal standardization of this variety, sometimes referred to as Standard Spoken Tamil (SST), and the circumstances in which it is written down are very restricted.

In the public domain, conversational portions of novels and short stories by modern social-realism writers have included spoken forms. Writers vary, however, in whether they use it for all dialogue or reserve it for particular characters, usually buffoons, who are then contrasted with more heroic characters speaking literary Tamil. There is no pressure for consistency in spelling, which varies considerably. As Schiffman comments:

> *'nobody will be offended if one sentence contains* mudalilee *'at first', and another sentence* modallee*'* (Schiffman 1998: 377).

Representations of spoken Tamil can also be found in pirated printed versions of the sound tracks of popular films (although the scripts that the actors work from are usually in literary Tamil), and also in cartoons. Privately Tamil speakers might represent their own dialect in Tamil, or even Roman, script when writing personal letters or, in some cases, diary entries (Schiffman, p.c.), but this is not common practice.

Various factors militate against written representations of spoken Tamil being purely phonetic, as Schiffman and Arokianathan (1986) note. One is the limitations imposed by the orthography, which is unable to represent directly certain features of the spoken language, such as distinctive voicing in stops (this will be discussed at greater length in section 5.2.1) and nasalized vowels. Scholars vary in the importance they attach to this: Britto, for example, comments:

> *'In contrast to spoken H,*[4] *it is extremely difficult to represent spoken L in writing'* (Britto 1986: 179).

and Zvelebil (1964: 261) argues that the graphemic system requires adjustment. Schiffman and Arokianathan (1986), in contrast, play down the technical difficulties, arguing that word pairs where voicing is minimally contrastive are rare, and that nasalization is always predictable from environment. They ascribe more importance to the desire for readability, a consideration which favours literary forms and spellings, so that texts do not look entirely unfamiliar to those who are literate in the formal variety. Another force for conservatism, at least in writing for the public domain, is the need for representations of spoken Tamil to be accessible to readers speaking a wide range of dialects. The result is that texts tend to include certain forms characteristic of the spoken language, but remain conservative in other aspects. As Schiffman and Arokianathan observe:

---

[3] See, for example, Zvelebil (1964: 257), Britto (1986: 130), Karunakaran and Jeya (1995: 142) and Schiffman (1999: 1).

[4] Britto uses H(igh) and L(ow) to refer to formal and colloquial Tamil respectively, following Ferguson's terminology for varieties in a diglossic situation (Ferguson 1959).

> *'Salting an utterance with these 'marked' spoken forms . . . insures that speakers will understand that the utterance is primarily a spoken one without jeopardizing the readability of the whole passage.'* (Schiffman & Arokianathan 1986: 374).

They also note some interesting asymmetries in representation: grammatical morphemes are represented as phonetically as possible, but lexical morphemes tend to be given in their literary forms, whilst verbs are more likely to be represented phonetically than nouns.

In the questionnaire, the echo forms clearly marked the sentences as belonging to colloquial Tamil, and they also contained various other 'marked' spoken features, predominantly in grammatical morphemes and verbs. For instance, the spoken form *varaar* 'he comes' was used in place of the formal equivalent *varukiraar*; a reduced marker was used for the negative of the infinitive (*-le*, rather than *-villai*) and the quotative particle *-ɳɳu*, which follows reported speech, was consistently used in place of the full participle form *enru*. Likewise, phonetic spellings intended to reflect spoken forms were employed, such as *kuʈukkaatee* 'do not give', in place of formal *koʈukkaatee*, *onnum* 'anything' for *onrum* and *patti* 'about' for *parri*. Subjects were encouraged to read out each of the sentences aloud before passing judgement, and had no difficulties in understanding them. Moreover, in a separate exercise involving a subset of the subjects, they were sufficiently comfortable with the notion of representing colloquial Tamil in orthography to help 'translate' a passage from literary into spoken Tamil.

3.2.3. Issues addressed by the questionnaire

In this section I shall survey the issues that the questionnaire was designed to address, identifying the pertinent sentences and reviewing any relevant comments in the literature. A summary table is provided in section 3.2.3.6, listing the main questions and corresponding sentences in the questionnaire.

3.2.3.1. Context

The negative, and also indefinite, connotations of echo forms have been discussed in section 2.3.2, and these determine, at least in part, the kinds of contexts in which they may appear. Steever discusses this explicitly, stating that echo expressions occur in rhetorically marked settings containing modal verb forms (in which he includes future tenses and conditional forms), or in negative and interrogative, but not indicative, contexts (Steever 1987: 745). He provides the minimal pair given in (1) and (2), in which the first sentence is said to be acceptable but not the second, as an illustration of his point.

(1) [*maaʈu*] [*kiiʈu*] *varum*
cow echo come.fut.3n
'Cows and such will come.'

(2) *[*maaʈu*] [*kiiʈu*] *vantatu*
cow echo come.past.3n
'Cows and such came.'

Examples given by other writers largely conform to Steever's conditions, and have predominantly negative contexts. In order to test whether subjects would indeed reject echo words in a positive context, the first two sentences of the questionnaire contrasted a negative with a positive context for the echoing of an infinitive.

(3) *avanaal* [*ooʈa*] [*kiiʈa*] *muʈiyaatu* (= Q1)[5]
he.instr run.infin echo can.neg.3sn
'He can't run and so forth.'

[5] Whenever a sentence in the text also appears in the questionnaire, the corresponding questionnaire (Q) number will be given afterwards in brackets, so that it can be identified in appendix D, where the full text of the questionnaire is given.

(4) *avanaal* [*ooʈa*] [*kiiʈa*] *muʈiyum* (= Q2)
he.instr run.infin echo can.3sn
'He can run and so forth.'

A couple of other positive indicative sentences were originally included in the initial pilot questionnaire, but were firmly rejected on the grounds that they were either too positive or too specific. Subsequently, therefore, in the interests of keeping the questionnaire to a manageable length, these extra examples were excluded in favour of sentences designed to investigate morphological constraints.

A further, more specific set of conditions for the echoing of tensed verbs is mentioned by Raghunathan (1995: 59)[6] and also Malten (1989: 44). Raghunathan states that they are restricted to embedded clauses and must be followed by an imperative carrying a sense of prohibition, giving the following pair of examples to demonstrate his point:

(5) *avanai aʈikkaaram* [*paɳɳalaam*] [*kiɳɳalaam*] *enru ninaikkaatee*
he.acc beating do.infin.hort echo say.vbp think.neg.imp
'Don't think that you can order him or do some such thing.'

(6) **avanai aʈikkaaram* [*paɳɳalaam*] [*kiɳɳalaam*] *enru ninai*
he.acc beating do.infin.hort echo say.vbp think
'Think that you can order him or do some such thing.'

Similar examples given by other writers adhere to this pattern, as do all the sentences in which tensed verbs are echoed in the questionnaire.

3.2.3.2. Which lexical categories can be echoed?

3.2.3.2.1. Nouns

It is uncontroversial that echo formations are built primarily on lexical or content words, and that grammatical or function words generally do not serve as bases. This makes sense in terms of the semantics of echoing, which would not be easily compatible with words that have no lexical content. Comments within the literature differ, however, on how many lexical categories can participate in these formations in Tamil. Nagarajan represents the permissive end of the spectrum, stating that:

> '*It is possible for any linguistic item, irrespective of its word class in Tamil, to have a corresponding echo form.*' (Nagarajan 1994: 165–166).

Others propose various restrictions, although all the writers surveyed mentioned expressions built on nouns and verbs, the categories typical of Dravidian languages (see section 2.4.1.1). Proper nouns present no problem, according to the literature, although Kothandaraman (1983: 167) notes that nominal predicates and individual words within a noun phrase cannot be echoed.

There are numerous examples of nominal expressions in the questionnaire, the majority of which are case-marked and so will be discussed in section 3.2.3.3.1. A few examples of common nouns without case-marking are given in (7)–(10).

(7) [*aaʈu*] [*kiiʈu*] *meeykka marakkaatee* (= Q19a)
goat echo graze.infin forget.neg.imp
'Don't forget to graze the goats and such like.'

(8) [*maaʈu*] [*kiiʈu*] *meeykka marakkaatee* (= Q19b)
cow echo graze.infin forget.neg.imp

[6] Raghunathan's article describes the echo words as surviving examples of the serial verb constructions that were used more frequently in Old Tamil. He assumes a very unconventional definition of serial verbs, which are usually taken to involve verbs that are lexically distinct, although their functions in the construction may be aspectual.

'Don't forget to graze the cows and such like.'

(9) *enakku un [pai] [kii] kuɻukkaatee* (= Q30a)
I.obl.dat you.obl bag echo give.neg.imp
'Don't give me your bag and so forth.'

(10) *innikki [appaa] [kippaa] varaar -ŋŋu collaatee* (= Q48c)
today father echo come.pres.3pl quote say.neg.imp
'Don't say "Father and so on are coming today".'

Various echo forms built on proper nouns, both names of places and of people, also feature in the questionnaire. The two examples not involving case-marking are both formed from the man's name 'Kumar' and are given in (11) and (12).

(11) *[kumaar] [kimaar] paakkale -ŋŋu poy collaatee* (= Q41)
Kumar echo see.infin.neg quote lie say.neg.imp
'Don't lie that Kumar or someone like that didn't see.'

(12) *[kumaar] [kimaar] onnum paattaan -ŋŋu poy collaatee* (= Q46b)
Kumar echo anything see.past.3sm quote lie say.neg.imp
'Don't lie that Kumar or someone like that saw anything.'

3.2.3.2.2. Pronouns

There is some disagreement over whether pronouns can be echoed. Abbi claims that such forms are possible, giving *avan kivan* 'he, etc.' (third person singular masculine) and *nii kii* 'you, etc.' (second person singular) as examples (Abbi 1992a: 21).[7] Kothandaraman (1983: 168) and Annamalai and Steever (1998: 125), however, specifically rule out the echoing of pronouns. Malten (1989: 44) agrees with this as a general principle, but offers one supposedly acceptable example, given in (13).

(13) *[naan] [kiin] enru peecikkiṯṯirukkaatee* 'Don't always speak only of yourself!'

A colloquial equivalent of this sentence was included in the questionnaire to test Malten's claim:

(14) *[naan] [kiin]-ŋŋu peecaatee* (= Q29)
I echo quote speak.neg.imp
'Don't speak only of yourself!'

3.2.3.2.3. Verbs

The clearest cases in the literature of verbs forming echo expressions involve participles. Such examples are given in various places: for instance, *vantu kintu* 'coming and so forth' (Abbi 1992a: 21), *viɻuntu kiɻuntu* 'falling or any such thing' (Raghunathan 1995: 59) and *pooyṯṯu kiiṯṯu*[8] 'going and other activities' (Schiffman 1999: 172). An infinitive also features in one of Malten's sentences, and Raghunathan's example reproduced in (5) above contains a modal form. As discussed in 3.2.3.1, it is generally recognized that tensed finite verbs can legitimately be echoed, but only under the specific conditions set out there. As with the nouns, most of the sentences involving verbal forms are designed to investigate specific morphological issues, and are therefore discussed in section 3.2.3.3.3 below. Examples of infinitives are included, both of the so-called 'weak' verbs, exemplified in (3) above, and also of the 'strong' verbs, which take a *-kka* infinitival marker.

[7] Abbi's transliterations have been slightly altered in line with the conventions set out in appendix A: in particular, the instances of <g> have been replaced with <k>. Section 5.2.1 describes the distribution of voiced and voiceless obstruents in Tamil, and section 5.5 presents experimental evidence for the acoustic nature of the echo /k/ segments.

[8] This is a complex participle form, containing a completive marker from the aspectual verb *(v)iṯu* (see section 3.2.3.2.3).

(15) *iŋkee [paṭikka] [kiṭikka] kuuṭaatu* (= Q33a)
here study.infin echo should.neg.3sn
'One should not study and so forth in this place!'

A negative infinitive is found in (16), and an imperative with negative marking in (17).

(16) *avanooṭu [caappiṭale] [kiippiṭale]* (= Q34a)
he.soc eat.infin.neg echo
'I don't eat and so forth with him.'

(17) *nii [peecaatee] [kiicaatee]* (= Q32a)
you speak.neg.imp echo
'Don't speak and so forth!'

Finally, there are several examples of tensed finite verbs, including the following:

(18) *enakku [teriyaatu] [kiriyaatu]-ŋŋu collaatee* (= Q39b)
I.obl.dat know.neg.3sn echo quote say.neg.imp
'Don't say "I don't know!".'

(19) *kumaarukku [kuṭutteen] [kiṭutteen]-ŋŋu poy collaatee* (= Q44)
Kumar.dat give.past.1s echo quote lie say.neg.imp
'Don't lie that you gave it or something to Kumar.'

(20) *innikki appaa [varaar] [kiraar]-ŋŋu collaatee* (= Q48b)
today father come.pres.3pl echo quote say.neg.imp
'Don't say "Father's coming today and so forth".'

(21) *kumaar onnum [paattaan] [kiittaan]-ŋŋu poy collaatee* (= Q46c)
Kumar anything see.past.3sm echo quote lie say.neg.imp
'Don't lie that Kumar saw anything and so forth.'

3.2.3.2.4. Adjectives and adverbs

There is only a handful of underived adjectives in Tamil: most are transparently denominal or deverbal. Moreover, adverbs are formed by the addition of *-aa(y)* or *-aaka* to a noun or denominal adjective,[9] leading Schiffman to comment that:

> '*There are no true adverbs in Tamil, i.e. none that I would list as such in the dictionary.*' (Schiffman 1999: 141).

Both Kothandaraman (1983: 167) and Malten (1989: 45) specifically exclude adjectives from the group of potential bases for echoing, the latter on the grounds that they are too closely connected to the following noun to be echoed alone. Abbi, in contrast, provides the pair of examples given in (22) and (23) (Abbi 1992a: 21).

(22) *pacce kicce* 'green and the like'

(23) *motti kitti* 'fat and the like'

The questionnaire includes four sets of sentences containing adjectives, with the option in each case of either echoing the adjective and noun together or the adjective alone.

(24) *enakku [civappu toppi] [kivappu toppi] piṭikkaatu* (= Q24b)
I.obl.dat red hat echo like.neg.3sn
'I don't like hats that are red etc.'

(25) *enakku [civappu] [kivappu] toppi piṭikkaatu* (= Q24a)
I.obl.dat red echo hat like.neg.3sn

[9] Schiffman (1999: 141) notes one exception: *nallaa* 'well', which is formed from the 'true' adjective *nalla* 'good'.

(26) [*vellai caṭṭai*] [*killai caṭṭai*] *vaaŋkaatee* (= Q25a)
white shirt echo buy.neg.imp
'Don't buy white shirts and such like.'

(27) [*vellai*] [*killai*] *caṭṭai vaaŋkaatee* (= Q25b)
white echo shirt buy.neg.imp

(28) [*keṭṭa kanavai*] [*kiṭṭa kanavai*] *keekkaatee* (= Q26a)
bad dream.acc echo listen.neg.imp
'Don't heed bad dreams and so forth.'

(29) [*keṭṭa*] [*kiṭṭa*] *kanavai keekkaatee* (= Q26b)
bad echo dream.acc listen.neg.imp

(30) *avan* [*nalla paiyan*] [*killa paiyan*]*-ŋŋu nampaatee* (= Q27a)
he good boy echo quote believe.neg.imp
'Don't believe that he's a good boy and so forth.'

(31) *avan* [*nalla*] [*killa*] *paiyan-ŋŋu nampaatee* (= Q27b)
he good echo boy quote believe.neg.imp

There is also one example containing the possessive adjective *un* 'your', testing whether or not it is possible to echo this in combination with the noun it qualifies.

(32) *enakku* [*un pai*] [*kii pai*] *kuṭukkaatee* (= Q30b)
I.obl.dat you.obl bag echo give.neg.imp
'Don't give me your bag and so forth.'

Adverbs are mentioned specifically only by Kothandaraman (1983: 166), who claims that they can be echoed, giving *metuvaa kituvaa* 'slowly and something like that', and *koopamaa kiipamaa* 'angrily and something like that' as examples. Echo forms based on one adverb are included in the questionnaire: in (33) the adverbial marker is included in the base for echoing, but in (34) it is the nominal stem that forms the base.

(33) [*cantoocamaaka*] [*kintoocamaaka*] *paaṭaatee* (= Q28a)
joy.adv echo sing.neg.imp
'Don't sing joyfully etc.'

(34) [*cantoocam*] [*kintoocam*]*aaka paaṭaatee* (= Q28b)
joy echo.adv sing.neg.imp

One further lexical category – postpositions – are discussed in section 3.2.3.3.2 below because of their close affinity with the case markers that form the subject of the next section.

3.2.3.3. Morphological constituency of echo bases

3.2.3.3.1. Case markers

The interaction of inflection with echoing is an important diagnostic of morphological constituency. Tamil is generally described as having two nominal stem forms or bases, the nominative and oblique (see, for instance, Zvelebil 1990: 21). Distinct oblique forms exist only for a subset of nouns, and are predictable from the phonological make-up of the segments at the end of the nominative stem.[10] If there is an oblique form, it is to this that the case markers are added; otherwise, markers are added directly to the nominative stem. Both stem forms, however, can appear independently of case markers, the oblique form being used adjectivally or with genitival function.

[10] Nouns ending in *-am* have corresponding oblique forms ending in *-attu*, those ending in *-ru* have oblique forms in *-ttu*, and certain nouns ending in *-ṭu* have obliques in *-ṭṭu*: full details can be found in Lehmann (1989: 14–17).

Traditional descriptions of Tamil, which are strongly influenced by the Sanskrit grammarians, identify eight cases, some with more than one form of the marker. Rhenius (1845: 16–17), for example, lists these as nominative, accusative, ablative (instrumental and social), dative, ablative of separation or motion, genitive, ablative of place[11] and vocative. There is considerable disagreement, however, in modern analyses of the language over how many cases should be recognized, and what they should be termed. In addition to the eight candidates given above, benefactive, purposive and sociative (or associative) are all to be found in different lists. This confusion is partly the result of using categories that were originally devised for an unrelated language, and is compounded by the differences between the modern spoken variety and the ancient formal language to which the case system was first applied.

Although there is little direct comment in the literature on how inflection interacts with echoing, the implication that can be drawn from the few examples given is that case markers are included within the part that is echoed. Arunachalam (1977: 9) articulates this view explicitly, as does Kothandaraman (1983: 167), although he makes an unexplained exception for the genitive. Malten, however, claims that both case suffixes and postpositions can appear either on base and echo portion, or only on the second (Malten 1989: 44). The two sets of examples that he gives to demonstrate this were included in the questionnaire (as Q8a and Q8b, and Q21a and Q21b), in order to test the claim. In all, examples of six different inflectional cases (accusative, dative, locative, sociative, genitive and ablative) were presented.

The first accusative pair are given below; the case marker is included in the base for echoing in (35) but not (36).

(35) [*puuvai*] [*kiivai*] *parikkaatee* (= Q5a)
flower.acc echo pluck.neg.imp
'Don't pick the flowers or anything else.'

(36) [*puu*] [*kii*]*vai*[12] *parikkaatee* (= Q5b)
flower echo.acc pluck.neg.imp

Further examples are given in (37) and (38).

(37) [*veelaiyai*] [*kiilaiyai*] *patti collaatee* (= Q22b)
work.acc echo about speak.neg.imp
'Don't talk to me about work and such like!'

(38) [*veelai*] [*kiilai*]*yai patti collaatee* (= Q22c)
work echo.acc about speak.neg.imp

There are several examples for the dative case: (39) and (40), and (41) and (42) involve place names, and (43) and (44) the name of a person. The order in which the two options are presented in the questionnaire varies, to avoid the sentences falling into an easily predictable pattern.

(39) [*cennaikku*] [*kinnaikku*] *pooka maaṭṭeen* (= Q6b)
Chennai.dat echo go.infin will-not.1s
'No way am I going to Chennai or anywhere near.'

(40) [*cennai*] [*kinnai*]*kku pooka maaṭṭeen* (= Q6a)
Chennai echo.dat go.infin will-not.1s

[11] Rhenius' *'ablative of place'* corresponds to what is widely known as the locative.
[12] The insertion of the *-v-* between the stem *puu* and the *-ai* accusative marker is a regular instance of sandhi following a back vowel: descriptions of this are to be found in all the grammars (e.g. Schiffman 1999: 21). After a front vowel *-y-* is inserted, as in (37) and (38).

(41) [*tiruppatikku*] [*kiruppatikku*] *pooka maaṭṭaan* (= Q7a)
Tirupati.dat echo go.infin will-not.3sm
'No way will he go to Tirupati or anywhere near.'

(42) [*tiruppati*] [*kiruppati*]*kku pooka maaṭṭaan* (= Q7b)
Tirupati echo.dat go.infin will-not.3sm

(43) [*kumaarukku*] [*kimaarukku*] *peṭṭiyai kuṭukkaatee* (= Q31a)
Kumar.dat echo box.acc give.neg.imp
'Don't give the box to Kumar or any of his type.'

(44) [*kumaar*] [*kimaar*]*ukku peṭṭiyai kuṭukkaatee* (= Q31b)
Kumar echo.dat box.acc give.neg.imp

In each of these examples the noun in question had no distinct oblique form, but where one does exist this opens up the possibility of having either the nominative or the oblique stem as the base for the echo formation. This was tested with further examples of the dative case in the following sets of three sentences, taking first the oblique stem and case marker together as the base for echoing, then just the nominative stem, and finally just the oblique stem. The first set uses the name of the town, Chidambaram, and the second a common noun meaning 'garden'.

(45) [*citamparattukku*] [*kitamparattukku*] *kumaar pooka maaṭṭaan* (= Q8b)
Chidambaram.obl.dat echo Kumar go.infin will-not.3sm
'No way is Kumar going to Chidambaram or any place like that.'

(46) [*citamparam*] [*kitampara*]*ttukku kumaar pooka maaṭṭaan* (= Q8a)
Chidambaram echo.obl.dat Kumar go.infin will-not.3sm

(47) [*citamparattu*] [*kitamparattu*]*kku kumaar pooka maaṭṭaan* (= Q8c)
Chidambaram.obl echo.dat Kumar go.infin will-not.3sm

(48) [*tooṭṭattukku*] [*kiiṭṭattukku*] *ooṭaatee* (= Q9a)
garden.obl.dat echo run.neg.imp
'Don't run to the garden or such like place.'

(49) [*tooṭṭam*] [*kiiṭṭa*]*ttukku ooṭaatee* (= Q9c)
garden echo.obl.dat run.neg.imp

(50) [*tooṭṭattu*] [*kiiṭṭattu*]*kku ooṭaatee* (= Q9b)
garden.obl echo.dat run.neg.imp

The locative case marker *-le* is used in the following pair of sentences, appearing twice in (51) but only once in (52).

(51) *kumaar* [*laṇṭanle*] [*kiṇṭanle*] *irukka maaṭṭaan* (= Q10a)
Kumar London.loc echo be.infin will-not.3sm
'Kumar won't stay in London or anywhere like that.'

(52) *kumaar* [*laṇṭan*] [*kiṇṭan*]*le irukka maaṭṭaan* (= Q10b)
Kumar London echo.loc be.infin will-not.3sm

The same pattern is seen again in (53) and (54) for the sociative marker *-ooṭu*.

(53) [*tiruṭanooṭu*] [*kiruṭanooṭu*] *peeca maaṭṭeen* (= Q11b)
thief.soc echo speak.infin will-not.1s
'I will not speak to thieves and such like.'

(54) [*tiruṭan*] [*kiruṭan*]*ooṭu peeca maaṭṭeen* (= Q11a)
thief echo.soc speak.infin will-not.1s

As mentioned above, the oblique stem can serve on its own in genitival function (see 101 and 102 below), but it is also possible to use a case marker. *-ooṭe*, the colloquial Tamil equivalent of formal *-uṭaiya*, is used in sentences (55) and (56), and *-in*[13] appears in sentences (57) and (58). This is associated with formal, rather than spoken Tamil (Schiffman 1999: 27), and it was included to see how subjects would react to the juxtaposition of a colloquial echo word with a formal marker.

(55) *itu [kumaarooṭe] [kimaarooṭe] peenaavaa* (= Q12a)
this Kumar.gen echo pen.qs
'Does this pen belong to Kumar or some such person?'

(56) *itu [kumaar] [kimaar]ooṭe peenaavaa* (= Q12b)
this Kumar echo.gen pen.qs

(57) *itu [kumaarin] [kimaarin] peenaavaa* (= Q12c)
this Kumar.gen echo pen.qs

(58) *itu [kumaar] [kimaar]in peenaavaa* (= Q12d)
this Kumar echo.gen pen.qs

The final case tested in the questionnaire, the ablative, differs from the others in having a composite marker formed from the locative *-ile* with the addition of *-eruntu*. The noun *tooṭṭam* 'garden' was used again, and subjects were presented with the two options in (59) and (60).

(59) *[tooṭṭattileruntu] [kiiṭṭattileruntu] puuvai parikkaatee* (= Q13a)
garden.obl.abl echo flower.acc pluck.neg.imp
'Don't pluck flowers from the garden or places like that.'

(60) *[tooṭṭattu] [kiiṭṭatt]ileruntu puuvai parikkaatee* (= Q13b)
garden.obl echo.abl flower.acc pluck.neg.imp

3.2.3.3.2. Postpositions

The distinction between case markers and postpositions in Tamil is hard to establish, given that they may be functionally very similar. Some postpositions follow the oblique stem of the noun, and are arguably bound forms. Others appear after particular case markers or even the nominative stem, and may have a more independent phonological status. Note, however, that this is arguably true of the ablative *-eruntu* case marker, which follows the locative case, and so it cannot be taken as a defining characteristic of postpositions. As Schiffman comments, writing of the modern language:

> *'Now we must admit to fuzzy boundaries between case markers and postpositions.'* (Schiffman 1998: 374).

Several examples of each were included in the questionnaire, to see whether they differed in their interaction with echoing.

The questionnaire contained postpositions following both case-marked and non-case-marked nouns. A clear instance of the first category was the postposition *munnaalee* 'before, in front of', which follows a dative case marker. Subjects were given the option of either echoing the postposition with the noun, as in (61), or placing it after the echo formation, as in (62).

(61) *[viiṭṭukku munnaalee] [kiiṭṭukku munnaalee] okkaaraatee* (= Q23b)
house.dat before echo sit.neg.imp
'Don't sit in front of the house or anywhere near it.'

[13] This is sometimes referred to as the euphonic increment (see, for example, Lehmann 1989: 12, 21), since it may be optionally inserted in nominal forms in several different contexts.

(62) [*viiṭṭukku*] [*kiiṭṭukku*] *munnaalee okkaaraatee* (= Q23a)
house.dat echo before sit.neg.imp

The postposition *patti* 'about, concerning the topic of, regarding' occurs after an accusative case-marked noun, according to Schiffman (1999: 43). It appears in the set of sentences (Q22), which test three dimensions of variation: whether or not the noun is actually case-marked, and, if it is, whether the case marker appears once or twice, and whether the postposition appears once or twice. Five of the six possible permutations are given in the questionnaire: the two testing whether the case-marking appears once or twice have already been given, in (37) and (38). The remaining three appear below as (63), (64) and (65).[14]

(63) [*veelaiyai patti*] [*kiilaiyai patti*] *collaatee* (= Q22e)
work.acc about echo speak.neg.imp
'Don't talk to me about work and such like!'

(64) [*veelai patti*] [*kiilai patti*] *collaatee* (= Q22d)
work about echo speak.neg.imp

(65) [*veelai*] [*kiilai*] *patti collaatee* (= Q22a)
work echo about speak.neg.imp

The postposition *meelee*, meaning 'above', may follow either the oblique stem (or the nominative stem if there is no distinct oblique form), in which case it refers to contact with the upper surface of the object in question, or the dative case, when it refers to a position above the object.

(66) *marattu meelee* 'on top of the tree'

(67) *marattukku meelee* 'above (but not touching) the tree'

meelee appears in the questionnaire in the first of these two functions, following a nominative stem,[15] and either being repeated as part of the base for echoing or appearing after the echoed noun.

(68) [*meejai meelee*] [*kiijai meelee*] *onnum vaikka kuuṭaatu* (= Q21a)
table on echo anything put.infin should.neg.3sn
'It is strongly forbidden to put anything on the table or nearby.'

(69) [*meejai*] [*kiijai*] *meelee onnum vaikka kuuṭaatu* (= Q21b)
table echo on anything put.infin should.neg.3sn

The final pair of sentences include *kiṭṭa*, which also occurs after a non-case-marked noun. It alternates with the dative case marker, and therefore has a particularly ambiguous status: Schiffman, for example, refers to it as *'the animate locative marker'* (Schiffman 1999: 31) but it may also be classed as a postposition (Asher 1985: 104).

(70) [*appaa kiṭṭa*] [*kippaa kiṭṭa*] *keekkaatee* (= Q20d)
father near echo listen.neg.imp
'Don't listen to your father and the older generation!'

(71) [*appaa*] [*kippaa*] *kiṭṭa keekkaatee* (= Q20c)
father echo near listen.neg.imp

[14] The sixth possibility, given below, was excluded since it was judged highly unlikely that the reduplicant could be interrupted by the case marker in this way.
[*veelai patti*] [*kiilaiyai patti*] *collaatee*
work about echo.acc echo speak.neg.imp

[15] Like *veelai*, *meejai* does not have a separate oblique form.

3.2.3.3.3. Verbal forms

As discussed in section 3.2.3.2.3 above, various non-finite verbal forms, notably infinitives and participles, participate in echo expressions. This section investigates the echoing of various morphologically complex forms, some involving modal auxiliaries and others aspectual verbs. In the spoken language, in particular, the elaborate sequences that accumulate are simplified, and some markers that were originally independent words have become phonologically reduced and bound to what precedes them. An obvious example is the negative marker *illai*, which is reduced to *-le* after the infinitive. Thus spoken *caappiṭale* corresponds to the full written form *caappiṭavillai*.

The simplest form of the verb, which also functions as the citation form, is referred to as the root and is used without the addition of further markers for the singular imperative, e.g. *paṭi* 'study!'. Tamil verbs then divide into two groups (traditionally known as 'weak' and 'strong' respectively), according to whether the infinitive simply has an *-a* in place of the stem-final vowel, as in *ooṭa* (see 3 and 4 above), or adds the marker *-kka*, as in (15). Sentence (72) was included in the questionnaire to investigate whether this marker can be added to the end of the reduplicated root, or has to be included in the base for echoing, as in (15).

(72) *iŋkee [paṭi] [kiṭi]kka kuuṭaatu* (= Q33b)
here study echo.infin should.neg.3sn
'One should not study and so forth in this place!'

The infinitival form also serves as the base onto which other morphological markers are added, such as the reduced negative marker *-le* exemplified in (16). Sentence (73) was paired with this in the questionnaire, to see whether it could be added to an echoed infinitive, rather than being included in the base for echoing.

(73) *avanooṭu [caappiṭa] [kiippiṭa]le* (= Q34b)
he.soc eat.infin echo.neg
'I don't eat and so forth with him.'

A parallel pair, in which a marker is added to an infinitive, involves negative imperative forms. Repetition of the whole form has already been seen in (17): the corresponding sentence, in which the marker follows the echoed infinitive is given in (74).

(74) *nii [peeca] [kiica]atee* (= Q32b)
you speak.infin echo.neg
'Don't speak and so forth!'

Free-standing infinitives are generally followed by modal verbs, such as *muṭiyaatu* 'cannot' in (3) above, and *kuuṭatu* 'should not' in (15). Sentences (75) and (76) test the possibility of including the modal verb *maaṭṭaan* 'will not', which is used to form the future negative, within the echo formation.

(75) *kumaar [paṭikka maaṭṭaan] [kiṭikka maaṭṭaan]* (= Q35b)
Kumar study.infin will-not.3sm echo
'Kumar will not study etc.'

(76) *kumaar [paṭikka] [kiṭikka] maaṭṭaan* (= Q35a)
Kumar study.infin echo will-not.3sm

As noted in section 3.2.3.2.3, participles in Tamil can be echoed straightforwardly and independently of any other verbal form in the sentence. They can also be expanded by the addition of aspectual material, such as the verb *(v)iṭu*,[16] which signals completion. Two

[16] The initial *v-* is frequently omitted, which sets it apart from its lexical counterpart *viṭu* 'leave, let', in which the *v-* is always pronounced.

examples of participles marked in this way are to be found in the questionnaire, with the option for each of either including the completive marker within the base for echoing, as in (77) and (79), or adding it to the end of the echo formation, as in (78) and (80).

(77) *kumaar [paaṭittu] [kiiṭittu] viiṭṭukku poonaan* (= Q37a)
Kumar sing.asp.vbp echo house.dat go.past.3sm
'After singing and so forth, Kumar went home.'

(78) *kumaar [paaṭi] [kiiṭi]ṭṭu viiṭṭukku poonaan* (= Q37b)
Kumar sing.vbp echo.asp.vbp house.dat go.past.3sm

(79) *[tuuŋkiṭṭu] [kiiŋkiṭṭu] caappiṭaatee* (= Q38a)
sleep.asp.vbp echo eat.neg.imp
'Don't eat after sleeping etc.'

(80) *[tuuŋki] [kiiŋki]ṭṭu caappiṭaatee* (= Q38b)
sleep.vbp echo.asp.vbp eat.neg.imp

In these examples the participles appear in non-finite subordinate clauses, but participles are also frequently combined with further verbal forms as part of a tensed verb. Sentence (81) is just such a case, comprising a participle of the verb meaning 'come', an aspectual participle *kiṭṭu*, from the self-benefactive verb *koo*,[17] the infinitive of the stative aspectual *iru*,[18] and finally the modal verb *maaṭṭeen*, inflected for person and number (first singular). This was used to test how the echo forms interact with a very complex verbal structure: informants were asked to comment on the sentence as it is given, with just the lexical participle repeated, and also suggest any alternative acceptable forms.

(81) *naaḷekki naan [vantu] [kintu]kiṭṭirukka maaṭṭeen* (= Q47)
tomorrow I come.vbp echo.dur.be.infin will-not.1s
'I won't be coming and so forth tomorrow.'

Another source of complexity in some Tamil verbs is the possibility of combining a noun with a verb to form a 'compound verb' or 'complex predicate'. These are referred to by Annamalai and Steever (1998: 115) as *'complex morphosyntactic vehicles'*, and Schiffman describes them as:

> *'on the borderline between being independent verbs, and being phrases containing a noun and a verb'* (Schiffman 1999: 107).

They are differentiated from normal verb phrases by a number of morphosyntactic properties, notably the absence of case marking on the noun, even though this may be the semantic object. Indeed, according to Corré, the compound verb may take an object of its own, which then bears an accusative marker: he gives the example in (82) (Corré 1962: 24).

(82) *avan peṇṇai peṇ paarttaan*
he girl.acc girl see.past.3sm
'He saw the girl with a view to marriage.'

Annamalai and Steever (1998: 124) mention various other properties of the noun showing its close association with the verb: it cannot be modified by adjectives, demonstratives or quantifiers, and cannot be extracted by relative clause formation, questioning, clefting or passivization. Particular verbs tend to recur in these compounds: Schiffman (1999: 109–113) discusses several *'common verbalizers'*.[19] One of his examples was included in the

[17] The formal equivalent of this participle form is *koṇṭu*.

[18] Note that the semantic force of the combination *kiṭṭiru* is durative.

[19] Similar constructions are found in Kannada (see section 2.4.1.3) and also the Indo-Aryan languages (see section 2.4.2.3). They all display some properties characteristic of incorporation (see Baker 1988), such as the absence of case marking on the noun. For Hindi, however, Mohanan (1994) systematically

questionnaire, and its integrity investigated, by giving the option of echoing either component separately or both together.

(83) [*patil collaatee*] [*kitil collaatee*] (= Q49c)
answer say.neg.imp echo
'Don't answer and so forth.'

(84) *patil* [*collaatee*] [*killaatee*] (= Q49a)
answer say.neg.imp echo

(85) [*patil*] [*kitil*] *collaatee* (= Q49b)
answer echo say.neg.imp

Finally, a considerable number of the sentences included occurrences of the quotative verb *-ɳɳu*, largely because reported statements are a favoured context for echo expressions in Tamil (see section 3.2.3.1). Phonologically, it seems to have lost its status as an independent word and now acts as a phrasal clitic. When a phrase is echoed, therefore, the question arises of whether the *-ɳɳu* will appear once, at the end, or twice, as part of the base for echoing. This was tested in (86), which differs from (18) only in the number of occurrences of *-ɳɳu*.

(86) *enakku* [*teriyaatu -ɳɳu*] [*kiriyaatu-ɳɳu*] *collaatee* (= Q39a)
I.obl.dat know.neg.3sn quote echo say.neg.imp
'Don't say "I don't know!".'

Another pair differentiated only by *-ɳɳu* are given in (87) and (88).

(87) [*kumaarukku kuʈutteen -ɳɳu*] [*kimaarukku kuʈutteen-ɳɳu*] *poy collaatee* (= Q43a)
Kumar.dat give.past.1s quote echo lie say.neg.imp
'Don't lie that you gave it to Kumar or some such nonsense.'

(88) [*kumaarukku kuʈutteen*] [*kimaarukku kuʈutteen*]*-ɳɳu poy collaatee* (= Q43b)
Kumar.dat give.past.1s echo quote lie say.neg.imp

3.2.3.4. Compounds

Compounds are very common in Tamil, although the degree of productivity varies according to which categories are being combined. Compounds consisting of verb+noun (e.g. *kuʈi-taɳɳiir* 'drinking water') and adverb+noun (e.g. *meelmaaʈi* 'upper storey') are possible, but rare: noun+verb compounds, discussed in the previous section, are more common. Binominal compounding, however, seems to be most productive, and all the examples in the questionnaire (except Q49) are of this type. As Annamalai and Steever comment, the prevalence of compounds may be connected with the relative scarcity of derivational morphology in Tamil:

> *'cross-categorial derivation is highly restricted; its function is instead borne by compounding'* (Annamalai & Steever 1998: 104).

The primary division in noun+noun combinations is between endocentric and exocentric compounds, i.e. those in which one element is semantically subordinate to the other, often referred to as the head, and those in which the two elements are coordinate. This second type, variously known as additive, coordinate or dvandva compounds (Tamil *ummaittokai*), are represented in the questionnaire by two examples, given in (89)–(92) below. Such pairs can be marked with the coordinative marker *-um* so that they function as conjoined nouns (see section 3.3.2.6.2), but it is also possible simply to juxtapose them, as here. The meaning is additive: in (89) hyponyms, terms for two kinds of domestic animal, are combined to give the sense of the superordinate term 'cattle'. In (91), *appaa* and *ammaa*,

differentiates complex predicates from instances of incorporation, using both syntactic and phonological criteria. Further work would be required to establish exactly how the Tamil examples should be categorized.

colloquial terms for 'father' and 'mother', which are antonyms, together mean 'parents'. A tendency to conjoin pairs of nouns, especially synonyms, has been seen as characteristic of the Indian subcontinent in general (Vacek 1994: 149), and is discussed in section 2.2.4. The questionnaire examples were designed to test whether such compounds can be echoed at all, and, if so, whether both elements are echoed together, as in (89) and (91), or just the second, as in (90), or just the first, as in (92).

(89) [*aaʈumaaʈu*] [*kiiʈumaaʈu*] *meeykka marakkaatee* (= Q19d)
goat-cow echo graze.infin forget.neg.imp
'Don't forget to graze the cattle and so forth.'

(90) *aaʈu*[*maaʈu*] [*kiiʈu*] *meeykka marakkaatee* (= Q19c)
goat-cow echo graze.infin forget.neg.imp

(91) [*appaa ammaa*] [*kippaa ammaa*] *keekkaatee* (= Q20a)
father mother echo listen.neg.imp
'Don't listen to your parents and their generation!'

(92) [*appaa*] [*kippaa*] *ammaa keekkaatee* (= Q20b)
father echo mother listen.neg.imp

The two elements of endocentric compounds in Tamil may bear a variety of semantic relationships to one another, but it is consistently the second element that forms the head, with the first element modifying it in some way. Three morphological types can be identified, varying according to whether the first element is a nominative or oblique stem, and whether or not the suffix *-am-* is inserted after an oblique stem. Sets of examples containing nominative stems are included in the questionnaire, with either the whole compound being echoed, as in (93), (94), (96) and (98), or just the first element, as in (95), (97) and (99).[20]

(93) [*talaivali*] [*kilaivali*] *irukkutaa* (= Q15a)
head-ache echo be.pres.3sn.qs
'Have you got a headache and so forth?'

(94) [*maamaram*] [*kiimaram*] *iŋkee naʈaatee* (= Q16a)
mango-tree echo here plant.neg.imp
'Don't plant mango-trees and so forth here.'

(95) [*maa*] [*kii*]*maram iŋkee naʈaatee* (= Q16b)
mango echo-tree here plant.neg.imp

(96) *enakku* [*maaŋkaay*] [*kiiŋkaay*] *piʈikkaatu* (= Q17a)
I.obl.dat mango-fruit echo like.neg.3sn
'I don't like mangoes and so forth.'

(97) *enakku* [*maa*] [*kiiŋ*]*kaay piʈikkaatu* (= Q17b)
I.obl.dat mango echo-fruit like.neg.3sn

(98) [*naaykkuʈʈi*] [*kiiykkuʈʈi*] *viiʈʈukkuḷḷee konʈuvaraatee* (= Q18a)
dog-calf echo house.inside bring.neg.imp
'Don't bring pups and so forth inside the house!'

(99) [*naay*] [*kiiy*]*kkuʈʈi viiʈʈukkuḷḷee konʈuvaraatee* (= Q18b)
dog echo-calf house.inside bring.neg.imp

The second morphological type is found in (100) and (101), where the first element is an oblique stem.

[20] It happens to be the case that, as with many Tamil nouns, none of the first elements have distinct oblique forms.

(100) [*tooʈʈattuppuu*] [*kiiʈʈattuppuu*] *parikkaatee* (= Q14b)
garden.obl.flower echo pluck.neg.imp
'Don't pick garden flowers or anything else.'

(101) *tooʈʈattup*[*puuvai*] [*kiivai*] *parikkaatee* (= Q14a)
garden.obl.flower.acc echo pluck.neg.imp

The third type, not exemplified in the questionnaire, contains a semantically vacuous *-am-* suffix, following an oblique stem, as in *aarr-aŋ-karai* 'river bank'.

There has been some interest in the literature (e.g. Christdas 1988, Bosch 1991, Wiltshire 1992, Bosch & Wiltshire 1993) in establishing the phonological constituency of Tamil compounds by looking at their interaction with certain morphophonemic processes. There are claims that compounds are in some respects phonologically unitary, and these will be discussed, together with new experimental evidence, in section 4.6.3. It is interesting to compare their structure with that of the echo formations, and also see whether compounds can be split by having a single element form the base for echoing.

3.2.3.5. Phrases

The implication in nearly all the descriptions of Tamil echo formations in the literature is that the base is restricted to a single word. Generally no explicit statement is made but all the examples given involve only single words. Malten is at first glance an exception to this, since he includes a sentence in which a noun and postposition are optionally echoed together (Malten 1989: 44):[21]

(102) [*meecaimeel*] [*kiicaimeel*] *onnum vaikkak kuuʈaatu*
table.on echo anything put.infin should.neg.3sn
'It is strongly forbidden to put anything on the table.'

(103) [*meecai*] [*kiicai*] *meel onnum vaikkak kuuʈaatu*
table echo on anything put.infin should.neg.3sn

Note that Malten writes the two elements together, without a space, in sentence (102), implying that they should be regarded as a single word.

A very different picture, however, is presented by Nagarajan, who makes the following statement:

> '*Besides echo formation or reduplication at the word level, it is possible (and in fact more natural) to repeat whole phrases or clauses in Tamil*' (Nagarajan 1994: 165).

She proceeds to provide a number of examples, concluding that the base for reduplication can be any major lexical item, any XP (i.e. syntactic phrase) or even an embedded S (sentence) constituent, if this has inherent contrastive focus and conveys a meaning of disbelief on the part of the speaker. In effect, this limits the occurrence of echoed clauses to the type of context mentioned in section 3.2.3.1 in connection with the echoing of finite verbs, i.e. subordinate clauses followed by the quotative particle and a prohibition in the main clause. Nagarajan also claims that the phrases that are echoed can be syntactically determined, and she provides examples of both adjectival and verbal phrases (Nagarajan 1994: 165).

(104) *avan* [*a.ɻakaana paiyan*] [*ki.ɻakaana paiyan*] *ennu colli mayaŋkaatee*[22]
he handsome boy echo say.vbp say.vbp be-charmed.neg.imp
'Don't be carried away by saying / thinking that he is a handsome boy.'

---

[21] An almost identical pair of sentences appear in the questionnaire, and have been discussed in section 3.2.3.3.2 above as (68) and (69).

[22] The transliterations and glosses used by Nagarajan have been altered slightly in line with the conventions set out in appendices A and C.

(105) *aŋkee [aaɻamaa illai] [kiiɻamaa illäi] ennu ninaiccu viɻuntu vaikkaatee*
there deep not echo say.vbp think.vbp fall.vbp asp.neg.imp[23]
'Don't trip up wondering if it is deep there or not.'

(106) *nii [raamanukku kuɻutteen] [kiimanukku kuɻutteen] ennu poy collaatee*
you Rama.dat give.past.1s echo say.vbp lie say.neg.imp
'Don't lie that you gave (it) to Rama.'

In (105), the echo phrase is based on a predicative adjective and negative marker: in (106) it is a tensed verb phrase containing an indirect object and verb. An overt subject appears in the following example (Nagarajan 1994: 166), which is the most complex that Nagarajan provides, containing a sequence of subject, object and compound verb.[24]

(107) *avan [raaman kurattiyai kaliyaaŋam paɳɳiŋaan] [kiiman kurattiyai kaliyaaŋam*
he Rama gypsy.acc marriage do.past.3sm echo

paɳɳiŋaan] ennu poy connaan
say.vbp lie say.past.3sm
'He lied that Rama had married a gypsy.'[25]

Although Nagarajan is the only writer to mention the possibility of phrasal reduplication in Tamil, similar data have been reported for Kannada by Lidz (2000) and for Bengali by Fitzpatrick-Cole (1994), as discussed in section 2.4.2.4.

A number of sentences containing echo phrases were included in the questionnaire, to investigate firstly whether the subjects would permit such phrases at all, and then to see what sort of sequences could be repeated. Several have been discussed already, those including echoed sequences of adjective+noun (24, 26, 28, 30, 32), and noun+postposition (61, 63, 64, 68, 70). Sentences (108) and (109) bear some similarity to (105), which also contains the negative particle and zero copula in its subordinate clause. They test the possibility of echoing noun and negative particle together, as opposed to just the noun.

(108) *avanukku [viiɻu illai] [kiiɻu illai]* (= Q36b)
he.dat house not echo
'He has no house and family etc.'

(109) *avanukku [viiɻu] [kiiɻu] illai* (= Q36a)
he.dat house echo neg

There are five sets of sentences in the questionnaire designed to test the possibility of echoing different verbal phrases. In each case, in addition to the sentence containing the echoed phrase, there are also sentences checking that different constituents of the phrase can be echoed individually. One set of examples is similar to Nagarajan's sentence in (106) above, containing an indirect object and a finite verb. Most of the examples have already been given: sentences (87) and (88) above echo the full phrase *kumaarukku kuɻutteen-ŋŋu*, with either one or two occurrences of the quotative particle. Sentence (19) echoes just the verb, and (110) echoes just the noun.

(110) *[kumaarukku] [kimaarukku] kuɻutteen -ŋŋu poy collaatee* (= Q45)
Kumar.dat echo give.past.3sm quote lie say.neg.imp
'Don't lie that you gave it to Kumar or anyone else!'

---

[23] The verb *vai*, which means 'take, put' as a lexical verb, is used here in aspectual function. As such, it can frequently be translated, as here, using the verb followed by 'up' (Schiffman 1999: 87).
[24] See section 3.2.3.3.3 above for discussion of compound verbs: *paɳɳu* 'make, do' is the most common of the verbs involved in these constructions (Schiffman 1999: 107–113).
[25] Note that Nagarajan provides no indication of the semantic force of echoing in this example.

Another set of examples that is morphologically exactly parallel, but syntactically rather different, involves a dative-stative construction (see Schiffman 1999: 105), in which a dative case-marked noun acts as the semantic subject of an impersonal verb. Echoing of the verb alone, with and without the quotative particle, is tested in sentences (86) and (18) respectively; the full phrase is echoed in (111).

(111) [*enakku teriyaatu -ŋŋu*] [*kinakku teriyaatu-ŋŋu*] *collaatee* (= Q39c)
I.obl.dat know.neg.3sn quote echo say.neg.imp
'Don't say "I don't know!".'

A nominative subject and negative infinitive are echoed together in (112). The corresponding sentence testing the verb only is given in (113), and the same sentence but with no overt subject appears in (114): echoing of the subject only was tested in (11).

(112) [*kumaar paakkale*] [*kimaar paakkale*]-*ŋŋu poy collaatee* (= Q40a)
Kumar see.infin.neg echo quote lie say.neg.imp
'Don't lie that Kumar didn't see or some such nonsense.'

(113) *kumaar* [*paakkale*] [*kiikkale*]-*ŋŋu poy collaatee* (= Q40b)
Kumar see.infin.neg echo quote lie say.neg.imp

(114) [*paakkale*] [*kiikkale*]-*ŋŋu poy collaatee* (= Q42)
see.infin.neg echo quote lie say.neg.imp
'Don't lie that you didn't see and so forth.'

A nominative subject and finite present tense verb (with quotative particle attached) make up the echo phrase in (115): the corresponding sentence echoing only the verb has been given above in (20) and that echoing only the subject in (10).

(115) *innikki* [*appaa varaar -ŋŋu*] [*kippaa varaar-ŋŋu*] *collaatee* (= Q48a)
today father come.pres.3pl quote echo say.neg.imp
'Don't say "Father's coming today and so forth".'

Finally, a sequence of subject, object and verb is echoed in (116), and the possibility of echoing the subject alone and the verb alone was tested in (12) and (21) respectively.

(116) [*kumaar onnum paattaan -ŋŋu*] [*kimaar onnum paattaan-ŋŋu*] *poy collaatee*
Kumar anything see.past.3sm quote echo lie say.neg.imp
'Don't lie that Kumar saw anything and so forth.' (= Q46a)

3.2.3.6. Summary

Two sentences of the questionnaire have thus far not received any comment in this section, since they were designed to elicit phonological, rather than morphological, information. The sentences are given in (117) and (118): both involve echoed nouns which are homonymous with forms that have undergone total reduplication. They received generally negative reactions from the subjects, which fits in with the tendency to avoid identity discussed in section 1.5.2.5.

(117) [*kiirai*] [*kiirai*] *caappiṭaatee* (= Q3)
greens echo eat.neg.imp
'Don't eat greens and so forth.'

(118) *enakku* [*kiḷi*] [*kiḷi*] *piṭikkaatu* (=Q4)
I.obl.dat parrot echo like.neg.3sn
'I don't like parrots and so forth.'

A summary of the issues discussed thus far, with the relevant sentences (numbered as in the questionnaire), is provided in table 2. In the interests of clarity, only the numbers of the sets of sentences, not all the individual sentences, are given. In a few cases, a sentence set

such as (Q19) or (Q30) is relevant to more than one issue, and so appears under more than one category, but the lists given are not exhaustive, e.g. the sentence sets listed for adjectives generally appear only there, although they are also relevant to the echoing of nouns and of phrases.

Table 2. Question summary

Questions		**Sentences**
Can echo words occur in positive contexts?		2
Which lexical categories can be echoed?	Nouns	19, 30, 41, 46, 48
	Pronouns	29
	Verbs	1, 32, 33, 34, 39, 44, 46, 48
	Adjectives	24, 25, 26, 27, 30
	Adverbs	28
What is the morphological make-up of the base?	Case markers	5, 6, 7, 8, 9, 10, 11, 12, 13, 22, 31, 45
	Postpositions	20, 21, 22, 23
	Verbal forms	32, 33, 34, 35, 37, 38, 39, 42, 43, 47, 49
How do compounds and echoing interact?		14, 15, 16, 17, 18, 19, 20
What kinds of phrase, if any, can be echoed?		36, 39, 40, 43, 46, 48
Can echo words involve total repetition?		3, 4

3.3. Data analysis

3.3.1. Distribution of scores

Each of the responses by the subjects (960 in total) was assigned a response score from 1 to 4: 1 indicated definite acceptance as a legitimate sentence of spoken Tamil, and 4 definite rejection. A score of 2 was allocated when acceptance was only implied or somewhat equivocal, and 3 when rejection was implicit or qualified in some way. Additional comments or suggestions by the subjects were also noted and have been taken into account in compiling the results. The full table of responses is given in appendix E: blank cells indicate absence of comment either way. These appear predominantly in the columns for speakers 11 and 12, who took part in the pilot interviews, and in the rows for set (Q49), since this set was added to the questionnaire after nine of the subjects had been interviewed.

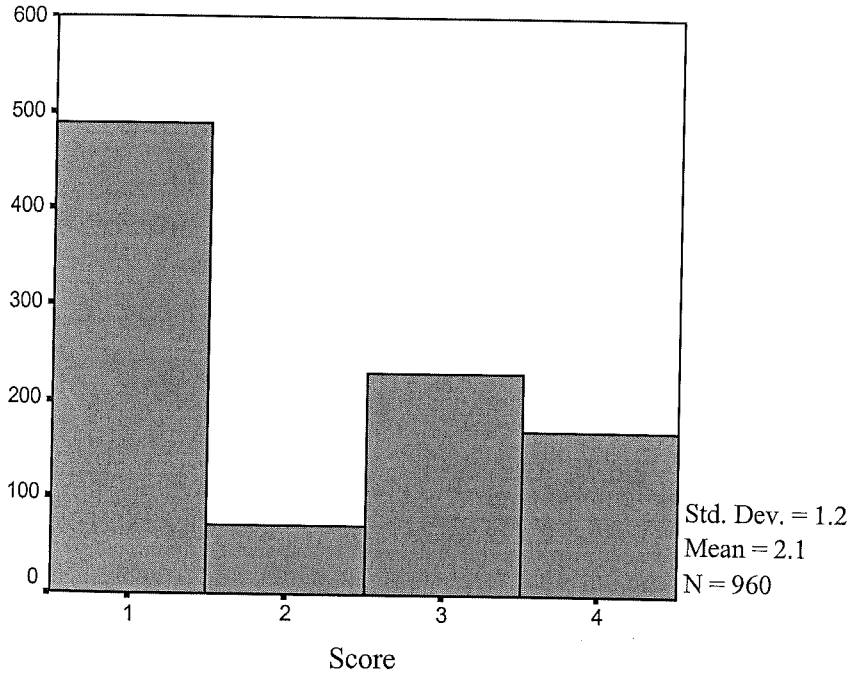
The design of the questionnaire was such that a bias towards positive responses was expected. Sentences were labelled for predicted response according to the claims in the literature: 398 were expected to be positive, and only 12 negative.[26] The remainder were either the subject of explicit disagreement in the literature, such as the pronoun in (Q29) and the adjectival examples in (Q24)–(Q27), or cases for which there was no comment either way, such as the compounds in (Q14)–(Q20), or the instances in which a case marker was not included in the base for echoing. These were all assigned a 50% probability of being positive, giving an overall figure of 70.1% for expected positive responses.

[26] Positive responses are predicted for the following sentences: (Q1), (Q5a), (Q6b), (Q7a), (Q8b), (Q9a), (Q10a), (Q11b), (Q12a), (Q13a), (Q19a), (Q19b), (Q20d), (Q21a), (Q22b), (Q22e), (Q23b), (Q28a), (Q30a), (Q31a), (Q33a), (Q34a), (Q35a), (Q36b), (Q37a), (Q38a), (Q39b), (Q40a), (Q40b), (Q41), (Q42), (Q43b), (Q44), (Q45), (Q46b), (Q46c), (Q48b) and (Q48c). A negative response is predicted for just one sentence, (Q2).

The histogram in figure 1 shows the distribution of responses for all subjects on all sentences, and reveals a slight bias towards positive responses, although not to the extent predicted: 58.3% of the responses were given a score of either 1 or 2. In cases where the sentence was rejected, there were more cases of qualified than firm rejection (i.e. scores of 3, rather than 4). A response pattern of this kind may mean that speakers are somewhat unclear about whether the sentences are entirely unacceptable, but it could also indicate a reluctance to criticize the interviewer too sharply. The overall distribution certainly reflects the tendency shown by the subjects to home in on what they perceived to be the correct option, and affirm that. Indeed, in some cases it was difficult to extract any clear opinion on the alternatives to the sentences that were favoured.

Individual patterns of response vary along two main dimensions: the distribution of scores and the extent of their conformity to those of the other subjects. Subject 10, for instance, proved unusually permissive, having a very high number of scores of 1 (61 in total). The distribution of responses for subject 5 was particularly polarized, with scores consisting almost exclusively of either 1 or 4.

Figure 1. Histogram of scores for all responses by all subjects



The standard deviation of each subject's set of responses from the mean scores for each sentence was calculated to give a measure of conformity, and the results are given in table 3.

The standard deviations range from 0.76 for subject 4 through to 1.26 for subject 5. As noted in section 3.2.1, there were sociolinguistic factors associated with three of the subjects that could potentially affect their response patterns: subject 4 had moved to Pondicherry at the age of fifteen, and her first language was not Tamil but Kannada. Her daughter (subject 8) also claimed Kannada as her first language, although she was a lifelong resident of Pondicherry. Interestingly, the mother's responses proved to have the lowest standard deviation of any of the subjects (0.76), suggesting that there was no significant interference from Kannada. The daughter, by contrast, had the second highest standard deviation (1.10), but it seems unlikely, given the results for her mother, that this should be attributed to her knowledge of Kannada. Both sets of responses were therefore included in the data set analyzed below. Subject 12, who differed from the others in having Tanjore, rather than Pondicherry, as his native place, proved to have a relatively high standard deviation (0.93), but again it was not judged sufficiently great to warrant exclusion of his responses.

Table 3. Standard deviation of subjects' responses

Subject	Gender	Age	Occupation	Languages	Standard Deviation
1	M	40	Academic	Ta, En, Fr	0.88
2	M	36	Librarian	Ta, En	0.76
3	M	45	Clerical worker	Ta, En	0.89
4	F	45	Sanitary assistant	Ka, Ta, En	0.76
5	M	28	Maintenance worker	Ta, En, Ma, Fr	1.26
6	F	26	Ph.D. student	Ta, En, Hi, Sa, Fr	0.82
7	M	27	Ph.D. student	Ta, En, Te, Fr	0.77
8	F	27	Sanitary assistant	Ka, Ta, En	1.10
9	M	36	Shopkeeper	Ta, En	0.81
10	F	26	Ph.D. student	Ta, En, Te	1.00
11	M	29	Ph.D. student	Ta, En, Fr	0.90
12	M	26	Ph.D. student	Ta, En, Hi, Ka	0.93

Subject 5 had the highest standard deviation (1.26), by a considerable margin. In part this was due to the polarized distribution of his responses, but it also reflects significant divergence from the overall trends for many of the sentences. No obvious reason for this idiosyncratic pattern of responses could be found. As the table shows, he was the only subject to have any knowledge of Malayalam, but since he did not claim fluency in the language it seems unlikely that this should have been responsible. Since there was no good reason on sociolinguistic grounds to exclude his responses, they were included in the results described below, and are the subject of occasional comment.

3.3.2. Results

3.3.2.1. Context

The pair of sentences included to test the possibility of echoing an infinitive in a positive context provided straightforward confirmation of the comments in the literature about echo words preferentially occurring in negative contexts. The relevant pair are repeated from (3) and (4) above in (119) and (120).

(119) *avanaal* [*ooṭa*] [*kiiṭa*] *muṭiyaatu* (= Q1)
he.instr run.infin echo can.neg.3sn
'He can't run and so forth.'

(120) *avanaal* [*ooṭa*] [*kiiṭa*] *muṭiyum* (= Q2)
he.instr run.infin echo can.3sn
'He can run and so forth.'

The first, negative example received a mean score of 1, indicating acceptance by all subjects, whereas responses to the corresponding positive sentence covered the entire range of scores, averaging 2.5,[27] and indicating very mixed opinions on the acceptability of echo words in a positive context. Although the majority of sentences in the questionnaire were negative, three further positive sets of sentences were included (Q12, Q15 and Q37). In no case did the favoured option receive unanimous approval (the scores were 1.9, 1.4 and 1.6 respectively),

[27] Where applicable, scores have been rounded to one decimal place.

whereas scores of 1 were obtained for some of the negative sentences: this provides further confirmation of the preference for negative contexts.

3.3.2.2. Which lexical categories can be echoed?

Subjects found echo forms based on non-case-marked nouns to be acceptable, with very few exceptions. Of the examples quoted in section 3.2.3.2.1 above, *aaṭu kiiṭu* 'goats and such like' and *maaṭu kiiṭu* 'cows and such like' in (7) and (8) received averages of 1.2 and 1.6[28] respectively, indicating general acceptance. Responses to *pai kii* 'bag and so forth' in example (9), were rather more equivocal, averaging 2.3, but the reason for this appears to be the absence of accusative case-marking.[29] The alternative sentence including case markers, given in (121), proved far more acceptable, receiving a mean score of 1.2.

(121) *enakku un [paiyai] [kiiyai] kuṭukkaatee* (= Q30c)
I.obl.dat you.obl bag.acc echo give.neg.imp
'Don't give me your bag and so forth.'

There was no objection to the echoing of proper names: sentences (11) and (12) containing *Kumaar kimaar* 'Kumar or someone like that' were given low mean scores (1.1 and 1.6 respectively), reflecting general agreement on their grammaticality. The sole example of an echoed pronoun (*naan kiin* in 14), which was borrowed from Malten (1989: 44), also proved to be generally acceptable (mean score 1.6), vindicating Malten's claim that pronouns can be echoed in certain contexts.

Of the various verbal forms tested, infinitives proved to be the least controversial: mean scores of 1, reflecting unqualified acceptance by all subjects, were given for both the 'weak' infinitive *ooṭa kiiṭa* 'to run and so forth' (in 3) and the 'strong' example *paṭikka kiṭikka* 'to study and so forth' (in 15). The negative infinitive *caappiṭale kiippiṭale* 'not to eat and so forth' (in 16) was also judged acceptable by all but one subject (the generally unorthodox subject 5), having a mean score of 1.4. There was some uncertainty, however, about the grammaticality of echoing a negative imperative form: *peecaatee kiicaatee* 'don't speak and so forth' (in 17) received mixed responses ranging from 1 through to 4, and averaging 2.2. This sentence differs from the others in being a direct form of address, and as such may be less well suited to echoing.

Four different tensed verbs were given in sentences (18)–(21) above, and overall appear to be acceptable bases for echo formations. Certainly *kuṭutteen kiṭutteen* 'I gave, presented, donated etc.' and the impersonal verb *teriyaatu kiriyaatu* 'not know and so forth' were straightforwardly acceptable, receiving mean scores of 1.1. Responses to *varaar kiraar* 'come and so forth' and *paattaan kiittaan* 'saw and so forth', by contrast, were rather more indeterminate, averaging 2.1 in both cases. It was notable, however, that there was only one definite rejection, for *paattaan kiittaan* from subject 5, whose responses were so irregular as not to carry much weight. In general, therefore, it appears that, within the contextual constraints outlined in section 3.2.3.1, finite verbs can be legitimately echoed.

The claim by Kothandaraman and Malten that adjectives can be echoed only in conjunction with the noun that they qualify was largely borne out by the data. In three of the four cases where there was the option of echoing either the adjective on its own, or adjective and noun together (sentences 24–31), the preference was for the latter. *veḷḷai kiḷḷai* 'white and such like', *keṭṭa kiṭṭa* 'bad and such like' and *nalla killa* 'good and such like' all received very mixed responses and mean scores of 2.8, 2.3 and 2.8 respectively, whereas *veḷḷai caṭṭai kiḷḷai caṭṭai* 'white shirts and such like' and *nalla paiyan killa paiyan* 'good boy and so forth' were

[28] The average here has been raised considerably by a response scoring 4 from the maverick subject 5.
[29] Inanimate objects in Tamil are not usually case-marked (hence the acceptability of examples 7 and 8), unless specificity or definiteness is intended (Schiffman 1999: 36).

clearly acceptable (averaging 1.3 and 1.2). *ketta kanavai kitta kanavai* 'bad dreams and so forth', however, scored only just better on average than its counterpart (2.3, as compared to 2.1). The reluctance to judge either option clearly acceptable may well be phonologically based, as subject 10 suggested when she commented that *ketta* and *kitta 'sound the same'*. The tendency for Indian languages in general to avoid echoing words already beginning with the relevant echo segment(s) was discussed in section 1.5.2.5, and was confirmed for Tamil by the subjects' responses to sentences (117) and (118) above, largely rejecting *kiirai kiirai* 'greens and so forth' and *kili kili* 'parrots and so forth' (mean scores of 2.9 and 2.8 respectively). *ketta kitta* 'bad and so forth' is not an exact parallel, but the acoustic differentiation between /i/ and /e/ in initial syllables is not particularly strong, and the distinction may be further confused by a lowering process said to affect /i/ vowels in certain contexts (see section 4.6.1).

The one exception to the preference for echoing adjective and noun together was the pair given in (24) and (25) above. *civappu kivappu toppi* 'hats that are red etc.' (mean score of 1.8) proved marginally preferable to *civappu toppi kivappu toppi* (mean score of 2.2), although in each case the full range of scores was represented in the responses. It is not clear why this pair of examples should go against the general trend. Phonologically *civappu* differs from the other three adjectives in its length, comprising three syllables, rather than two. One might speculate that this added phonological substance allows it to be echoed independently of the noun, but there is insufficient evidence to establish this as a general principle. The possessive adjective example given in (32) above (*un pai kii pai* 'your bag and so forth') was judged to be unacceptable by the vast majority of subjects, having an average score of 3.2. Semantic factors may explain this, as the meaning associated with echoing does not combine easily with this class of adjectives. Overall, the results lend some credence to Malten's remarks about adjectives being closely associated with the nouns that they qualify: there is a clear preference for echoing both together, which thus provides the first evidence for phrasal, rather than just word-based echo formations in Tamil.

The single adverbial example in the questionnaire: *cantoocamaaka kintoocamaaka*[30] 'joyfully etc.' (example 33) was given a mean score of 1.6, indicating general acceptance. Moreover, this was preferred to the alternative in (34), *cantoocam kintoocamaaka* (mean score of 2.4), in which the first element is simply the noun from which the adverbial form is derived, and the derivative *-ka* suffix is added to the end of the echo expression. Kothandaraman's claim that adverbs can be echoed therefore receives some support from the responses to the questionnaire.

In general, there appears to be little constraint in terms of lexical category on the bases for Tamil echo formations. Nouns seem to be the prototypical case but various verbal forms were undisputed, and some examples involving adverbs and pronouns proved widely acceptable. Indeed, a general bias against echoing adjectives independently was the only category-based restriction to emerge from the questionnaire results.

3.3.2.3. Morphological constituency of echo bases

3.3.2.3.1. Case markers

The general pattern was clear acceptance of sentences in which the case marker appeared on both base and echoed part, but mixed reactions to sentences in which the marker appeared only at the end of the whole echo formation. So, for instance, in (35) and (36) *puuvai kiivai* 'flowers and such like', in which the accusative marker *-ai* appears twice, received strong approval (mean score of 1.1), whereas *puu kiivai* was decisively rejected overall (mean score of 3.4). In another pair of examples involving the accusative case marker

[30] The echo expression given here is the direct transliteration of what appeared on the questionnaire. Most subjects, however, preferred the alternative form *cantoocamaa kintocamaa*, and did not pronounce the final *-ka* when reading the sentence.

(37 and 38), *veelai kiilaiyai* 'work and so forth' proved the least popular of five different options (mean score of 2.7). *veelaiyai kiilaiyai* was generally acceptable (mean score of 1.6), although the uninflected form *veelai kiilai* (see 65) was marginally preferable for most speakers (mean score of 1.5).[31]

Responses to the sentences involving dative cases revealed a consistent preference for echoing the case marker with the stem, although the alternative, i.e. adding the case marker to the end of the echoed stem, was disfavoured rather than definitively rejected. Thus *cennaikku kinnaikku* 'to Chennai and such places' (39) scored 1.4 overall, whereas *cennai kinnaikku* (40) had a score of 2; *tiruppatikku kiruppatikku* 'to Tiruppati and such places' (41) scored 1.3, whilst its counterpart *tiruppati kiruppatikku* (42) scored 2.3, and *kumaarukku kimaarukku* 'to Kumar or some such person' (43) scored 1.4 on average, as compared to 2.3 for *kumaar kimaarukku* (44). The same preference for echoing stem and marker together emerged from the sets of three sentences in (45)–(50): *citamparattukku kitamparattukku* 'to Chidambaram and such places' (45) received an average score of 1.4, and the common noun *tooṭṭattukku kiiṭṭattukku* 'to the garden and such places' (48) a score of 1, indicating unanimous approval. The options of echoing either the nominative stem on its own, i.e. *citamparam kitamparattukku* in (46), or oblique stem on its own, i.e. *citamparattu kitamparattukku* in (47), met with equally mixed responses, both scoring 2.4 on average. *tooṭṭam kiiṭṭattukku* (49), however, scored considerably higher than *tooṭṭattu kiiṭṭattukku* (50), with an average of 3.4 as opposed to 2.4. It seems reasonable to attribute the rejection of *tooṭṭam kiiṭṭattukku* to the disruption in segmental identity between the base and the reduplicant. Considerations of correspondence would certainly favour a sequence of [oblique stem+echo+marker] over [nominative stem+echo+marker], and this seems to be reflected, at least to some degree, in subjects' reactions.

The locative examples follow the pattern established for the accusative and dative, with *laṇṭanle kiṇṭanle* 'in London and such places' (51) being the preferred option (mean score of 1.5), whilst *laṇṭan kiṇṭanle* (52) was judged acceptable by some but not others, having a mean score of 2.3. For the sociative *-ooṭu* and genitive *-ooṭe* markers there was the same overall trend in the responses, although approval of the sentences with repeated case marker was more equivocal: *tiruṭanooṭu kiruṭanooṭu* 'with thieves and such people' (53) received a mean score of 1.8, and *kumaarooṭe kimaarooṭe* 'belonging to Kumar or some such person' (55) 1.9 (the positive context may be responsible for the relatively high score in this example). The corresponding formations without the first occurrence of the marker in (54) and (56), i.e. *tiruṭan kiruṭanooṭe* and *kumaar kimaarooṭe*, scored 2.3 and 3 respectively. This seems to provide some confirmation of Kothandaraman's claim that the genitive case exceptionally does not permit the echoing of a case marker (Kothandaraman 1983: 167), although it is unclear why this should be so. The alternative form of the genitive (with the *-in* suffix sometimes described as the euphonic increment) used in sentences (57) and (58) was generally rejected: *kumaarin kimaarin* (57) had a mean score of 2.8 and *kumaar kimaarin* (58) 2.9. A plausible explanation, suggested by one of the subjects, is that the *-in* suffix is too formal to be compatible with the colloquial echo expressions.

The sentences containing examples of the final case tested, the ablative, received rather mixed judgements, although *tooṭṭattu kiiṭṭattileruntu* 'from the garden or places like that' (60) was marginally preferred to *tooṭṭattileruntu kiiṭṭattileruntu* (59), with mean scores of 1.9 and 2.3 respectively. This runs counter to the trend seen in all the other examples, where repetition of the fully inflected form is favoured. As noted in section 3.2.3.3.1, the

[31] The echoed noun is followed in this context by a postposition *patti* 'about'. This is classed by Schiffman (1999: 43) as occurring with the accusative, but, as he notes (Schiffman 1999: 38), the behaviour of postpositions in Tamil is generally somewhat unpredictable, and case-marking seems to be optional before certain postpositions.

ablative differs from the other cases in having a composite marker, so this morphological complexity may explain subjects' preference for echoing only the stem. A tendency to avoid elaborate bases for echoing has already been observed in connection with the adjective *civappu*, which is exceptionally echoed alone, without its accompanying noun. In that case the number of syllables, i.e. phonological complexity, is arguably responsible for the difference, since all the adjectives tested were monomorphemic. In this instance *tooṭṭattileruntu* is indeed phonologically complex, comprising six syllables, but this is matched by *citamparattukku*, which is acceptable as a base for echoing. Morphologically, however, the two differ: *tooṭṭattileruntu* contains three morphemes, but *citamparattukku* only two, so this suggests that morphological complexity may be the crucial difference here. Some support for this comes from a comment by one of the subjects that *tooṭṭattile kiiṭṭattileruntu* is well-formed, since the base of this formation is bimorphemic. The complex ablative marker is here split into its two component parts, with the base containing the first part (and thus being identical to a stem with locative case-marking), and *-eruntu* appearing at the end of the echoed form. As noted in section 3.2.3.3.2 above, the division between case markers and postpositions is somewhat blurred: the behaviour of *-eruntu* may therefore align it with postpositions, rather than being evidence against the overall pattern of including case markers within echo formations.

3.3.2.3.2. Postpositions

The examples of postpositions preceded by case markers show a preference for echoing the case-marked noun on its own, with the postposition added at the end, a pattern that is in line with the behaviour of the ablative in the previous section. It is particularly clear for sentences (61) and (62), in which *viiṭṭukku kiiṭṭukku munnaalee* 'in front of the house or anywhere near it' (62) received general approval (mean score of 1.5), whilst *viiṭṭukku munnaalee kiiṭṭukku munnaalee* (61) was accepted by some but firmly rejected by others (mean score of 2.3). There was a marginal preference for echoing just the case-marked noun in *veelaiyai kiilaiyai patti* 'about work and such like' (37), rather than noun and postposition together in *veelaiyai patti kiilaiyai patti* (63), but both seemed to be generally acceptable (mean scores of 1.6 and 1.9 respectively). Reactions to the non-case-marked alternatives (see footnote 14) were more sharply divided, with *veelai kiilai patti* (65) acceptable (mean score of 1.5), but *veelai patti kiilai patti* (64) rejected by a majority of subjects (mean score of 2.6).

The reverse pattern, i.e. preference for the postposition to be included in the base for echoing, is found in the other two sets of examples, in (68)–(71): *meejai meelee kiijai meelee* 'on the table or nearby' (68) is preferred to *meejai kiijai meelee* (69), with average scores of 1.8 and 2.8 respectively, and *appaa kiṭṭa kippaa kiṭṭa* 'to father and the older generation' (70) is preferred to *appaa kippaa kiṭṭa* (71), with mean scores of 1.6 and 2.4. Both of these postpositions are preceded by oblique stems, as seen in *marattu meelee* (66), for example, although this is not apparent from either *meejai* or *appaa*, since their nominative and oblique stems are not distinct. These sequences of oblique stem+postposition, which are bimorphemic, therefore pattern with the bimorphemic case-marked nouns in their interaction with echoing. Case-marked noun+postposition sequences, however, which are trimorphemic, are not echoed together, and in this regard behave like the ablative case-marked nouns, which also contain three morphemes. This issue is discussed further in section 3.4.

3.3.2.3.3. Verbal forms

The integrity of various complex verbal forms was tested by considering their interaction with echoing, as described in section 3.2.3.3.3. The infinitival forms of 'strong' verbs, which contain a *-kka* marker, could only be echoed as a whole: *paṭi kiṭikka* 'to study and so forth' (72) was overwhelmingly rejected (mean score of 3.6), whereas *paṭikka kiṭikka* (15) was unanimously approved (mean score of 1). The two examples of forms built on the infinitive could also only be echoed in their entirety. Thus the negative infinitive *caappiṭale kiippiṭale* 'not to eat and so forth' (16) is an acceptable echo form (mean score of 1.4), but

caappiṭa kiippiṭale (73) was generally rejected (mean score of 3), even though the infinitive forms an acceptable base when no further markers are added. The partially echoed negative imperative *peeca kiicaatee* 'do not speak and so forth' (74) was decisively rejected overall (mean score of 3.2), although there was some controversy over the acceptability of the fully echoed alternative *peecaatee kiicaatee* (17): responses were split fairly evenly between 1 (firm acceptance) and 4 (firm rejection), producing an average of 2.2.

In each of the cases considered thus far, the final morphological marker has been written as part of the same word as its base: when infinitives are followed by auxiliary modal verbs, however, there is an orthographic word break between the two, as seen in (75). Echoing this sequence together, i.e. *paṭikka maaṭṭaan kiṭikka maaṭṭaan* 'will not study etc.', proved generally unacceptable (mean score of 3.1), whereas echoing only the infinitive, i.e. *paṭikka kiṭikka maaṭṭaan*, was perfectly acceptable (mean score of 1). A distinction thus seems to be drawn between markers which are too closely bound to the root to be separated from it by an echo, and are written as part of a single word, and more independent parts of the verbal complex, which are written separately, and cannot be included in the base for echoing. Note, however, that the positioning of word breaks in written representations of colloquial Tamil is not fixed. There is some flexibility in how sequences of verbal forms are written, with words more likely to be written as one when a final vowel is elided, e.g. *vantukiṭṭu irukka* vs. *vantukiṭṭirukka*.

Two examples of verbal forms built on the participle, rather than the infinitive, involve the completive verb *viṭu*. In both cases the option of including the aspectual marker in the base for echoing was preferred: *paaṭiṭṭu kiiṭiṭṭu* 'after singing and so forth' (77) scored 1.6 overall, whereas *paaṭi kiiṭiṭṭu* (78) scored 2.5, and responses were even more clearly differentiated for *tuuŋkiṭṭu kiiŋkiṭṭu* 'after sleeping etc.'(79) and *tuuŋki kiiŋkiṭṭu* (80), which scored 1.2 and 3.7 respectively. Example (81), in which an echoed participle precedes aspectual material, as part of an extended verbal complex *vantu kintukiṭṭirukka maaṭṭeen* 'will not be coming and so forth', was rejected by all of the subjects, with varying degrees of firmness (mean score of 3.1). For this particular sentence they were specifically asked to suggest alternative forms: some responded that there was no acceptable echo formation, and one gave the fact that three different words are involved as the reason. Echoing the base participle *vantu* with the aspectual participle *kiṭṭu*, i.e. *vantukiṭṭu kintukiṭṭu irukka maaṭṭeen*, was, however, acceptable to one subject. Two others suggested essentially the same form, but with *kinu* as the aspectual participle, in place of *kiṭṭu*, i.e. *vantukinu kintukinu irukka maaṭṭeen*. *kinu* is a dialectal variant of *kiṭṭu* regularly associated with northern Tamil (see, for example, Zvelebil 1964: 242). Overall, therefore, the preference seems to be for echoing sequences of participles together, regardless of whether these are followed by further verbal forms.

The compound verb that appears in examples (83)–(85) was tested on only a subset of the subjects, and this may be responsible for the polarized nature of the responses. A very clear pattern emerges: the option of echoing just the noun, i.e. *patil kitil collaatee* 'don't answer and so forth' (85) was fully acceptable (mean score of 1), whereas echoing just the verb, i.e. *patil collaatee killaatee* (84), or the noun and verb together, i.e. *patil collaatee kitil collaatee* (83), was rejected by all (mean scores of 3.7 and 3 respectively). The readiness of the subjects to introduce echoed material between the noun and verb is interesting in the light of the restrictions listed in 3.2.3.3.3, and may count against an incorporation analysis for the Tamil compound verbs.

Finally, there was no strong reaction to the issue of whether the quotative particle *-ŋŋu* should be included in an echo formation, or just added to the end. Overall, there was a slight preference for not repeating it: thus *teriyaatu-ŋŋu kiriyaatu-ŋŋu* 'I don't know and so

forth' (86) scored 2.1, whereas *teriyaatu kiriyaatu-ŋŋu* (18) scored 1.1, and *kumaarukku kuʈutteen-ŋŋu kimaarukku kuʈutteen-ŋŋu* 'gave it to Kumar or some such nonsense' (87) fared slightly worse than *kumaarukku kuʈutteen kimaarukku kuʈutteen-ŋŋu* (88), with mean scores of 1.9 and 1.5 respectively. There seems to be some room for flexibility and personal choice here, but the tendency overall is not to repeat the quotative particle.

3.3.2.4. Compounds

Subjects responded in clearly different ways to the three kinds of compounds included in the questionnaire. The first group, the coordinate or dvandva compounds, seemed to resist echoing altogether, whether of the whole compound or just one element, although the first option was marginally preferable. Thus *aaʈumaaʈu kiiʈumaaʈu* 'cattle and so forth' (89) scored 3.3 overall, being rejected by all except the unconventional fifth subject, whilst *aaʈumaaʈu kiiʈu* (90), with only the second element repeated, was even worse (3.5 on average). Either element of the compound, however, can be echoed when occurring as an independent word: both *aaʈu kiiʈu* 'goats and such like' (7) and *maaʈu kiiʈu* 'cows and such like' (8) have low average scores (1.2 and 1.6 respectively). The same pattern was seen in *appaa ammaa kippaa ammaa* 'parents and their generation' (91), which scored 2.3 on average, whereas *appaa kippaa ammaa* (92), with the first element repeated, was markedly worse, scoring 3.3 on average. Again echoing one of the components on its own is unproblematic: *appaa kippaa* (10) received an average score of 1.3. The general reluctance to echo these forms at all is in line with the observations made for Hindi (see section 2.4.2.3).

The endocentric compounds containing two nominative stems evoked a consistent pattern of responses, with repetition of the whole compound proving uncontroversially acceptable in each case: *talaivali kilaivali* 'headache and so forth' (93) scored 1.4, *maamaram kiimaram* 'mango-trees and so forth' (94) 1.2, *naaykkuʈʈi kiiykkuʈʈi* 'pups and so forth' (98) 1.1, whilst *maaŋkaay kiiŋkaay* (96) received a score of 1 from each of the subjects. Reactions to the alternatives, in which the first element only was echoed, were generally negative: both *maa kiimaram* (95) and *naay kiiykuʈʈi* (99) scored 3.6 overall and *maa kiiŋkaay* (97) 3.7, indicating clear rejection. This response pattern contrasted with subjects' reactions to the third type of compound, in which the first element is an oblique stem. Repetition of the whole compound was mostly judged unacceptable, with *tooʈʈattuppuu kiiʈʈattuppuu* 'garden flowers and such like' (100) receiving an average score of 3.1. However, echoing the second element only, i.e. *tooʈʈattuppuuvai kiivai* (101), was generally admissible, the mean score being 1.5. Once again, therefore, a bimorphemic base is preferred over a trimorphemic base. One of the subjects commented that the fully echoed compound was too much of a mouthful to say comfortably: this may well be due to the sequence of eight stops in this particular echo formation, with no two consecutive stops produced at the same place of articulation. Given this lexical peculiarity, it would be unwise to place too much weight on this example, since it is the sole representative of its class in the questionnaire. Further discussion of compounds and how they compare in structure with echo formations is to be found in section 4.6.3.

3.3.2.5. Phrases

Some evidence of the possibility of echoing adjectival and postpositional phrases in Tamil has already been seen: section 3.3.2.2 described the preference for echoing adjectives together with the nouns that they qualify, and in section 3.3.2.3.2 there were examples of postpositions following oblique stems that were echoed as one. The various other examples designed to investigate the echoing of phrases were set out in section 3.2.3.5 above; several of them were modelled on examples claimed to be acceptable by Nagarajan (1994). The first, containing noun and negative particle *viiʈu illai kiiʈu illai* 'no house and family etc.' (108), which is similar to an example of Nagarajan's given in (105), proved unacceptable to all the subjects, receiving an average score of 3.5. By contrast, example (109), in which only the noun is echoed, proved perfectly acceptable (mean score of 1). Tamil thus diverges from

several of the other Dravidian languages: parallel sentences to (108), containing noun and negative particle echoed together, are acceptable in Telugu, Malayalam and Kolami, as noted in section 2.4.1.4.

Responses to the various echoed verb phrases were mixed. The example that received the lowest average score and was therefore judged most acceptable involved a sequence of indirect object+verb. It is repeated from (88), and corresponds closely to Nagarajan's *raamanukku kuṭutteen* 'gave (it) to Rama' (106).

(122) [*kumaarukku kuṭutteen*] [*kimaarukku kuṭutteen*]-*ŋŋu poy collaatee* (= Q43b)
Kumar.dat give.past.1s echo quote lie say.neg.imp
'Don't lie that you gave it to Kumar or some such nonsense.'

This scored an average of 1.5, receiving 3 from one subject, but otherwise only scores of 1 or 2. The parallel sentence in which the quotative particle is repeated twice (87) was not as popular, but still scored 1.9 overall. In contrast, the examples in which only the verb (19) or only the noun (110) were echoed received lower scores, averaging 1.1. Even in a case like this, therefore, where echoing of the phrase is clearly acceptable, there is still a slight preference for the echoing of individual words instead. Example (123), repeated from (111) above, contains a very similar morphological sequence, i.e. a dative case-marked pronoun followed by a verb, although the syntax and semantics are rather different, since the pronoun acts as the semantic subject, rather than the indirect object, and the verb form is impersonal.

(123) [*enakku teriyaatu -ŋŋu*] [*kinakku teriyaatu-ŋŋu*] *collaatee* (= Q39c)
I.obl.dat know.neg.3sn quote echo say.neg.imp
'Don't say "I don't know!"'

Unlike (122), this proved to be unacceptable, receiving an average score of 3.3. In part this may be due to the bias against repetition of the quotative particle (see section 3.3.2.3.3), but the overall score is so high that this is unlikely to be the only reason. Moreover, the problem does not seem to rest with the component parts: the verb was echoed on its own, with and without the quotative particle, in (86) and (18), which received average scores of 2.1 and 1.1 respectively. There is no parallel sentence echoing only the pronoun, but the positive responses to (14), in which the nominative form was echoed, show that there is no general restriction against echoing pronouns. The problem thus seems to lie with the particular construction and, given that it is morphologically congruent to (122), this suggests that syntactic factors may be at work.

Responses to the other three echoed phrases considered fall somewhere between those for (122) and (123). The most acceptable contains a nominative subject echoed with a negative infinitive, and received equal numbers of scores of 1 and 3. This is repeated from (112) above as (124).

(124) [*kumaar paakkale*] [*kimaar paakkale*]-*ŋŋu poy collaatee* (= Q40a)
Kumar see.infin.neg echo quote lie say.neg.imp
'Don't lie that Kumar didn't see or some such nonsense.'

Again echoing the component parts of the phrase individually was acceptable: echoing the subject only (11) received an average score of 1.1, and echoing of the verb only (113) scored 1.3, and 1.1 when the subject was omitted (114).

Echoing of a subject and finite verb together was tested in (125), repeated from (115) above, and proved to be generally unacceptable (mean score of 2.6), although there was only one definite rejection, i.e. score of 4.

(125) *innikki* [*appaa varaar -ŋŋu*] [*kippaa varaar-ŋŋu*] *collaatee* (= Q48a)
today father come.pres.3pl quote echo say.neg.imp
'Don't say "Father's coming today and so forth".'

Echoing of the subject alone (10) was acceptable (average score of 1.3), but there was some uncertainty about the validity of echoing the verb (20), which is reflected in the overall score of 2.1. A similar pattern of responses was seen for the set of sentences associated with (126), which is repeated from (116) above.

(126) [*kumaar onnum paattaan -ŋŋu*] [*kimaar onnum paattaan-ŋŋu*] *poy collaatee*
Kumar anything see.past.3sm quote echo lie say.neg.imp
'Don't lie that Kumar saw anything and so forth.' (= Q46a)

Again echoing of the subject (12) was fine, scoring an average of 1.6, whereas echoing of just the verb (21) produced mixed responses (average of 2.1). The whole phrase (subject, object and verb) was also given the whole range of possible scores, averaging out at 2.3. For these last two examples, therefore, there appears to be some uncertainty about whether the tensed verb can legitimately be echoed, and this uncertainty is compounded when further items are included in the echo formation.

Overall, therefore, the questionnaire results provide only limited support for Nagarajan's claim that echoing of phrases is positively preferred to that of single words in Tamil:

> '*Besides echo formation or reduplication at the word level, it is possible (and in fact more natural) to repeat whole phrases or clauses in Tamil'* (Nagarajan 1994: 165).

Whilst this was found to be true of adjectival and some postpositional phrases, in none of the other examples tested were phrases judged to be more acceptable than echoing their component parts individually. Nevertheless, the questionnaire results do provide empirical confirmation that echoing at the phrase level is permitted in certain restricted contexts in Tamil.

3.3.2.6. Elaborated echo expressions

Although responses to the questionnaire were mostly restricted to comments on the acceptability of the sentences given, in a few cases subjects also suggested elaborations on the basic echo expressions that were acceptable, or even preferable, in their view. Four of them suggested the addition of *ellaam* in various contexts, and a couple also proposed making the echo formations into explicitly coordinated structures.

3.3.2.6.1. Echo expressions and *ellaam*

ellaam is a quantificational element which can occur in various syntactic configurations in Tamil (see, for example, Lehmann 1989: 109–111). When it functions as the modifier of a noun, it can appear either prenominally or postnominally, and means 'all'. Plural markers are more variable in colloquial Tamil than the formal language, and appear also to be less frequent: one of the subjects claimed that they would hardly ever use one. In some cases, *ellaam* fulfils this role, as Schiffman comments:

> *'Many Tamil speakers substitute* ellaam *'all, everything' as a postposed marker of plural, sometimes with other plural markers'* (Schiffman 1999: 28).

Semantically it has an obvious affinity with the generalizing sense of the echo formations, and can be used either as an alternative to echoing or in addition to it. For instance, subjects 6 and 12 both suggested the addition of *ellaam* to *kiirai* 'greens' (117), since echoing would produce a form homonymous with total reduplication. These two proposed that it should be added after the echoed portion in a number of other instances, whilst subject 11 favoured the addition of *ellaam* to almost every echo expression. He commented himself, however, that this preference was a matter of personal taste, not necessarily shared by other speakers.

The responses of subject 11 included cases where *ellaam* was added to various nouns, including case-marked examples, compounds, and even adjectival and postpositional phrases. In general, *ellaam* follows immediately after the echoed material, as in (127).

(127) *avan* [*aɻakaana paiyan*] [*kiɻakaana paiyan*] *ellaam ninaiccikinu mayaŋkaatee*
he handsome boy echo all think.vbp.dur enchant.neg.imp
'Don't get carried away by thinking that he is a boy who is handsome and has other good qualities!'

However, there were a couple of sentences in which the *ellaam* was attached at the end of the phrase containing the echoed material, such as (128) (cf. 29 above), an example offered by subject 6, who, contrary to the overall trend, preferred to take the adjective on its own as the base for echoing, rather than adjective and noun together.

(128) [*keṭṭa*] [*kiṭṭa*] *kanavai ellaam keekkaatee*
bad echo dream.acc all listen.neg.imp
'Don't heed bad dreams and so forth.'

Subject 11 also gave an example of this type, based on (37) above, in which a postposition intervenes between the echoed material and *ellaam*.

(129) [*veelaiyai*] [*kiilaiyai*] *patti ellaam collaatee*
work.acc echo about all speak.neg.imp
'Don't talk to me about work and such like!'

Phonologically, therefore, *ellaam* need not be directly adjacent to the echo formation, although it is grammatically closely associated with it. It produces no significant difference in the meaning, simply reinforcing the generalizing semantics of echoing. Whether or not it is used, and the extent of its use, seem to be matters of personal choice.

3.3.2.6.2. Echo expressions and coordination

The other elaboration, suggested by subjects 6 and 11, involves the introduction of the *-um* clitic, which marks coordination, into the echo formation. Both proposed this as a way of making example (108) acceptable, by adding *-um* to the noun and its echo, as in (130).

(130) *avanukku* [*viiṭum illai*] [*kiiṭum illai*]
he.dat house.co not echo
'He has no house and family etc.'

Subject 11 also suggested the insertion of *-um* into example (75). This involves an infinitive and auxiliary verb being echoed together, and was judged unacceptable without the clitic.

(131) *kumaar* [*paṭikkum*[32] *maaṭṭaan*] [*kiṭikkum maaṭṭaan*]
Kumar study.infin.co will-not.3sm echo
'Kumar will not study and so forth.'

The addition of the *-um* clitic to the coordinated constituents is the standard means of expressing conjunction in Tamil. They may be of any lexical category, as long as the conjuncts are of the same type, and it is also possible for phrases and even clauses to be conjoined, but not sentences.[33] Example (132), from Schiffman (1999: 191), shows a similar pattern to (130) and (131), with infinitives followed by the negative particle forming the conjuncts.

(132) *naan reṇṭu naaɻaa caappiṭavum illai tuungavum illai*
I two day eat.infin.co not sleep.infin.co not
'I didn't eat or sleep for two days.'

The semantics of coordination are implicit in echoing, which adds to the base a meaning along the lines of 'and other related things/properties/activities' (see section 2.3.2). Attachment of *-um* to the base and its echo, rather like the addition of *ellaam*, thus serves to

[32] Note that *paṭikkum* is a reduced form of *paṭikkavum*, in which the final vowel of the infinitive has been elided before the clitic.
[33] See, for example, Asher (1985: 67–73), Lehmann (1989: 239–242) and Schiffman (1999: 190–191).

reinforce a notion already inherent in the echo formation. The suggestion that the echoed part of the formation can host the *-um* clitic is of interest, given that it is otherwise only lexical items that can serve this function. Semantically the echoed part is in some respects parasitic upon the base, which might suggest a hierarchical structure for the echo words, with the base as the head. The fact that the lexical category of the whole is determined by that of the base might be taken to support this hypothesis. However, the option of conjoining the two parts instead implies a coordinate structure, with the echoed part having some degree of syntactic and semantic independence from the base. The possibility that an echoed part might take on a life of its own, as it were, is not unparalleled: there are examples in the literature from other Dravidian languages, as noted in section 2.3.2. For instance, Bhaskararao (1977: 10) observes that younger speakers of Telugu may give a separate meaning to the second half of the echo formation. He gives the example in (133), in which the echoed part is clearly syntactically independent of the base, and semantically distinct.

(133) *[pe[[i]* *ayindi kaani* *[gi[[i]* *avaleedu*
'The marriage took place but not *gi[[i* (the consummation).'

In other cases the echoed part seems to function as a derogatory sobriquet, as in (134), from Malayalam (Malten 1989: 44).

(134) *ummaan koʈuttaal* [*ammaavan*] *alleŋkil* [*kimmaavan*]
'As long as he feeds me, he's called "Uncle", but otherwise "Kincle".'

Such sentences show the potential for the reduplicant of the echo formations to develop syntactic and semantic independence. The conjoined examples suggest that there may be some progression in this direction in Tamil, although it is unclear how much can be deduced from this about the structure of echo words that are not explicitly conjoined. One remaining puzzle is why the addition of *-um* should make the echo formations acceptable when subjects reject the sentences without them. Possibly the introduction of an explicit structure of conjunction provides some kind of support for the second element, which is lacking in the echo expression.

3.4. Conclusion

A summary of the findings from the questionnaire is presented in table 4: the questions set out in table 2 of section 3.2.3.6 are repeated, and answers provided from the data that have been analyzed.

Table 4. Questions and answers

Questions		**Answers**
Can echo words occur in positive contexts?		Yes, but preferably in negative contexts
Which lexical categories can be echoed?	Nouns	Yes
	Pronouns	Yes
	Verbs	Yes, especially infinitives
	Adjectives	Only together with nouns
	Adverbs	Yes
What is the morphological make-up of the base?	Case markers	Stem+case preferred over stem alone, except for the ablative
	Postpositions	Non-case-marked stem+postposition
	Verbal forms	Infinitives and participles: auxiliaries not included
How do compounds and echoing interact?		Exocentric compounds not echoed: endocentric compounds echoed as a whole, unless first element is an oblique stem; then second element only is echoed
What kinds of phrase, if any, can be echoed?		Adjectival, postpositional and some verbal phrases
Can echo words involve total repetition?		Preferably not

As discussed in the preceding sections, there is considerable individual variation in the responses, so the results in the table represent overall tendencies, rather than categorical answers. The constituency of the base clearly varies, including not only individual words but also sequences of words in certain restricted contexts. Unlike Bengali and Kannada, which also admit phrasal echo expressions, subparts of words seem to be generally disfavoured as bases in Tamil, as evidenced by the preference for case markers to be included within the base. This could be motivated either by a well-formedness requirement on the base, or on the constituent that is separated from the base by the reduplicant. The latter seems to be more likely in Tamil, given examples like (16) and (73) (Q34), where **caappiṭa kiippiṭale* is rejected, even though a positive infinitive such as *caappiṭa* is perfectly acceptable as a base elsewhere. The problem therefore seems to lie with breaking the boundary between the infinitive and the negative marker *-le*, and could be formulated in terms of failure to meet the marker's subcategorization requirements.

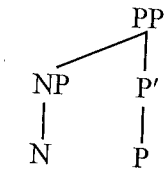
If a more precise analysis is to be formulated, the constituency that is relevant for determining the shape of the echo expressions, whether prosodic, morphological or syntactic, must be established. Interestingly, Lidz and Fitzpatrick-Cole come to very different conclusions on this issue in their studies of echo expressions in Kannada and Bengali. Lidz argues at length that echo reduplication respects morphological rather than phonological constituency, and that only syntactic constituents can be repeated in cases of phrasal echoing. Fitzpatrick-Cole, by contrast, adopts a prosodic analysis, arguing that the base for echoing corresponds to a well-formed prosodic domain in non-phrasal examples. In cases of phrasal echoing, she claims that in general any string of words within a default[34] phonological phrase

[34] Fitzpatrick-Cole adopts the intonational analysis of Hayes and Lahiri (1991), distinguishing between

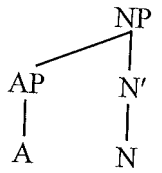
may be echoed, and that the reduplicated form as a whole must be contained within a single phonological phrase. She specifically argues against a syntactic or phonological characterization of the base, giving the example of *kʰub patla šari* 'very thin sari' (Fitzpatrick-Cole 1996: 334), where any individual word (except the adverb) or sequence of words may be reduplicated, including *patla šari*, which is not a syntactic constituent. She also notes that sequences of object+verb cannot be echoed together unless they form a complex predicate, attributing this to the phrase boundary said to precede verbs in Bengali (Hayes & Lahiri 1991: 92).

The Tamil data, by contrast, are largely consistent with a syntactic analysis, with echo bases restricted to syntactic phrases headed by lexical constituents. The postpositional examples form PPs and the adjective-noun sequences NPs, with the adjectival phrase in the specifier position of an X-bar tree, as in (135).

(135)



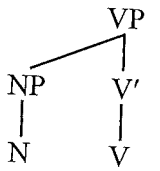
appaa *kiʈʈa*
father near



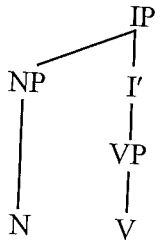
nalla *paiyan*
good boy

The sequence of indirect object and verb that proved widely acceptable forms a verb phrase, and differs crucially in its syntactic representation from a sequence of subject and verb, which together form a functional maximal projection in the version of X-bar theory given by Haegeman (1994). Such sequences were rejected as bases for echoing, even when the subject is marked with the dative case, and is thus morphologically indistinguishable from an indirect object. The two contrasting structures are illustrated in (136).

(136)



kumaarukku *kuʈutteen*
Kumar.dat give.past.1s



**enakku* *teriyaatu*
I.obl.dat know.neg.3sn

Another example rejected by the subjects, the sequence of noun plus *illai* 'not', is also not contained within a lexical maximal projection. Since all the acceptable instances of phrasal echoing do meet this criterion, it seems that this is a necessary, though not sufficient, condition for phrasal bases.

As noted at several points in section 3.3.2, a further restriction may be at work, favouring bimorphemic bases for echoing. This is seen most clearly in the interaction between echoing and postpositions, where postpositions are included in the base after a monomorphemic non-case-marked noun, but not when the noun is marked for case and so already bimorphemic. This would also explain the exceptional behaviour of the ablative case marker: since it is morphologically complex its inclusion would produce a trimorphemic base, and this is accordingly rejected by most subjects. The behaviour of the endocentric compounds provides further evidence for the same tendency: both elements are included when they are monomorphemic, and only one when the second element is inflected. In these instances the putative constraint seems to act as an upper bound on the size of the base; in

---

phonological phrases that are formed by default rules of phrasing and restructuring, and those that are the result of fast-speech restructuring (Hayes & Lahiri 1991: 92).

others it arguably acts as a lower bound, motivating the inclusion within the base of material beyond a single word if it is sufficiently closely associated syntactically. This may be true of some of the phrasal examples, such as the adjective and noun sequences, since the adjectives involved are monomorphemic and indeclinable. It could also underlie the choice of *vantukiṭṭu* or *vantukinu*, rather than just *vantu*, as the base in the extended verbal complex of sentence (81), although the nature of the boundary between the two participles may be the determining factor, and not just the morphological make-up of the base. The tendency for the base to be bimorphemic is, however, not adhered to without exception, as examples of polymorphemic bases such as *kumaarukku kuṭutteen* demonstrate.

The question of exactly how the syntactic and morphological restrictions on the base that have been discussed could be incorporated into a formal analysis will be discussed in detail in the concluding chapter. First, however, the issue of phonological constituency needs to be explored, to establish whether there are any prosodic restrictions on the base, as Fitzpatrick-Cole claims for Bengali. The next two chapters will therefore investigate the phonological structure of the echo expressions, through acoustic analysis.

Chapter 4: Tamil stress

4.1. Introduction

In this chapter and the next detailed acoustic analysis will be applied to the Tamil echo expressions to investigate their phonological constituency. Attention will be focussed on two properties: stress and the nature of the fixed echo /k/ segment. Since there has been very little experimental study of Tamil phonetics, much of these chapters will be devoted to background work, seeking to identify the phonetic manifestations of Tamil stress and the different realizations of velar obstruents respectively. The echo expressions will then be examined in the light of these findings.

Section 4.2 reviews prior work on stress in Tamil, introducing the claims about stress-related vocalic differences that will be tested in the rest of the chapter. Section 4.3 presents evidence for vowel reduction in non-initial syllables, and section 4.4 investigates the possibility that differences in vowel quality and duration are correlated with dynamic rather than fixed stress. Data on the monophthongization of /ai/ and its relation to syllable position are presented and analyzed in section 4.5. Finally, section 4.6 discusses what the vocalic qualities found in the echo words may reveal about their structure, comparing them with expressives and compounds.

4.2. Overview of the literature

Stress refers to the relative prominence of certain syllables, which provides speech with its rhythmic structure. It is a phonological property, and one that has no predictable phonetic manifestation. Hayes, for instance, introduces his book on metrical stress theory with the pronouncement:

> *'A body of careful experimental work has established that no one physical correlate can serve as a direct reflection of linguistic stress levels'* (Hayes 1995: 5).

Phonetic correlates of stress must therefore be sought on a language-specific basis, and may involve several different dimensions of variation.

The nature of stress in Tamil has yet to be satisfactorily addressed: references in the literature tend to be both cursory and cautious, and frequently call for further research. Stress is not lexically distinctive in Tamil and, in comparison with many other languages, it is not strongly marked, leading some to deny its existence altogether. Arden's grammar, for instance, states simply:

> *'There is no accent in Tamil. All syllables are pronounced with the same emphasis.'* (Arden 1934: 59).

This is by no means an isolated opinion: for comparable pronouncements on Tamil see Pope (1859: 22),[1] Jones (1967: 138), and more recently Arokianathan (1981: 5).

There is similar uncertainty over the existence and nature of stress in most other Dravidian languages: the literature on stress has almost entirely ignored the family as a whole. Steever offers the following generalization:

> *'Stress occurs in many Dravidian languages, but is not distinctive. Typically, stress falls on the first syllable of a word.'* (Steever 1998: 18).

He also claims fixed initial stress specifically for Gonḍi (Steever 1998: 274) and Kannada (Steever 1998: 131). An analysis of stress in Malayalam by Mohanan (1986: 111–115),

[1] Note, however, that Pope (1859: 22) qualifies his statement that Tamil in general admits of no accent with the comment:

> *'The root syllable will however be distinguished by something akin to accent'.*

however, claims that it is quantity-sensitive and thus dynamic, varying between the first and second syllable of a word. His basic generalization is that the primary stress falls on the first syllable, unless it has a short vowel and is followed by a long vowel, in which case the stress falls on the second syllable. A similar pattern is given for Brahui, a Dravidian isolate in the north of the Indian subcontinent. Elfenbein (1998: 394) notes that the first of any long vowels in a Brahui word attracts the stress, whilst a polysyllabic word consisting entirely of short syllables has initial stress.[2] The stress system of Koṇḍa is also described in comparable terms by Krishnamurti and Benham (1998: 244), who state that an initial syllable receives stress if the vowel is long but, if it is short, stress falls on the second syllable. Overall, therefore, two main possibilities have been suggested for Dravidian languages: fixed initial stress and some kind of quantity-sensitive system. Lack of evidence makes it impossible to establish at this stage whether either pattern is dominant.

Similar questions are debated for Tamil: there is disagreement over the location of stress, and whether it is fixed, i.e. tied to a particular syllable within the word, or dynamic, i.e. varies in its location. Andronov (1973: 114), for instance, claims explicitly that stress is free or dynamic in modern Tamil, and that this constitutes a change from the classical language. He presents various examples of the reduction or loss of vowels in initial syllables as evidence, but makes no suggestions as to the principles underlying stress placement in the modern language. Section 4.4 will return to this question, considering the possibility that Tamil stress is quantity-sensitive.

Amongst those who argue for a fixed accent in Tamil, the initial syllable is the preferred location.[3] Caldwell was an early advocate of this view: he states that the accent in Dravidian languages is uniformly acute and describes its locus as follows:

> *'The first syllable of every word may be regarded as the natural seat of accent; but if the word is compounded, a secondary accent distinguishes the first syllable of the second member of the compound.'* (Caldwell 1856: 142).

Trubetzkoy (1969: 297) also recognizes an initial fixed accent, but claims that a secondary accent is found on the final syllable of longer words. Caldwell comments on how well a fixed initial accent accords with the agglutinative structure of Dravidian words, in which the initial syllable or syllables form the lexical root (nominal or verbal) whilst later syllables carry inflectional and derivational information. Beythan too makes this connection, claiming that clear articulation differentiates the sense-bearing root syllables from the formative syllables:

> '*Die sinntragende Stammsilbe wird besonders deutlich artikuliert; sie wird sich dann gegen die formativen Silben behaupten.*' (Beythan 1943: 21).

Such comments suggest that Tamil may be an example of what Hayes terms a 'morphological' stress system (1995: 31). In a discussion of the typology of stress systems, he proposes a distinction between rhythmic and morphological stress. In the second category stress is said to elucidate the morphological structure of words, often with a particular syllable of the root bearing stress whilst affixes are either stressless or only weakly stressed. This description accords well with a system of fixed initial stress in Tamil.

In the modern literature, Balasubramanian lends some tentative support to the first syllable being the locus of stress, at least for words in isolation, on the basis of his own

---

[2] Note, however, that a large proportion of Brahui tribesmen are bilingual, also speaking the Iranian language Balochi. Close contact with Balochi has been ongoing since at least the twelfth century, and it has had a considerable effect on the sound system, eliminating almost all oppositions foreign to Balochi (Elfenbein 1998: 391). It is possible, therefore, that the quantity-sensitivity of the modern Brahui stress system is an innovation due to interference from Balochi, and not part of the language's Dravidian inheritance.

[3] Andronov (1973: 112) makes a passing comment about the final syllable being favoured by some, but gives no supporting references.

intuitions as a native speaker (Balasubramanian 1980: 453). Marthandan, whose judgements are based largely on proprioception, also comments that it is predominantly the first syllable that receives stress:

> *'Position of stress in the word is by no means fixed to any syllable of the individual word; yet in connected speech it is found more often than not in the initial syllable of the word except in cases of emphasis'* (Marthandan 1983: 309).

The phonetic correlates of stress in Tamil are the subject of some comment in the literature, although these are for the most part impressionistic, and sometimes contradictory. Christdas (1988: 188), for instance, comments that pitch, intensity and duration, prime exponents of stress cross-linguistically, are all unaffected. Claims to the contrary include Nagarajan's statement that extra low pitch is a correlate of prominence in Tamil (Nagarajan 1994: 82), and references by Andronov (1973: 113) and Asher (1985: 231) to intensification of the voice on a stressed syllable. Duration has also been implicated: Andronov (1973: 115), for instance, comments that the length of vowels may be in some way dependent on accent in the colloquial language, but gives no supporting evidence. Balasubramanian (1980) claims that vowel duration varies according to the prosodic weight of the syllable in which it occurs. He proposes a four-level hierarchy, which will be discussed in more detail in section 4.4, but makes no connection with stress placement.

Acoustic evidence has been brought to bear on stress in only a limited number of studies, and the results have not been conclusive. For example, Balasubramanian undertook some investigation of intensity, using mingograms to plot expiratory air-flow. Although differences in intensity could be linked to both vowel quality and length, there were no clear results with regard to stress in normal utterances (Balasubramanian 1972: 529). He also conducted a study of emphatic stress (briefly reported in Balasubramanian 1980: 456), finding that this was uniformly manifested in the initial syllable, which supports the case for this being the locus of normal stress. Clear differences in the acoustic signal were identified for the syllables bearing emphatic stress.[4] They included various lengthening effects: of an initial long vowel (by a factor of as much as two), of a geminate consonant, of the closure phase of a voiceless plosive and of a word-initial liquid or approximant.[5] Balasubramanian also found glottal onsets for emphasized words beginning with short vowels. Pitch rises were reported in earlier work for syllables marked with emphatic stress (Balasubramanian 1972: 544), and there is some suggestion that this also extends to non-emphatic stressed syllables (Balasubramanian 1980: 456). Overall, however, Balasubramanian concluded that there was no solid experimental evidence for any phonetic correlates of non-emphatic stress (Balasubramanian 1980: 453).

Phonological diagnostics of various kinds have been used to argue that initial syllables are prominent. This is the approach taken by Christdas, who proposes that stressed syllables are marked out by not being subject to certain rules of reduction, and defines accent as *'the property that blocks the application of such reductions'* (Christdas 1988: 196). She gives four rules, the first two of which apply to all non-initial syllables. The second two, however, are properties of final syllables that do not apply in monosyllables, and as such might be attributable to a 'minimal word' constraint.[6]

---

[4] Note that Tamil also uses particles, notably the clitic *taan*, to mark emphasis.
[5] This last finding is disputed by Marthandan (1983: 311), although his objections are based on intuition rather than evidence to the contrary.
[6] For discussion of minimal word constraints see, for example, McCarthy and Prince (1986, 1990), and Itô (1990) on Japanese.

1. vowel reduction, whereby underlying /a/ and /u/ in non-initial syllables are centralized to [ɜ] and [ɯ][7] respectively, e.g. *irukkiratu* 'it is' is realized as [iɾɯkɪɾəðɯ].
2. nasal assimilation, whereby nasals in a syllable coda assimilate in point of articulation to a following stop in any but the initial syllable.
3. nasal deletion – in spoken Tamil final nasals are deleted with concomitant nasalization of the preceding vowel, e.g. *maram* 'tree' [mɑɾɜ̃]. This is said not to apply, however, to monosyllables with nasal codas, e.g. *naan* 'I' [nɑːn], not *[nɑ̃ː].
4. glide deletion – according to Christdas, final palatal glides are deleted in non-initial syllables, and so maintained in monosyllables, e.g. *kaay* 'fruit' [kɑːj].

Christdas' analysis is set within Lexical Phonology, and so she captures the special status of the initial accented syllables by specifying a non-accented feature [–*] in the environment of the relevant rules. A reworking of the same data in the framework of Harmonic Phonology is to be found in the work of Bosch and Wiltshire (Bosch 1991, Wiltshire 1992 and Bosch & Wiltshire 1993). They relate the differences between accented and unaccented syllables to their licensing properties, arguing, for instance, that unaccented syllables cannot license [round] and [low], and hence cannot contain [u] and [ɑ], which are specified for these features. Unaccented syllables also license only a restricted set of features in their codas, and crucially not point of articulation features, which accounts for the assimilation of place that applies to nasals, and the absence of heterorganic stop-stop clusters from non-initial syllables. Distributional arguments of this kind certainly suggest that the initial syllable is in some ways phonologically different from the rest.

The phonetic effects of reduction applying to vowels in non-initial syllables are described in terms of centralization (Christdas 1988: 193), and also a tendency for the durations of /i/, /a/ and /u/ to reduce as they move further from the initial accented syllable (Christdas 1988: 175). Schiffman (1999: 17) comments on the fronting of /a/ in non-initial syllables, mentioning the final /a/ of infinitives as particularly likely to undergo fronting. He also refers to the unrounded and somewhat fronted quality of non-initial /u/, describing it as '*more like IPA [ɯ] or [ʉ]*'.[8] As he remarks, the [ʉ] symbol is also appropriate in some cases for the realization of a short /i/, suggesting some merger between the two high vowels. Dialectal variation may well affect the range and distributions of vocalic realizations. For his own dialect, Balasubramanian proposes an underlying inventory of seven, rather than the usual five, short vowel phonemes: /ɪ/, /ɛ/, /o/, /ʊ/,[9] /ɑ/, /ə/, and /ɨ/ (Balasubramanian 1972: 102–103). He associates both /ɪ/ and /ɨ/ with orthographic <i>, and /ʊ/ and /ɨ/ with orthographic <u>. /ə/ is said to be represented by either <a> or <e>, although no attempt is made to define the relevant environments.

Another process widely reported in the literature as affecting only non-initial syllables is the monophthongization of the diphthong /ai/ to [ɛ]. Schiffman (1999: 23), for example, lists it in his summary of phonological rules applying in standard spoken Tamil and notes that it does not occur in monosyllables or the initial syllables of polysyllabic words. Internal syllables, however, are said to be affected, along with final /ai/, whether that forms an accusative ending or part of a nominal root.[10] There are some indications that the

[7] Christdas' use of IPA symbols is somewhat misleading here since [ɯ] is no more centralized than [u], the difference between them being one of rounding only.

[8] The IPA symbol [ɯ] denotes a back, unrounded vowel, and [ʉ] a more central, rounded vowel.

[9] The current equivalent IPA symbol has been substituted for the closed omega used by Balasubramanian.

[10] For similar comments see Britto (1986: 198) and Steever (1998: 14). Balasubramanian (1972: 218) and Asher (1985: 220) also mention the same process, although they both describe it as being restricted

monophthongization may be dialectally restricted: Asher notes that word-final /ai/ in the formal variety corresponds to *-e* or *-a* in some dialects (Asher 1985: 219). Moreover, the absence of any mention of monophthongization in Christdas' work (although it would appear useful further evidence for her claims about stress), might be due to its absence from the Kanniyakumari dialect on which her account of Tamil phonology is based.

The following sections bring acoustic analysis to bear upon the question of Tamil stress, examining phonetic properties that may differentiate stressed from unstressed syllables. Vowel reduction is investigated in section 4.3 and monophthongization of /ai/ in section 4.5.

4.3. Vowel reduction

4.3.1. Data collection

A series of recordings were made in Pondicherry, using a portable DAT recorder with lavalier microphones (Audio-technica AT803b). For this purpose three subjects were used: two male speakers (BB and RP), and one female (AL). They were asked to read out five times a set of 79 Tamil sentences, which are listed in full with glosses and translations in appendix F. The sentences were designed with the aim of investigating the various distinctions that have been associated with stress. They are relatively simple in structure and the language is colloquial, although they were presented in written form. For instance, they contain a number of echo words and expressives, both characteristic of spoken Tamil, and a number of reduced verb forms, e.g. *irukku* 'it is' (written *irukkiratu*) and *pooraan* 'he goes' (written *pookiraan*). The recorded data were digitized at a rate of 16,000 samples per second (16 bit resolution) and segmented into sound files, one for each sentence, so that acoustic analysis could be performed.

In order to investigate the acoustic realization of vowels in the recorded sentences, sets of words containing /i/, /a/ and /u/ vowels were chosen. As far as possible, these sets were balanced with regard to adjacent segments, and also syllable position. For each example fifteen tokens were recorded (five for each of the three speakers), although a few tokens were excluded because of interference from background noise, accidental slips by the speaker or the absence of identifiable formants in the vocalic portions of the signal. The sets of words are listed below, with the vowels of interest emboldened: /a/ in table 1, /i/ in table 2 and /u/ in table 3. The numbers in brackets after each allow the wider context to be identified, by indicating the sentence in which the word occurred (see appendix F). They are arranged in columns according to whether the relevant vowel falls in an initial, medial or final syllable, and the rows are roughly sorted by phonetic environment, with the preceding and/or following segment being matched across columns.

---

to word-final position in polysyllabic words, rather than affecting all non-initial syllables.

Table 1. Words containing /a/

Context	Initial	Medial	Final
v_n	*vantaa* (20)	*avanukku* (57)	*avan* (15)
v_	*varaar* (61)	*koɳʈuvaraatee* (44)	*kiɻʈavan* (12)
l_	*laɳʈan* (66)	*uḷḷaŋkai* (33)	*koolam* (50)
_n	*anpu* (75)	*mookanin* (32)	*paiyan* (28)
r_	*rattam* (67)	*paattirattai* (19)	*citamparam* (41)
t/ʈ_	*talai* (36)	*paattatu* (43)	*eŋkiʈʈa* (53)
t_m	*tampi* (32)	*citamparam* (41)	*rattam* (67)
p_ra	*parantatu* (65)	*citamparam* (41)	
_nta	*anta* (69)	*parantatu* (65)	
p_ʈi	*paʈiccaan* (1)	*ippaʈi* (1)	
k_	*kattatu* (7)		*uʈaikka* (19)
_ll	*nallatalla* (10)	*nallatalla* (10)	*aɻʈakaana* (49)
n_		*enakku* (4)	*ena* (77)
y_n		*paiyan-ŋŋu* (15)	*mookan* (12)
ʈ_		*paʈattai* (40)	

Table 2. Words containing /i/

Context	Initial	Medial	Final
t_	*tiruʈan* (52)	*paattirattai* (19)	*patti* (75)
p_	*piʈikkaatu* (3)	*caappiṭṭaan* (8)	*tampi* (32)
_pp, l_	*ippaʈi* (1)	*kelippaay* (48)	*eli* (43)
t/r_	*tinnaatee* (6)	*ciriyatu* (22)	*kari* (30)
_n	*inta* (22)		*mookanin* (32)
ʈ_		*piʈikkaatu* (3)	*ippaʈi* (1)
tt_		*kiṭṭiyatu* (14)	*naaykkuṭṭi* (44)
k_	*kili* (10)		
_r	*irukkaan* (48)	*irukkiraan* (51)	

Table 3. Words containing /u/

Context	Initial	Medial	Final
k_ʈ	*kuʈutteen* (63)	*naaykkuṭṭi* (44)	
	muʈiyum (71)	*viiṭṭukkuḷḷee* (52)	
k_m	*kumaar* (34)		*piʈikkum* (67)
k_	*kukai* (22)		*enakku* (14)
tt_		*marattukku* (65)	*neettu* (37)
tt_		*viiṭṭukkuḷḷee* (52)	*kaaṭṭu* (55)
ʈ_		*iʈuveen* (50)	*eʈu* (40)
r_		*kumaarukku* (23)	*kiŋaru* (16)
r_		*irukku* (57)	*oru* (43)
ʈ_		*koɳʈuvaraatee* (44)	*maaʈu* (31)

4.3.2. Description and analysis of data

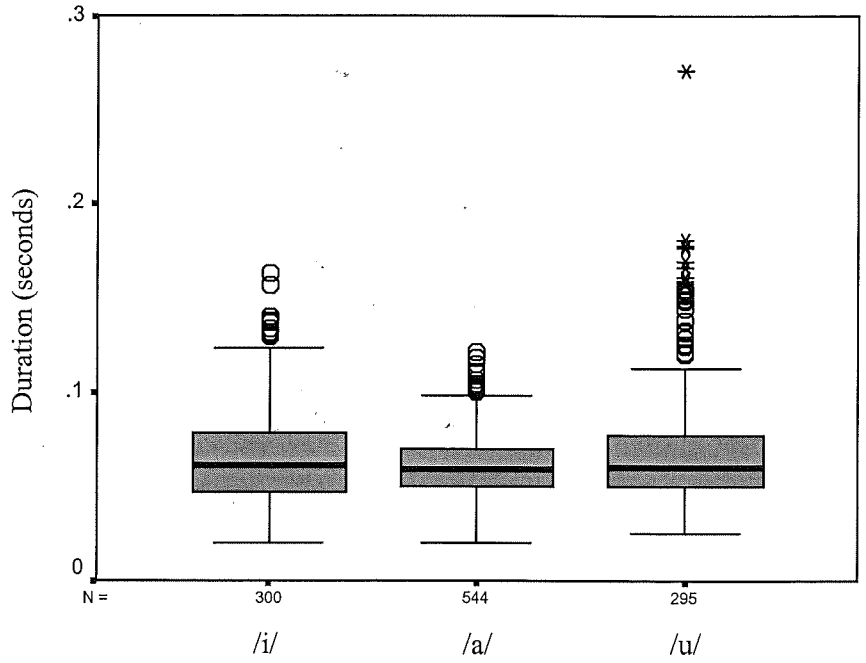
For each vowel token cursors were set at the boundaries using the software package ESPS/*xwaves*™, and several measurements were recorded automatically. The duration of each vowel was measured, and the values of the first and second formants at the mid-point of the vowel, in order to reduce the effects of transitions to and from the neighbouring segments as far as possible. Linear predictive coding analysis was used, applying autocorrelation within a pitch period to identify broad spectral peaks corresponding to vocal tract resonances. Twelve coefficients were used in the linear predictive equation, and the window duration was 49 milliseconds, with 5 millisecond steps between analysis frames.

Although only the first and second formants were analyzed, the formant tracking was initially set to locate three formants. This worked well for the two male speakers, but some difficulties were encountered with the female speaker, particularly for high F2 values. The problem was largely resolved by resetting the tracking to search for four formants, and so different settings were employed during the data collection, according to the gender of the informant. The measurements were carefully checked for plausibility by looking at the frequency values and marking any that were significantly out of line with other values for the same speaker. In the few cases where the formant tracking was clearly picking up the wrong frequencies the correct values were found using spectra at the mid-point.

Since the data involved five repetitions of the sentences by each speaker, a repeated measures general linear model was employed where possible in the statistical analysis. Subsets of the data for each vowel type were arranged with speaker defining the rows, taking a mean over the five different repetitions of the relevant segment. Phonetic environment and syllable position (initial, medial or final) were defined as the within-subjects factors. The model therefore tested for a gradient effect determined by syllable position in the variable under investigation (duration, F1 or F2). In most cases, however, the question of interest involved a two-way distinction, between initial and non-initial, and so there were insufficient degrees of freedom for a repeated measures analysis to be conducted. In these instances, the data were checked for homogeneity of variance using the Levene test, and if homogeneity was found an analysis of variance was carried out. In some cases performing a log transformation on the data reduced the inhomogeneity sufficiently, and this is noted in the text where relevant. Otherwise, non-parametric tests, which do not assume homogeneity of variance, were employed: the Mann-Whitney test for two independent samples, and the Kruskal-Wallis test for three or more samples.

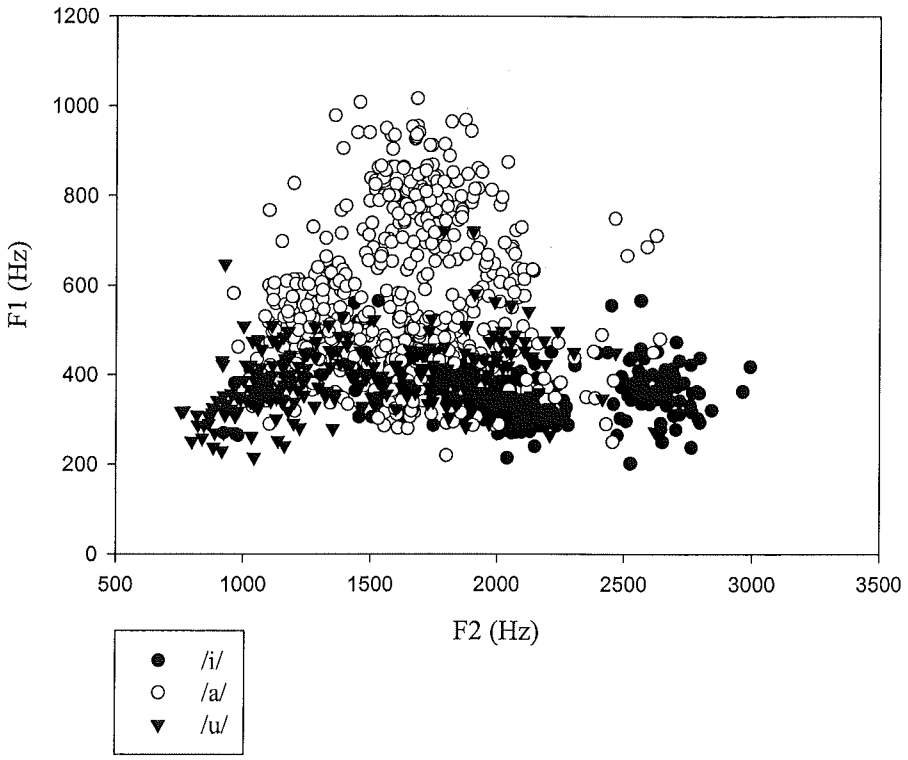
The total number of tokens measured was 1139, divided between the three vowels as follows: /a/ 544, /i/ 300, /u/ 295. Overall there appear to be no significant differences in duration between the three vowels: the boxplots in figure 1 of duration against vowel show no discernible pattern. Statistical testing confirms this: the Kruskal-Wallis test applied to the three vowel samples shows differentiation by duration to be insignificant ($p < .163$).

Figure 1. Boxplots of duration of /i/, /a/ and /ù/



The nature of the vowel does, however, prove highly significant ($p < .0005$) in determining the values of both F1 and F2, according to the Kruskal-Wallis test. The scatterplot in figure 2 shows F1 (on the vertical axis) against F2 (on the horizontal axis) for the full set of vowels. Values for the three different vowels are differentiated by shape: tokens of /a/ being represented by empty circles, tokens of /i/ by filled circles and tokens of /u/ by triangles.

Figure 2. Scatterplot of F1 against F2 for /a/, /i/ and /u/



There is considerable overlap between the realizations of the different vowels, and particularly between /i/ and /u/, which is in line with comments in the literature about a

partial merger between the two. Moreover, this is not simply an artefact of taking all three speakers together: the overlapping patterns are replicated in individual scatterplots for each speaker, and particularly for speaker BB.

Some idea of the interspeaker variation in the data can be gained from figures 3, 4 and 5, which show the median values of F1 and F2 for each speaker for the vowels /a/, /i/ and /u/ respectively.

Figure 3. Line diagram for /a/

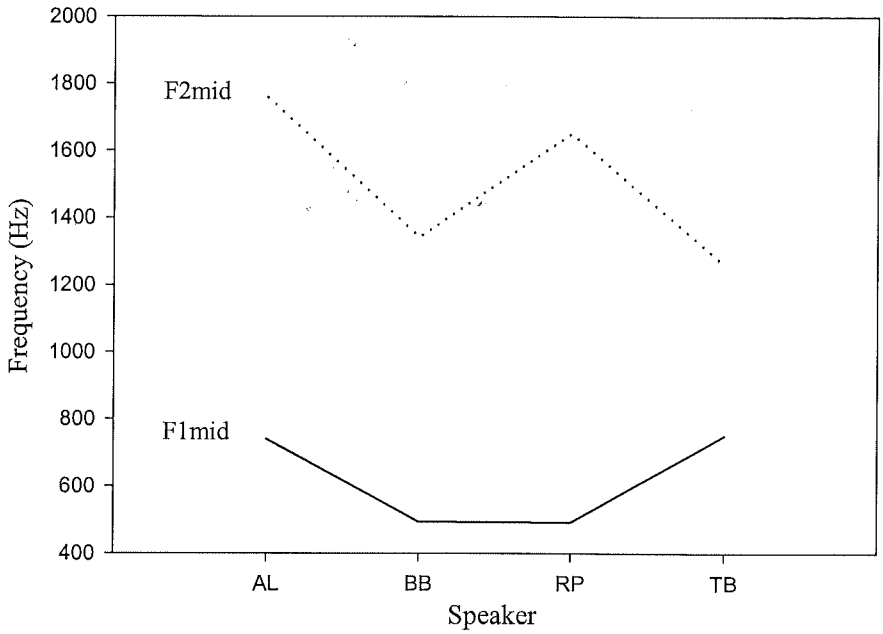


Figure 4. Line diagram for /i/

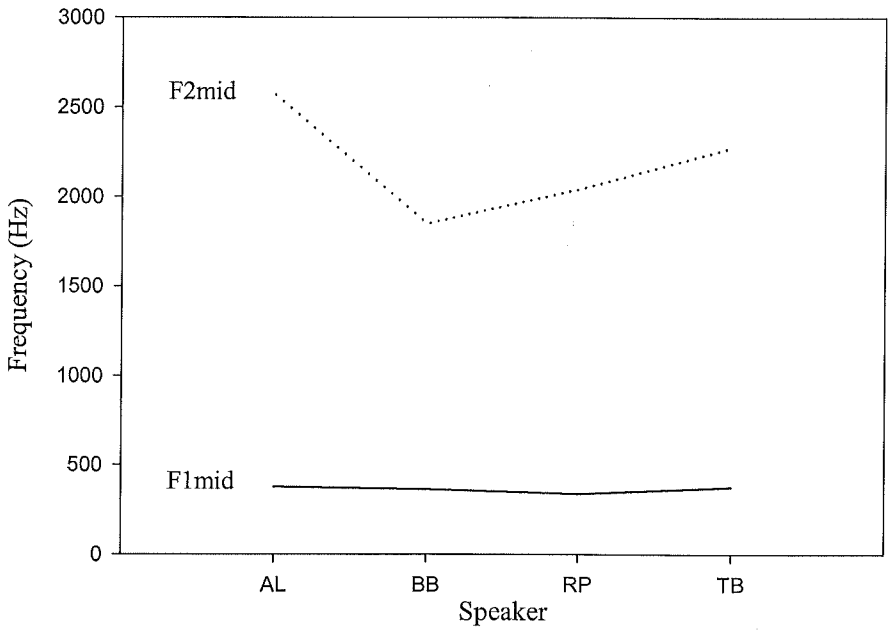
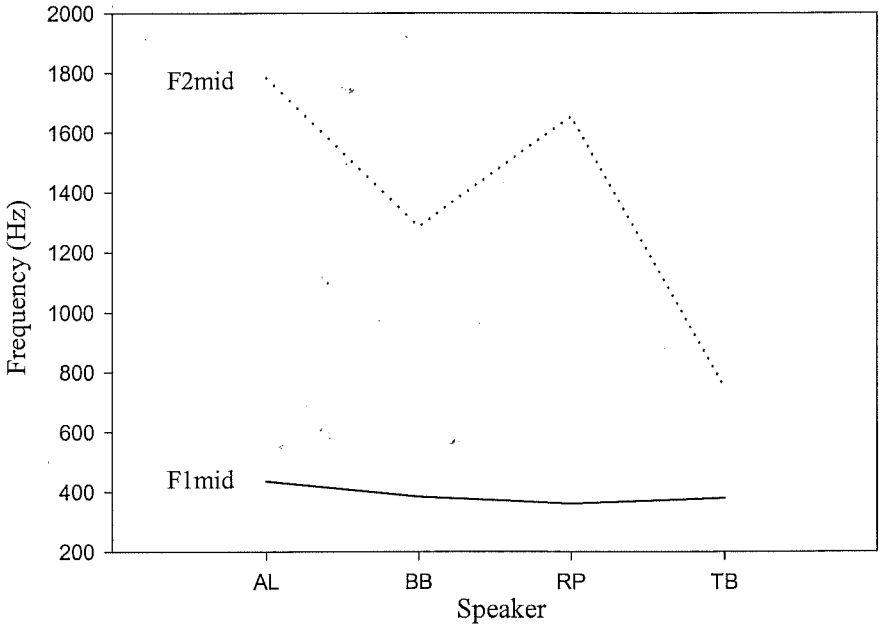


Figure 5. Line diagram for /u/



The median values are used because they give a more realistic indication of the average than the mean when the overall distribution of values is skewed. Moreover, the medians of the three values given by Balasubramanian (1972: 89) for /ɑ/, /ɪ/ and /ʊ/ in his own speech are also included as a comparison (labelled TB).[11] These prove to be relatively close to those of the three informants, falling within the range covered by the values for /i/ and just outside those for /a/. The only notably discrepant value is the second formant frequency for /u/: Balasubramanian's median value of 750 Hz is much lower than those of all three subjects.

As the figures demonstrate, there are consistent differences between the speakers that hold for each of the vowels (and, indeed, for each subset when the vowels are divided according to whether they occur in initial, medial or final syllables). The values of the female speaker are highest on average, for both F1 and F2. Speaker BB has marginally higher values for F1 than speaker RP, but the situation is reversed for F2, with RP showing higher values on average than BB. These differences are statistically significant: when the data were grouped by speaker and submitted to the Kruskal-Wallis test for independent samples both F1 and F2 proved highly significant for all speakers ($p < .0005$). There is no significant interspeaker variation for duration, however, according to an analysis of variance taking duration as the independent variable and speaker as the only fixed factor ($p < .417$).

The effect of the surrounding phonetic environment on the formant frequency values is potentially a confounding factor in the analysis. Efforts were made to minimize this effect by taking values at the mid-point, and also balancing the data sets as far as possible. Nevertheless, the influence of neighbouring segments on the quality of the preceding or following vowel might still be affecting the data. Two environments that are quite widely represented in the data and are likely to produce a high degree of coarticulation involve neighbouring nasal and retroflex segments. Impressionistically, anticipatory lowering of the velum before a nasal segment results in some nasalization of preceding vowels. Acoustically this has the effect of introducing nasal formants and antiformants into the signal, which typically reduce the overall energy of the vowel and widen the formant bandwidths, thereby masking the pattern of the oral formants (Kent, Dembowski & Lass 1996: 196–197). To test whether this is the case here, each set of vowels was divided according to whether or not the

[11] Balasubramanian's vowel tokens were all uttered in monosyllables of the form /pV/ to reduce the effects of phonetic context on the formant frequencies.

vowel was followed by a nasal segment and the two subsets compared. /i/ tokens followed by a nasal proved to have slightly lower F1 values than those not followed by a nasal ($p < .04$). Conversely, in /a/ and /u/ tokens it was F2 that was affected, again being lowered before a nasal ($p < .023$ for /a/ and $p < .001$ for /u/ on Mann-Whitney tests).

The characteristic curled posture of the tongue in retroflex articulations might also be expected to influence neighbouring vowels. However, previous research has concluded that the effect is largely restricted to the third and fourth formants. Ramasubramanian and Thosar (1971: 73), for instance, noted a lowering of F3 in vowels adjacent to retroflex stops in Tamil. Stevens and Blumstein, in a theoretical discussion of the acoustic consequences of retroflexion note that:

> *'the overall acoustic pattern is characterized by a clustering of formants F2, F3 and F4 in a relatively narrow frequency region . . . and a consequent accentuation of the spectrum in this region'* (Stevens & Blumstein 1975: 219).

Spectrographic analysis of data from Hindi speakers confirmed this prediction, and perception experiments using synthesized stimuli showed that the frequency of the burst for a retroflex stop should correspond to that of F4. As Ladefoged and Maddieson note, the articulatory configuration for Hindi is less extreme than for the Dravidian languages, with the apical edge rather than the underside of the tongue making contact with the hard palate (Ladefoged & Maddieson 1996: 27). The acoustic effects of the two articulations, however, differ only in degree. The only reference to any effect on F2 in Tamil is found in Walldén (1974–1975: 201), who comments that F2 is usually lower in a vowel followed by a retroflex consonant than in one before a non-retroflex consonant.[12]

The expectation, then, is that the F1 and F2 frequencies under analysis should not be significantly affected by adjacency to retroflex segments, although F2 may be lowered slightly. This was largely borne out when the formant frequencies of those vowel tokens followed by a retroflex consonant were compared with the rest. A significant difference in F1 was found only in /u/ ($p < .01$ on a univariate analysis of variance), with those tokens preceding a retroflex consonant having slightly lower first formant frequencies overall. F2 values were significantly different for all three vowels, showing lowering in /i/ and /u/ tokens ($p < .026$ on an analysis of variance and $p < .0005$ on a Mann-Whitney test respectively), but higher F2 values for tokens of /a/ before retroflex than non-retroflex consonants ($p < .004$ on an analysis of variance). Overall, no consistently strong effects arising from proximity to either a nasal or retroflex consonant can be identified. Hence there is no need to exclude such cases from the data set: coarticulatory effects, although detectable, are not sufficiently strong to confound the analysis.

4.3.2.1. Analysis of /a/ tokens

The data sets for each vowel type were analyzed with particular attention to any differences that might correlate with the position of the syllable containing the vowel. Accordingly, the duration, F1 and F2 values for all the /a/ vowels were grouped according to whether they occurred in an initial or non-initial syllable. The non-parametric Mann-Whitney test produced highly significant results for all three sets of measurements ($p < .0005$ for duration and F1, $p < .001$ for F2). The direction of the differences can be seen in the boxplots in figures 6 to 8, which show the measurements for initial, medial and final syllables. For each speaker the first syllable is characterized by longer duration, higher F1 and lower F2 values, with the differentiation being clearest for the female speaker, AL. The magnitude of the differences is, however, low, as shown by the degree of overlap on the boxplots.

---

[12] This claim is made on the basis of spectrographic analysis, following on from an impressionistic study reported at much greater length in Walldén (1973). Her concern is with the precise environments in which /i/ and /e/ are centralized in Tamil, and thus covers both the coarticulatory effect of a following retroflex consonant and the lowering of /i/ discussed in section 4.6.1.

Figure 6. Boxplots of duration of /a/ against syllable position

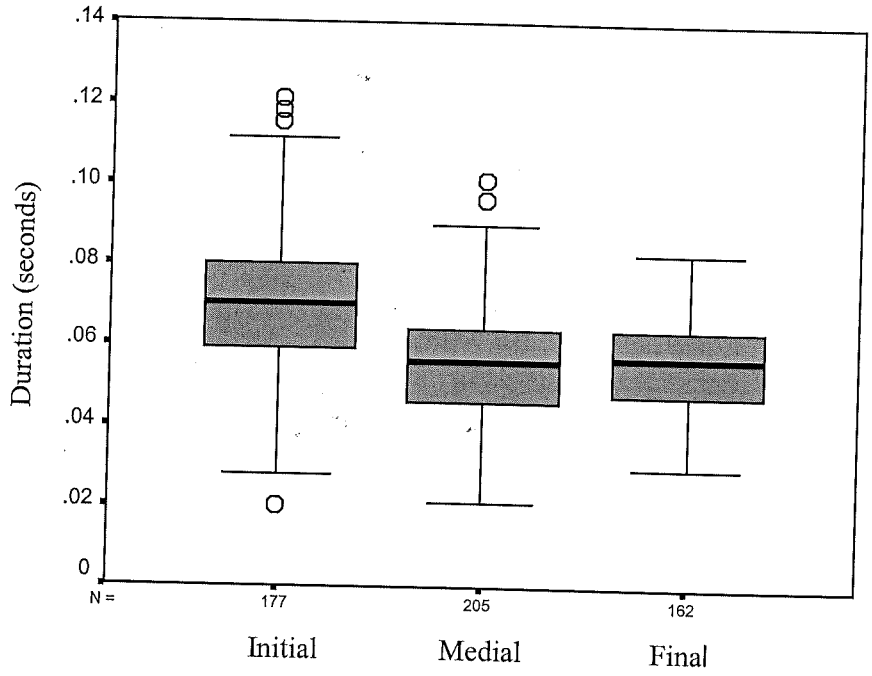


Figure 7. Boxplots of F1 values of /a/ against syllable position, clustered by speaker

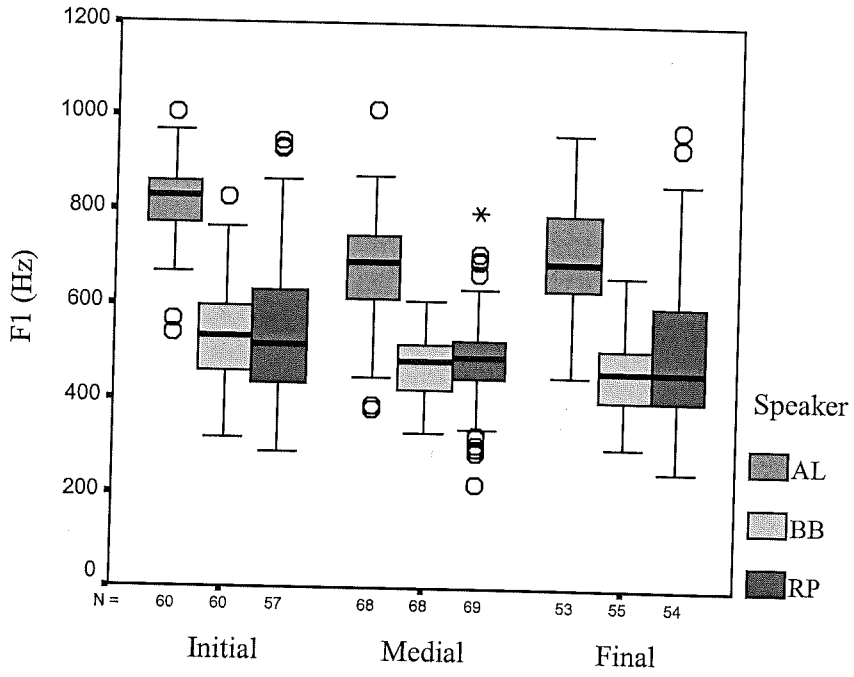
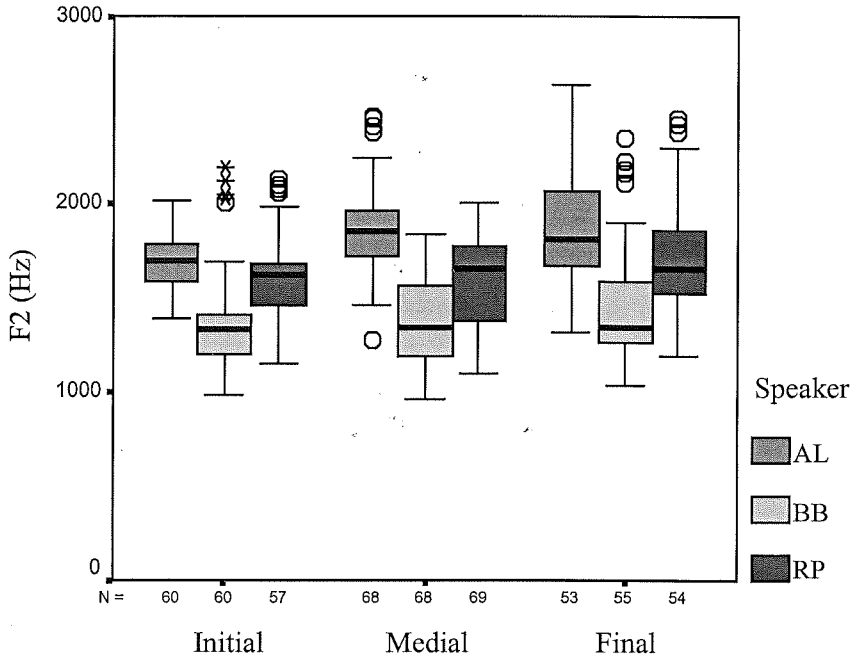
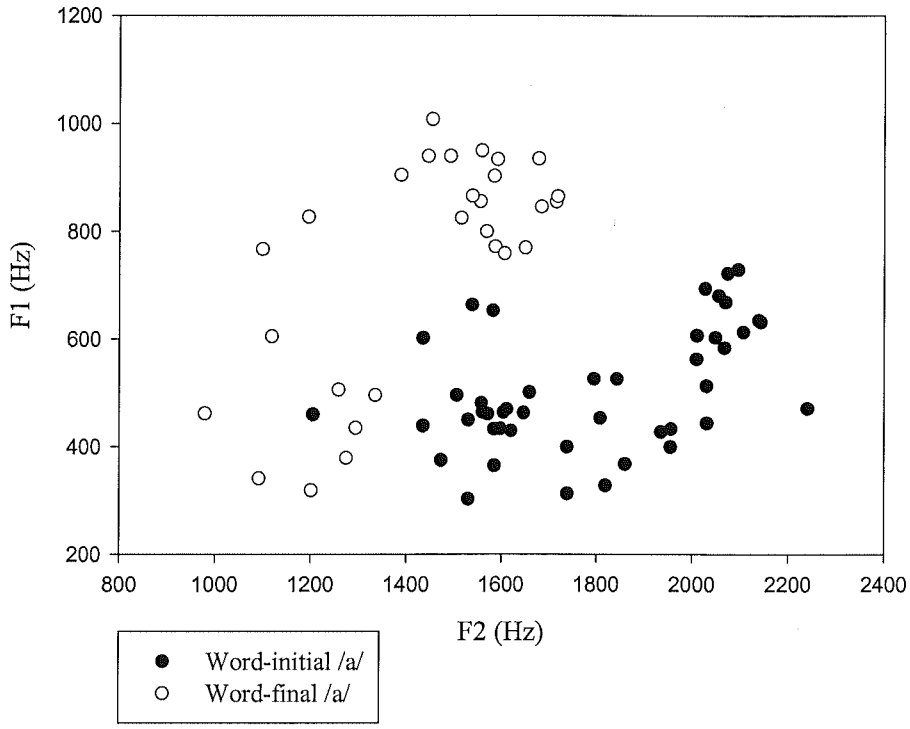


Figure 8. Boxplots of F2 values of /a/ against position, clustered by speaker



A clearer picture is gained by looking at a subset of the data and contrasting word-initial /a/ with word-final /a/ in the following words: ***a***npu, ***a***nta, en***a***, eŋkiʈʈ***a***, uʈaikk***a*** and aʈakaan***a***. For each token of these vowels the F1 values have been plotted against the F2 values in the scatterplot in figure 9, with the filled circles corresponding to word-initial and the empty circles to word-final /a/ tokens.

Figure 9. Scatterplot of F1 against F2 for word-initial and word-final /a/ tokens



The separation between the two groups demonstrates a clear difference in acoustic quality, the word-final cases being more centralized. This is reflected by significant statistical results

for the subset: an analysis of variance taking F1 as the dependent variable shows speaker, position and the interaction between them to be highly significant ($p < .0005$), with an R^2 value of .637. Mann-Whitney tests showed the F2 values to be significantly affected by position ($p < .0005$), whilst the effect of position on duration falls a little below significance ($p < .067$).

Since the trends in formant frequency values seen in the data set as a whole are accentuated in the subset, the suspicion arises that it is a property of word-final /a/ vowels, rather than of /a/ vowels in non-initial syllables, that is responsible for the results. This was tested by excluding /a/ tokens from final syllables and performing a Mann-Whitney test on the rest of the data, grouped according to whether the syllable is initial or medial. Although the distinction was not significant for F2 ($p < .079$), both duration and F1 remained highly significant ($p < .0005$). The reverse pattern was seen when medial syllables were compared with final syllables: Mann-Whitney tests showed the distinction not to be significant for duration ($p < .297$) or F1 ($p < .674$), whereas F2 was significantly affected by position according to an analysis of variance ($p < .017$).

These results, indicating that the relevant distinction is between initial and non-initial position for duration and F1, but between final and non-final position for F2, were confirmed by a repeated measures analysis. This was conducted on the /a/ tokens from the words listed in the first six rows of table 1, taking environment and position as within-subjects factors. A reduced data set was used so that the phonetic environments were better balanced, and instances where more than one token were taken from the same sentence were excluded. For duration the three-way distinction in syllable position was insignificant ($p < .11$), and the same was true of F1 ($p < .057$). The results of Mauchly's test indicated that the assumption of sphericity did not hold for either F1 or F2, and so the Greenhouse-Geisser adjusted figures were used in these cases. A significant result was found for F2 ($p < .017$), suggesting that degree of fronting varies in a gradient fashion, decreasing over the course of the whole word. Note that this runs directly counter to Schiffman's observation (Schiffman 1999: 17) that non-initial /a/, and especially word-final /a/ in infinitives, is susceptible to fronting.

Overall, therefore, F1 values prove the most robust in differentiating initial from non-initial /a/ vowels, and there is supporting evidence from differences in duration. F2 also seems to play some role, although it varies in gradient fashion, rather than being straightforwardly correlated with the initial vs. non-initial distinction.

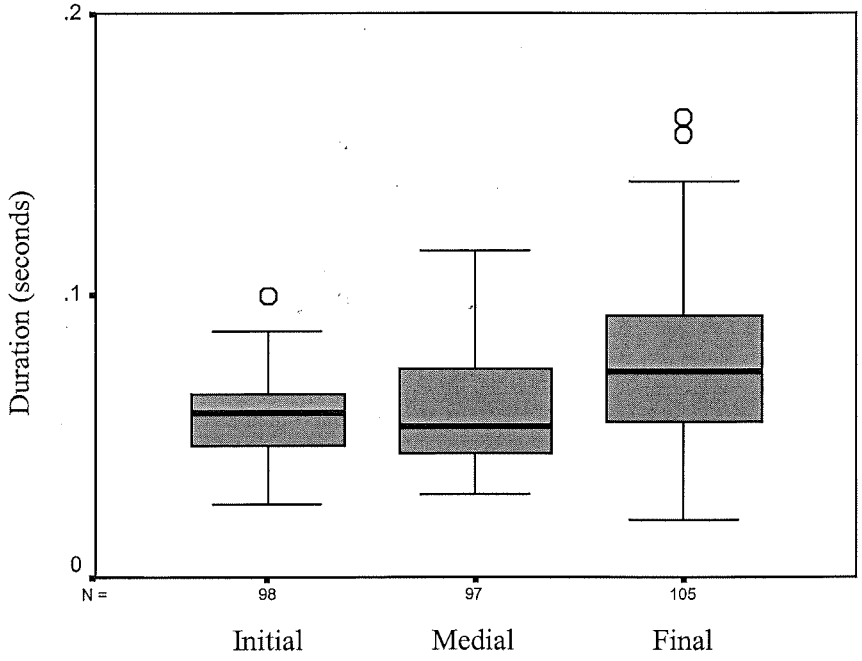
4.3.2.2. Analysis of /i/ tokens

The data for the /i/ vowels showed very little difference in formant structure between initial and non-initial syllables, although lowering of F2 and raising of F1 would be expected in reduced syllables. Mann-Whitney tests were performed on the formant frequency data, grouped according to whether or not the vowel token was in an initial syllable: in neither case did this prove to be significant ($p < .129$ for F1 and $p < .226$ for F2). Duration, however, was significantly affected by vowel position ($p < .001$), but not in the direction that would be expected for reduction. Figure 10 shows the duration values for initial, medial and final syllables in turn, and it is the final syllables that are markedly longer than the others.

There appears to be very little difference in duration between initial and medial syllables, and this was confirmed by a Mann-Whitney test in which grouping according to whether syllables were initial or medial proved insignificant ($p < .881$). The most plausible explanation of the results for the data set as a whole, therefore, is word-final lengthening of /i/ (in all but one of the words in which /i/ is in the final syllable it is also word-final). The lengthening of a word-final syllable at a syntactic boundary was documented by Cooper and Paccia-Cooper (1980: 46), who referred to it as cumulative phrase-final lengthening. Later studies have also associated lengthening effects with prosodic boundaries, such as the analysis by Byrd et al. (2000) of the interaction between duration and prosodic boundaries in Tamil. They found evidence of a lengthening effect in word-final nasals preceding a range of

prosodic boundary types, including the word. Such an effect may well account for the pattern observed here, although there is no obvious explanation for why it is not also seen in the /a/ and /u/ tokens.

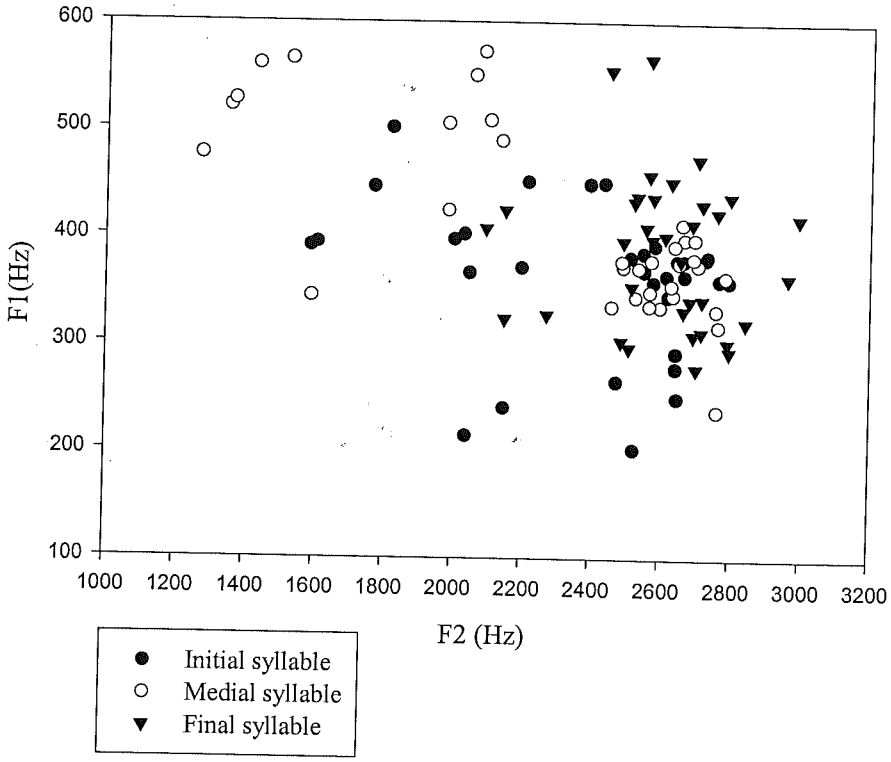
Figure 10. Boxplots of duration of /i/ against syllable position



Since the final syllable is exceptional in its duration, it is possible that the formant frequency values may also be affected. Accordingly, these cases were excluded from the data set and the initial and medial syllables compared. An analysis of variance of the F1 values, taking syllable position (i.e. initial or medial) as the only fixed factor, showed it to be significant ($p < .047$, and $R^2 = .02$). Closer inspection shows this to be largely due to one speaker (AL), who has higher F1 values overall in non-initial syllables. Her formant frequency values for the three different syllable positions are shown in the scatterplot in figure 11, with filled circles indicating initial, empty circles medial and triangles final syllables. There is no significant difference between the F2 values of initial and medial /i/ vowels. The repeated measures analysis conducted on the /i/ tokens from words in the first four rows of table 2 confirmed that the values of F1 and F2 are not significantly correlated with a three-way distinction in syllable position ($p < .971$ for F1 and $p < .248$ for F2). There is a significant result for duration, however: the Greenhouse-Geisser adjusted figure is $p < .023$, and this reflects the pattern already established.

The general tendency for F1 values to be raised in medial syllables, giving the /i/ vowel a more centralized quality, lends some support to the claims about reduction. However, this effect is only clearly manifested in one of the speakers, and seems to be somewhat masked by the tendency for final /i/ to be more closed. Clearer evidence for reduction will emerge in subsequent studies, notably comparisons between /i/ tokens in base words and echoes in section 4.6.2, and between the first and third syllables of expressives in section 4.6.4.

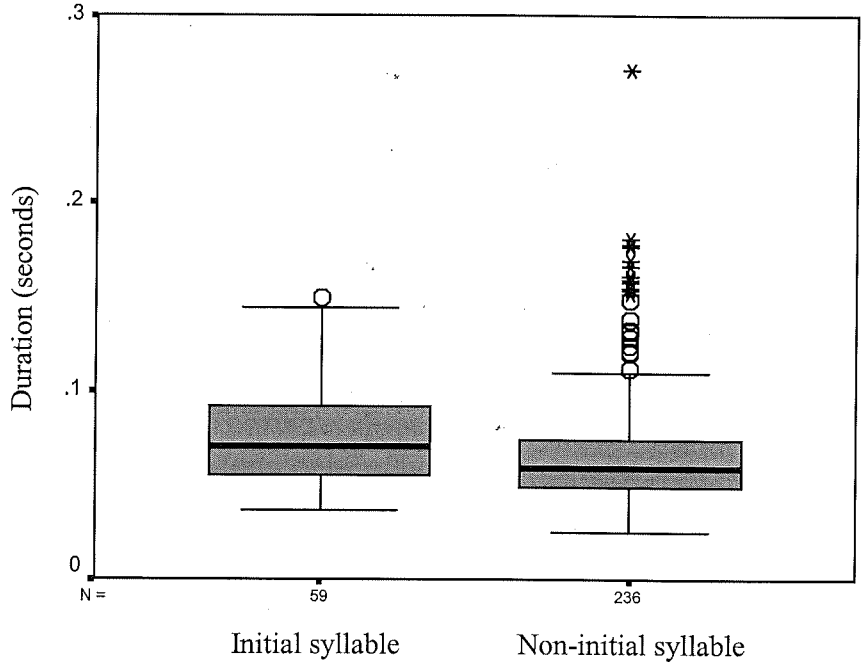
Figure 11. Scatterplot of F1 against F2 for /i/ for speaker AL



4.3.2.3. Analysis of /u/ tokens

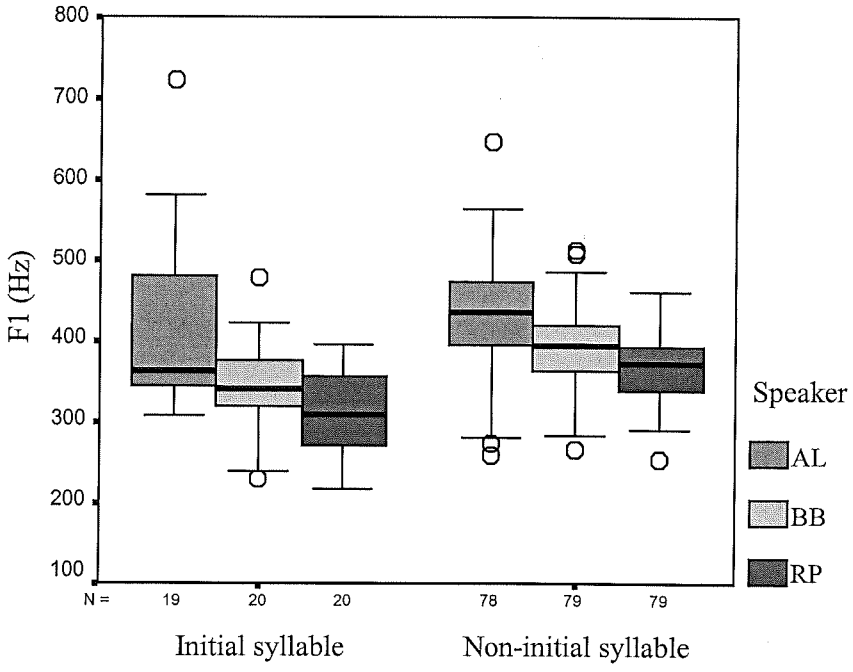
The data for the set of /u/ tokens provide the clearest evidence for the reduction of non-initial syllables. In fact, for all three of the parameters measured (i.e. duration, F1 and F2) the non-parametric Mann-Whitney test shows the position of the syllable to be highly significant ($p < .004$ for duration, $p < .0005$ for F1 and F2). Initial syllables appear to be longer overall than non-initial, although the degree of differentiation is not great. An analysis of variance taking duration as the dependent variable, and speaker and position of syllable (i.e. initial or non-initial) as fixed factors, showed neither factor to be significant ($p < .147$ for position and $p < .337$ for speaker). The low level of differentiation is illustrated by the boxplots in figure 12, which show values of duration grouped according to whether the syllable is initial or non-initial.

Figure 12. Boxplots of duration of /u/ against syllable position



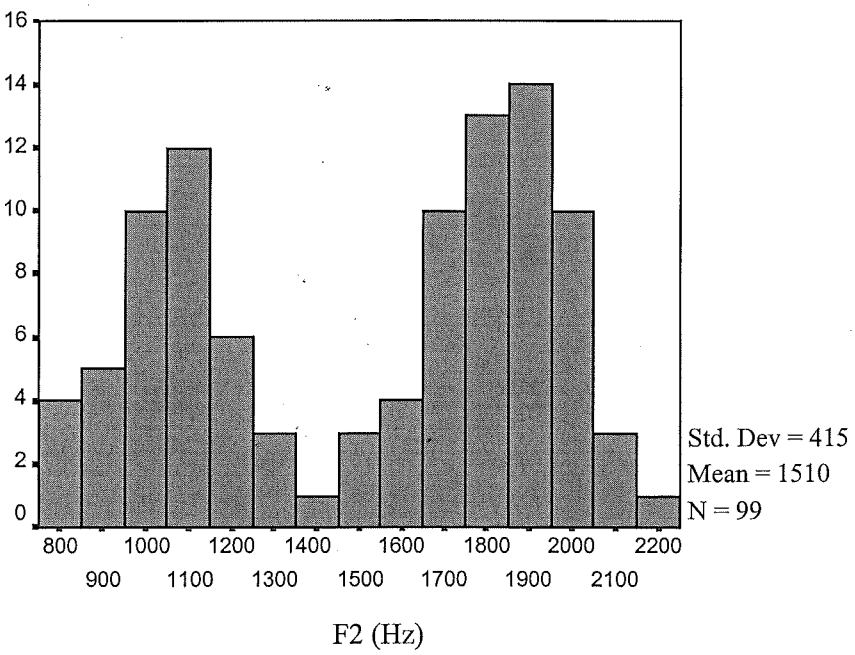
The data for F1 show a somewhat clearer distinction, with values being lower overall in the initial syllable for all speakers. However, there is still considerable overlap in the range of values covered, as the set of boxplots in figure 13 demonstrates. A clear pattern of interspeaker variation emerges, and the significant effect of the speaker's identity on F1 was confirmed by the non-parametric Kruskal-Wallis test ($p < .0005$).

Figure 13. Boxplots of F1 values of /u/ against syllable position, clustered by speaker



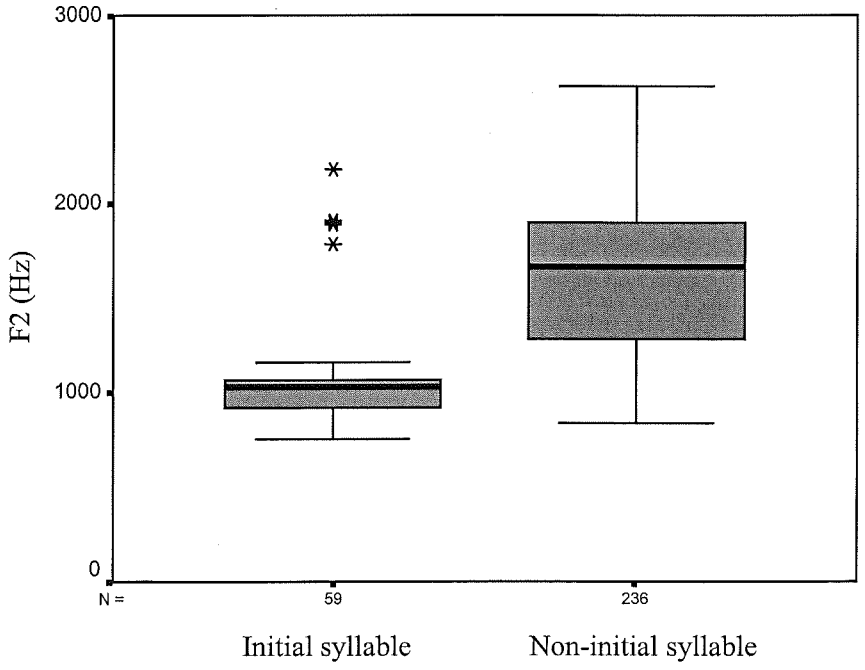
The distribution of values of the second formant frequency shows signs of bimodality: this is clearest in the case of speaker RP, as the histogram in figure 14 illustrates.

Figure 14. Histogram of F2 values of /u/ for speaker RP



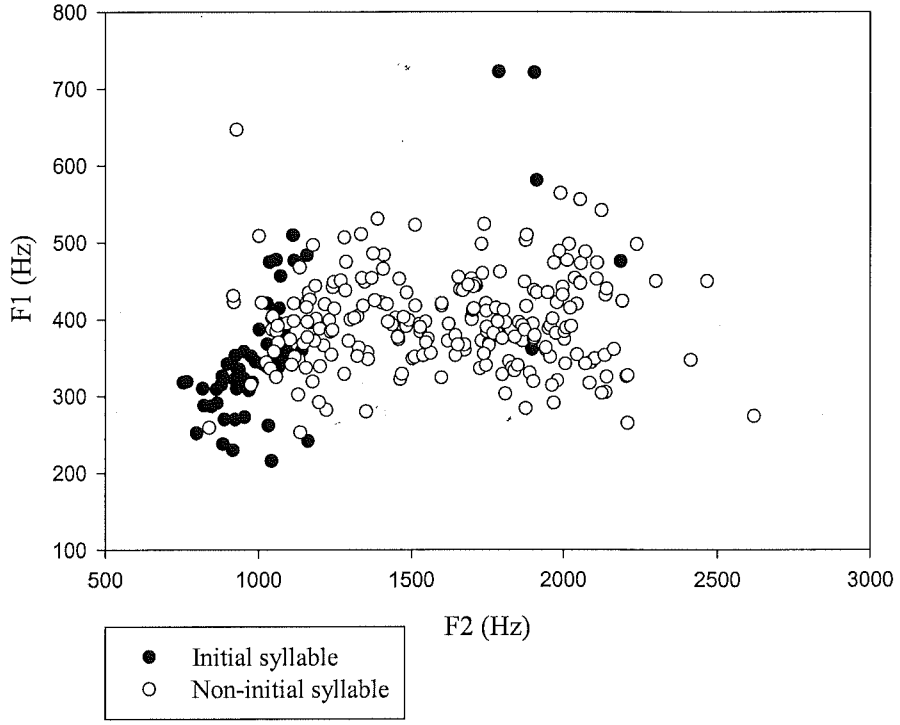
The clustering of the data around two modes correlates with the position of the syllable containing the vowel. The values of F2 for initial and non-initial syllables are quite clearly differentiated, as can be seen from the boxplots in figure 15. The interquartile ranges for initial and non-initial syllables are separate for the two male speakers, and there is minimal overlap for the third speaker.

Figure 15. Boxplots of F2 values of /u/ against syllable position for all speakers



The scatterplot in figure 16, which combines data points from all three speakers, shows a clear contrast between initial and non-initial /u/, the latter having markedly higher F2 values.

Figure 16. Scatterplot of F1 against F2 for /u/ in initial and non-initial syllables for all speakers



Higher F1 and F2 would be consistent with a more centralized articulation of the vowel, but lip-rounding may also play a part. Protrusion of the lips, which extends the vocal tract, is known to lower all formants, so the low formant frequency values of /u/ in initial syllables could reflect a correlation between stress and lip-rounding in Tamil.

These results provide strong evidence for an initial vs. non-initial distinction, and the results of repeated measures analysis confirm that there is no gradient effect applying over the word as a whole. This was conducted on /u/ tokens from the words in table 4, which represents a balanced subset of the data in table 3. None of the three parameters (duration, F1 or F2) were found to be significantly correlated with the three-way distinction in syllable position.

Table 4. Words containing /u/ used in the repeated measures analysis

Initial syllable	**Medial syllable**	**Final syllable**
kuɻutteen (63)	*iɻuveen* (50)	*piɻikkum* (67)
muɻiyum (71)	*kumaarukku* (23)	*enakku* (14)
kumaar (34)	*irukku* (57)	*eɻu* (40)
kukai (22)	*marattukku* (65)	*oru* (43)

4.3.3. Conclusions

The data provide some experimental confirmation of the vowel reduction that has been reported to affect non-initial syllables in Tamil, although the strength of the evidence varies between the three vowels. Duration turns out to be significantly affected by position for all three of the vowels studied. However, this result should be treated with some caution as duration is potentially affected by several different factors. The data have not been adjusted for overall speaking rate, for example, and, despite attempts to balance the data sets as far as possible, there are considerable differences in syllabic structure, which are also known to have an effect on vowel duration (Balasubramanian 1980). Moreover, it has already

been noted that the result for /i/ runs counter to the expected direction for reduction, and may be largely due to a lengthening effect in word-final position.

Factors affecting formant frequencies are more limited (although phonetic environment is obviously a major consideration), and so consistent differences firmly correlated with syllable position may be a more reliable indicator of reduction than duration. For /i/ it is only F1 that is significantly differentiated, and this rather subtle effect is found only between initial and medial syllables (/i/ in final syllables being characterized by greater length and a less open quality). /a/ and /u/, however, show significant differences for both first and second formants between initial and non-initial syllables.

The results for the section as a whole are summarized in table 5: each cell gives the expected relationship between vowels in initial and non-initial syllables if the latter are reduced in both quality and duration,[13] and the symbols (explained in the key underneath) indicate how far these expectations have been met in the data. Given that vowel reduction has been consistently associated with absence of stress cross-linguistically, these results constitute good evidence that Tamil stress is realized on an initial syllable.

Table 5: properties of vowels in initial syllables, compared with non-initial syllables

	/i/		/a/		/u/	
Duration	longer	××	longer	✓✓	longer	✓✓
F1	lower	✓	higher	✓✓	lower	✓✓
F2	higher	–	lower	✓✓	lower	✓✓

Key:
- ✓✓ highly significant result
- ✓ significant result
- – no significant result
- × significant result in the opposite direction
- ×× highly significant result in the opposite direction

4.4. Is Tamil stress quantity-sensitive?

4.4.1. Nature of the system

This section returns to the possibility that stress placement is not fixed on the initial syllable but is quantity-sensitive, considering whether this provides a more convincing interpretation of the data. The issue of prosodic weight, and specifically its relation to duration, is addressed by Balasubramanian (1980). On the basis of measurements of vowel durations in syllables of different structure, he proposes the four-tier system given in (1) (Balasubramanian 1980: 466).

(1)

{V}	{CV}	{VC, CVC, V:, CV:}	{V:C, CV:C}
lighter	light	heavy	heavier

The distinction between light and heavy syllables is uncontroversial, and the claim to a further contrast between heavy and super- or extra-heavy is not unparalleled.[14] The unusual

[13] The relationship between the formant frequencies of reduced and non-reduced variants of the peripheral /i/ and /u/ vowels can be derived from basic phonetic principles and hold good cross-linguistically. The situation with /a/, which already has a more central articulation, is less obvious. The prediction that F1 should be higher and F2 lower in the non-reduced variant is based on comparison between Balasubramanian's formant frequency figures for his /ɑ/ vowel, 747 and 1262 Hz respectively (Balasubramanian 1972: 89), and the 500 and 1500 Hz values of a canonical schwa (see, for example, Johnson 1997: 70).

[14] For instance, a three-way weight contrast is used in analyses of Hindi by Pandey (1989), Hayes (1995: 162–167) and Pierrehumbert and Nair (1996). Superheavy syllables also appear in accounts of Arabic stress (e.g. McCarthy 1985), whilst 'overlong' syllables are proposed for Estonian (e.g. Prince 1980).

feature of the hierarchy, one that Balasubramanian claims to have taken directly from the Sanskrit work *Ṛk Praatiśaakhya*, lies in the distinction between lighter and light. It is a central premise of syllable theory that prevocalic segments in the syllable are prosodically inert, and have no influence on prosodic weight.[15] In practice, however, the 'lighter' and 'heavier' syllable types occur so infrequently in this particular data set that they can be ignored for the purposes of this analysis. The key distinction here is that between light and heavy, with both (C)V: and (C)VC syllables counting as heavy.

As mentioned in section 4.2, there is no detailed description in the literature of how a quantity-sensitive system might work in Tamil. Some attention has, however, been paid to this issue for Malayalam, with discussions by Mohanan (1986: 111–115) and Hayes, based on Mohanan's data (Hayes 1995: 92–93). The two languages are closely related, and the phonetic correlates of stress claimed by Mohanan for Malayalam are very similar to the findings in section 4.3 for Tamil. He describes vowel reduction in unstressed syllables as the primary phonetic manifestation of stress, noting that it involves both centralization of the vowel and a reduction in duration (Mohanan 1986: 112). The placement of stress is restricted to the first two syllables of the word, with the primary stress falling on the first syllable unless it has a short vowel and is followed by a long vowel, in which case the stress falls on the second syllable. Schematically this can be represented as in (2), which shows the four logically possible combinations: asterisks mark the placement of stress, while ∪ indicates a light syllable and – a heavy syllable.

(2a) * (b) * (c) * (d) *
∪ ∪ ∪ – – ∪ – –

Hayes accounts for this by the construction of left-strong unbounded feet, with a prohibition against degenerate syllables bearing stress to account for the pattern in (b). This case, the only one where the stress falls on a non-initial syllable, crucially differentiates the system from one with fixed initial stress. Interestingly, the few examples that Andronov (1973: 114) gives to support his contention that Tamil stress is no longer fixed on the initial syllable all involve this pattern, i.e. a sequence of a light and then a heavy syllable. He argues that there are cases of widespread reduction and even loss of initial vowels over time, which must have been preceded by the stress moving away from the first syllable, giving the examples *iraṇṭu* vs. *reṇṭi* 'two', *iraa* vs. *raa* 'night' and *enakku* vs. *neekki* 'to me', the second alternant in this last case being a specifically Brahmin variant. In the analysis that follows, therefore, it is assumed that Tamil follows the same pattern as Malayalam, as a first attempt at investigating the possibility of quantity-sensitivity.

4.4.2. Data analysis

It is conceivable that the significant correlation found in section 4.3 between vowel reduction and non-initial position is a by-product of a more important relationship, between vowel quality and the placement of stress in a quantity-sensitive system. Since in many cases stress will be placed on the initial syllable in a quantity-sensitive system there will be a considerable degree of overlap, which could be responsible for the results of section 4.3. The aim of this analysis, therefore, is to test whether there is a closer correlation between vowel reduction and the absence of dynamic stress than between vowel reduction and non-initial syllable position.

The words containing vowels that were analyzed in section 4.3 (see tables 1, 2 and 3 of section 4.3.1) were all marked for stress according to the quantity-sensitive system outlined above. This involved making certain decisions about syllabification: intervocalic

[15] Convincing counterexamples to the observation that onset segments never contribute to weight have not been found, although Hayes (1995: 306) mentions Madimadi as a possible case, since certain syllable onsets in this language cause a preceding syllable to count as heavy.

nasal-stop clusters and geminates were assumed to straddle the syllable boundary, with the nasal and first half of the geminate forming the coda of the previous syllable, and the stop and second half of the geminate the onset of the next. This is in line with the practice of Balasubramanian in examples such as *kam.bi* 'wire' and *am.maa* 'mother' (1972: 497, 496). He also justifies an ambisyllabic representation for geminate stops on the basis of their duration (Balasubramanian 1980: 453),[16] motivating the position of the second syllable boundary in a word like *e.nak.ku* 'to me'. The words containing vowels that were measured in section 4.3 and are judged to be stressed on this system are listed in table 6 below.

Table 6. Words containing vowels that might be stressed in a quantity-sensitive system

Vowel	**Initial syllable**	**Second syllable**
/a/	*kattatu* (7)	*enakku* (4)
	nallatalla (10)	*avan* (15)
	vantaa (20)	*paṭattai* (40)
	tampi (32)	*citamparam* (41)
	laṇṭan (66)	*parantatu* (65)
	rattam (67)	
	anta (69)	
	anpu (75)	
/i/	*ippaṭi* (1)	*piṭikkaatu* (3)
	tinnaatee (6)	*kelippaay* (48)
	kili (10)	
	inta (22)	
	tiruṭan (52)	
/u/	*muṭiyum* (71)	*irukku* (57)

The proportion of 'stressed' to 'unstressed' vowels is similar under the two systems: of the 79 vowels measured, 23 were in initial syllables, and so assumed to be stressed under the fixed stress system, as compared to 22 marked as stressed under the quantity-sensitive system. These are divided between 14 in initial syllables and 8 in second syllables, so there is, as expected, a sizeable overlap in the predictions of the two systems.

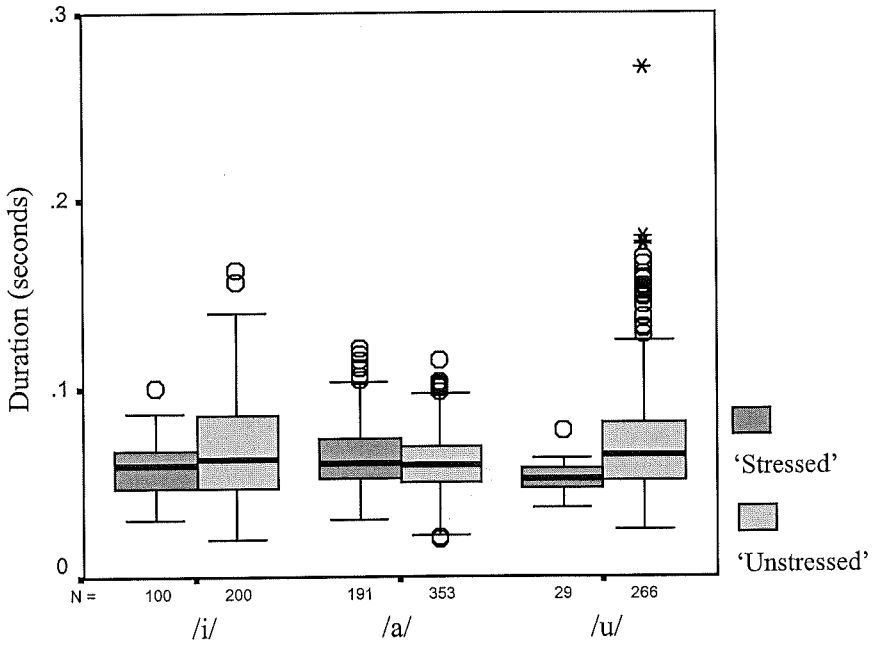
Statistical tests were performed on the same three parameters measured in section 4.3, i.e. the duration of the vowel and the values of the first and second formant frequencies at its mid-point. First the whole data set of 1139 tokens was analyzed, to investigate whether the division between quantity-sensitive 'stressed' and 'unstressed' was a significant factor, and how it compared with the division between initial and non-initial syllables. Each vowel was taken in turn, with the numbers of 'stressed' vs. 'unstressed' tokens being 100 vs. 200 for /i/, 191 vs. 353 for /a/ and 29 vs. 266 for /u/. The difference here is expected to be one of degree, given the overlap in predictions. Just as the results in section 4.3 could be a by-product of a stronger correlation with quantity-sensitive stress, so the converse is also true, i.e. if stress is fixed on the initial syllable, some degree of correlation will probably emerge for the quantity-sensitive factor. Attention was therefore confined next to vowels in initial syllables (334 tokens), comparing those that would be stressed on a quantity-sensitive system with those that would not. Again the three vowels were considered in turn, with 70 'stressed', as opposed to 28 'unstressed' for /i/, 117 'stressed', as opposed to 60 'unstressed' for /a/, and 14 'stressed', as opposed to 45 'unstressed' for /u/. If stress falls uniformly on the initial syllable, there should be no significant differences between them: if it is quantity-sensitive, however, differences between the two categories are expected.

[16] See sections 5.3 and 5.4 for experimental data on the duration of geminate obstruents in Tamil.

4.4.3. Results

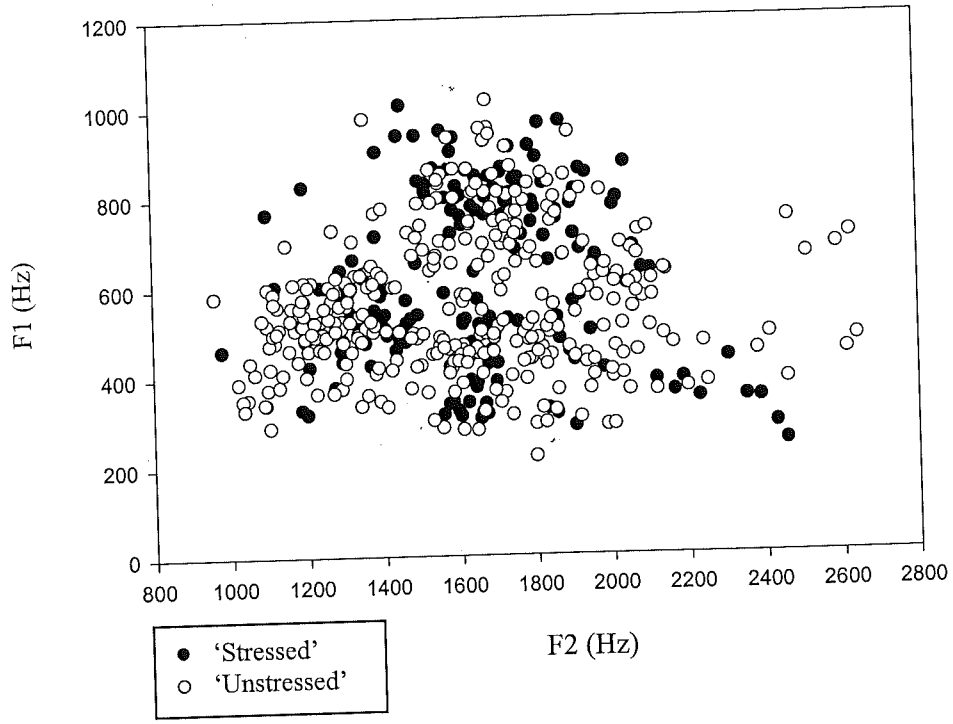
Analysis of the duration data for each of the three vowels revealed no consistent pattern in the differences between those cases marked as 'stressed' and those assumed to be 'unstressed'. The /a/ tokens proved to be slightly longer in 'stressed' syllables ($p < .017$ on an analysis of variance taking the division between 'stressed' and 'unstressed' as the fixed factor), as compared with the highly significant result ($p < .0005$) for the difference between initial and non-initial syllables. Both /i/ and /u/ vowels proved to be shorter on average in the 'stressed' syllables than the 'unstressed' ones, which is precisely the opposite pattern to that expected. The non-parametric Mann-Whitney test was used, and showed the distinction to be significant for /i/ ($p < .035$) and highly significant for /u/ ($p < .0005$). The absence of any clear pattern overall (illustrated in the boxplots of figure 17) contrasts with the findings of section 4.3, where /a/ and /u/ were clearly longer in initial syllables than non-initial, and the anomalous result for /i/ was accounted for by word-final lengthening.

Figure 17. Boxplots of duration of /a/, /i/ and /u/; clustered according to quantity-sensitive stress



The difference in the first formant frequencies of the vowel tokens were all consistent with the 'unstressed' vowels being centralized, which is also what was found for the non-initial syllables in section 4.3. For /a/ and /u/, which had F1 values that were respectively higher and lower in the 'stressed' syllables, the difference was highly significant ($p < .007$ and $p < .008$ on the Mann-Whitney test). This compares with $p < .0005$ for both vowels in the corresponding results for the initial vs. non-initial distinction. In the case of /i/, which had lower F1 values in the 'stressed' syllables, the result was highly significant ($p < .0005$ on an analysis of variance, as compared to $p < .047$ for the initial vs. non-initial difference). The F2 values showed little correlation with the division between 'stressed' and 'unstressed' syllables: there was no significant relationship between F2 and the quantity-sensitive stress factor for either /a/ or /u/. The absence of any clear distinction in formant structure between 'stressed' and 'unstressed' vowels is illustrated by the scatterplot in figure 18, showing the F1 and F2 values for the /a/ tokens, with filled circles for the 'stressed' tokens and empty circles for 'unstressed'. This represents a considerable proportion of the data (544 of the 1139 tokens), and also has the most balanced ratio of 'stressed' to 'unstressed' vowels (191: 353).

Figure 18. Scatterplot of F1 against F2 for /a/ tokens



/i/, however, is again exceptional: the 'stressed' vowels proved to have significantly higher F2 values ($p < .012$ on a Mann-Whitney test), which is consistent with the 'unstressed' vowels being centralized, and compares with no significant result at all for the initial vs. non-initial distinction.

Table 7 summarizes the results in the same format as table 5 in section 4.3.3, to facilitate comparison between the two proposed systems of stress placement.

Table 7. Properties of 'stressed' vowels, compared with 'unstressed' vowels

	/i/		/a/		/u/	
Duration	longer	×	longer	✓	longer	××
F1	lower	✓✓	higher	✓✓	lower	✓✓
F2	higher	✓	lower	–	lower	–

Key:
- ✓✓ highly significant result
- ✓ significant result
- – no significant result
- × significant result in the opposite direction
- ×× highly significant result in the opposite direction

It is clear from the higher number of ticks in table 5 (13, as opposed to 8) and lower number of crosses (2, as opposed to 3), that the division between initial and non-initial syllables correlates much more closely overall with the difference between reduced and unreduced vowels than the 'stressed' vs. 'unstressed' distinction of table 7. This is particularly clear for the /a/ and /u/ vowels, which show highly significant results for all three parameters in table 5. These are only replicated for the F1 values in table 7, and even then the levels of probability are lower than for table 5. The /i/ results run counter to the general trend, correlating more closely with reduction in table 7 than table 5, but this may again be at least partially attributable to the confounding effect of word-final lengthening discussed in section 4.3.2.2. If vowel reduction can reliably be used as a diagnostic of the absence of stress in

Tamil, there is clearly better evidence for fixed initial stress than the quantity-sensitive system assumed here.

This conclusion is supported by the comparison between 'stressed' and 'unstressed' initial syllables. Testing each vowel for each of the three parameters produced significant results for the /i/ tokens, which were significantly longer in 'stressed' syllables ($p < .026$ on an analysis of variance taking quantity-sensitive stress as the fixed factor), and also had lower F1 values ($p < .0005$ on the Mann-Whitney test) and higher F2 values ($p < .004$ on an analysis of variance) than in 'unstressed' syllables. All these differences are precisely what would be expected if the 'unstressed' vowels are reduced. As before, however, the results for /i/ prove to be exceptional: there are no significant results at all for /a/, and /u/ has only one significant result, in the opposite direction: the 'stressed' vowels are significantly shorter than those marked as 'unstressed' ($p < .0005$ on a Mann-Whitney test). These two vowels between them account for more than two-thirds of the data (236 out of 334 tokens in total). Since they provide no evidence at all for differentiation according to the quantity-sensitive system of stress that has been assumed, this supports a fixed system in which stress falls uniformly on the initial syllable.

4.4.4. Stress-timing vs. syllable-timing

A related issue, and one that has received some attention in the literature, is where Tamil fits into the traditional dichotomy between stress- and syllable-timed languages. This is particularly associated with Abercrombie, who proposed that it was a categorical distinction based on physiological differences (Abercrombie 1965: 17). The underlying premise was that languages contain isochronous units, either at the level of the foot, in which case they are classified as stress-timed, or at the level of the syllable, in which case they are described as syllable-timed. A classic example of the contrast, which features in Abercrombie's study, is the difference between English, which is taken to be a stress-timed language, and French, which supposedly exhibits syllable-timing.

The assumption of isochrony in either stress- or syllable-timed languages has since been discredited by numerous empirical studies (e.g. Roach 1982, Wenk & Wioland 1982 on French). Evidence against isochrony in Tamil is to be found in Balasubramanian (1980). Assuming for the purposes of the experiment that stress is fixed on the initial syllable of polysyllabic words, Balasubramanian measured interstress intervals, concluding that the time taken to utter a foot was variable but roughly proportionate to the number of syllables it contained (Balasubramanian 1980: 456). An extensive study of syllable duration, involving measurements from about 3000 Tamil words, also produced no evidence for isochrony, but rather motivated the construction of the hierarchy given in (1) above. On the basis of duration data, therefore, Balasubramanian concluded that Tamil could properly be classified as neither stress-timed nor syllable-timed.

Despite the durational basis of the classification being disproved, it has continued to attract interest as a way of describing cross-linguistic rhythmic differences. The strictly categorical nature of the distinction has largely been abandoned: Roach (1982), for example, argued that no language is wholly stress-timed or syllable-timed, but that languages differ in which type of rhythm predominates. Roach also suggested that various prosodic properties are correlated with the distinction, and similar proposals appeared in Dauer (1983). Syllable structure is seen as a key factor influencing rhythm, with languages that have been classed as syllable-timed tending to have simpler syllable structures. Dauer also implicated quantity-sensitivity, commenting that stress in stress-timed languages generally falls on prosodically heavy syllables, whilst light syllables are unstressed, and that syllable structure and stress therefore *'reinforce each other'* in stress-timed languages (Dauer 1983: 56). The results of the previous section suggest that this is not the case in Tamil, which also has fairly simple syllable structure. As noted above, codas are largely confined to the first half of a geminate or a homorganic nasal-stop cluster, and the composition of complex onsets is also heavily

restricted.[17] If Dauer's generalization is correct, it may be the simple structure of Tamil syllables that has caused the language to be generally classed as syllable-timed in the literature. Corder (1973: 253) makes this claim for Tamil, as does Ravisankar (1994: 337), although they give no grounds for their classification. Asher also comments that Tamil is *'certainly more akin'* to a syllable-timed than a stress-timed language (Asher 1985: 230). Marthandan (1983: 308) is the one exception, claiming that Tamil has a stress-timed rhythm, but again offers no experimental evidence to support this.

Roach and Dauer both propose that vowel reduction is characteristically associated with stress-timed languages. Dauer, for instance, states that:

> *'syllable-timed languages do not regularly have reduced variants of vowels in unstressed position'* (Dauer 1983: 57).

Given the empirical evidence for vowel reduction in Tamil presented in section 4.3, this is clearly problematic for the classification of Tamil as a syllable-timed language. Once again, therefore, Tamil seems to evade any straightforward categorization.

Two recent papers (Ramus, Nespor & Mehler 1999 and Grabe & Low (to appear)) have re-examined possible durational correlates of the distinction between stress-timing and syllable-timing. In each case, passages of speech from a number of languages were segmented into acoustically defined vocalic and intervocalic portions, thereby avoiding the need to make any judgement about stress placement. These data were then used to calculate a measure of rhythmic diversity. Grabe and Low investigated 18 languages, including Tamil, making recordings of single speakers reading the 'North Wind and the Sun'. They applied a formula to both vocalic and intervocalic data known as the pairwise variability index (PVI), which involves computing the difference between successive intervals.[18] Since data came from a single speaker only, the consistency of the PVI was checked by dividing the values into three equal subsections. The degree of correlation between sections was found to be highly significant in each case, confirming the stability of the PVI for individual speakers.

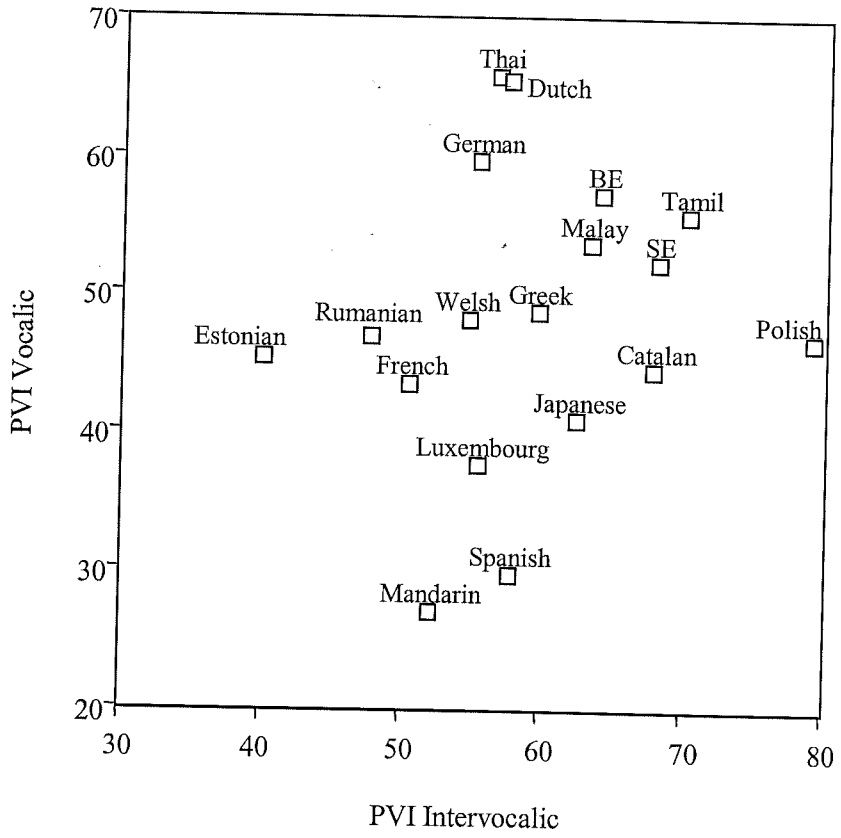
Those languages within the sample traditionally classified as stress-timed, such as English, Dutch and German, were found to have high PVI values for vocalic and intervocalic intervals, whereas those described as syllable-timed, such as French and Spanish, had lower values for both intervals. This is seen in figure 19, which plots vocalic variability on the vertical axis against intervocalic variability on the horizontal axis, with the stress-timed languages clustered towards the top right of the graph. Interestingly, Tamil is among them, with high values for both vocalic and intervocalic PVI (55.8 and 70.2 respectively). The high variability in Tamil vowel durations reflects the existence of both vowel reduction in the language and a distinction between long and short vowels. In Tamil phonological length seems to be realized by significant differences in acoustic duration, and not just quality distinctions. Evidence for this is found in Balasubramanian (1972: 179), which presents durational data for seven minimal pairs differentiated only by vowel length. Ratios of short to long range from 1:1.6 for [ɑ]:[ɑː] to 1:2.4 for [ʊ]:[uː]. Mean duration values from the data in section 4.3 for non-initial (i.e. reduced) vs. initial (i.e. unreduced) /a/ and /u/ show ratios of 1:1.3 and 1:1.1 respectively, so if these differences are combined a long /a/ vowel might be twice as long and a long /u/ vowel 2.6 times as long as the reduced variant of the equivalent short vowel. The high intervocalic variability is rather more puzzling, given Tamil's

[17] For comprehensive descriptions of Tamil syllable structure and phonotactics see Balasubramanian (1972: 491–502), Marthandan (1983: 273–299) and Asher (1985: 224–228).

[18] The PVI was normalized for speaking rate for the vocalic data, since a significant correlation between interval duration and speaking rate (defined as the average interval duration produced by a speaker) was found. No normalization was applied to the intervocalic data, however, on the grounds that intervocalic intervals show considerable cross-linguistic differences in their segmental composition, whereas vocalic intervals mostly consist of single segments.

relatively simple syllable structure. However, in colloquial Tamil,[19] vowels that would simply be reduced in the formal language may be omitted altogether, thereby producing more complex consonant clusters (Asher 1985: 226). This would extend the range of intervocalic intervals, and so increase variability.

Figure 19. Rhythmic variability profiles for data from 18 languages. Reproduced, with permission, from Grabe and Low (to appear).



Grabe and Low also applied the variability measures suggested by Ramus et al. to their data as a comparison, namely %V, the proportion of time devoted to vowels, and ΔC, the standard deviation of the consonantal (i.e. intervocalic) intervals. Tamil had a very high ΔC score, being surpassed only by Polish, which is in line with its high intervocalic PVI and thus aligns it with the traditional stress-timed languages. Unlike these languages, however, Tamil also had the second highest %V score, a characteristic shared with the syllable-timed languages. Taking both measures together, therefore, as in figure 20, Tamil ends up in a rather isolated position.

Tamil's peculiarity lies in the fact that, unlike the other languages, it does not exhibit a complementary relationship between %V and vocalic variability (those languages classified as stress-timed have a high vocalic PVI but low %V, whereas syllable-timed languages have a low vocalic PVI but high %V). As Grabe and Low comment:

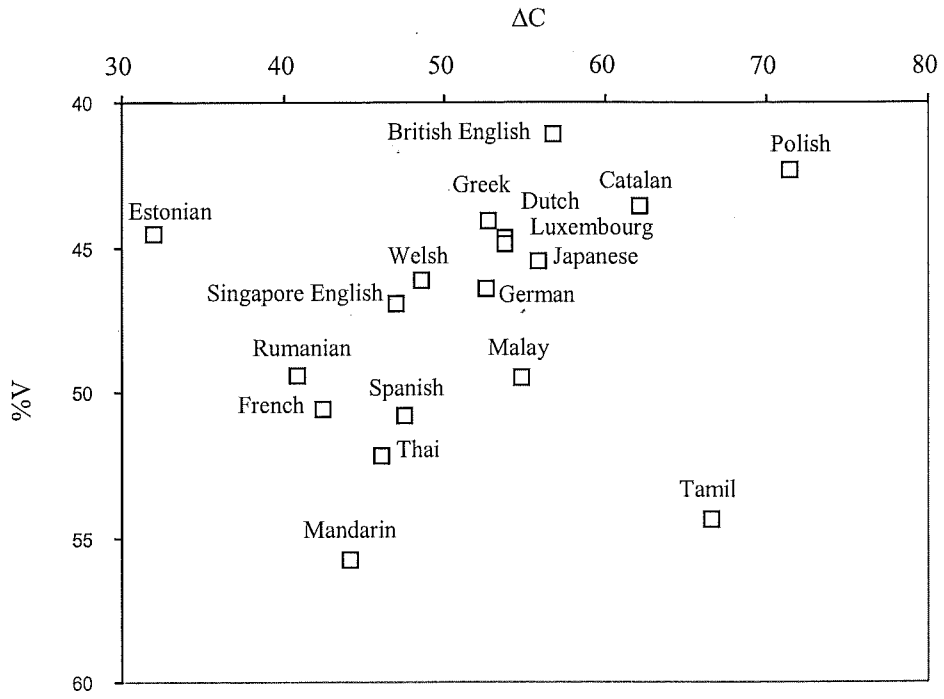
> *'If the relationship between the two measures provides the acoustic basis for an impression of stress- or syllable-timing, . . . Tamil would not be classifiable.'* .

Since this study involved only a single speaker, its results should be treated with some caution. Nevertheless, if Tamil lacks a relationship that contributes to the impression of

[19] The recording was made in Singapore by a native Tamil speaker. Schiffman (1998: 364) notes that SST (standard spoken Tamil) is widely spoken in Singapore, and there are no grounds to believe that this dialect differs in any crucial respect from the data analyzed in this thesis.

rhythmicity in other languages, it may provide some explanation for why native speakers, even those with phonetic training, lack intuitions about the placement of stress.

Figure 20. Rhythmic variability profiles, with %V plotted on the vertical axis in reverse order and ΔC on the horizontal axis. Reproduced, with permission, from Grabe & Low (to appear).



4.5. Monophthongization of /ai/

As noted in section 4.2, another property said to distinguish initial from non-initial syllables is the realization of the diphthong /ai/, which is apparently monophthongized when non-initial (see, for example, Asher 1985: 219, Schiffman 1999: 23). In order to investigate how far this phenomenon could be detected in the speech of the informants the following set of words, all to be found in sentences from appendix F, were analyzed.

Table 8. Words containing /ai/

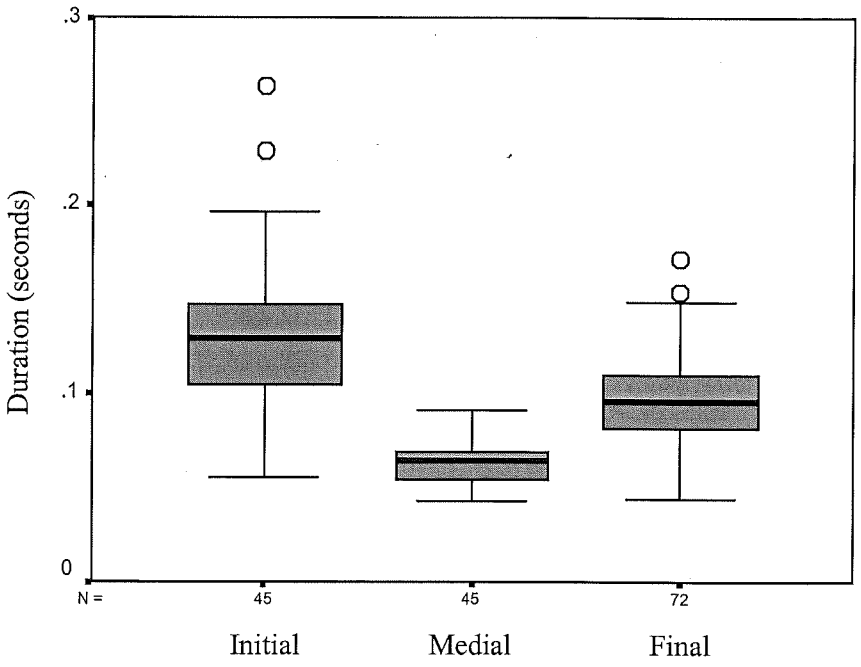
Initial	**Medial**	**Final**
pai (4)	*uʈaikka* (19)	*kuk**ai*** (22)
kai (55)	*cenn**aikku*** (34)	*ullaŋk**ai*** (33)
taikkiraa (72)	*nin**aikkaatee*** (15)	*marat**tai*** (31)
		*paattirat**tai*** (19)
		aaʈai (72)

As the arrangement in columns demonstrates, the data set contains instances of /ai/ in initial, medial and final syllables. As a result of the phonotactic constraints holding in Tamil all the cases of /ai/ in final syllables are also word-final. The total duration of the diphthong was measured in each case, and formant frequencies were measured at points a quarter, half and three quarters of the way through the segment. The difference between the quarter and three quarter points was then calculated to give some measure of the change in vowel quality. In a fully diphthongal realization the first formant frequency should fall markedly and the second rise, so the three quarters value was subtracted from the quarter value for F1 (ΔF1), whilst the quarter value was subtracted from the three quarters value for F2 (ΔF2), so that the signs of

the difference should both be positive. If /ai/ is monophthongized, however, the formants should be in a relatively steady state, and so the values of the differences should be low.

The overall duration of the segments proved to be notably longer in the initial syllables, as the boxplots in figure 21 illustrate. This was confirmed statistically by an analysis of variance of the duration values in which the syllable position (initial vs. non-initial) was a highly significant factor ($p < .0005$, $R^2 = .311$). The identity of the speaker, however, was not significant ($p < .909$ on an analysis of variance).

Figure 21. Boxplots of duration of /ai/ against syllable position



It is also noticeable from figure 21 that final /ai/ is longer overall than /ai/ in a medial syllable, a pattern already seen in the /i/ vowels (see figure 10 of section 4.3.2.2). This suggests that the word-final lengthening hypothesized for /i/ above may be a general tendency affecting word-final vowels articulated in the same general region of the vocal tract. Unlike the /i/ vowels, however, duration is not significantly correlated with a three-way distinction in syllable position, as shown by a repeated measures analysis performed on the /ai/ tokens from the words in the first three rows of table 8 ($p < .077$).

The measures of formant frequency change for initial vs. non-initial syllables produced significant results for both F1 and F2. The Mann-Whitney test was performed on the values of the difference between the quarter and three quarters points (ΔF1 and ΔF2), grouped according to syllable position, and this proved highly significant in each case ($p < .0005$). Speaker was also a highly significant factor ($p < .004$ for ΔF1 and $p < .0005$ for ΔF2) according to the Kruskal-Wallis test. For each speaker the degree of formant change was markedly higher in the initial than the final syllable, although the relation between the values for medial and final syllables showed some variation. The boxplots in figures 22 and 23 show the overall distribution of values for initial, medial and final syllables, taking all speakers together. The differences between initial and medial syllables only, and initial and final syllables only were also found to be significant ($p < .001$ and $p < .0005$ respectively for F1 and F2). The differentiation between medial and final only was less strong (proving insignificant for ΔF1 and with a significance level of $p < .009$ for ΔF2), suggesting that the primary distinction is indeed between initial and non-initial. This was confirmed by repeated measures analyses, which gave no significant correlation between either ΔF1 or ΔF2 ($p < .677$ and $p < .297$ respectively) and a three-way distinction in syllable position.

Figure 22. Boxplots of ΔF1 of /ai/ against syllable position

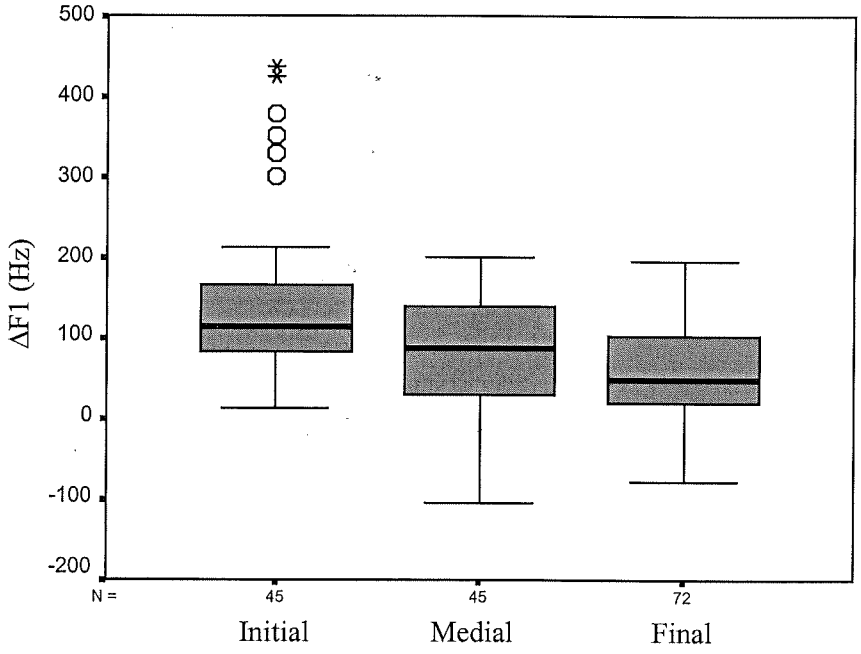
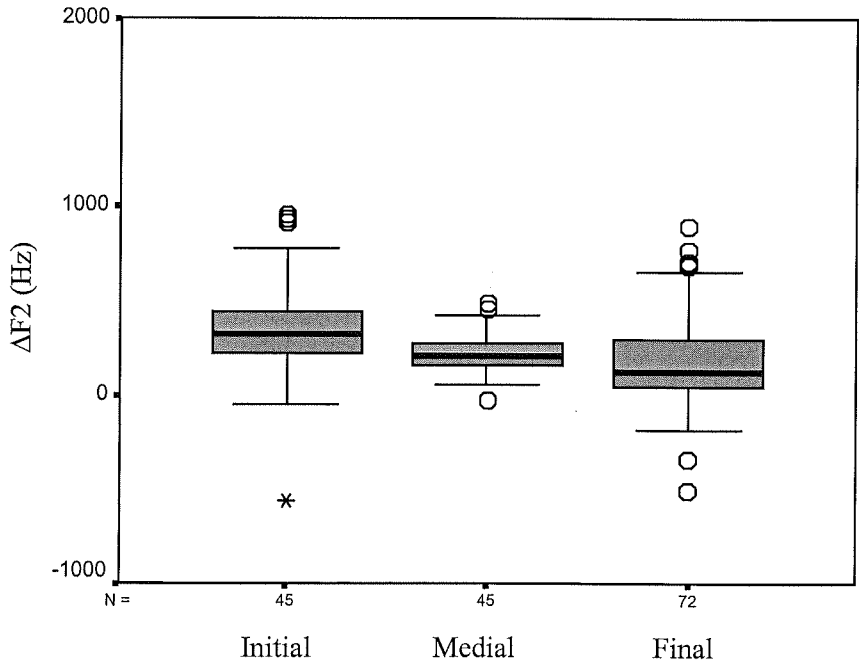


Figure 23. Boxplots of ΔF2 of /ai/ against syllable position



The line charts in figures (24) and (25) show the median values at the quarter, mid and three quarters points for /ai/ in initial, medial and final syllables to give some indication of the degree of formant frequency change in each.

Figure 24. Line chart showing change in F1 in initial, medial and final syllables

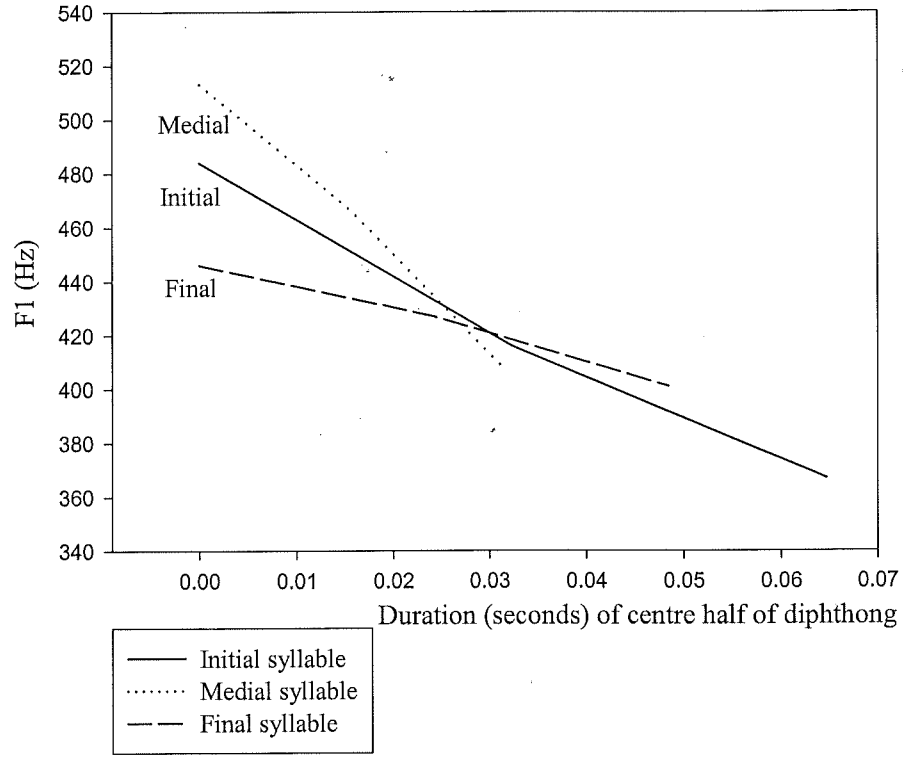
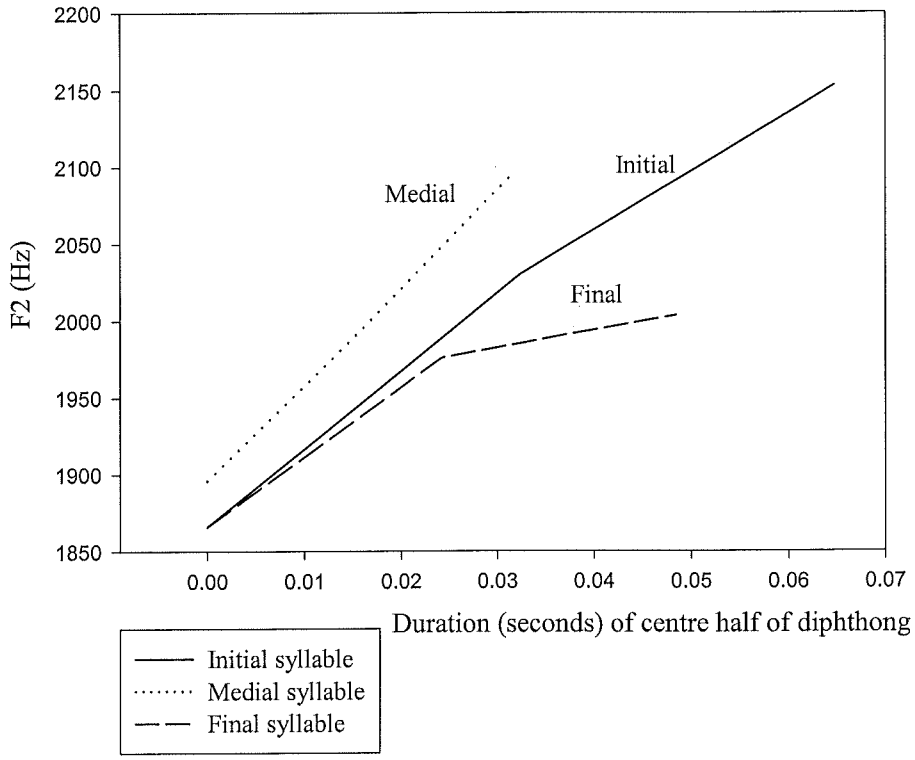


Figure 25. Line chart showing change in F2 in initial, medial and final syllables



Some similarities emerge: in each case the steepest gradient and thus the fastest rate of change is seen in the medial syllable, followed by the initial syllable, with the formant frequencies of word-final /ai/ showing the slowest rate of change. Moreover, for both F1 and F2 the rate of change is slower in the third quarter of the diphthong than the second in initial

syllables (the rate of change in medial syllables is fairly constant). In general, however, the rates of change in frequency reflected by the gradients are not strongly differentiated. This suggests that the difference between initial and non-initial /ai/ is mainly a matter of duration, rather than a contrast between a dynamic and a steady-state segment. There is no clear consensus in the phonetics literature on the relative salience in the perception of diphthongs of dynamic information, such as the direction and slope of formant change, and the values of the formant frequencies at the onset and offset. Gottfried, Miller and Meyer (1993), for instance, test three different hypotheses, which they term 'onset and offset', 'onset and slope' and 'target plus direction', concluding that all three distinguish effectively between different American English diphthongs. The structure of diphthongs does, however, show cross-linguistic variation (see, for example, Lindau, Norlin & Svantesson 1985), and further research is needed to establish the crucial parameters for Tamil.

These results show that the realization of the /ai/ diphthong is clearly influenced by syllable position, with duration being affected most strongly. These experimental findings thus provide further evidence that initial and non-initial syllables are differentiated phonetically in Tamil.

4.6. Analysis of echo words

If, as the results of the preceding sections seem to indicate, stress in Tamil is uniformly realized on the initial syllable of a word, it should act as a diagnostic for the internal phonological structure of the Tamil echo words. If they form a single word phonologically, the echo form as a whole should receive only one main stress: if they act as two words, the echoed portion would also be expected to receive stress. With this issue in mind, the phonetic properties of the echo /i/ vowel were investigated. First, however, attention was paid to a lowering process affecting /i/ vowels in initial syllables that could have potentially confounding effects.

4.6.1. Lowering

4.6.1.1. Background

The nature of /i/ vowels in different syllables is potentially affected not only by reduction but also by lowering in certain environments. This is mentioned repeatedly in the literature[20] as affecting the high vowels /i/ and /u/ in an initial syllable before the low back vowel /a/ and diphthong /ai/, and is described by Asher (1985: 229) in terms of vowel harmony. He comments that the tendency for the vowel to be lowered is less strong if a long consonant intervenes, as in *illai* 'no, not'. Schiffman (1999: 19) also states that the lowering is restricted to cases where the high vowel is followed by a single consonant, i.e. occurs in an open syllable. The process is said to be entirely regular, even affecting some borrowed words, although the proximate demonstrative *ivan* 'this man', where the initial /i/ is arguably a morpheme, is noted by both Asher and Britto as an exception. In this instance, the existence of the corresponding interrogative *evan* 'who' means that the /i/~/e/ distinction is lexically contrastive, and this may well be responsible for the /i/ not being lowered.

Since the lowering process seems to be restricted to the initial syllable of a word, the key question that arises in connection with the echo words is whether or not the /i/ of the echo sequence is lowered in the relevant environment. If it is, this would suggest that the echoed portion is being treated as a word in its own right: if not, it is presumably phonologically subordinate to the base. The only specific reference to this issue found in the literature is in connection with echo words where the base begins with /ki-/ or /kii-/.[21] Steever

[20] See, for example, Trubetzkoy (1969: 285), Asher (1985: 229), Britto (1986: 198), Steever (1987: 744) and Schiffman (1999: 19, 22).

[21] See section 1.5.2.5 for discussion of this constraint, which is widely reported for echoing in Indian languages, and section 3.3.2.2 for evidence collected from Tamil informants on this point.

(1987: 744) comments that such words cannot be echoed, even though in some instances vowel lowering would convert the base, but not echoed part, into /ke/, as in the hypothetical example *keŋaru kiŋaru* 'well'. However, words with initial syllables /ke-/ or /kee-/ can be echoed. Steever's assumption therefore seems to be that the first syllable of the echoed part does not count as word-initial, and so is not subject to lowering.

The first step in investigating this was to establish the acoustic differences between /i/ and /e/ vowels in environments where lowering is not an issue. Accordingly, the vowels in the initial syllable of the following pairs of words were measured for duration and the values of the first and second formant frequencies recorded at the mid-point of the vowel. Three cases were discarded because of background noise when the data were checked for consistency, leaving 147 tokens to be analyzed.

Table 9. Words containing /i/ and /e/

Context	**/i/**	**/e/**
_l	*illai* (39)	*eli* (43)
k_li	*kili* (10)	*kelippaay* (48)
_n	*inta* (22)	*enakku* (14)
_ŋk	*iŋkee* (16)	*eŋkiʈʈa* (17)
_ʈu	*iʈuveen* (50)	*eʈu* (40)

4.6.1.2. Results

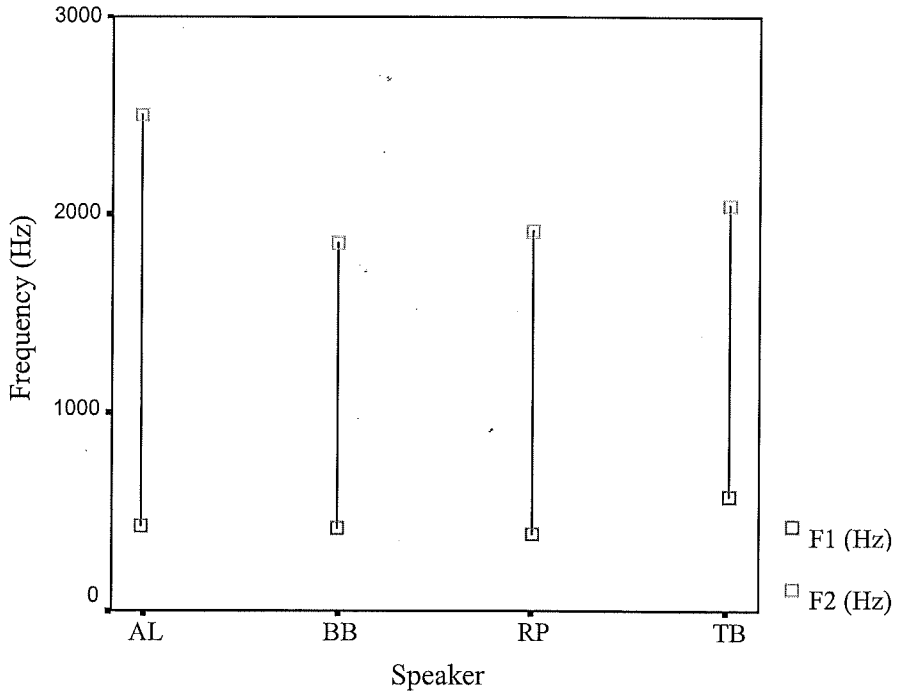
Histograms of the values for the three parameters measured revealed some skewing, notably for F2, and so median, rather than mean values, were used as an average. The values for the /i/ vowel tallied reasonably well with the cases of /i/ in initial syllables used in the reduction study, as the following table demonstrates. The values for the /i/ tokens in non-initial syllables are shown in brackets.

Table 10. Median formant frequencies for /i/

Speaker	**AL**	**BB**	**RP**
F1	351 (375)	333 (346)	316 (329)
F2	2676 (2537)	1874 (1833)	2097 (2051)

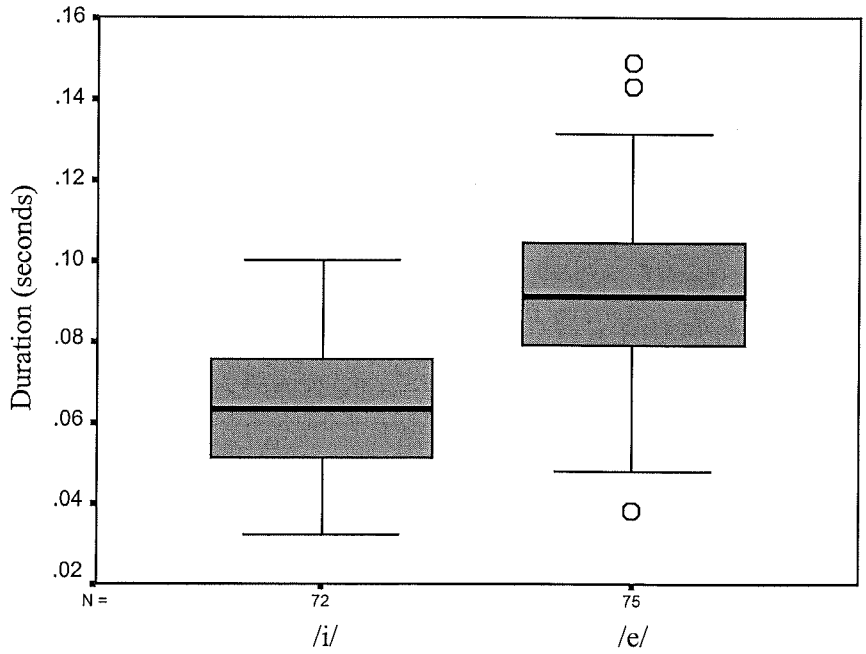
The median values of both formants for /e/ are displayed in the drop-line chart for each speaker, together with the values that Balasubramanian gives for his vowel [ɛ]. Although Balasubramanian's F2 fits in well with the medians of the three subjects, his F1 is markedly higher than the averages recorded here. The interspeaker variation follows the same trends noted for the /a/, /i/ and /u/ vowels, with speaker AL having the highest values for both formants, and BB having higher F1 but lower F2 values than RP.

Figure 26. Drop-line chart of median F1 and F2 values of /e/ for four speakers



The boxplots in figure 27 show that the interquartile ranges for the duration of /i/ and /e/ are quite clearly separated, with the /e/ vowels being consistently longer.

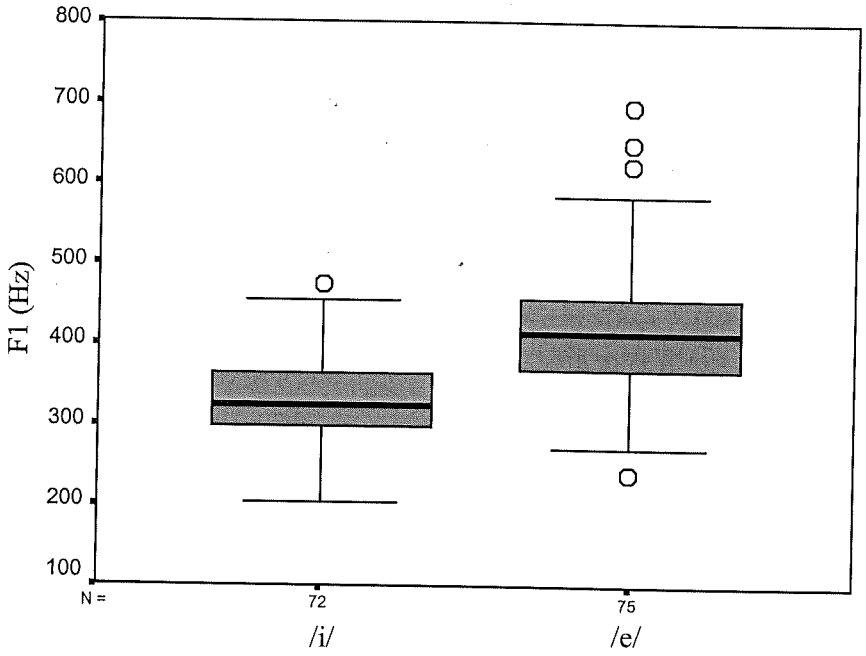
Figure 27. Boxplots of duration of /i/ and /e/



The statistical significance of this distinction was confirmed by an analysis of variance, taking duration as the dependent variable and speaker and vowel as the fixed factors. The assumption of equality of error variances was just met ($p < .05$ on a Levene test), and the value of R^2 was .414. The identity of the vowel proved to be highly significant ($p < .0005$), and that of the speaker only just significant ($p < .047$), although there was also significant interaction between the factors ($p < .011$). In addition to the length distinction, the first formant frequencies are significantly different, with /e/ showing consistently higher values,

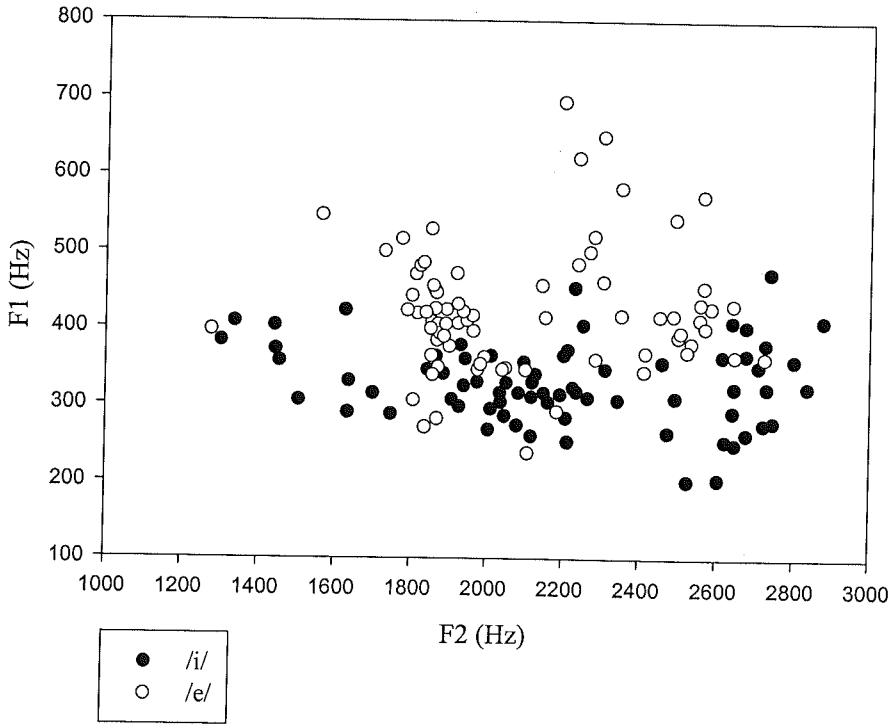
as would be expected. The non-parametric Mann-Whitney test confirmed this: grouping by vowel proved to be highly significant ($p < .0005$). Second formant frequencies, however, are not significantly differentiated: an analysis of variance taking vowel as a fixed factor showed it to be insignificant ($p < .139$). The identity of the speaker is a significant factor, however, due to the high F2 values that are a distinguishing feature of AL's speech.

Figure 28. Boxplots of F1 values of /i/ and /e/



Significant differences in both duration and F1 thus characterize the two vowels, although the scatterplot in figure 29 illustrates that the distinction between them is not particularly sharp.

Figure 29. Scatterplot of F1 against F2 for /i/ and /e/



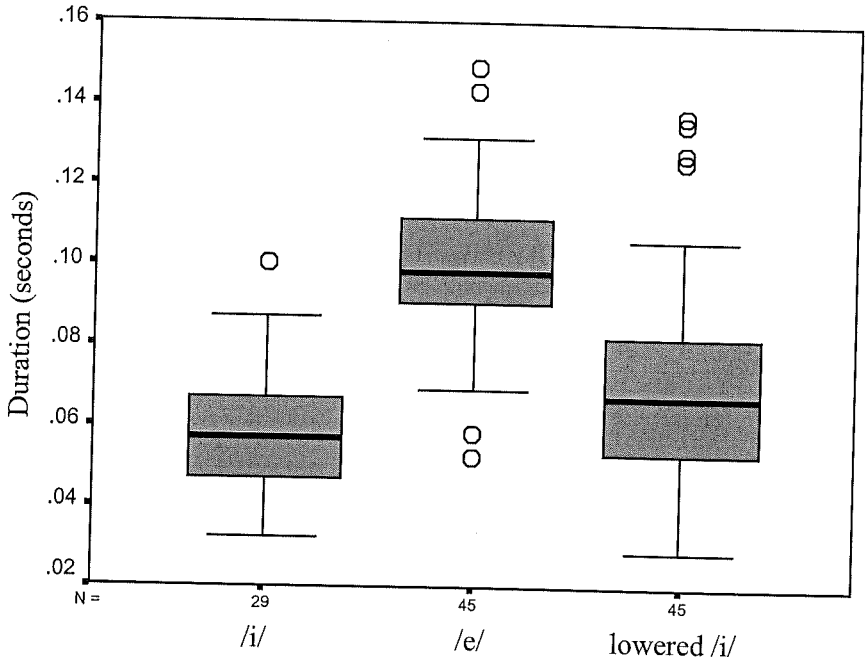
The next stage in the analysis was to compare the cases where lowering is thought to occur with counterparts containing undisputed instances of both /i/ and /e/. For this purpose the same series of measurements (duration, mid F1 and mid F2) was considered for the following eight words, with the third column containing the expected cases of lowering:

Table 11. Words containing /i/, /e/ and lowered /i/

Context	**/i/**	**/e/**	**lowered /i/**
_l	*kili* (10)	*eli* (43)	*ilaiyai* (68)
k_ɳ[22]		*keɳɟi* (25)	*kiɳaru* (16)
_ʈ	*iʈuveen* (50)	*eʈu* (40)	*iʈam* (47)

The durations of lowered /i/ turned out to be intermediate between those of /i/ and /e/, as the boxplots in figure 30 show. The /i/ and /e/ values here are quite clearly separated, with /e/ being characteristically longer, and lowered /i/ somewhere between the two.

Figure 30. Boxplots of duration of /i/, /e/ and lowered /i/



A repeated measures analysis was conducted on the data to establish whether duration or formant frequencies are significantly correlated with the three-way distinction under consideration. None of the three parameters gave significant results for either segment type or phonetic environment, so the significance of the differences between pairs of vowels was then tested (i.e. /i/ vs. lowered /i/, /e/ vs. lowered /i/ and /i/ vs. /e/). Unsurprisingly, the /i/ vs. /e/ distinction proved to be highly significant for duration ($p < .0005$ on an analysis of variance), which is in line with the results for the data set in table 9. Moreover, lowered /i/ was significantly different in duration from /e/ ($p < .0005$ for the difference from /e/ on an analysis of variance), although not from /i/ ($p < .097$ on the Mann-Whitney test).

[22] Although *kiɳɟan*, the echoed reduplicant of *laɳɟan* 'London', would fill the gap in this row, it was not included in the data set as the /i/ is arguably not in an initial syllable. Its inclusion had no effect upon the analysis of F1and F2, but it slightly distorted the results for duration, producing a significant difference between /i/ and lowered /i/, when there was none for the data set in table 11. This can be plausibly attributed to a decrease in duration arising from reduction. *kiɳɟan* was, however, included in the repeated measures analysis to satisfy the requirement that each within-subjects factor have at least three levels.

As with the data set in table 9, the first formant frequencies of /e/ are significantly higher than those of /i/ ($p < .0005$). The frequency values for lowered /i/ are sufficiently ambiguous that they are not significantly different from /i/ (an analysis of variance shows $p < .058$ for the difference between /i/ and lowered /i/), but $p < .05$ for the difference between lowered /i/ and /e/.[23] No distinction can be made in the F2 values at all, which is unsurprising, given that the F2 values of /i/ and /e/ are not significantly different. In conclusion, then, there is only slight evidence of the lowering process mentioned in the literature, indicating that suggestions of a merger between lowered /i/ and /e/ have been overstated. Lowered /i/ vowels are not significantly differentiated from their non-lowered counterparts on any of the measurements tested. However, they differ markedly from /e/ in their duration, and their F1 values are also significantly different.

4.6.2. Analysis of echo word structure

The results of section 4.3 indicated that the duration of /i/ is not significantly affected by the position of the syllable in which it occurs (aside from the issue of word-final lengthening). Formant frequencies were only marginally affected, with non-initial /i/ having slightly higher F1 values than /i/ in initial syllables. Thus the /i/ in the first syllable of an echoed portion would be expected to show a slightly higher first formant frequency if the echo word is treated as a single phonological word, as in (3a). If the two parts are treated as independent words, however, with a structure as in (3b), there should be no such reason for any differentiation between the two /i/ vowels.

(3a) $[piccai\ kiccai]_{\omega}$ (b) $[piccai]_{\omega}\ [kiccai]_{\omega}$

In an echo word where the environment for lowering is met, however, this may also be responsible for a distinction between the two /i/ vowels. If the structure is as in (3a) the first one should be lowered (i.e. have a raised F1 and reduced duration, in comparison with non-lowered /i/ vowels), but not the second. If the structure is as in (3b), however, the lowering should affect both /i/ vowels. Combining the effects of reduction and lowering, it is expected that there will be no significant differentiation between the two /i/ vowels if the echo portion forms its own independent word, as in (3b), both being equally affected by lowering and with reduction irrelevant. If the structure of the echo words is as in (3a), however, there is potentially a conflict between the lowering effect, which would result in the F1 value of the initial syllable being raised with respect to other /i/ vowels, and reduction, which would raise the F1 values of all the non-initial /i/ tokens. These two effects could conceivably cancel out so that there was no differentiation; otherwise, a distinction reflecting the stronger effect would be found.

The data set given in table 12 was used to test these predictions, with the environment for lowering being met in all the base words (in the left-hand column).

Table 12. Base and echo words

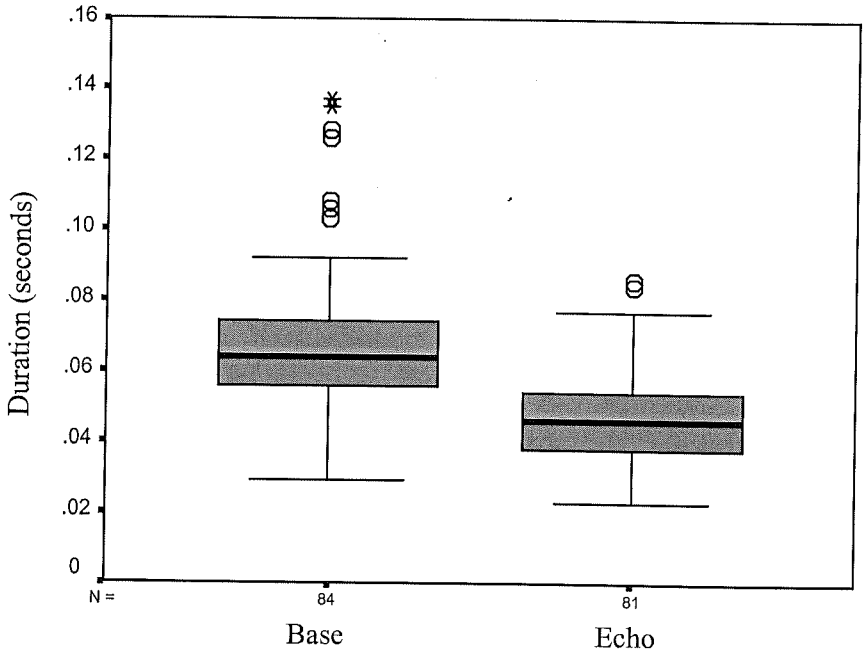
Base	**Echo**
ilaiyai (6)	*kilaiyai* (6)
civappu (9)	*kivappu* (9)
citamparam (41)	*kitamparam* (41)
iṭam (47)	*kiṭam* (47)
miḷaku (59)	*kiḷaku* (59)
kiṭavan (12)	*kiṭakaana* (54)

[23] This was further investigated by including another two pairs of words: *miḷaku* and *ninaikkaatee*, in which the environment for lowering is met, counterbalanced by *kiḷi* and *tinnaatee* respectively. Again, however, the first formant frequency values proved statistically indistinguishable.

In the first five pairs of words, the two parts belonged to the same echo formation but in the last case the vowel in the initial syllable of the base word was not /i/ but /a/ (the base being *aɻakaana*), and so an unrelated but phonetically similar lexical item was used.

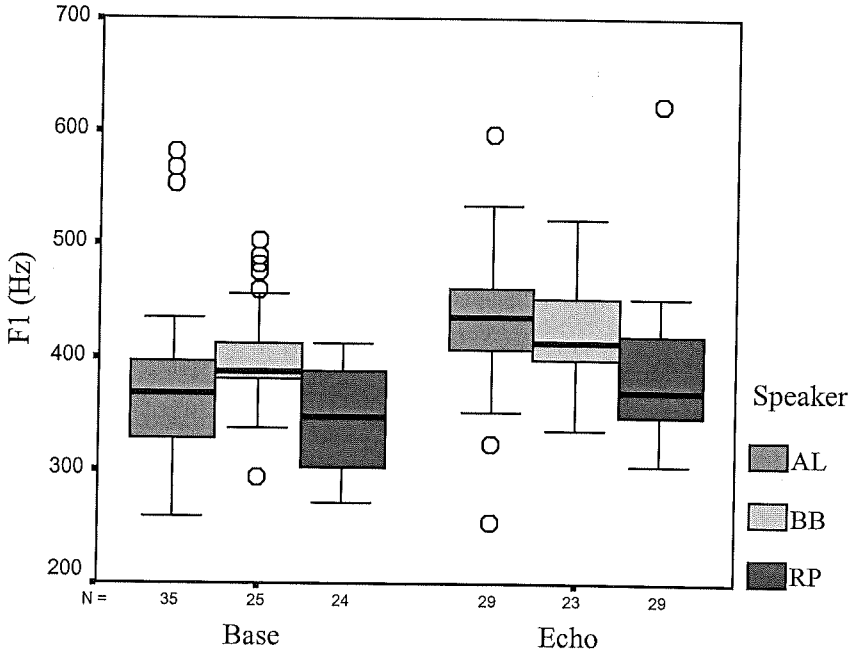
Comparison of the two sets of data produced some significant results, pointing to structure (3a) for the echo words. /i/ is markedly longer overall in the first set of words, the bases for echoing, as the boxplots in figure 31 show.

Figure 31. Boxplots of duration of /i/ in bases and echoes



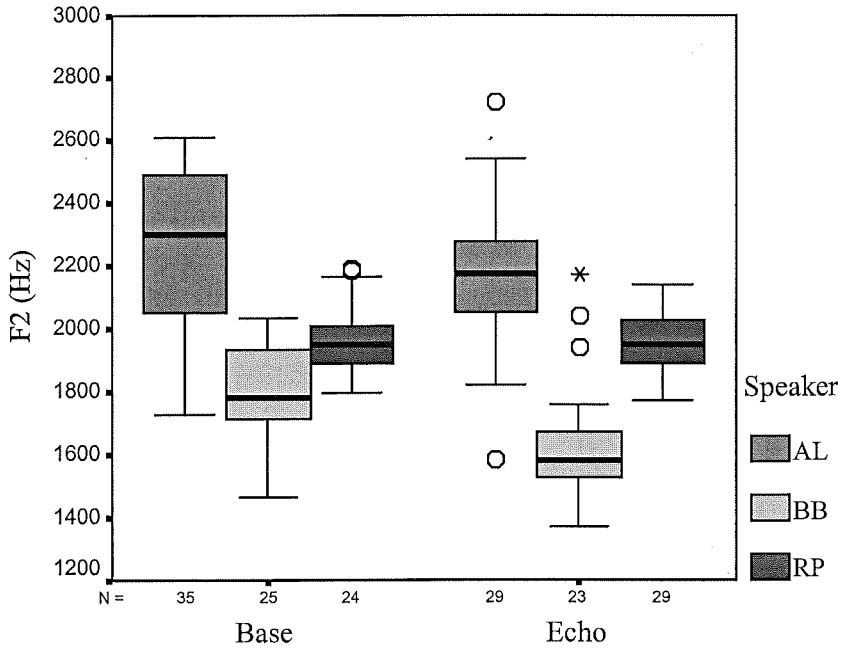
The statistical significance of the difference was confirmed by a Mann-Whitney test on the duration values of the two groups ($p < .0005$). Speaker, however, does not prove to be a significant factor ($p < .792$ on an analysis of variance).

Figure 32. Boxplots of F1 values of /i/ in bases and echoes, clustered by speaker



As the boxplots in figure 32 show, there is also differentiation of F1, with higher values occurring in the echoes than the base words. Both the base vs. echo distinction and the identity of the speaker are highly significant factors ($p < .0005$), and there is no significant interaction between them. F2 is also significantly affected, according to whether the /i/ vowel occurs in the base or the echo ($p < .031$ on an analysis of variance), with lower values in the echo. Speaker identity is again highly significant ($p < .0005$ on the Kruskal-Wallis test), as demonstrated by the separation of the frequency ranges, particularly in the echoes, displayed in figure 33.

Figure 33. Boxplots of F1 values of /i/ in bases and echoes, clustered by speaker



The fact that significant differences have been found between base and echo is evidence against the structure in (3b), in which the base and echo form independent phonological words. Moreover, the nature of the differences is consistent with a process of reduction, in which reduced tokens are shorter and more centralized in quality, and so have higher F1 values and lower F2 values. As noted in the previous section, the claims made about lowering in the literature appear somewhat exaggerated. If it has any effect here, it is outweighed by the reduction which affects the echo vowels.

4.6.3. Comparison with compounds

4.6.3.1. Background

The semantic and morphological properties of different types of Tamil compound have been discussed in section 3.2.3.4. There has been some interest in the literature on their phonological properties, particularly the application or non-application of various morphophonemic processes at the boundary between the two elements, and the implications for their structure. Annamalai and Steever, for instance, discussing the scope of such morphophonemic changes, offer the following generalization:

> *'They are obligatory with a bound morpheme, less frequent between members of a compound and least frequent when the combination does not result in a compound'* (Annamalai & Steever 1998: 103).

Consideration of similar issues is found in Mohanan (1986), which applies the theory of Lexical Phonology, recently developed at that time, to Malayalam. He argues that there are consistent phonological differences between endocentric and exocentric compounds, or sub-

and co-compounds. For example, gemination[24] applies within sub-compounds, but not co-compounds, and he also observes that:

> *'A subcompound with a single primary stress and a single word melody acts as a unit, irrespective of the number of stems in it, while each stem in a co-compound is a separate unit, with stress and tone assignment taking place on every stem.'* (Mohanan 1986: 33).

Mohanan's explanation of the data involves dividing the lexicon into four ordered strata,[25] with the key morphological processes distributed between them. The domain of application of phonological rules is then restricted to particular strata, as shown in (4).

(4)

Stratum I	Derivations	
Stratum II	Sub-compounding	Gemination
Stratum III	Co-compounding	Stress and tone assignment
Stratum IV	Inflections	

Furthermore, a principle of opacity (Mohanan 1986: 15) is assumed to render the internal structure of words at a particular stratum invisible at all following strata. Thus stress assignment, for instance, is taken to be blind to the internal structure of sub-compounds and so treats them as a single unit. Mohanan also argues for a loop between the second and third strata to account for certain complex compounds in which a co-compound forms a constituent element of a sub-compound. The introduction of such a device is a departure from the theory of Lexical Phonology as it was originally set out in Kiparsky (1982), and has been avoided in later versions of the theory (e.g. Booij & Rubach 1987). It certainly weakens the theory considerably, and has limited motivation, given that the compounds in question show the stress pattern of a co-compound, and thus do not appear to undergo the phonological rules of stratum III for a second time.

Christdas (1987) presents a simplified version of the model for Malayalam, eliminating the loop and reducing the number of strata to two. She also considers the situation in Tamil, examining the behaviour of binominal compounds in relation to various phonological rules. Three groups are compared: exocentric and endocentric compounds comprising two stems, and those endocentric compounds containing an *-am-* affix.[26] According to Christdas, exocentric compounds are differentiated from the others by not having their first stem marked with an oblique feature, and hence do not undergo two rules supposedly triggered by its presence.[27] Both of these (apical gemination and *m*-deletion), however, are directly concerned with deriving the oblique stem, rather than establishing how the internal boundary of the compound interacts with phonological processes of more general application. Unlike Mohanan, Christdas does not assign endocentric and exocentric compounds to separate strata, claiming that the differences between the three types of

---

[24] As in Tamil, geminates may appear either stem-finally, i.e. at the end of the first element of the compound, or stem-initially, i.e. at the beginning of the second element: Mohanan subsumes both of these within one general rule.

[25] A fifth stratum is proposed by Mohanan and Mohanan (1984: 581) to accommodate the concatenation of serial verbs and the auxiliary system.

[26] Note that Christdas' typology of compounds fails to include examples such as *tooṯṯattuppuu* 'garden flower', in which the first element is morphologically oblique, but there is no *-am-* affix.

[27] Christdas' theory rests upon the assumption that the first element of a compound like *marakkatavu* 'wooden door' is marked as an oblique stem. In her analysis such a form is subject to *m*-deletion, but has not undergone the rule of *tt*-insertion which applies in the compounds containing *-am-*, and also case-marked forms, e.g. *maratt-ukku* 'to the tree'. An alternative hypothesis, and one that avoids decomposing the oblique stem in this way, is that the first element in such cases is a nominative stem form. The absence of the final *-m* could be the result of cluster simplification: Tamil does not permit clusters containing nasal consonants and geminates.

compounds can be derived from the presence or absence of the morphological feature [oblique], together with the notions of head and percolation.

Discussion of compounds in Christdas' thesis (Christdas 1988) is limited to endocentric examples. She argues that these have a stem-stem structure, with each stem being assigned an accent feature prior to compounding. Her evidence comes from consideration of those rules that she takes to be blocked by the presence of accent or stress (see section 4.2). For instance, she states that there is no vowel reduction in the initial syllable of the second element of a compound and that final nasals and glides in a monosyllabic second element are not deleted. She also claims that there is only partial voicing of a post-nasal obstruent when it forms the first segment of a second element, as in *kooɭam toonʈi,* for which she gives the phonetic representation [koɭɜnTooɳɖi] 'pond dredge'. The capitalization is supposed to indicate partial voicing, due to an *'optional phonetic implementation rule'* (Christdas 1988: 192). Likewise, intervocalic obstruents are said to be only partially lenited when they belong to the first syllable of the second element of a compound, e.g. *muuʈupani* 'fog' [muuɖuPɑni]. These claims all motivate the assignment of stress to the first syllable of the second element of a compound. The only exception that Christdas mentions (to be discussed at greater length in section 4.6.3.2.3 below) concerns the occurrence of *kaay* as the second element of a compound. This reputedly shows full voicing of the initial obstruent after a nasal, lenition when intervocalic and deletion of the final glide, and thus is taken to act as an affix, rather than an independent stem.

Discussion of the phonological structure of endocentric compounds in Tamil is also to be found in work by Bosch and Wiltshire, although they rely almost exclusively upon Christdas for their data. Bosch (1991: 105), for instance, gives the structure of the compounds as [[word] [word]]. If correct, this represents a significant difference between compounds and the echo words, which appear to constitute a single phonological word. The following sections therefore test these claims, by comparing various compounds with occurrences of the second element as an independent lexical item.

4.6.3.2. Results

4.6.3.2.1. Vowel reduction

Christdas' claim about the absence of vowel reduction in the second element of compounds was investigated by looking at tokens of /i/ and /a/ in the initial syllables of *piɭɭai* 'child' and *kari* 'curry' and the corresponding vowels in the compounds *kilipiɭɭai* 'parakeet' (lit. parrot-child) and *kaaykari* 'vegetable-curry'. The /i/ vowels show little in the way of differentiation: neither duration nor F2 differ significantly, although F1 is significantly lower in *piɭɭai* than *kilipiɭɭai* ($p < .047$). This largely replicates the results of section 4.3.2.2, as summarized in table 5, where /i/ tokens in initial and non-initial syllables showed significant differences associated with reduction for F1 only. The /a/ tokens analyzed in section 4.3.2.1, however, proved significantly different in all three parameters measured. This finding is also repeated in the /a/ tokens from *kari*, which are significantly longer than those of *kaaykari* ($p < .0005$ on the Mann-Whitney test), with higher F1 values ($p < .0005$ on the Mann-Whitney test) and lower F2 values ($p < .024$ on an analysis of variance). These results suggest that the third /i/ of *kilipiɭɭai* and the /a/ of *kaaykari* share the phonetic properties of non-initial syllables, which have been associated with the absence of stress. This casts doubt upon a word-word structure for the compounds, suggesting that they may instead, like the echo words, form a single phonological word, with a single stress.

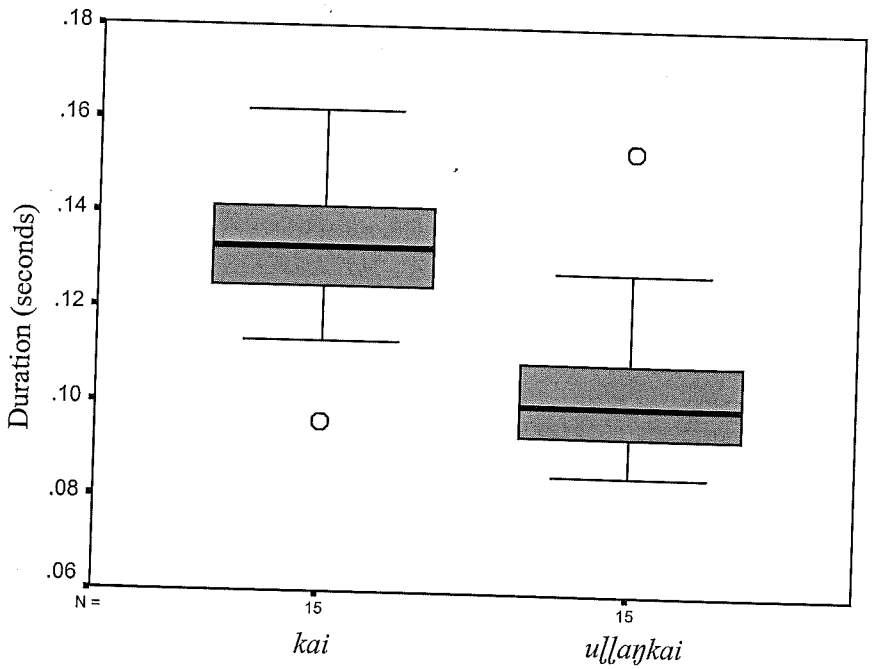
4.6.3.2.2. Monophthongization of /ai/

Section 4.5 revealed highly significant differences in both the duration and the formant structure of the diphthong /ai/ in initial and non-initial syllables. Parallel differences are to be expected between the vocalic portions of *kai* 'hand' and the second part of the

compound *uɭɭaŋkai* 'elbow', if the compound forms a single phonological unit. If it comprises two phonological words, however, each bearing their own stress, then the expectation is that the two occurrences of *kai* should not differ in any significant respect.

The results again support the first possibility, i.e. that the *kai* of *uɭɭaŋkai* behaves like any other non-initial syllable. The duration of /ai/ is significantly shorter than in *kai*, when it occurs as an independent lexical item, as the boxplots in figure 34 demonstrate.

Figure 34. Boxplots of duration of /ai/ in *kai* and *uɭɭaŋkai*



An analysis of the variance of duration showed the difference between the two words to be a highly significant factor ($p < .0005$, $R^2 = .403$), but not the identity of the speaker ($p < .558$). The change in F1 over the centre half of the vocalic portion was also significantly greater in *kai* than *uɭɭaŋkai* ($p < .017$, $R^2 = .188$), and speaker proved to be a significant factor in this case ($p < .002$). There was no significant result, however, for the corresponding difference in F2. Overall, this provides further evidence in favour of the compounds bearing a single stress, on the initial syllable of the first element only.

4.6.3.2.3. Glide deletion

One further potential diagnostic of phonological structure is the deletion of a word-final palatal glide following a long vowel. This is mentioned repeatedly in the literature[28] as one of the differences between formal and colloquial Tamil applying, for instance, to the adverbial ending, which is *-aay* in the formal language but pronounced simply as [aː] in colloquial Tamil. The general consensus is that deletion occurs in the final syllable of a polysyllabic word, but that it is blocked in monosyllables.[29] Christdas (1988: 195–6), indeed, includes it in her list of differences between initial and non-initial syllables (see section 4.2), as evidence for the accent falling uniformly on the initial syllable. Moreover, both Asher and Schiffman refer to the addition of a vowel to the final glide in such cases: Asher gives [ʋaːji]

[28] See, for instance, Asher (1985: 259), Britto (1986: 199) and Schiffman (1999: 16).

[29] Note, however, that Zvelebil (1970: 181) claims that glide deletion applies even to monosyllables, giving [nɑː] as the regular form in spoken Tamil of *naay* 'dog'. Dialectal variation is mentioned by Marthandan (1983: 56), who comments that in northern Tamilnadu words containing *-aay* end in a short front central vowel, whereas the final [j] is dropped entirely in the south of the state.

as the colloquial form of *vaay* 'joint', and Schiffman gives *ceyyi* as the equivalent in standard spoken Tamil of *cey* 'do, make'. In acoustic terms, therefore, monosyllables and non-final syllables are expected to have a dynamic structure with the first and second formants diverging clearly at the end, whereas non-initial final syllables should show no significant formant movement. There are obvious similarities with the monophthongization of /ai/ investigated in section 4.5, but the formant transitions are expected to start at a later stage of the vocalic portion in *-aay*.

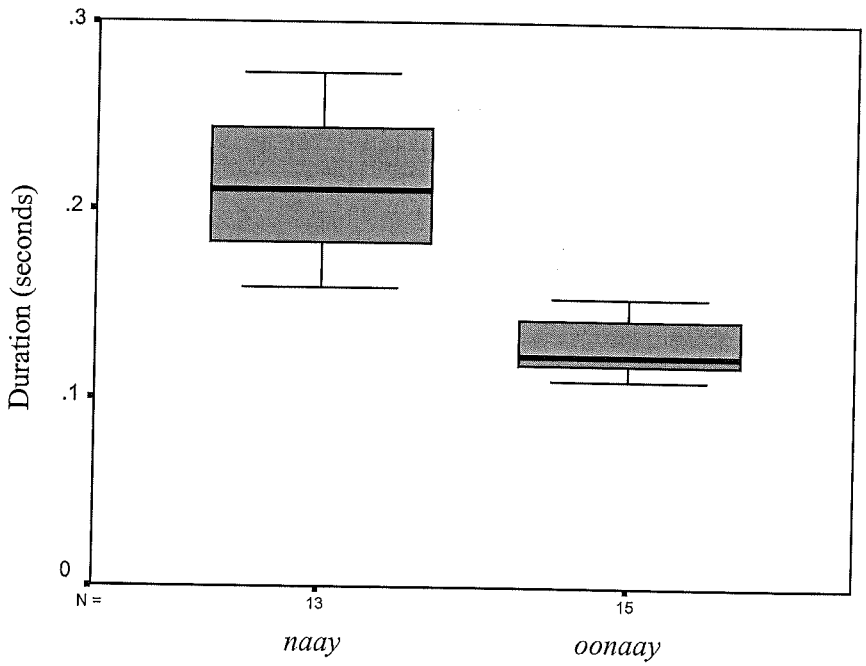
The relevant examples from the sentence set, including compounds, are given in table 13 below.

Table 13. Words containing *-aay*

Non-compounds	**Compound first element**	**Compound second element**
naay (18)	*naaykkuṭṭi* (44)	*oonaay* (35)
kaay (64)	*kaaykari* (78)	*maaŋkaay* (73)
taay (74)		*maattaantaay* (20)

Their durations were measured, and also the differences in frequency value between the quarter and three quarters point of the first two formants (ΔF1 and ΔF2). The prediction was that there should be no significant differences between the *-aay* of *kaay* and *kaaykari* and of *naay* and *naaykkuṭṭi*, and this was borne out by the results: analyses of the variance of duration, ΔF1 and ΔF2 revealed no significant differences between the monosyllabic and bisyllabic words. The *-aay* of *naay* and *oonaay*, however, are expected to differ significantly in duration, and this expectation is met in the data. The boxplots in figure 35 show the duration values to be clearly separated, and syllable position (but not speaker) proved to be highly significant ($p < .0005$) in an analysis of the variance of log duration values.

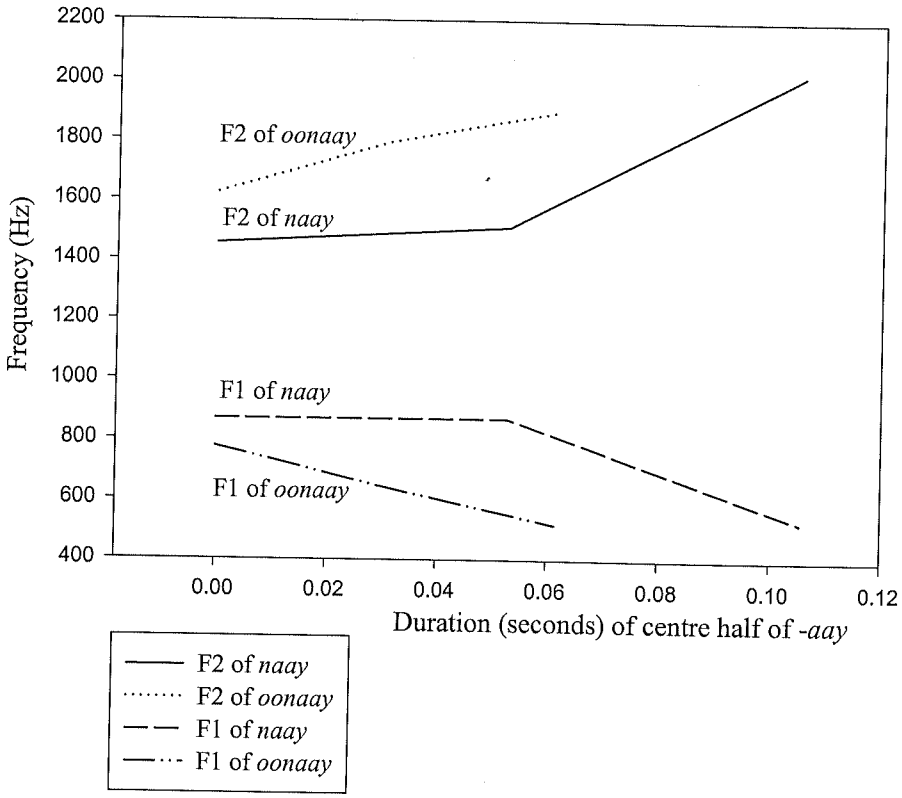
Figure 35. Boxplots of duration of *-aay* in *naay* vs. *oonaay*



Reduced duration in a non-initial syllable is consistent with glide deletion, but it seems also to be generally true of Tamil: all previous results have shown initial syllables to be markedly longer than their non-initial counterparts. The formant patterns are consequently important evidence: if glide deletion has occurred, the differences in formant frequencies during the vocalic portion should be minimal, whereas significant drops in F1 and rises in F2

are expected between the quarter and three quarters points if the final glide is still present. These were found in both *naay* and *oonaay*, strongly suggesting that deletion of the glide had occurred in neither word. The overall change in the formant frequencies is comparable in the two words, and analyses of the variance of ΔF1 and ΔF2 revealed no significant differences between them. There were, however, interesting differences in the shape of the formant patterns, illustrated in figure 36.

Figure 36. Line diagram illustrating formant movement in the centre half of *-aay* in *naay* and *oonaay*



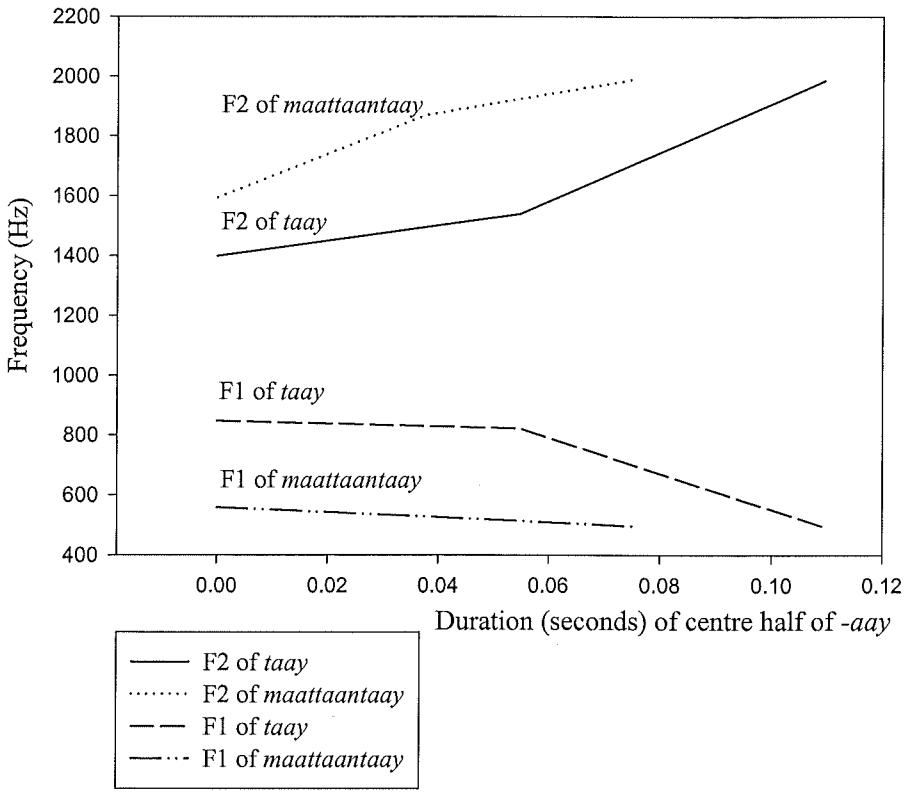
In *naay* both F1 and F2 remain fairly constant during the second quarter of the vowel, but then diverge sharply during the third quarter. In *oonaay*, by contrast, there is a steady rate of change throughout. The distinction between initial and non-initial syllables, therefore, seems to be not so much between [aːj] and [aː], as between [aːj] and [aj].

The two compounds in table 13 are the subject of specific comment by Christdas (1988: 404). She claims that there is no glide deletion in compounds such as *maattaantaay* 'stepmother', but that exceptionally compounds with *kaay* as the second element, e.g. *teeŋkaay* 'coconut', do show deletion. Some support for this comes from Asher (1985: 259), who gives [maːŋgaː] as the colloquial version of *maaŋkaay* 'mango'. The general tendency for glides not to be deleted is reiterated in Bosch and Wiltshire (1993: 12), who give the example of *puttunooy* 'cancer' (lit. anthill disease) not undergoing glide deletion. Christdas argues that the difference in interaction with glide deletion is evidence of two types of compound structure in Tamil, with the first syllable of both elements receiving stress in examples like *maattaantaay*, but only the first syllable of the first element in exceptional cases like *maaŋkaay*, in which the *kaay* element behaves like a suffix.

Both *maaŋkaay* and *maattaantaay* were included in the data set, and compared with cases where the second element occurred as an independent item. The *-aay* of *kaay* and *maaŋkaay* were found to differ significantly on all three parameters tested, i.e. duration, ΔF1

and ΔF2. Analyses of variance were performed on each, and position within the word proved to be highly significant for duration and ΔF1 ($p < .0005$), and significant for ΔF2 ($p < .019$). (Speaker is a significant factor only for ΔF2: $p < .005$ on an analysis of variance.) The formant pattern of the *-aay* of *maaŋkaay* was closely comparable to that of *oonaay*; the first and second formant frequencies diverge clearly but at a fairly constant rate.[30] *kaay*, like *naay*, however, showed a greater rate of change in the third than the second quarter of the vocalic portion. The findings are therefore consistent with the whole compound forming a single phonological unit, like *oonaay*, with the first syllable more prominent than the second. Interestingly, the same general pattern is found when *taay* and *maattaantaay* are compared. Duration and ΔF1 vary significantly between the two ($p < .002$ and $p < .012$ respectively on analyses of variance), although there is no significant result for ΔF2. Formant frequency movement, as shown in figure 37, again follows the same pattern, with *taay* showing sharp divergence of the first and second formant frequencies in the third quarter, and *maattaantaay* an even rate of change.

Figure 37. Line diagram illustrating formant movement in the centre half of *-aay* in *taay* and *maattaantaay*



This is at odds with the claims of Christdas and others, for whom the two elements of the compound form separate phonological words. According to them, neither *taay* as an element of a compound nor *taay* as an independent lexical item undergo glide deletion, so there should be no significant differences between them. The acoustic data do, however, show evidence of glide deletion, in the evening out of formant movement and reduction in duration. All the evidence for this relates to the distinction between initial and non-initial syllables, with the second elements of both *maaŋkaay* and *maattaantaay* behaving as non-initial syllables. These findings are entirely in line with those of the previous two sections, and

[30] There was one notable exception to this amongst the fifteen tokens: the female speaker (AL) produced one instance of *maaŋkaay* with no discernible divergence of the formants.

suggest that, contrary to the claims of Christdas and others, the endocentric compounds of Tamil form a single phonological unit bearing one main stress.

4.6.4. Comparison with expressives

As noted in section 2.2, echo expressions belong to a continuum of reduplicated forms, which also includes the category termed 'expressives' (see section 2.2.3). Those that are most directly comparable to the echo words comprise a reduplicated stem followed by various inflectional endings, depending on category. For example, *kiɻukiɻukkiraan* 'he is trembling' consists of a repeated stem *kiɻu*, and *-kkiraan*, which contains the verbal affix indicating present tense and the third person singular masculine ending. As with the echo forms and compounds, therefore, the question arises of whether they form a single complex phonological word, or whether they divide into two, with the inflectional morphology phonologically dependent only on the second half. These two possibilities are given for *kiɻukiɻukkiraan* in (5).

(5a) $[kiɻukiɻukkiraan]_{\omega}$ (b) $[kiɻu]_{\omega}\ [kiɻukkiraan]_{\omega}$

If the structure is as in (5a), the expectation is that the third vowel should show evidence of vowel reduction when compared to the first vowel. If the structure is as in (5b), however, there should be no significant differences between the two. This was tested by taking duration and formant frequency measurements for the expressives given in table 14 (149 tokens in all). The numbers in brackets refer to the sentence in appendix F in which the word appears, and the glosses indicate both meaning and syntactic category.

Table 14. Expressive examples

Example	**Gloss**
kirukiruŋŋu (26)	'giddiness'$_{N}$
kirukiruppu (76)	'giddiness'$_{N}$
kiɭukiɭuppaa (60)	'laughing loudly'$_{ADV}$
kiɻukiɻukkiraan (45)	'tremble.pres.3sm'$_{V}$
paɭapaɭakkum (25)	'shine.fut.3sn'$_{V}$

The first four examples were tested first: each contains the *ki-* syllable also found in the echo forms. The difference in duration is demonstrated in the boxplots in figure 38, with the first syllable being markedly longer than the third. The statistical significance of the difference was confirmed by an analysis of variance taking duration as the dependent variable and syllable and speaker as the fixed factors. Both proved to be significant (although not their interaction): for speaker $p < .021$ and for syllable $p < .013$, whilst the value of the squared correlation (R^2) was .133.

Differences in formant frequency consistent with vowel reduction were also found: according to the non-parametric Mann-Whitney test, F1 was significantly lower in the first than the third syllable ($p < .0005$) and F2 significantly higher ($p < .0005$). The scatterplot in figure 39 shows the formant frequencies plotted against each other, with different tokens generally falling into one of two overlapping clusters based on syllable: filled circles denote /i/ tokens occurring in initial syllables and empty circles /i/ tokens in third syllables. The final expressive involving /a/, *paɭapaɭakkum*, also had significantly longer tokens of /a/ in the initial syllable than the third syllable ($p < .003$ on an analysis of variance), but there were no significant results associated with formant frequencies. As they derive from a single example, however, the results do not warrant much weight being placed upon them.

Figure 38. Boxplots of duration of /i/ in first and third syllables, clustered by speaker

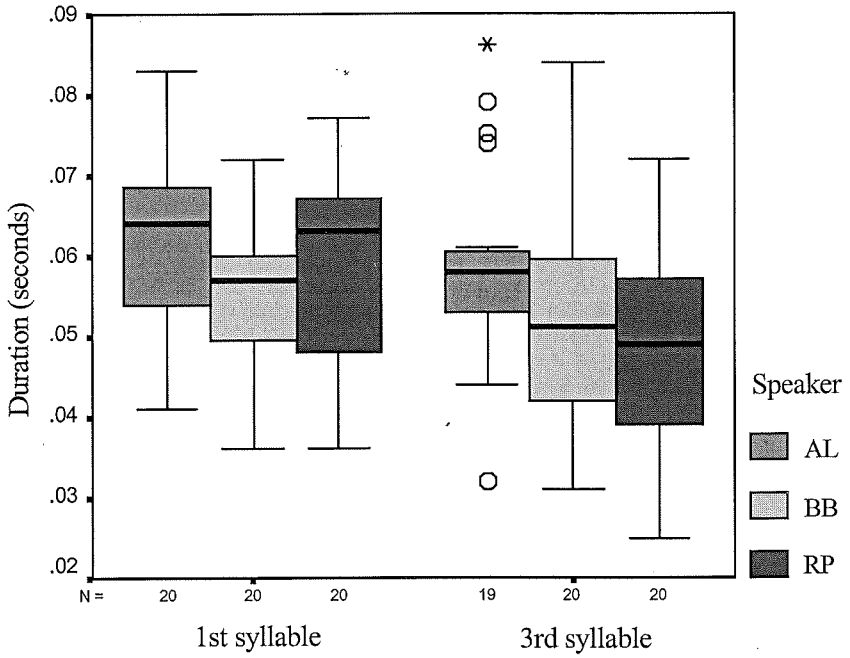
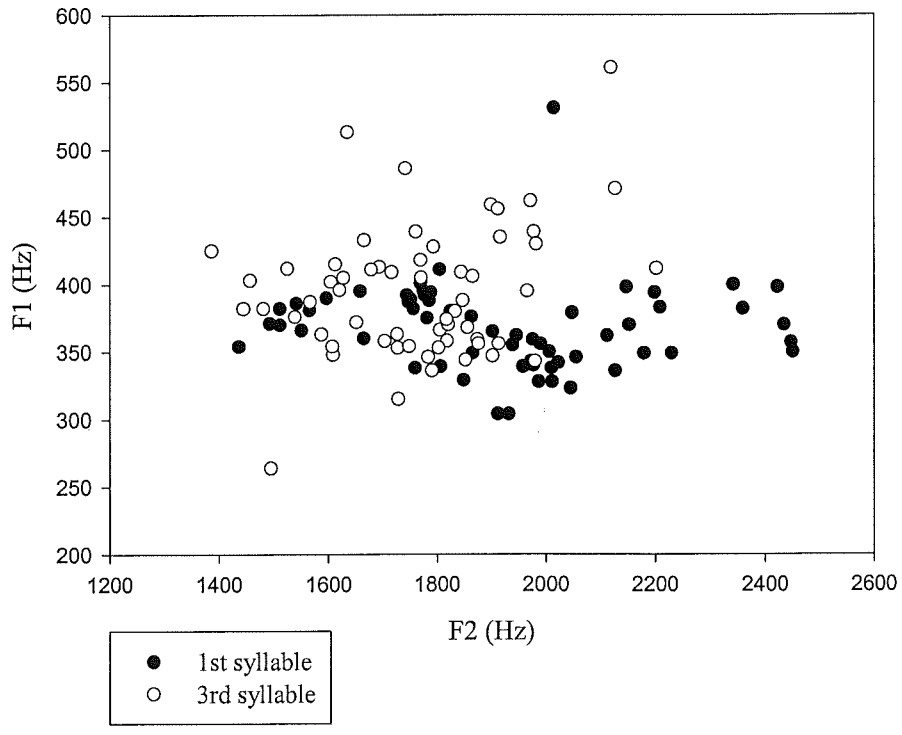


Figure 39. Scatterplot of F1 against F2 for /i/



Overall, these findings clearly point to the expressives forming single complex phonological words, i.e. the structure given in (5a). In this respect, therefore, the evidence from stress patterns indicates that expressives are comparable both to the echo words and also the compounds investigated in section 4.6.3. A further source of acoustic evidence for the phonological structure of these expressions will be brought to bear in the next chapter.

Chapter 5: Word-initial and word-internal /k/ and /kk/

5.1. Introduction

This chapter considers another potential diagnostic of echo word structure, the nature of the echo /k/ segment. As described in section 5.2, the phonetic nature of the Tamil obstruents has long been recognized to vary widely, according to position within the word and whether the segments are single or geminate. Each of these factors is explored in the chapter, beginning in section 5.3 with the question of whether or not colloquial Tamil makes any phonetic distinction corresponding to the orthographic differentiation between single and geminate obstruents in word-initial position. The phonetic correlates of the word-internal distinction between /k/ and /kk/ are then examined in section 5.4. Section 5.5 uses the results established in the preceding sections to investigate the structure of echo words and phrases, comparing them, as in chapter 4, with expressives.

5.2. Overview of the literature

5.2.1. Phonology, phonetics and orthography of the Tamil obstruents

Five places of articulation are phonemically distinguished in the series of Tamil obstruents: bilabial /p/, dental/alveolar /t/, retroflex /ʈ/, palatal /c/ and velar /k/. Their phonetic realizations show considerable variation, depending upon their position in the word, and other contextual factors. The nature and distribution of the realizations have been variously documented: an early and detailed phonetic description is given by Firth in an appendix on Tamil pronunciation at the end of Arden's grammar (Firth 1934). More recent publications devoted to the phonetics of Tamil include a monograph by Rajaram, which is a systematic phonetic description of the language (Rajaram 1980), and also the doctoral theses of Balasubramanian (1972) and Marthandan (1983), the first of which is supported by extensive instrumental data.[1] Whilst the general distribution is common to all the obstruents, there are some further peculiarities specific to particular members of the series. For instance, the retroflex sounds are limited to word-internal position, and the distinction between dental and alveolar place of articulation is generally made only in the formal language[2] and restricted to intervocalic geminates.

The general pattern is that obstruents are realized as voiceless plosives initially and medially when members of a cluster, and voiced stops when following a nasal. Intervocalically, however, single stops are lenited, so that they are realized as voiced fricatives. Balasubramanian (1972: 50) gives the following description of the distribution of realizations for the bilabial stop symbol <ㅂ>:

(1) [pʻ][3] initially in a word

(2) [β] intervocalic single consonant

(3) [pˑ] intervocalic geminate preceded by a long vowel in disyllabic words and in polysyllabic words, regardless of the length of the preceding vowel

(4) [pː] intervocalic geminate preceded by a short vowel in disyllabic words

(5) [b] medially after a nasal consonant

---

[1] See also Lisker (1958, 1972), Balasubramanian and Asher (1984), Asher (1985), Christdas (1988) and Zvelebil (1990).

[2] As mentioned in footnote 43 of chapter 1, Christdas (1988) reports that the Kanniyakumari dialect on which her research is based unusually makes a distinction between dental and alveolar stops (as does Jaffna Tamil).

[3] The diacritic represents slight aspiration.

Parallel lists are given for the other consonants, with the exception that the retroflex obstruent is a flap [ɽ] in intervocalic position, rather than a fricative, and Balasubramanian gives the lenited form of /k/ as a voiced glottal fricative [ɦ].[4]

There is a single orthographic symbol corresponding to each of the five places of articulation, with geminates written double. This is suited to the pattern of complementary distribution outlined in (1)–(5), but holds only for a subset of the lexicon – the native Dravidian vocabulary. Borrowed words, including those from Sanskrit and English, may have voiced stops in initial (and also intervocalic) position, thereby introducing a distinctive contrast in voicing. As Balasubramanian (1972: 30–39) points out in his introductory chapter, the adoption of vocabulary that does not conform to the native phonological system is neither a recent nor a marginal phenomenon. The *Tolkaappiyam*, an ancient grammatical treatise, contains hundreds of Sanskrit loan words. He also reports an informal survey of *'working class'* Tamil speakers conducted in 1970 which found that loan words accounted for 45% of the vocabulary used in colloquial utterances. Despite the wholesale incorporation of such borrowed material into the lexicon, the writing system remains conservative, and opinions differ over the need for reform (see section 3.2.2).

The orthographic system is regarded as basically phonemic in the literature on Tamil linguistics.[5] As Annamalai and Steever comment:

> *'Lacking an adequate phonology of modern Tamil, linguists take the transcription of the written language as the underlying phonological representation – simultaneously the output of the syntax and the input to the phonology – and the corresponding spoken form as the surface representation.'* (Annamalai & Steever 1998: 101).

Consequently there is usually considerable overlap between the phonological derivational rules proposed and the steps given for converting literary into colloquial Tamil. Christdas is something of an exception: in her thesis she makes the case for more abstract underlying representations, but it is only at morphological boundaries that her lexical representations depart consistently from the orthographic forms.[6] Balasubramanian (1982) also notes some respects in which the orthography fails to correspond in a one-to-one fashion to underlying phonological distinctions. For example, he notes that two symbols are used in the pairs <ந> and <ன>, and <ர> and <ற> to represent a single underlying segment. The symbols <ந> and <ன> are both of restricted, though overlapping distribution, and represent between them dental, alveolar and postalveolar nasals, which occur in complementary distribution. There are also some constraints on the environments in which each of the symbols <ர> and <ற> can occur, but Balasubramanian concludes that both are pronounced as either of two freely varying alternants, [r] or [ɾ]. One further departure from a one-to-one correspondence between symbols and underlying segments occurs in the vowel system. As noted in section 4.2, according to Balasubramanian (1972: 102–103), his dialect contains seven underlying short vowel phonemes, but these are represented by only five orthographic symbols. With these exceptions taken into consideration, however, it is a reasonable working hypothesis that the phonological and orthographic inventories are isomorphic.

---

[4] Note that there are some discrepancies in the literature (presumably due to differences between dialects) over the nature of a few of the phonetic realizations: Christdas (1988: 143), for example, gives voiceless [x] as the usual intervocalic realization of /k/, with further weakening to [h] in fast speech, especially in the final syllable of a polysyllabic word. She also gives the labiodental approximant [ʋ] as a weakened variant of /p/ in intervocalic position (Christdas 1988: 136).

[5] One exception to this is Marthandan (1983: 81), who posits an additional five phonemes: /b/, /d/, /dʒ/, /ɖ/ and /g/.

[6] In particular, Christdas (1988: 279) argues that all nominal stems are underlyingly consonant-final.

5.2.2. Gemination in Tamil

The contrast between single and geminate stops in word-internal intervocalic position may be lexically distinctive, as in the following minimal pair of nouns:

(6) *kaṭai* 'end' *kaṭṭai* 'tree trunk'

The double representation in *kaṭṭai* is taken to reflect an underlying geminate, an inherent property of the lexical item: other instances of double consonants, however, are attributed in the literature to various rules of gemination operating in the morphology. These cover a whole range of cases: doubling of stem-final consonants before case endings, as in (7) and (8), double *-tt-* for oblique forms of noun stems ending in *-m*, as in (9), doubling of the /k/ of the plural suffix, as in (10), doubling of the initial /t/ of the emphatic clitic *taan*, as in (11), doubling of the initial stop of the second element of certain compounds, as in (12), and even the contrast between single and double stops in some related pairs of verbs, as in (13). Finally, (14) gives an example of word-initial gemination, showing doubling of the initial segment of the verb.

(7)	*pal*	'tooth'		*pallai*	'tooth.acc'
(8)	*viiṭu*	'house'		*viiṭṭukku*	'house.dat'
(9)	*maram*	'tree'		*marattai*	'tree.acc'
(10)	*puukkaḷ*	'flower.pl'	cf.	*viiṭukaḷ*	'house.pl'
(11)	*aarraittaan*	'river.acc.emph'		*taan*	emphatic clitic
(12)	*naaykkuṭṭi*	'puppy'	cf.	*kuṭṭi*	'calf'
(13)	*meeykiratu*	'graze.pres.3sn (intrans.)'			
	meeykkiratu	'pasture.pres.3sn (trans.)'			

(14) *paampai ppaartteen*
snake.acc see.past.1s
'I saw the snake.'

Some have regarded all of these as reflexes of a single process whereby consonants are doubled at boundaries (see, for example, Annamalai & Steever 1998: 103).[7] However, there are some fairly major distinctions to be made, in terms of whether the segment doubled precedes or follows the relevant boundary, the nature of the boundary itself and the range of segments affected. The remainder of this section will be devoted to the final kind of example, seen in (14).

Word-initial doubling of /c/, /k/, /p/ and /t/ is reflected orthographically by writing another instance of the stop symbol at the end of the preceding word, as in (15), which gives the orthographic representation of (14).

(15) பாம்பைப் பார்த்தேன் *paampaip paartteen*

Traditional grammars categorize such instances as external sandhi, giving long lists of seemingly arbitrary environments in which the doubling occurs. For instance, the eighteenth century grammar of the '*common dialect*' by Beschi (translated 1831) provides some sixteen rules on the issue, with the cautionary remark:

> '*And let not any despise them, nor blame me as if I did hunt after trifles and waste my time with impertinences: nor let them shun the loss of time in reading these rules,*

[7] This is also the approach taken by Mohanan (1986) to a range of similar gemination processes in Malayalam.

> *as if it were lost labour, but rather rest assured, that it will be amply repaid: especially as by this very circumstance the meaning of the Speech is materially altered'* (Beschi 1831: 14).

His comments certainly imply that the distinction is auditorily, rather than just orthographically, significant. Arden (1934: 67–68) has a slightly more concise treatment of the issue, specifying that words ending in a vowel cause an initial /k, p, c, t/ to be doubled after:

(16) *anta* 'that', *inta* 'this', *enta* 'which', *appaṭi* 'in that way', *ippaṭi* 'in this way', *eppaṭi* 'how', *marra* 'other', but not *paṭi* 'so, thus'

(17) an accusative or dative case

(18) an inflectional base of a noun when the first of two words is combined with the second to form a compound

(19) nouns ending in *-y*, *-r*, and sometimes *-ṛ*, which qualify a following noun in an adjectival sense

(20) verbal participles *pooy* and *aay*, those ending in *-i* and those ending in *-u* preceded by a double consonant

(21) an infinitive ending in *-a*

No doubling occurs, however, if the first word is a relative participle, finite verb, vocative case, ends in the emphatic marker *-ee*, or the interrogative markers *-oo* or *-aa*, or the postpositions[8] *uṭaiya, ooṭu, iliruntu* or *ininru*.

A more recent discussion of the phenomenon, described as gemination in syntax, is to be found in Christdas (1987). She attempts to give a principled explanation for the contexts in which it is found, concluding that the doubling is triggered by inflected nouns (those built on the oblique stem), and applies only within the syntactic constituent V'. There are a few discrepancies between the conditions for gemination described by Christdas and Arden's list. For example, Arden states that doubling occurs after any dative case, whereas Christdas excludes dative case-marked nouns that function as the subject (Christdas 1987: 140). There is, indeed, ongoing debate amongst Tamil scholars over the rules defining exactly where word-initial gemination occurs, although there is a broad measure of agreement over the main environments.

The degree to which these 'rules' are observed may depend, at least in part, upon the variety of the language being used. As noted in section 3.2.2, Tamil has both colloquial and formal varieties, with written texts almost exclusively in formal Tamil.[9] Britto (1986), in a book devoted to the issue, proposes one further variety or diasystem, distinguishing between Classical, High (or Literary) and Low (or Colloquial) Tamil. Each is said to serve a different function, the first being a *'dead'* variety preserved in ancient literary texts, and acting now as a symbol of Tamil culture. High Tamil is described as a simplified and modernized version of this, used for writing and for high register functions, whereas Low Tamil is the colloquial variety, used by all Tamil speakers for low functions, but not written. Britto comments that the external sandhi rules are observed more rigidly in Classical than Literary Tamil (Britto 1986: 130), whilst Anton and Hellmann (1975: xlviii), in a survey of Tamil usage in mass media, note a tendency towards ignoring these same rules. Such remarks raise the possibility

[8] Note that I follow Arden's terminology here: elsewhere these 'postpositions' are referred to as case markers (see section 3.2.3.3.2).

[9] The main linguistic treatments of Tamil all make reference to these discrepancies: see, for instance, Asher's list of correspondences (Asher 1985: 259–262), also Balasubramanian (1972: 9–27) and Annamalai and Steever (1998: 125–127).

that phonetic correlates of word-initial gemination may have weakened or even disappeared entirely from the modern colloquial variety.

The thesis of Balasubramanian (Balasubramanian 1972) is the most detailed investigation of the phonetic nature of word-initial gemination to date, and the only one to employ instrumental analysis. In his fifth chapter, which is devoted to the double 'stops' of Tamil, he presents duration measurements for stops in both connected speech and words spoken in isolation, using spectrographic and kymographic analysis. In the former category he gives values for the duration of stop gaps in word-initial position, for both single and double orthographic representations (see table 2 below for the actual figures). As he comments, the durational differences between single and double are not great and there is considerable inconsistency between samples (partly attributable, he suggests, to differences in the degree of emphasis given to the relevant words). His conclusion concerning the phonetic distinction between single and geminate is as follows:

> *'If the word initial stop is not orthographically doubled in writing down a bit of connected prose, it is not usually a voiceless stop in speech, unless a pause intervenes between this word and the one preceding it. Otherwise it is a voiced fricative (or flap). Another striking factor is that what is orthographically doubled is always a voiceless stop in speech, however short the duration of closure may be.'* (Balasubramanian 1972: 266).

Effectively, Balasubramanian is ascribing to word-initial position the same distinction that has been consistently reported for word-internal single and geminate segments in Tamil. Voicing and frication are therefore taken to be the main correlates of gemination, with duration subordinate.

Durational differences have also been discussed by others, with regard to whether they justify a double representation for the geminates. Marthandan (1983: 39), for instance, takes the position that orthographic doubling does not represent phonological gemination, apparently on the grounds that orthographically double segments are not sufficiently long. Several writers comment on the variability of closure durations: Firth, for instance, in his appendix to Arden's grammar (Arden 1934), notes that the '*double stops . . . vary considerably in length*'[10]. He does, however, state that their length is determined in part by environment, with the longest realizations following short vowels in prominent syllables and shorter realizations after long vowels. This claim was investigated experimentally by Lisker (1958), who concluded that:

> *'there is no very striking relation between the position of a voiceless stop and the duration of its closure interval'* (Lisker 1958: 296).

He recognized a durational distinction between the closure durations of orthographically single and geminate segments but regarded this as merely a *'concomitant'* of the other phonetic differences. Daniel Jones commented briefly on the subject, comparing the Tamil double obstruents with geminates characterized primarily by length in other languages. He implied that duration played some role in distinguishing between single and geminate in Tamil, arguing that the geminates were:

> *'at times shorter than completely doubled consonants, but never shortened to such an extent as to make it impossible to regard them as long on the whole'* (Jones 1967: 173).

Like Firth and Lisker, Balasubramanian (1972) considered the influence of environment on closure durations (supporting Firth's claims contra Lisker), but made no comparisons between geminates and single segments word-internally, presumably on the grounds that

[10] This quotation appears on page iii of the appendix. For similar comments on durational variability see also Lisker (1958: 301) and Balasubramanian (1972: 241).

single segments are realized as fricatives and hence have no identifiable stop gap. Balasubramanian and Asher (1984) did conclude, however, that the double representation of word-internal obstruents was supported by the similarity of their durations to geminate nasals and laterals, and homorganic nasal and plosive sequences, for which a double phonological representation is uncontroversial.

5.2.3. Correlates of gemination cross-linguistically

The claim that duration may not be the primary phonetic correlate of gemination (contrary to the expectations generated by many phonological representations[11]), finds ample support in the phonetics literature. Acoustic analyses of geminates frequently implicate duration, which may be a sufficient cue on its own,[12] but a wide range of other phonetic correlates have also been noted, including differences in peak amplitude, fundamental frequency, and the quality of adjacent vowels. Indeed, the specific claim that the lenition of single segments plays a crucial role in the single/geminate distinction is not without parallel: for instance, Cohn, Ham and Podesva (1999) describe a similar phenomenon in the Indonesian language Toba Batak. According to Louali and Maddieson (1999), geminates in some Berber dialects have undergone a spirantizing process, which has given rise to a fricative/plosive distinction, where historically the difference was purely durational. The purpose of their investigation was to establish whether or not the durational difference, which is now functionally redundant in these dialects, is being maintained: comparison with non-spirantizing dialects suggests that it is. Investigation into the phonetic nature of geminates thus indicates that it may be misguided to focus upon one correlate to the exclusion of all others, rather that different acoustic properties may combine to cue the distinction.

Furthermore, work by Local and Simpson (1988, 1999) claims that the phonetic correlates of gemination may not be confined to one point in the utterance, but involve long domain effects. The detailed impressionistic transcriptions upon which their claims are based are of particular interest, in that their data is taken from Malayalam, which is closely related to Tamil. Their first study examined transitive/intransitive pairs of verbs distinguished by gemination, as in the pair of past forms given in (22).

(22) *aaṭi* 'swing' (intrans.) *aaṭṭi* 'swing' (trans.)[13]

The consonants themselves are described as differing in terms of voicing, the non-geminates being voiced and the geminates voiceless, and also degree of occlusion, the geminates being characterized by complete occlusion. However, Local and Simpson claim that the differences extend far beyond the geminate consonants themselves, and that the gemination alternation involves:

> *'a complex combination of temporal, articulatory and phonatory features extending over a number of syllables'* (Local & Simpson 1999: 595).

They observed similar long domain effects in nominal pairs containing intervocalic sonorants lexically distinguished by gemination. In addition to varying duration, which affects not only the sonorants but also vocalic portions of the nouns, they found differences in consonantal resonances and the quality of vowels, illustrated in formant plots (Local & Simpson 1999). Whilst many current theories of phonology do not mesh easily with this kind of phonetic analysis, there is a body of earlier work with a much broader concept of gemination. This

[11] Quantity is the only aspect of gemination to be depicted in autosegmental phonology: geminates are doubly linked, having one unit at the segmental or melodic level, associated with either two timing units or two moras, just like distinctively long vowels. See, for instance, Hyman (1985), McCarthy and Prince (1986, 1988) and Hayes (1989): all are proponents of a bimoraic surface representation for geminates, although they disagree about the underlying forms.

[12] For example, Lahiri and Hankamer (1988) claim that duration of stop closures is the critical acoustic feature differentiating geminates from non-geminates in both Turkish and Bengali.

[13] Tamil has a very similar phenomenon, exemplified in (13) above.

draws on the ideas of Firth and defines gemination as a prosody, in opposition to the prosody of non-gemination, with varied phonetic exponents that may extend beyond the segment in question. Palmer's analysis of gemination in Tigrinya plural forms (Palmer 1957), for example, defines gemination as *'a prosody of the entire word'*, with phonetic exponents including consonantal duration, tenseness and position of articulation, and also vowel quality. There is even an application of the theory to the Malayalam transitive/intransitive alternations (Nayar 1972), although this recognizes only punctual differences. Parallel long domain effects to those observed by Local and Simpson may well also exist in Tamil: this study, however, confines itself to the segments in question, investigating the relative contributions made by duration, voicing and frication to the single vs. geminate distinction.

5.3. Investigation of word-initial obstruents

5.3.1. Data collection

5.3.1.1. Subject profile

The data were obtained by recording a native Tamil speaker, a man in his late twenties, in a sound-proofed room in the Phonetics Laboratory of Oxford University. Like the majority of subjects interviewed for the questionnaire of chapter 3, he had grown up in the union territory of Pondicherry, although he had subsequently studied in Madras, and then in the U.K. He had received no tuition in written Tamil since the age of six, and hence disclaimed command of the written language, although entirely fluent in the spoken variety. Any phonetic correlates of gemination found in his speech were therefore unlikely to be spelling pronunciations.

5.3.1.2. Design and presentation of sentences

The test data comprised 59 different sentences, each containing between one and four occurrences of the segments of interest, i.e. those in word-initial position following a voiced segment. These are listed in full with glosses and translations in appendix G: the segments that were measured are emboldened, and the Tamil convention of representing geminates orthographically by writing an additional copy of the stop symbol at the end of the preceding word has been followed. Brackets indicate that the subject varied between repetitions over whether he included or omitted the bracketed word. The sentences were designed to contain a wide range of environments, including most of those listed by Arden. All four of the obstruent segments that can appear in word-initial position are represented, in varying degree, as table 1 illustrates. Due to the marked imbalance in the ratio of <c> to <cc> tokens, these were mostly excluded from the results reported in section 5.3.2. Moreover, they proved to be markedly different in their acoustic properties from the other segment types, showing the steady-state structure of a fricative rather than a stop: it was only possible to identify a stop gap in 10 of the tokens.

Table 1. Numbers of segment types in word-initial position

Segment representation	Number	Segment representation	Number
<c>	44	<cc>	5
<k>	55	<kk>	56
<p>	113	<pp>	149
<t>	25	<tt>	19

As noted in section 5.2.2, there is continuing disagreement amongst Tamil scholars over the exact set of environments in which word-initial gemination occurs. The description given in Christdas (1987: 139–140) has been followed wherever applicable in appendix G, so doubling is assumed to occur after a case-marked noun on a postposition, verb or noun

phrase, but not after a dative subject or a non-case-marked noun.[14] In cases where Christdas gives no direct judgment either way, other writers have been consulted. For instance, in sentences including certain adverbs and participle forms, Arden's conventions, as described in (16)–(21) above, have been used, and Balasubramanian's comment that doubling does not occur after *oru* 'one' (Balasubramanian 1972: 256) has also been taken into account.

In order to avoid any possibility of spelling pronunciations, the test sentences were not presented in written form. Rather, equivalent English sentences were given to the subject through headphones, and he was asked to translate into colloquial Tamil. Before the recording was made there was one complete run-through of the sentences to ensure that he was comfortable with the translations. The recorded data contained five repetitions of each sentence, which were presented to the subject in randomized order. A few cases where the subject had obviously paused or stumbled were eliminated from the data set, leaving a total of 466 word-initial tokens.

5.3.1.3. Parameters measured

The data were digitized at a rate of 16,000 samples per second (16 bit resolution), and the following set of three measurements were made for each word-initial segment:

a) duration

The total duration of each segment was measured, from the point where the broad band of voicing energy ended to where the following vowel began. In all cases where a stop gap could be identified the total duration was recorded in two parts, with the stop gap forming the first measurement. This facilitated comparison with Balasubramanian's measurements of closure duration, and also allowed the ratio of stop gap to total duration to be calculated (cases where no distinct stop gap could be identified were assigned a ratio of zero). Such a measure reflects differences in the internal temporal structure of the segment, and is unlikely to be greatly affected by variation in speaking rate.

b) minimum RMS amplitude

The point of minimum amplitude during the duration of the segment was identified, and the RMS (root mean square) value recorded. A voiceless stop would have a minimum of zero, during the stop gap, whilst frication (and also continuous voicing) would raise the RMS amplitude level.

c) probability of voicing

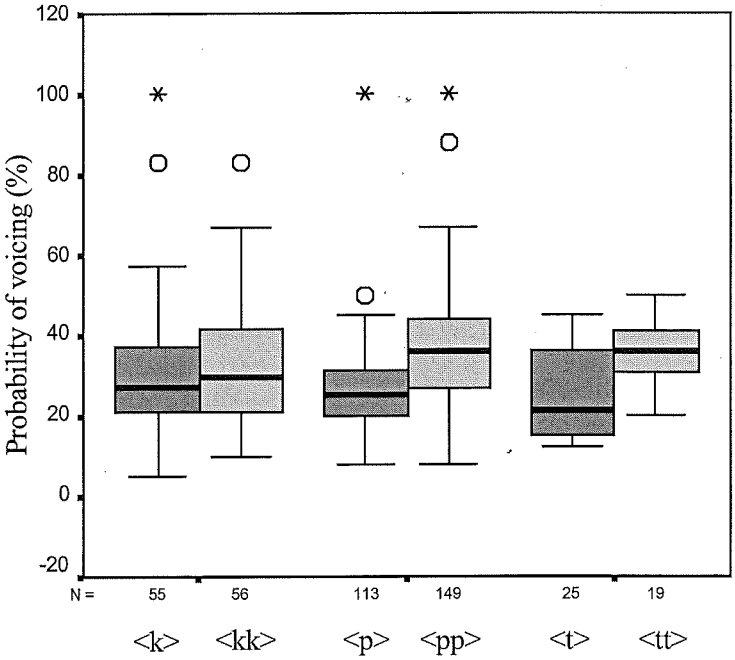
Since the minimum RMS amplitude value distinguishes only between fully voiced and voiceless stops, a more direct measure of voicing was taken. The probability of voicing was therefore calculated for the segment as a whole, i.e. the percentage of frames (each of ten milliseconds duration) containing at least eighty per cent voicing.

5.3.2. Results

If the orthographic distinction between single and geminate segments in word-initial position is matched by the kind of phonetic differences proposed for word-internal segments, this should be reflected in statistically significant differences consistent with lenition, both in the data set as a whole and also the individual segment types. In particular, higher values of both voicing and minimum RMS would be expected in the single segments, as compared to the geminates. Neither are found in the data: an analysis of variance of the minimum RMS values shows the distinction between single and geminate to be a factor of no significance ($p < .462$). In contrast, the single vs. geminate difference is a highly significant factor ($p < .0005$) in the distribution of voicing values, but it is the segments represented orthographically as single that have the lower values overall, contrary to expectation.

[14] Note, however, that Christdas does qualify this by remarking that '*not all case-inflected nouns trigger I(nitial) G(emination)*' (1987: 139), and gives the specific example of the locative as an exception.

Figure 1. Boxplots of the probability of voicing for single and geminate obstruents



The boxplots in figure 1 show the probability of voicing values for each of the obstruents, with the boxplot on the left of each pair corresponding to the segments represented orthographically as single and those on the right to the orthographic geminates. The non-parametric Kruskal-Wallis test on the voicing data showed the segment type to be significant ($p < .01$), and so the Mann-Whitney test was performed on each segment type individually, with the data grouped according to whether the segments are orthographically single or geminate. The results were somewhat inconsistent: the difference between <k> and <kk> is not significant, whereas that between <t> and <tt> is significant ($p < .035$ and $p < .04$ respectively), and that between <p> and <pp> is highly significant ($p < .0005$).

Since no evidence of the lenition of single segments was found, the duration data were examined to see if the single vs. geminate distinction was instead cued by durational differences. The durations of the stop gaps (where these were identifiable) were first compared with the figures given by Balasubramanian (1972: 258). The range of values for each of the segment types, single and geminate, are set out below in table 2, together with the number of tokens and the standard deviation of the distribution. Mean values are included to allow comparison with Balasubramanian's measurements, which are given underneath in bold, where applicable.

Table 2. Stop gap durations of word-initial obstruents

Segment representation	Range	Mean	Number	Standard deviation
<k>	0–94 **50–100**	35 **64**	55 **9**	26
<kk>	0–90 **60–72**	35 **66**	56 **3**	20
<p>	0–69 **45–120**	26 **79**	113 **30**	13
<pp>	0–55 **30–140**	24 **96**	149 **18**	13
<t>	12–45	23	25	10
<tt>	14–76 **40–105**	24 **82**	19 **18**	14

The general pattern found by Balasubramanian was that the word-initial geminates had longer stop gaps. They were, however, only marginally longer, and there was no statistical confirmation that the difference is significant. The measurements in this study, which covered a greater number of tokens than Balasubramanian's data, show only negligible differences in the mean values. The absence of statistical significance was confirmed by applying the Mann-Whitney test to the whole data set: the single vs. geminate distinction is not a significant factor ($p < .771$). The same is true for <p> vs. <pp>, and <t> vs. <tt>: only in <k> vs. <kk> is there a significant difference in the duration of the stop gap ($p < .015$ on an analysis of variance), with the segments represented orthographically as single proving longer, i.e. the reverse trend to that seen in Balasubramanian's data.

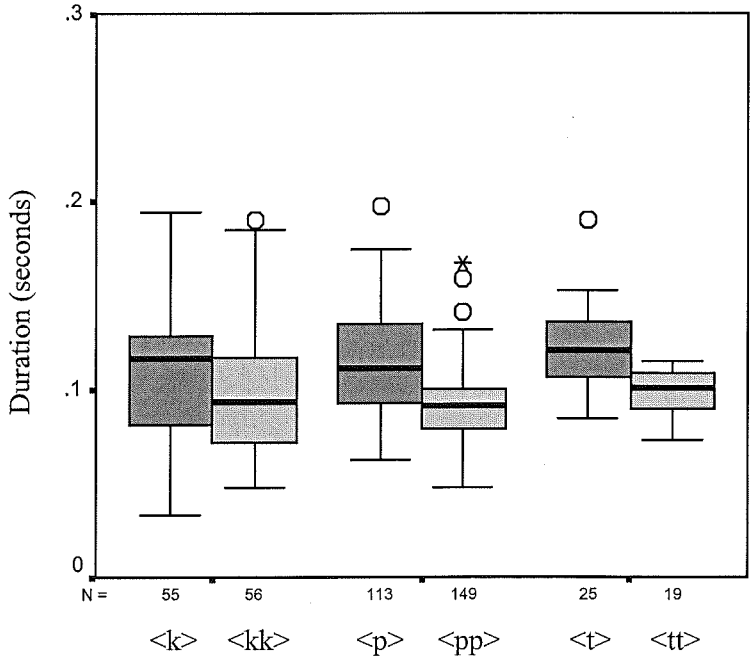
The total durations of the segments were analyzed in similar fashion: table 3 below follows a similar format to that of table 2, but the duration measurements relate to the whole segment, rather than just the stop gap.

Table 3. Total durations for word-initial obstruents

Segment representation	Range	Mean	Number	Standard deviation
<c>	74–200	124	44	31
<cc>	89–115	99	5	10
<k>	33–194	105	55	38
<kk>	48–190	98	56	34
<p>	63–198	115	113	28
<pp>	48–167	91	149	18
<t>	84–190	123	25	22
<tt>	73–115	98	19	12

Examination of the mean values reveals a consistent trend towards orthographically single segments being longer than their geminate counterparts. This is illustrated in figure 2, which shows boxplots of the total duration values for three segment types, with the single obstruents on the left side of each cluster and the geminates on the right.

Figure 2. Boxplots of the total duration values of word-initial obstruents

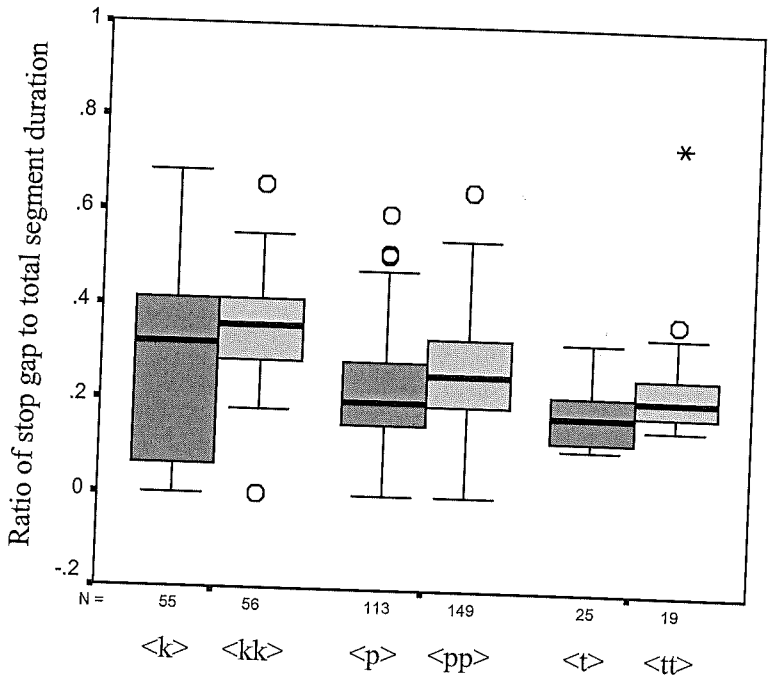


Applying the Mann-Whitney test to the data set as a whole showed the single vs. geminate distinction to be highly significant ($p < .0005$); segment type also proved to be significant on the Kruskal-Wallis test ($p < .013$). Consideration of individual segment types confirmed that <p> and <t> are significantly longer than <pp> and <tt> ($p < .0005$ for both), but interestingly the difference in overall duration between <k> and <kk> is not statistically significant ($p < .162$), despite the stop gap being significantly longer.

This raises the possibility that the orthographic differentiation between single and geminate may correlate with differences in the internal acoustic structure of the segments. The ratios of stop gap to total segment duration were therefore investigated. These ranged from zero in cases where no distinct stop gap could be identified to .761, approximately three-quarters of the total duration. A univariate analysis of the variance of ratio values showed the single vs. geminate distinction to be a highly significant factor ($p < .0005$), although the squared correlation was not high ($R^2 = .019$). Segment type also had an effect: an analysis of the variance of ratios under a log transformation showed it to be a highly significant factor ($p < .0005$). Examination of individual segment types revealed a consistent pattern: the stop gap accounts for a higher proportion of the total duration in segments represented orthographically as geminate than in their single counterparts. The extent of the difference, however, is relatively minor, as the boxplots in figure 3 demonstrate.

Statistical testing for individual segments produced a significant result only for <t> vs. <tt>, and even here the level of significance was not high ($p < .036$ on an analysis of variance). There is some evidence, therefore, of a tendency for the single vs. geminate distinction to be reflected in the internal acoustic structure of segments, but it is not strong enough to produce consistently significant results.

Figure 3. Boxplots of the ratio of stop gap to total segment duration



5.3.3. Conclusion

The interactions of three different (though not entirely independent) phonetic properties with the single vs. geminate distinction were investigated. The possibility that differences in the internal durational structure of the segments are pertinent was also tested using the ratio of stop gap duration to total segment. Probability of voicing and minimum RMS were chosen as indicators of lenition, both being raised in lenited realizations. There is general agreement that this distinguishes word-internal single stops from the corresponding

geminates, and it has also been connected with word-initial cases by Balasubramanian. Minimum RMS, however, proved to be of no significance in differentiating between orthographically single and geminate segments, whilst the significant results associated with voicing involved higher levels in the geminates than the singles, contrary to expectation. Likewise, in those cases where duration was found to be significantly affected, the geminates proved to be shorter than the single segments. This is again the reverse pattern to that expected, and contrasts with the results of the next section, where geminates are found to be consistently significantly longer than single segments word-internally.

None of the properties investigated, therefore, support the notion that the orthographic distinction between single and geminate in word-initial position is phonetically relevant in colloquial Tamil. It is in principle possible that, although the properties considered here were not strong enough to cue the single vs. geminate distinction on their own, they may do so in combination with other, longer domain effects. Certainly the work by Local and Simpson on Malayalam (discussed in section 5.2.3 above) suggests that the phonetic cues may extend far beyond those considered here, including, for instance, differences in the quality of the adjacent vowels. Nevertheless, given the consensus that voicing and frication are crucial in differentiating word-internal single stops from geminates, and given the results of the next section implicating duration, it is still reasonable to expect that these would be among the primary correlates of a word-initial distinction. The absence of any positive results strongly suggests that, for this speaker at least, there is no such distinction.

There is some conflict, therefore, with the findings of Balasubramanian, who maintained that orthographic geminates are always produced as voiceless stops (see section 5.2.2). This is not true of the data examined here: orthographic geminates have been shown to have higher levels of voicing, and may even be fully voiced. Interspeaker variation is one possible explanation, since in each case the data come from a single speaker, and there are considerable differences in their linguistic backgrounds. Balasubramanian gathered most of his data by recording himself and he classified his own speech as *'non-literary, colloquial, brahmin (Iyer) Tamil'* (Balasubramanian 1972: 27). The speaker recorded in this study, in contrast, spoke a non-brahmin variety, which he described as Madras Tamil. Moreover, his lack of education in the written language may also be relevant. It is conceivable that speakers such as Balasubramanian, who are familiar with the orthographic conventions, reflect these in their speech, but that those who are not educated in the written language lack the distinction altogether.

Another possibility is that the differences between these data and Balasubramanian's, which was recorded almost thirty years ago, may reflect a historical change, with the earlier data showing a distinction that has now been lost. This would fit with the observations made by Britto and Anton and Hellmann (see section 5.2.2 above) that the external sandhi rules are increasingly being ignored, particularly in informal Tamil. If this is true, the environments in which the doubling occurs may have been restricted, but it should still be possible to divide the data into the two categories on acoustic grounds. The measurements taken here, however, offer no obvious basis for any differentiation, and it seems reasonable to conclude that the modern colloquial Tamil spoken by this subject lacks a single vs. geminate distinction in word-initial position.

5.4. Investigation of word-internal /k/ and /kk/

5.4.1. Design of data set

A second study was conducted, looking specifically at the velar obstruents, and comparing them with instances of echo /k/, to see what light this could throw on the structure of the echo words. Duration measurements were taken from the set of sentences analyzed in

chapter 4 and listed in appendix F.[15] As described in section 4.3.1, three speakers (two male, one female) were recorded, repeating each sentence five times, and so the data set contained up to fifteen tokens of each word. These are given in table 4 below, with the segments of interest emboldened; the numbers in brackets indicate the sentence containing each word in appendix F. The set is divided equally between instances of /k/ in the four different environments listed at the top: word-internal single and geminate, word-initial segments and echo /k/. Given that the study reported in the previous section found no consistent and significant differences in word-initial position, no distinction is made between those initial segments that are represented by single symbols and those represented by double symbols.

Table 4. Instances of /k/ and /kk/

Internal /k/	**Internal /kk/**	**Initial /k/**	**Echo /k/**
*ku**k**ai* (22)	*uʈai**kk**a* (19)	***k**ari* (30)	*ki.ɻa**k**aana* (15)
*a.ɻa**k**aana* (54)	*iru**kk**utaa* (27)	***k**elippaay* (48)	*kiʈutteen-ŋŋu* (23)
*miʆa**k**u* (59)	*paa**kk**umaa* (31)	*ku**k**utteen-ŋŋu* (58)	***k**iccai* (24)
*a.ɻa**k**aana* (49)	*tai**kk**iraa* (72)	*ki**k**iyatu* (14)	***k**imaarukku* (63)
*moo**k**an* (60)	*ena**kk**u* (3)	***k**aay* (64)	***k**ilaiyai* (6)
*kiʆa**k**u* (59)	*ninai**kk**aatee* (15)	***k**ukai* (22)	***k**illa* (28)
*ki.ɻa**k**aana* (15)	*kumaaru**kk**u* (58)	*kiʆippiʆʆai* (79)	*kiʆa**k**u* (59)
*a.ɻa**k**aana* (15)	*piʈi**kk**aatu* (3)	*kiŋaru* (16)	***k**iimpu* (3)

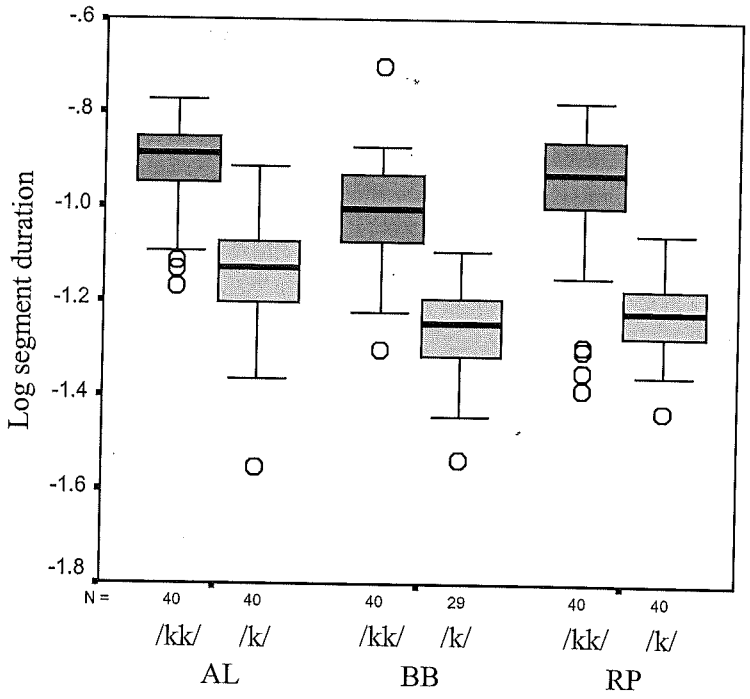
A repeated measures analysis was performed on the tokens from the first five rows of the table, taking the four-way division reflected in the columns as a within-subjects factor. The Greenhouse-Geisser adjusted figures (employed because sphericity could not be assumed according to Mauchly's test) showed no significant correlation between this factor and duration ($p < .191$), or log duration ($p < .136$). The following sections will therefore examine subsets of the data, seeking to establish differences between individual columns. Section 5.4.2 focusses on the first three columns, and section 5.5 investigates how the echo examples compare with these. It has already been seen in section 4.6.4 that expressives bear some structural similarities to the echo words: section 5.5.3 will therefore consider whether these are reflected in the nature of their velar obstruents.

5.4.2. Results

The first set of tests compared the total durations of /k/ and /kk/ in word-internal position, and revealed a clear difference, with the geminates being significantly longer than the single segments. Since the distribution of the duration values was slightly skewed to the right, a log transformation was performed on the data. Analysis of the variance of these values showed both speaker and the difference between single and geminate segments to be highly significant ($p < .0005$ in each case). There was no significant interaction between the two factors ($p < .802$) and the squared correlation of the model (R^2) was .575. The clustered boxplots in figure 4 show the differences in log duration between single and geminate for each speaker: all three of these differences prove to be highly significant. Duration, therefore, does seem to be a consistent and significant correlate of the single vs. geminate distinction, although this may well be subordinate to other more salient phonetic differences.

[15] Since these sentences were recorded in the field, in quiet but not sound-proofed conditions, the quality of the recordings was not sufficient to allow reliable voicing and amplitude measurements to be taken.

Figure 4. Boxplots of the log duration of word-internal /kk/ vs. /k/ for each speaker



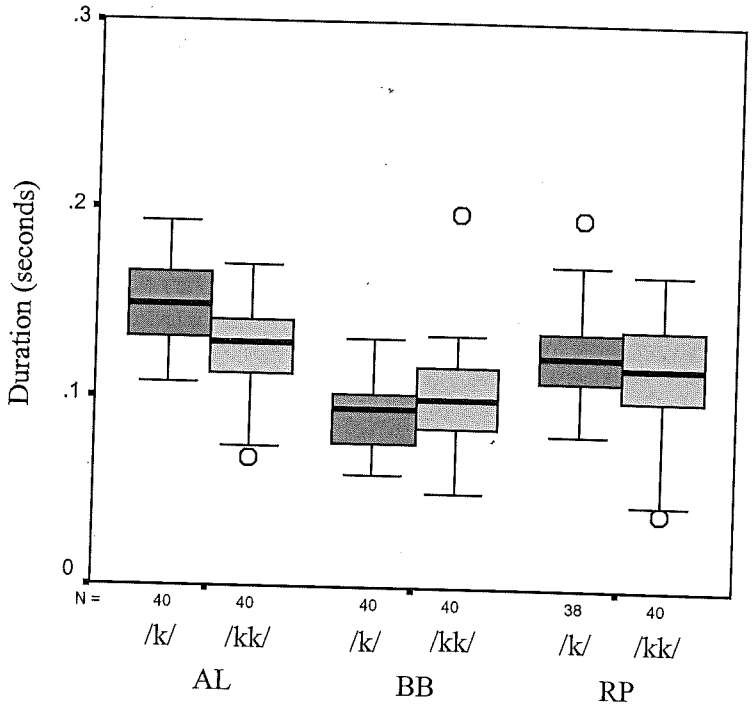
The internal geminates were next compared with single word-initial /k/, since both are described in the literature as voiceless stops. According to Balasubramanian, there is no significant durational distinction between the two. He measured the stop gaps of geminates in words produced in isolation and also word-initial single segments preceded by *oru* 'one', and concluded that:

> *'there is no connection between the orthographic doubling of the stop symbol and the duration of the stop element of the stops'* (Balasubramanian 1972: 254).

Balasubramanian and Asher (1984: 54), in contrast, state that the closure durations of the second group are always shorter, giving 117 milliseconds as the average for word-initial /k/ after *oru*, whereas average values for intervocalic /kk/ range from 124 to 190 milliseconds.

As the boxplots in figure 5 demonstrate, the data analyzed here show minimal differentiation between the two groups. Overall, indeed, the examples of word-initial /k/ prove marginally longer than word-internal /kk/, the opposite pattern to that observed by Balasubramanian and Asher. An analysis of variance showed the single vs. geminate distinction to be significant ($p < .007$), but speaker is also a highly significant factor ($p < .0005$), and there is significant interaction between them ($p < .0005$). Looking at the speakers on an individual basis, the single vs. geminate distinction proves to be significant only for speaker AL, so there are no good grounds for assuming this to be a consistent trend.

Figure 5. Boxplots of the duration of word-initial /k/ vs. word-internal /kk/ for each speaker



5.5. Echo expressions

A clear durational distinction has been established between word-initial /k/ and word-internal /kk/ on the one hand, and word-internal /k/ on the other. This section therefore considers the echo words listed in the fourth column of table 4, to see whether their durational characteristics align them with one group rather than the other. The relevant comparisons are with instances of word-initial /k/ and word-internal /k/: there is no reason to believe that a geminate representation of the echo segments is justified.[16] The results have some bearing on the question of the phonological structure assigned to the echo words. Evidence of vowel reduction in the initial syllables of the echoed portion (see section 4.6.2) indicated that this may be phonologically subordinate to the base, with the structure as in (23a), rather than (23b).

(23a) $[piccai\ kiccai]_{\omega}$ (b) $[piccai]_{\omega}\ [kiccai]_{\omega}$

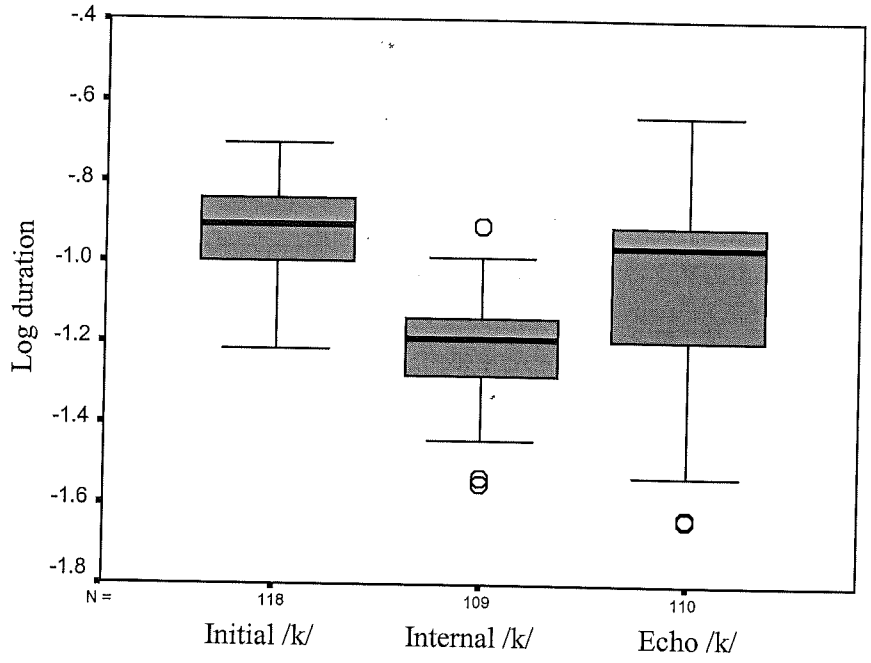
If (23a) is the correct representation, and the distribution of obstruent realizations is determined by position in the phonological word, then the echo /k/ is expected to show the characteristics of a word-internal /k/. If the structure represented in (23b) is correct, however, the echo /k/ should be like a word-initial /k/.

5.5.1. Results

The boxplots in figure 6 show the range of log duration values for initial /k/, internal /k/ and instances of echo /k/, and it is immediately striking that echo /k/ shows considerable variation in its duration. Speaker proves to be a highly significant factor in an analysis of the variance of log durations ($p < .0005$), and some interesting differences emerge when the three speakers are considered individually.

[16] Transcriptions of the echo expressions in the literature show only a single /k/, without exception.

Figure 6. Boxplots of the log duration of initial /k/, internal /k/ and echo /k/ for all speakers



For speaker BB, echo /k/ is significantly shorter than initial /k/ ($p < .0005$ in an analysis of variance taking /k/ type as the fixed factor), but its duration does not differ significantly from internal /k/. This is illustrated in figure 7 and suggests that echo /k/ may behave as if it were word-internal for this speaker, which would support the structure represented in (23a). Speaker RP, however, shows a rather different pattern, illustrated in figure 8.

Figure 7. Boxplots of the log duration of initial /k/, internal /k/ and echo /k/ for speaker BB

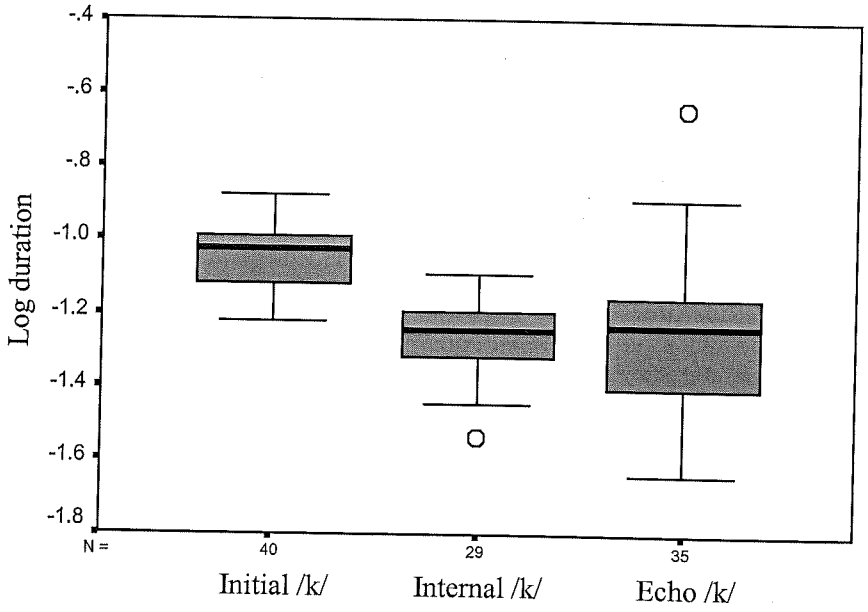
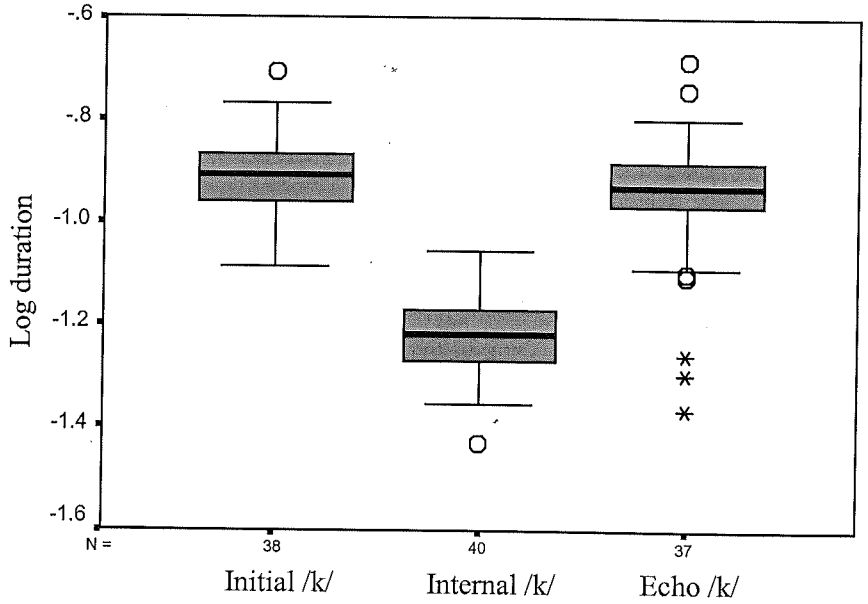


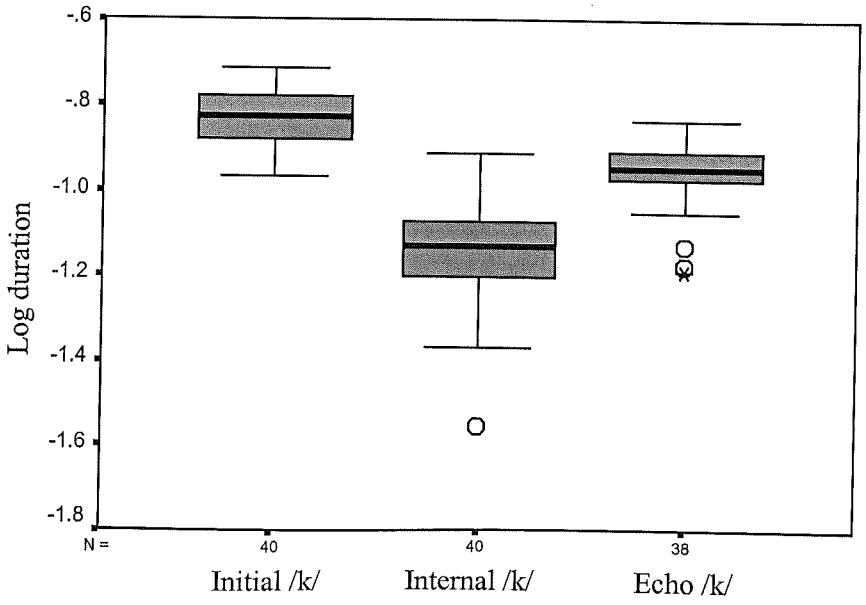
Figure 8. Boxplots of the log duration of initial /k/, internal /k/ and echo /k/ for speaker RP



Here there is no significant difference between initial /k/ and echo /k/, but the echo /k/ values are significantly longer than those of internal /k/ ($p < .0005$ on the non-parametric Mann-Whitney test). This would suggest that speaker RP was treating the echoed portion phonologically as a separate unit, as in (23b) above. He was certainly more prone to leaving short pauses before the echoed part than the other speakers. This may have been partly due to the fact that he was reading out the echo words from a written representation, rather than producing them spontaneously. As is conventional in the literature, the echo words were written with short spaces between the two halves, and this may have contributed to a pronunciation as two separate words, although it is unclear why it should have affected this speaker more than the other two.

Finally, speaker AL showed a slightly different pattern again: as figure 9 demonstrates, the duration of her tokens of echo /k/ fell between those of initial /k/ and internal /k/.

Figure 9. Boxplots of the log duration of initial /k/, internal /k/ and echo /k/ for speaker AL



Moreover, in both cases the distinction was highly significant ($p < .0005$ on analyses of variance comparing first initial /k/ and echo /k/, and then internal /k/ and echo /k/). Such data provide clear support for neither structure (23a) nor structure (23b). Overall, therefore, the results are very mixed, and do not constitute consistent evidence for either of the proposed phonological representations.

The difference between the two male speakers might suggest individual variation in the phonological structure of the echo words, with both (23a) and (23b) as possibilities. If this is the case, the nature of the echo /k/ should covary with the presence or absence of vowel reduction in the /i/ vowel. Specifically, it might be expected that there would be reduction in the /i/ tokens of speaker BB, but not RP. The data analyzed in section 4.6.2 were therefore re-examined, to see whether the significant differences between the /i/ tokens in echo words and bases were consistently present in all the speakers.

There was no evidence for the hypothesized distinction: highly significant durational differences were found in the /i/ vowels of all three speakers ($p < .0005$). AL and RP also had significantly lower first formant frequencies in the base /i/ vowels ($p < .003$ and $p < .021$ respectively), whilst BB had significantly higher second formant frequencies ($p < .006$). The vocalic properties investigated in chapter 4, therefore, do not support the notion that speakers can structure the echo words as either one or two phonological units, with speaker BB choosing the former option and RP the latter: instead, the evidence from all three speakers converges on a representation as a single phonological unit. It is possible, therefore, that RP's /k/ results do not reflect the echoed portion forming an independent word, and so beginning with an initial /k/, but rather a representation of the echo /k/ as a word-internal geminate, rather than single segment. The occurrence of pauses preceding the echoed portion rather counts against such an explanation, however. Given the variation found here, data from further speakers would clearly be needed to identify which, if any, represents the typical pattern of results.

5.5.2. Echo phrases

All the cases of echo /k/ investigated thus far involve single words being repeated. As established in section 3.3.2.5, however, Tamil allows short sequences of words to be repeated in some circumstances. The structure of these echo expressions is of particular interest: do they form single phonological units, and if so, of what type? Is the echoed part of the expression phonologically independent of the first half? The echo /k/ tokens from phrasal examples were compared with the /k/ of single words to see whether any significant differences in their durational properties could be found. The relevant examples are given in table 5, with the full phrases given on the left and corresponding single words, matched as closely as possible for phonetic environment, on the right.

Table 5. Phrasal examples

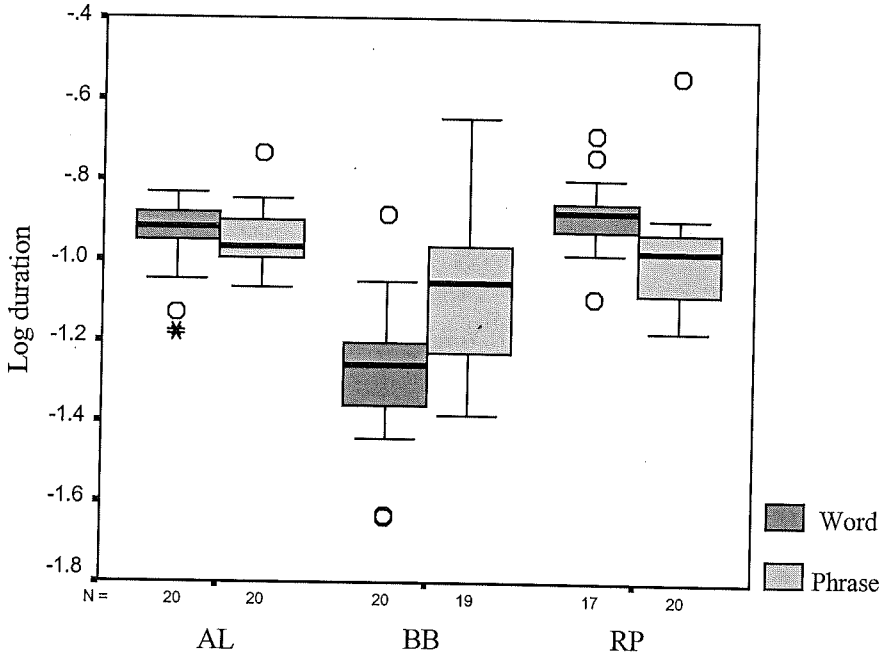
Phrasal echoes	**Single word echoes**
killa paiyan (28)	*kilaiyai* (6)
kiɻakaana paiyan-ŋŋu (49)	*kiɻakaana* (15)
kimaarukku kuʈutteen-ŋŋu (11)	*kimaarukku* (63)
kippaa varaar-ŋŋu (61)[17]	*kiimpu* (3)

As with the examples of echo /k/ analyzed in the previous section, speaker is a highly significant factor in an analysis of the variance of the duration values ($p < .0005$). The boxplots in figure 10 show that BB differs from the other speakers in having longer tokens of

[17] When presented to subjects as part of the questionnaire, this particular phrase produced mixed responses, with an average of 2.6, but only one definite rejection. No objections were raised by the three subjects who were asked to read it out for the recordings.

echo /k/ in the phrasal examples, and this difference proves to be significant ($p < .002$ on an analysis of variance). Speakers RP and AL, by contrast, have marginally shorter tokens of /k/ in the phrases than the words, but in neither case is the difference statistically significant.

Figure 10. Boxplots of the log durations of echo /k/ in echo words vs. echo phrases



The result for BB suggests that his echo /k/ tokens in phrases can be classified with the word-initial cases of /k/, whilst his echo /k/ tokens in single words behave as if they are word-internal. Presumably, therefore, once an echo expression contains more than a single word its phonological structure is such that the echo /k/ begins a new phonological word. As with BB, the significant difference between initial /k/ and echo /k/ found in AL's speech disappears when the echo /k/ belongs to a phrase. For speaker RP, however, echo /k/ tokens are indistinguishable from initial /k/ in terms of their length, whether they belong to an echo phrase or an echo word.

5.5.3. Comparison with expressives

Section 4.6.4 established that the vocalic properties of certain expressives pointed clearly to the kind of phonological structure exemplified in (24a), rather than (24b), and were therefore parallel to the echo expressions in this respect.

(24a) $[kiɻukiɻukkiraan]_{\omega}$ (b) $[kiɻu]_{\omega}$ $[kiɻukkiraan]_{\omega}$

This section examines the properties of the /k/ segments in each, comparing the word-initial examples with the word-internal tokens of /k/. The examples considered are listed in table 6: there is a considerable degree of overlap with table 14 in section 4.6.4, but the last word (*paɭapaɭakkum*) has been replaced by an example containing a long vowel (*kiiccukiiccuŋŋu*).

Table 6. Expressive examples

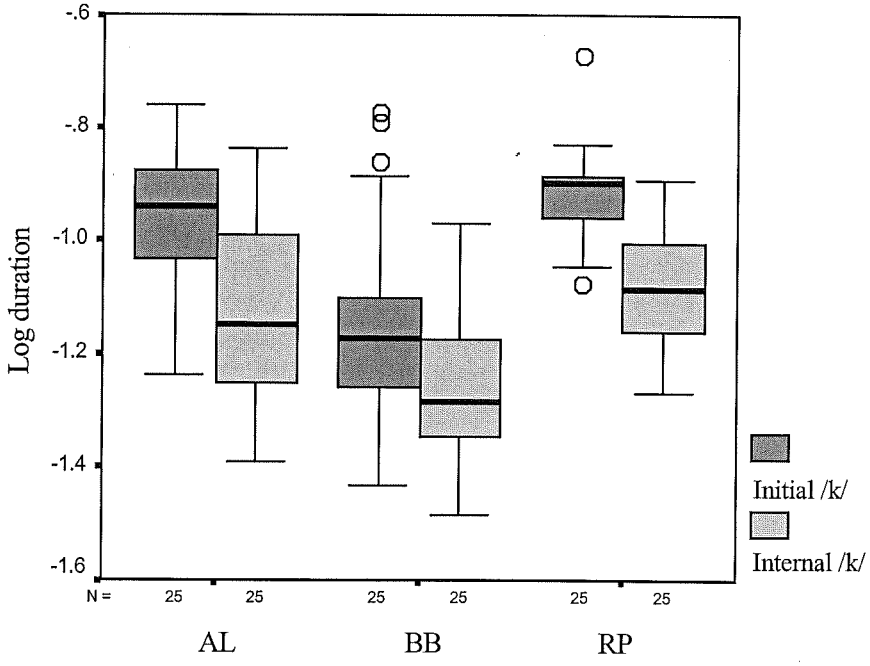
Example	**Gloss**
kirukiruŋŋu (26)	'giddiness'
kirukiruppu (76)	'giddiness'
kiɭukiɭuppaa (60)	'laughing loudly'
kiɻukiɻukkiraan (45)	'tremble.pres.3sm'
kiiccukiicuŋŋu (7)	'chirping'

The results are entirely consistent with the second /k/ of each word behaving as a single word-internal segment, contrary to the claim made by Emeneau and Hart that the copy of the initial consonant:

> *'is not subject to any difference of allophony that a single stop might show when intervocalic within a word or a compound.'* (Emeneau & Hart 1993: 78).

In these data the medial /k/ tokens are shorter overall than the initial /k/ tokens, as the boxplots in 11 clearly demonstrate.

Figure 11. Boxplots of the log duration of initial and internal /k/ in expressive words



This was confirmed statistically by univariate analyses of the variance of duration: position within the word and speaker proved highly significant ($p < .0005$ for both), although not their interaction.[18] Moreover, consideration of individuals revealed significant distinctions in the log duration values of all three speakers.

5.6. Conclusion

As section 5.4 established, duration is a consistent and significant correlate (though not necessarily the primary one) of both the single vs. geminate distinction and also the difference between word-initial and word-internal /k/. It can therefore be used as a diagnostic of phonological structure, and has been straightforwardly applied in this way to the expressives analyzed in section 5.5.3. The results of so doing are entirely in line with the findings of section 4.6.4, and point to the expressives behaving as single phonological units, with the reduplicated section phonologically subordinate to the first half. Analysis of the same measurements for the echo expressions produced somewhat inconclusive results, due to the wide variation between individual speakers. There was, however, some convergence with the findings of chapter 4, indicating that echo expressions based on one lexical item have the structure of a single phonological word, whereas echo phrases have a more complex structure. The final chapter considers how these findings can best be accommodated by reduplicative theory.

[18] The interaction was shown not to be significant ($p < .463$) on an analysis of the variance of duration, taking both speaker and position as fixed factors. This just failed the Levene test of homogeneity of variance ($p < .047$).

Chapter 6: Discussion and conclusions

6.1. Introduction

This final chapter draws together the findings of chapters 3 to 5, and explores some of the theoretical issues posed by the Tamil data. The first section summarizes what has been established concerning both the restrictions on forming echo expressions and their constituency. The different kinds of linguistic information involved are considered, particularly for the cases of phrasal echoing. The theoretical implications of such data for the structure of the grammar are discussed in section 6.3, comparing the Tamil examples with other morphological processes that are sensitive to syntactic structure. Finally, section 6.4 presents a comprehensive analysis of the Tamil data, and discusses issues that might be pursued in future research.

6.2. Discussion of findings

The colloquial nature of echo words is widely recognized for all the languages in which they occur, and they are consequently limited to conversational contexts. In the responses to the questionnaire described in chapter 3 this was reflected by subjects' rejection of echo words involving the formal genitive marker *-in* (see section 3.3.2.3.1), presumably because features of formal and informal speech are not expected to co-occur. In addition to the overall speech register, there are also some quite specific syntactic restrictions on the occurrence of echo words. Negative contexts are favoured, a tendency reported in the literature and confirmed by responses to the questionnaire (see section 3.3.2.1), and phrasal examples are mostly confined to subordinate clauses followed by a negative imperative. A complete formalization of echoing would need to incorporate such co-occurrence restrictions, although they are not the primary concern of this study.

Another type of restriction concerns the category of the base, which appears to be limited to words that have lexical content, possibly as a consequence of the semantic properties associated with echoing. Unlike some of the other Indian languages, however, Tamil seems not to place further restrictions upon the kinds of lexical categories involved (see section 3.3.2.2): nouns, verbs, adverbs, adjectives and postpositions can all appear in the base, although the latter two categories tend to be accompanied by further material.

The constituency of the echo expressions was investigated in detail, and certain generalizations emerged (see section 3.4). As noted in chapter 3, there was some variation in the responses of different individuals, so the generalizations have the status of overall tendencies, describing common patterns, rather than any one idiolect. In each case where subjects were required to consider different forms with the same meaning, the average results indicated that only one of the options was judged acceptable: scrutiny of individual speaker's responses (see appendix E), however, reveals many instances where an individual judged two options to be equally valid. This is in line with the descriptions of echoing in Kannada and Bengali: Lidz (2000), for example, claims that the pairs of echo forms given in examples (85) and (86) of section 2.4.1.2 are equally valid, and not semantically differentiated. Fitzpatrick-Cole (1994) notes several areas of flexibility in the formation of Bengali echo expressions, including binominal compounds, which can take either the whole compound or just the second element as the base (see example 123 of section 2.4.2.3). Neither Lidz nor Fitzpatrick-Cole, however, give any indication of whether this represents interspeaker differences or variation within a single idiolect. Fieldwork conducted on Hindi did not reveal many instances where a single meaning could be expressed by multiple echo expressions, but in those cases where it was possible (echoed compounds and complex predicates) more than one option was judged to be acceptable (see example 122 of section 2.4.2.3). A descriptively adequate account of echoing should therefore recognize the possibility of multiple echo forms corresponding to a single meaning, although further research would be required to establish any factors conditioning their distribution. In a constraint-based framework this could be

achieved by formulating constraints that were not sufficiently restrictive to choose between the different possibilities. An analysis of echoing as a rule or process, however, would require a disjunct formulation, with appropriate conditioning or probabilistic weighting associated with the different options.

Despite the variation admitted by individual speakers, the overall tendencies are clear, and it is on these that attention will be focussed. As in other analyses of reduplication, a primary concern is to establish how base and reduplicant are to be defined, so the evidence for different types of constituency will be reviewed first, considering both the base itself, and the echo form as a whole. The morphological make-up of the base for echoing shows considerable variation, both in terms of the number of morphemes involved and their relation to one another. As noted in section 3.4, however, a significant proportion of cases is bimorphemic. Indeed, a putative constraint requiring that the base contain two morphemes was found to be consistent with the behaviour of postpositions and case markers (especially the exceptional treatment of the ablative), verbal forms, compounds and even adjective and noun sequences. There are, however, clear counter-examples involving both monomorphemic and polymorphemic bases, so neither an inviolable bimorphemic constraint nor a process that consistently identifies and copies a sequence of two morphemes can account for all the data.

A further possibility is that the base coincides with a specific complex morphological constituent, rather than being defined simply by the number of morphemes. Unlike Kannada, where, according to Lidz (2000: 158), echo reduplication can apparently break a word at any of its morpheme boundaries, there seems to be a restriction against echoing subparts of words in Tamil. The 'word' is, of course, notoriously difficult to define, but a working hypothesis for Tamil is that the morphological word constitutes a lexical root followed by all associated suffixes or markers. In many cases, this is a sufficient definition of the base, accounting, for instance, for the general tendency to include nominal case markers within the base. It also explains why **caappiṭa kiippiṭale* is not an acceptable echo expression, even though *caappiṭa kiippiṭa* is (see section 3.3.2.3.3): the negative marker *-le* belongs to the same morphological word as the infinitive in *caappiṭale*, and so cannot be excluded from the base, even though the infinitive is a perfectly acceptable base when it forms an independent morphological word.

There are, however, two significant groups of exceptions: the first includes many of the examples that the putative bimorphemic constraint seemed to explain, such as the unusual behaviour of the ablative case marker. Compounds also pose a problem for this kind of account, since they contain two lexical roots and so comprise two morphological words. Echoing of a single element only would therefore be expected, and is found in cases where the first element is an oblique stem, e.g. *tooṭṭattuppuuvai kiivai*, but not in the other endocentric compounds, which are echoed as a whole, e.g. *maamaram kiimaram* (see section 3.3.2.4). The other set of exceptions also involve the base exceeding a single morphological word (or indeed any definition of the word), and contain phrases headed by lexical constituents.

As argued in section 3.4, syntactic constituency crucially limits the base of echo phrases to material dominated by a lexical maximal projection. Attempts to extend a syntactic definition to the non-phrasal cases fail, however, in cases like *tooṭṭattuppuuvai kiivai*, where the base for echoing does not form a whole syntactic constituent. The readiness of several subjects to accept examples where case markers were excluded from the base also counts against a purely syntactic analysis. Echoing therefore seems to have access both to sublexical information about morphological structure and supralexical syntactic constituency, which raises theoretical issues addressed in the next section.

A further possibility is that the morphological and syntactic restrictions discussed thus far are not of any explanatory value in themselves, but rather represent a by-product of restrictions based on prosody. As noted in section 2.4.3, this is said to be the situation in Bengali by Fitzpatrick-Cole (1996: 330), who claims that both the base and the echo

expression as a whole are prosodically limited: the base being minimally a phonological stem and the whole expression maximally a phonological phrase. Her analysis takes prosodic lexical phonology as its framework, and employs a particular conception of the prosodic hierarchy, which owes much to Inkelas (1990). This recognizes two parts to the hierarchy, both shown in (1), a metrical hierarchy which extends from the metrical word down to the mora, and a prosodic domain hierarchy extending from the utterance level down to the phonological stem.

(1)

Metrical Hierarchy		Prosodic Domain Hierarchy	
W	Metrical Word	U	Utterance
|		|	
F	Foot	I	Intonational Phrase
|		|	
σ	Syllable	ϕ	Phonological Phrase
|		|	
μ	Mora	ω	Phonological Word
		|	
		ψ	Phonological Stem

At the lexical level, therefore, there is some overlap between the two hierarchies, with the possibility of divergence between the metrical structure and the constituents that form phonological rule domains.[1]

The two constraints proposed by Fitzpatrick-Cole are only bounds, a lower bound on the constituency of the base and an upper bound on the constituency of echo phrases, so echo phrases are contained within the phonological phrase, but do not necessarily exhaust it. Indeed there are cases where the base is said to correspond to neither a syntactic nor a prosodic constituent, as in example (2) (Fitzpatrick-Cole 1996: 334).

(2) *k^{h}ub* [*patla šari*] [*ṭatla šari*] 'sari that is very thin and so forth'

Fitzpatrick-Cole's grounds for basing her account on prosody rather than syntax concern cases such as verb phrases, which cannot be echoed, even though they form a syntactic constituent. This is attributed by Fitzpatrick-Cole to an obligatory phonological phrase boundary preceding verbs in Bengali, a principle that she adopts from the account of Bengali prosodic structure by Hayes and Lahiri (1991). Even her prosodic analysis, however, lacks predictive power: it offers some constraints but does not specify precisely which echo form or forms correspond to a particular meaning. The Hindi echo expressions, by contrast, are amenable to a systematic prosodic characterization, as argued in section 2.4.3, since the base can be identified in each case as a prosodic word.

Is a prosodic analysis of echoing workable for Tamil? The main difficulty in assessing this is the lack of research on the prosody of Tamil,[2] (or indeed any of the Dravidian languages). There is, for instance, nothing comparable to the work of Hayes and Lahiri on Bengali, which uses intonational patterns and segmental rules to analyze the formation and restructuring of phonological phrases. The basic principles of Tamil prosody have yet to be firmly established, as illustrated by the review of work on stress in section 4.2. The experimental work reported in that chapter found significant durational and qualitative differences correlated with syllable position in both monophthongs and diphthongs, which were interpreted as evidence of stress on the initial syllable. Application of these results to the echo expressions revealed only one identifiable stress in non-phrasal examples, suggesting that the expression as a whole forms a single phonological word (see section 4.6.2).

[1] The phonological stem (or PStem) is a sublexical phonological rule domain, corresponding to one of the cycles of lexical phonology. It features in a number of other analyses of reduplication (some of which have been mentioned in chapter 1), such as Downing (1994, 1999), Myers and Carleton (1996) and Hyman and Mtenje (1999), all on Bantu languages, and also Itô (1990) on Japanese.

[2] Ravisankar (1994) is the most notable exception, but his analysis is geared towards the classification of suprasegmental phonemes, rather than hierarchical prosodic structure. Marthandan (1983: 318–339) also offers a brief descriptive account of Tamil intonation.

Moreover, this structure proves to be directly parallel to that of endocentric compounds and expressives (see sections 4.6.3 and 4.6.4).

Further experimental confirmation of the proposed phonological structures emerged from the investigation of word-initial and word-internal /k/ and /kk/ in chapter 5. Their various phonetic properties were analyzed in sections 5.3 and 5.4, and the relevant /k/ segments in the echo words and expressives compared to establish their prosodic position. Although there was considerable variation amongst individual speakers for the echo words, the results were generally consistent with a representation as a single phonological word. In phrasal examples, however, the echo /k/ segment had the phonetic characteristics of a word-initial rather than word-internal segment, suggesting that there is a phonological word boundary before the echoed portion.

It is not possible, therefore, to identify every echo expression with a single prosodic constituent: there is at least a two-fold division between non-phrasal examples, which form a single prosodic word, and phrasal examples. It may be that phrasal examples correspond consistently to some unit higher up the prosodic hierarchy, such as the phonological phrase. However, the absence of any independent criteria for establishing prosodic structure in Tamil above the level of the word renders this hypothesis effectively untestable in the current state of affairs. Moreover, higher level prosodic constituents are at least partly parasitic upon syntactic structure, and so in many cases prosodic and syntactic constituency coincide. Instances in which the two are known to diverge would therefore be needed to determine which is relevant to the echo phrases.

The acoustic evidence clearly points to the non-phrasal echo expressions forming a single phonological word. Beyond this, however, nothing has been established about the internal structure of the echo expressions, such as the nature of the prosodic boundary between base and reduplicant. Given that the echo formation involves total reduplication of the base, modulo the introduction of the fixed segments, the expectation would be that base and reduplicant are of the same prosodic category. In the dual description analysis this would be reflected by the reduplicant inheriting the topmost prosodic node of the base and all its subordinate structure. In the majority of cases the base corresponds to a phonologically independent item, so the prosodic word is the obvious candidate as the topmost node: this is what was assumed for the Kolami echo words given in section 1.5.2.6, for example, and is also consistent with the evidence from compounding. The data on stress patterns suggest that endocentric compounds have the structure of a single phonological word, and so echo expressions in which the whole compound is repeated are consistent with structure (3).

(3) $[[\text{base}]_{\omega}[\text{reduplicant}]_{\omega}]_{\omega}$ e.g. $[[\textit{piccai}]_{\omega}[\textit{kiccai}]_{\omega}]_{\omega}$

Note that this structure involves recursion of the prosodic word node, and is therefore inconsistent with the constraints on the prosodic hierarchy encapsulated in the Strict Layer Hypothesis (Selkirk 1984: 26). This requires that prosodic constituents be grouped into a strict hierarchy, with no overlapping domains and with every utterance exhaustively parsed for each category. There has, however, been criticism of the prohibition on recursion (see, for example, Ladd 1986), and limited recursion is generally permitted in more recent work (see, for example, Selkirk 1995 and Peperkamp 1997). The main objection to (3) does not therefore concern its configuration, but rather the problems associated with defining the reduplicant as a phonological word when it bears no stress, and the acoustic structure of its initial segment most closely resembles a word-internal, not word-initial, /k/. One approach would be to stipulate that stress is assigned only to the outermost prosodic word in recursive structures such as (3), and would thus be found on the initial syllable only. A distinction could also be made between metrically strong and metrically weak prosodic words: the base being strong and the reduplicant weak. This is represented in (4), where subscript ωs marks a strong prosodic word and ωw a weak one.

(4) $[[\text{base}]_{\omega s}[\text{reduplicant}]_{\omega w}]_{\omega s}$ e.g. $[[\textit{piccai}]_{\omega s}[\textit{kiccai}]_{\omega w}]_{\omega s}$

This strategy can account for the single stress, on the grounds that the reduplicant is metrically subordinate to the base, but there is still no explanation of why the fixed /k/ segment differs from other word-initial tokens of /k/. One alternative is to abandon the nested prosodic word configuration and propose that base and reduplicant each correspond to lower level prosodic constituents, and only together comprise a prosodic word, as in (5). The challenge is then to identify the relevant lower level prosodic category, marked in (5) with a subscript ψ.

(5) $[[\text{base}]_{\psi}[\text{reduplicant}]_{\psi}]_{\omega}$ e.g. $[[piccai]_{\psi}[kiccai]_{\psi}]_{\omega}$

One option would be to follow Fitzpatrick-Cole in arguing for a non-metrical sublexical prosodic constituent known as the phonological stem. Note, however, that this may be morphologically complex, containing case markers, for instance, and should not be assumed to coincide with the morphological stem. Alternatively, the possible terminological confusion could be avoided by making a distinction between a lexical prosodic word, ψ, and a postlexical prosodic word, ω. In either case both the single stress and the nature of the /k/ can be accounted for on the assumption that ω, and not ψ, forms the domain for stress assignment and determines the distribution of obstruent realizations. A further consideration is the similarities noted between the echo formations and the endocentric compounds and expressives, which seem to point to parallel prosodic structure for the three formations (see sections 4.6.3, 4.6.4 and 5.5.3). A stem-stem, rather than word-word structure, is consistent with these data, providing further evidence in support of structure (5).

The main tendencies to have emerged from the non-phrasal examples are thus captured by a prosodic structure in which the echo expression as a whole forms a single phonological word, comprised of two lower level phonological constituents. A purely prosodic analysis, however, fares less well than the proposed bimorphemic constraint in handling some of the more peripheral examples. Syntactic constituency must also be taken into account in the phrasal examples, whether it directly determines the nature of the base, or is mediated by postlexical prosodic structure. The next section will therefore consider how the interaction between different kinds of linguistic information can be handled.

6.3. Theoretical implications

The theories reviewed in chapter 1 take various different approaches to the morphological and phonological aspects of reduplication. All the examples considered, however, involve reduplication of either lexical or sublexical strings: none of the mainstream theories analyze cases of phrasal reduplication. The possibility that reduplication may be sensitive to syntactic structure, whether directly or through supralexical prosodic constituency, therefore needs to be addressed. Incorporating reference to syntactic structure has implications for the architecture of the grammar, although the precise details will vary according to the framework adopted.

For derivational models of the grammar in which different components are linearly ordered, the main challenge posed by phrasal reduplication is that a morphological process requires access to information about syntactic structure. This raises problems for theories in which reduplication is treated as a process of word formation occurring in the lexicon prior to the syntactic component. Several of the theories of reduplication reviewed in chapter 1 fall into this category, including Aronoff (1976) and Carrier-Duncan (1984), and analyses framed in terms of lexical phonology, such as Uhrbach (1987). Indeed, by having the same kind of process manipulate both morphological and syntactic structure phrasal reduplication is a challenge to lexicalist theories in general, whether the radical lexicalism advocated by Di Sciullo and Williams (1987), or proposals of a split morphology divided between derivational morphology in the lexicon and post-syntactic inflectional morphology in the phonology (e.g. Anderson 1992). In this respect phrasal reduplication is by no means unique: various instances of word formation interacting with syntax have been noted in the literature. Examples of phrasal compounds which appear to have become lexicalized are cited from

English, e.g. *off-the-rack dress*, and other Germanic languages (see, for example, Lieber 1992: 11–12), and also Romance languages, e.g. French *trompe-l'oeil* 'illusion' (literally 'deceive the eye') (Di Sciullo & Williams 1987: 83). Hebrew construct state nominals share characteristics with both syntactic constructions and lexical compounds according to Borer (1988), and Shibatani and Kageyama (1988) claim that there are certain noun-verb compounds in Japanese that are syntactically constructed. They argue that their formation is not only post-syntactic but also, given their patterns of accentuation, follows the post-lexical phonology. In this respect they may be comparable with the echo phrases of Bengali, which are constrained by post-lexical prosodic constituency in Fitzpatrick-Cole's analysis. Pulleyblank and Akinlabi also present examples of post-syntactic word formation from Yoruba, including cases of reduplication such as (6) (Pulleyblank & Akinlabi 1988: 155). The inclusion of the dummy case marker *ní* is taken as evidence of syntactic, rather than lexical, construction.

(6) *fẹ́nilọ́mọfẹ́nilọ́mọ* 'somebody who takes people's daughters and marries them'
(*fẹ́* 'marry', *ẹni* 'person', *ní* dummy case marker, *ọmọ* 'child')

Even in the specific area of reduplication, therefore, it seems that the Indian phrasal examples are not unparalleled.

Various approaches are possible for handling cases of this kind: reduplication may be placed in a post-syntactic morphological component, as in Lidz's study of echo reduplication in Kannada, to be discussed below. Alternatively, morphology may be regarded as an autonomous module running in parallel with the rest of the grammar, from the lexicon through the syntax to the phonology, and being responsible for word formation at any of these levels. Proposals of this kind are to be found in the articles on post-syntactic word formation by Borer (1988) and Shibatani and Kageyama (1988), also in Spencer (1991: 455). In a framework of this type, echoing can apply both in the syntax and the lexicon, and be monitored in each by the same principles of morphological well-formedness.

Another approach is to split reduplication into two processes applying at different levels of the grammar. This is the line taken by Fitzpatrick-Cole (1996: 342), who suggests that a feature like [+REDUP] is *'somehow'* assigned in the syntax and interpreted elsewhere. The tactic of divorcing the marking of the base from the copying process is reminiscent of the proposals by Aronoff (1976) and Carrier-Duncan (1984) mentioned in section 1.3.1. There the motivation was accounting for interaction with the phonology by allowing certain phonological processes to intervene between the assignment of [+reduplication] features in the morphology, and the copying process. Here the effect is to assimilate reduplication to inflection, being marked by a morphosyntactic feature in the syntax and matched with the appropriate phonological substance at a later stage of the derivation.

The analysis of phrasal reduplication in Kannada by Lidz (Lidz 2000) takes distributed morphology as its theoretical framework (see Halle & Marantz 1993). This theory proposes that morphology forms an autonomous component ordered between the syntax and the phonology, and that word formation is syntactic and post-syntactic, not lexical. Indeed, the lexicon itself is replaced by the *'narrow lexicon'* (Marantz 1997: 203), which supplies the terminal elements manipulated by the syntax, and the *'vocabulary'*, which assigns phonological features to the morphosyntactic feature bundles after the syntax, at the level of morphological structure. Vocabulary insertion may be followed by various morphologically conditioned *'readjustment rules'* (Halle & Marantz 1993: 124), and reduplication is identified as one of these (Raimy 2000: 4). As a result it has access not only to supralexical, i.e. syntactic, structural relations, but also to sublexical structure. Lidz's article argues that echo reduplication in Kannada applies equally to syntactic and sublexical constituents, and thus directly supports such a theory over lexicalist approaches in which sublexical structure is erased on insertion into the syntax.

Lidz notes, however, that there are aspects of sublexical structure to which echo reduplication appears to be blind, in that it cannot apply inside certain affixes, such as the gerund marker *-ike* in (7) (Lidz 2000: 156).

(7)	*tooru-vike*	*tooruvikẹ giiruvike*	**tooru-giiru-vike*
	show-gerund	show-gerund echo	show echo gerund
	'showing'	'showing and related activities'	

He demonstrates that such affixes cannot be distinguished from those inside which echo reduplication can apply by a distinction between derivation and inflection, or semantic idiosyncrasy versus compositionality, or stem-level versus word-level affixation. He concludes that the difference lies in the nature of their representation, and that their interaction with echo reduplication shows:

> *'which affixes are present because they correspond to independent heads in the phrase structure and which are present because of categorical properties of the context'* (Lidz 2000: 163).

However, Lidz suggests no principle underlying the division, and leaves the search for other properties converging on the same distinction to further research, so this particular aspect of his hypothesis remains unsupported conjecture at this stage.

To what extent do Lidz's arguments hold for Tamil? The need for the echoing process to access supralexical syntactic structure clearly applies to Tamil, and, as seen above, in non-phrasal examples the base may be defined as a sublexical prosodic constituent. Furthermore, the responses of some individual speakers do permit echoing of sublexical morphological constituents, notably stems without case markers. Some Tamil examples could therefore be accounted for by an echo rule applying after the syntax, without word-internal structure needing to be visible, but this is true of only a limited number of cases. This approach also suffers from the disadvantage that echoing is separated from other morphological processes. A theory in which all morphology is post-syntactic is therefore preferable, whilst also providing a straightforward account of sublexical echoing.

Cases of phrasal echoing, therefore, have direct implications for the relative ordering of morphology and syntax, counting against theories in which all instances of reduplication are ordered prior to syntax. If, as is claimed to be the case in Kannada and Bengali, echoing applies to supralexical, lexical and sublexical constituents, the Tamil echo data are evidence in favour of theories in which syntactic and morphological constituency are not sharply distinguished.

6.4. Conclusion

This final section incorporates all the different aspects of echoing investigated in this thesis to give a full theoretical treatment of the phenomenon in Tamil. As noted in chapter 3, echo expressions are characteristic of colloquial Tamil, and are restricted in their occurrence to certain syntactic contexts. A wide range of lexical categories participates in echoing, and the constituency of the base varies, including both phrasal and non-phrasal examples. It was argued in chapter 1 that an explicit definition of the base is needed in accounts of reduplication, and that this can involve reference to either morphological or prosodic constituency. The existence of echo phrases, not only in Tamil but also Kannada and Bengali, shows that echoing can also be sensitive to syntactic structure.

Analysis of the phrasal examples from Tamil revealed that the base consistently coincided with a syntactic phrase, although the possibility was also raised that postlexical prosodic structure may determine what can serve as the base. In the absence of any solid work on higher level prosodic constituency in Tamil, this question cannot be resolved in the current state of knowledge. Sensitivity to syntactic or postlexical prosodic structure has certain implications for the place of reduplication (and indeed, morphology, in general) in the grammar, an issue that was discussed in section 6.3. The Tamil data were taken to support a

model in which reduplication is post-syntactic, and able to manipulate both sublexical and postlexical structure.

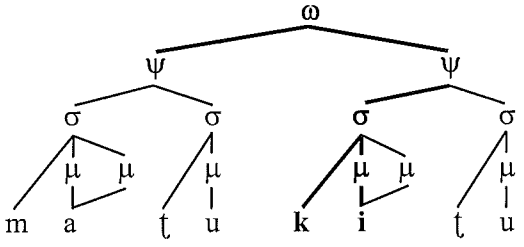
The constituency of the base in the non-phrasal examples was investigated from two perspectives, morphological in chapter 3 and prosodic in chapters 4 and 5, and using two methodologies, a questionnaire and acoustic analysis. The two acoustic studies pointed clearly towards the expression as a whole forming a single domain for stress assignment, and the distribution of obstruent realizations. This was identified as the phonological word, and was argued in section 6.2 to be comprised of two sublexical prosodic constituents, labelled ψ. Various possibilities for their prosodic status were considered, with the favoured option being a non-metrical phonological stem constituent.

The dual description analysis can be applied to Tamil echo reduplication in the following way. As described in sections 1.4 and 1.5.2.6, the reduplicant is defined by two descriptions, one being the description of the base, including all of its segmental and prosodic structure, and the other a partial description of the reduplicant. Any conflict is resolved in favour of the description of the reduplicant, and there are various possible means of implementing this, such as the use of a default inheritance system (see section 1.4). This involves all the properties of the reduplicant being inherited from the base by default, i.e. insofar as they do not conflict with the description of the reduplicant given in (8). In instances of total reduplication all that needs to be specified as the description of the reduplicant is the topmost prosodic node of the base, ψ in the case of the Tamil echo words. The segment-changing dimension of the echo expressions is straightforwardly handled by incorporating the fixed segments within the description of the reduplicant. This is shown on the left in (8): the right side gives the structure of the reduplicated expression as a whole, prior to lexical insertion. When this occurs the properties of the lexical description are distributed over the two ψ nodes, percolating downwards from the topmost node. The first ψ, the base, is supplied with the full lexical description; the second inherits these same properties only to the extent that they do not conflict with the description of the reduplicant.

(8)

For instance, when the noun *maaʈu* 'cow' is inserted into this structure, the result is *maaʈu kiiʈu*, as illustrated in (9).

(9)



As this example demonstrates, the transfer of distinctive vowel length from the base is unproblematic, with the fixed vowel of the reduplicant taking over the double moraic association inherited from the base. Cases of non-transfer can also be accommodated, as described in section 1.4. The dual description proposal is also superior to other constraint-based analyses in making segment-changing reduplication a straightforward subcase of reduplication in general. This contrasts with optimality theoretic analyses of reduplication (e.g. McCarthy & Prince 1995), in which segmental and prosodic requirements on the reduplicant are handled by completely different mechanisms.

As noted in section 6.2, the responses to the questionnaire seemed to display a general bias in favour of bimorphemic echo expressions, in both the phrasal and non-phrasal examples. In particular, it seemed to prove a uniting factor behind cases where it is possible to envisage echo expressions that are formally different but semantically equivalent. For example, all the instances in the questionnaire concerning the inclusion or exclusion of case markers from the base fall into this category. The overall preference was for including the case marker, so that the base was made up of two morphemes, a root and an affix. The ablative proved the one exception, and is also the only example of a morphologically complex case marker: its inclusion would therefore have produced an exceptional trimorphemic base. Arguments that morphological criteria play some part in determining the make-up of the base can also be made for postpositional phrases, verbal forms, compounds and even adjective and noun sequences. If morphological constraints do prove to be relevant, they could be incorporated into the dual description analysis by allowing the description of the reduplicant to be specified for morphological structure, in addition to prosodic structure. However, more data are needed to establish exactly what principles are at work in these examples, and to tease apart the relative influence of morphological and prosodic structural considerations. This last proposal must therefore remain speculative, but suggests a possible extension of the proposal, and an avenue for further research.

Finally, the data reported in this thesis raise questions about the nature of echo reduplication cross-linguistically. Throughout the thesis, the findings on Tamil have been compared with the small body of research on echo expressions in other languages, particularly Kannada, Hindi and Bengali. As noted in section 2.3.4, echo expressions are found in all the languages of the Indian subcontinent, and beyond. There are numerous dimensions of variation: from a phonological viewpoint, for instance, the nature of the fixed segments varies, and also the strategies employed in identity avoidance. The lexical categories which can participate in echo expressions also differ, as do the constraints on the syntactic, morphological and prosodic constituency of the base. At this stage the sample of languages for which data are available is too limited to be able to deduce much about the different parameters of variation, or to associate particular properties with genetic affiliation. The main contribution made by this thesis is defining the key issues posed by echo reduplication, establishing detailed answers for Tamil, and providing a model for research in other languages.

Appendix A: Tamil transliteration conventions

As discussed in section 5.2.1, the Tamil orthography makes two distinctions that do not correspond to a phonemic difference in most modern dialects (see Balasubramanian 1982). The symbols concerned are <ந> and <ன>, and <ர> and <ற>: the first pair are accordingly transliterated simply by <n>, and the second pair by <r>. The IPA symbol for a retroflex approximant, i.e. [ɻ], is used to transliterate the symbol <ழ>. This is a departure from the common practice of using either <z> or <zh>, but gives a better indication of its articulatory properties, as reported in Naryanan, Byrd and Kaun (1999). The grantha letters are not included in the list below as they do not occur in the thesis.

அ	a	க	k	ந	n
ஆ	aa	ச	c	ன	n
இ	i	ட	ʈ	ங	ŋ
ஈ	ii	த	t	ஞ	ñ
உ	u	ப	p	ண	ɳ
ஊ	uu	ம	m	ல	l
எ	e	ய	y	ள	ɭ
ஏ	ee	வ	v	ழ	ɻ
ஐ	ai			ர	r
ஒ	o			ற	r
ஓ	oo				
ஔ	au				

Appendix B: Hindi transliteration conventions

The conventions adopted to transliterate the Devanagari script are set out below, in dictionary order. These are all fairly standard within the literature and are not intended to be accurate phonetic representations. For instance, the short vowel transliterated <a> has a schwa-like quality and is accordingly represented <ə> by some scholars (e.g. Abbi 1992a). However, <a> has been used here, in order to make its synchronic relationship with the corresponding long vowel, represented <aa>, more transparent. This is seen, for example, in the relationship between *nikalnaa* 'to be out', and the causative form *nikaalnaa* 'to take out'. Double letters are also used in <ii> and <uu>, the lengthened counterparts of <i> and <u>. For retroflex sounds, the relevant IPA symbols are used, as in Tamil, rather than the underdot.

अ	a	क/क़	k / q	ठ	t^h	ब	b
आ	aa	ख / ख़	k^h / $\underline{k^h}$	ड/ड़	ɖ / ɽ	भ	b^h
इ	i	ग / ग़	g / g	ढ/ढ़	$ɖ^h$ / $ɽ^h$	म	m
ई	ii	घ	g^h	ण	ɳ	य	y
उ	u	ङ	ŋ	त	t	र	r
ऊ	uu	च	c	थ	t^h	ल	l
ऋ	r̥	छ	c^h	द	d	व	v
ए	e	ज / ज़	j / z	ध	d^h	श	ś
ऐ	ai	झ	j^h	न	n	ष	ṣ
ओ	o	ञ	ɲ	प	p	स	s
औ	au	ट	ʈ	फ/ फ़	p^h / f	ह	h

Appendix C: Glossing conventions

abl	ablative case marker
acc	accusative case marker
adj	adjectival suffix
adv	adverbial suffix
asp	aspectual verb
caus	causative marker
co	coordinate clitic
conc	concessive marker
cond	conditional marker
conj	conjunctive marker
cont	contingent marker
dat	dative case marker
dir	direct form
dis	disjunctive clitic
dur	durative marker
echo	echo reduplicant
emph	emphatic clitic
fpl	feminine plural
fs	feminine singular
fut	future tense
gen	genitive case marker
hort	hortative marker
imp	imperative
infin	infinitive
instr	instrumental case marker
loc	locative case marker
mpl	masculine plural
ms	masculine singular
neg	negative marker
nom	nominalizing suffix
obl	oblique form
past	past tense
pres	present tense
prog	progressive verb
prohib	prohibitive marker
qs	interrogative clitic
quote	quotative particle
refl	reflexive marker
rel	relative participle
soc	sociative case marker
suff	suffix
vbp	verbal participle
voc	vocative case marker
1pl	first person plural
1s	first person singular
2s	second person singular
3pl	third person plural
3s	third person singular
3sf	third person singular feminine
3sm	third person singular masculine
3sn	third person singular neuter

Appendix D: Questionnaire

1. *avanaal* [*ooɽa*] [*kiiɽa*] *muɽiyaatu*
   he.instr run.infin echo can.neg.3sn
   'He can't run and so forth.'

2. *avanaal* [*ooɽa*] [*kiiɽa*] *muɽiyum*
   he.instr run.infin echo can.3sn
   'He can run and so forth.'

3. [*kiirai*] [*kiirai*] *caappiɽaatee*
   greens echo eat.neg.imp
   'Don't eat greens and so forth.'

4. *enakku* [*kiɭi*] [*kiɭi*] *piɽikkaatu*
   I.obl.dat parrot echo like.neg.3sn
   'I don't like parrots and so forth.'

5a. [*puuvai*] [*kiivai*] *parikkaatee*
    flower.acc echo pluck.neg.imp
    'Don't pick the flowers or anything else.'

5b. [*puu*] [*kii*]*vai parikkaatee*
    flower echo.acc pluck.neg.imp

6a. [*cennai*] [*kinnai*]*kku pooka maaʈʈeen*
    Chennai echo.dat go.infin will-not.1s
    'No way am I going to Chennai or anywhere near.'

6b. [*cennaikku*] [*kinnaikku*] *pooka maaʈʈeen*
    Chennai.dat echo go.infin will-not.1s

7a. [*tiruppatikku*] [*kiruppatikku*] *pooka maaʈʈaan*
    Tirupati.dat echo go.infin will-not.3sm
    'No way will he go to Tirupati or anywhere near.'

7b. [*tiruppati*] [*kiruppati*]*kku pooka maaʈʈaan*
    Tirupati echo.dat go.infin will-not.3sm

8a. [*citamparam*] [*kitampara*]*ttukku kumaar pooka maaʈʈaan*
    Chidambaram echo.obl.dat Kumar go.infin will-not.3sm
    'No way is Kumar going to Chidambaram or any place like that.'

8b. [*citamparattukku*] [*kitamparattukku*] *kumaar pooka maaʈʈaan*
    Chidambaram.obl.dat echo Kumar go.infin will-not.3sm

8c. [*citamparattu*] [*kitamparattu*]*kku kumaar pooka maaʈʈaan*
    Chidambaram.obl echo.dat Kumar go.infin will-not.3sm

9a. [*tooʈʈattukku*] [*kiiʈʈattukku*] *ooɽaatee*
    garden.obl.dat echo run.neg.imp
    'Don't run to the garden or such like place.'

9b. [*tooʈʈattu*] [*kiiʈʈattu*]*kku ooɽaatee*
    garden.obl echo.dat run.neg.imp

9c. [*tooʈʈam*] [*kiiʈʈa*]*ttukku ooʈaatee*
garden echo.obl.dat run.neg.imp

10a. *kumaar* [*laɳʈanle*] [*kiɳʈanle*] *irukka maaʈʈaan*
Kumar London.loc echo be.infin will-not.3sm
'Kumar won't stay in London or anywhere like that.'

10b. *kumaar* [*laɳʈan*] [*kiɳʈan*]*le irukka maaʈʈaan*
Kumar London echo.loc be.infin will-not.3sm

11a. [*tiruʈan*] [*kiruʈan*]*ooʈu peeca maaʈʈeen*
thief echo.soc speak.infin will-not.1s
'I will not speak to thieves and such like.'

11b. [*tiruʈanooʈu*] [*kiruʈanooʈu*] *peeca maaʈʈeeɳ*
thief.soc echo speak.infin will-not.1s

12a. *itu* [*kumaarooʈe*] [*kimaarooʈe*] *peenaavaa*
this Kumar.gen echo pen.qs
'Does this pen belong to Kumar or some such person?'

12b. *itu* [*kumaar*] [*kimaar*]*ooʈe peenaavaa*
this Kumar echo.gen pen.qs

12c. *itu* [*kumaarin*] [*kimaarin*] *peenaavaa*
this Kumar.gen echo pen.qs

12d. *itu* [*kumaar*] [*kimaar*]*in peenaavaa*
this Kumar echo.gen pen.qs

13a. [*tooʈʈattileruntu*] [*kiiʈʈattileruntu*] *puuvai parikkaatee*
garden.obl.abl echo flower.acc pluck.neg.imp
'Don't pluck flowers from the garden or places like that.'

13b. [*tooʈʈattu*] [*kiiʈʈatt*]*ileruntu puuvai parikkaatee*
garden.obl echo.abl flower.acc pluck.neg.imp

14a. *tooʈʈattup*[*puuvai*] [*kiivai*] *parikkaatee*
garden.obl.flower.acc echo pluck.neg.imp
'Don't pick garden flowers or anything else.'

14b. [*tooʈʈattuppuu*] [*kiiʈʈattuppuu*] *parikkaatee*
garden.obl.flower echo pluck.neg.imp

15a. [*talaivali*] [*kilaivali*] *irukkutaa*
head-ache echo be.pres.3sn.qs
'Have you got a headache and so forth?'

15b. [*talai*] [*kilai*]*valikkutaa*
head echo ache.pres.3sn.qs

16a. [*maamaram*] [*kiimaram*] *iŋkee naʈaatee*
mango-tree echo here plant.neg.imp
'Don't plant mango-trees and so forth here.'

16b. [*maa*] [*kii*]*maram iŋkee naʈaatee*
mango echo-tree here plant.neg.imp

17a. *enakku* [*maaŋkaay*] [*kiiŋkaay*] *piʈikkaatu*
I.obl.dat mango-fruit echo like.neg.3sn
'I don't like mangoes and so forth.'

17b. *enakku* [*maa*] [*kiiŋ*]*kaay piʈikkaatu*
I.obl.dat mango echo-fruit like.neg.3sn

18a. [*naaykkuʈʈi*] [*kiiykkuʈʈi*] *viiʈʈukkuḷḷee koɳʈuvaraatee*
dog-calf echo house.inside bring.neg.imp
'Don't bring pups and so forth inside the house!'

18b. [*naay*] [*kiiy*]*kkuʈʈi viiʈʈukkuḷḷee koɳʈuvaraatee*
dog echo-calf house.inside bring.neg.imp

19a. [*aaʈu*] [*kiiʈu*] *meeykka marakkaatee*
goat echo graze.infin forget.neg.imp
'Don't forget to graze the goats and such like.'

19b. [*maaʈu*] [*kiiʈu*] *meeykka marakkaatee*
cow echo graze.infin forget.neg.imp
'Don't forget to graze the cows and such like.'

19c. *aaʈu*[*maaʈu*] [*kiiʈu*] *meeykka marakkaatee*
goat-cow echo graze.infin forget.neg.imp
'Don't forget to graze the cattle and so forth.'

19d. [*aaʈumaaʈu*] [*kiiʈumaaʈu*] *meeykka marakkaatee*
goat-cow echo graze.infin forget.neg.imp

20a. [*appaa ammaa*] [*kippaa ammaa*] *keekkaatee*
father mother echo listen.neg.imp
'Don't listen to your parents and their generation!'

20b. [*appaa*] [*kippaa*] *ammaa keekkaatee*
father echo mother listen.neg.imp

20c. [*appaa*] [*kippaa*] *kiʈʈa keekkaatee*
father echo near listen.neg.imp
'Don't listen to your father and the older generation!'

20d. [*appaa kiʈʈa*] [*kippaa kiʈʈa*] *keekkaatee*
father near echo listen.neg.imp

21a. [*meejai meelee*] [*kiijai meelee*] *onnum vaikka kuuʈaatu*
table on echo anything put.infin should.neg.3sn
'It is strongly forbidden to put anything on the table or nearby.'

21b. [*meejai*] [*kiijai*] *meelee onnum vaikka kuuʈaatu*
table echo on anything put.infin should.neg.3sn

22a. [*veelai*] [*kiilai*] *patti collaatee*
work echo about speak.neg.imp
'Don't talk to me about work and such like!'

22b. [*veelaiyai*] [*kiilaiyai*] *patti collaatee*
work.acc echo about speak.neg.imp

22c. [*veelai*] [*kiilai*]*yai patti collaatee*
work echo.acc about speak.neg.imp

22d. [*veelai patti*] [*kiilai patti*] *collaatee*
work about echo speak.neg.imp

22e. [*veelaiyai patti*] [*kiilaiyai patti*] *collaatee*
work.acc about echo speak.neg.imp

23a. [*viiṯṯukku*] [*kiiṯṯukku*] *munnaalee okkaaraatee*
house.dat echo before sit.neg.imp
'Don't sit in front of the house or anywhere near it.'

23b. [*viiṯṯukku munnaalee*] [*kiiṯṯukku munnaalee*] *okkaaraatee*
house.dat before echo sit.neg.imp

24a. *enakku* [*civappu*] [*kivappu*] *toppi piṭikkaatu*
I.obl.dat red echo hat like.neg.3sn
'I don't like hats that are red etc.'

24b. *enakku* [*civappu toppi*] [*kivappu toppi*] *piṭikkaatu*
I.obl.dat red hat echo like.neg.3sn

25a. [*veḷḷai caṭṭai*] [*kiḷḷai caṭṭai*] *vaaŋkaatee*
white shirt echo buy.neg.imp
'Don't buy white shirts and such like.'

25b. [*veḷḷai*] [*kiḷḷai*] *caṭṭai vaaŋkaatee*
white echo shirt buy.neg.imp

26a. [*keṭṭa kanavai*] [*kiṭṭa kanavai*] *keekkaatee*
bad dream.acc echo listen.neg.imp
'Don't heed bad dreams and so forth.'

26b. [*keṭṭa*] [*kiṭṭa*] *kanavai keekkaatee*
bad echo dream.acc listen.neg.imp

27a. *avan* [*nalla paiyan*] [*killa paiyan*]*-ŋŋu nampaatee*
he good boy echo quote believe.neg.imp
'Don't believe that he's a good boy and so forth.'

27b. *avan* [*nalla*] [*killa*] *paiyan-ŋŋu nampaatee*
he good echo boy quote believe.neg.imp

28a. [*cantoocamaaka*] [*kintoocamaaka*] *paaṭaatee*
joy.adv echo sing.neg.imp
'Don't sing joyfully etc.'

28b. [*cantoocam*] [*kintoocam*]*aaka paaṭaatee*
joy echo.adv sing.neg.imp

29. [*naan*] [*kiin*]-*ŋŋu* *peecaatee*
I echo quote speak.neg.imp
'Don't speak only of yourself.'

30a. *enakku un* [*pai*] [*kii*] *kuɽukkaatee*
I.obl.dat you.obl bag echo give.neg.imp
'Don't give me your bag and so forth.'

30b. *enakku* [*un pai*] [*kii pai*] *kuɽukkaatee*
I.obl.dat you.obl bag echo give.neg.imp

30c. *enakku un* [*paiyai*] [*kiiyai*] *kuɽukkaatee*
I.obl.dat you.obl bag.acc echo give.neg.imp

31a. [*kumaarukku*] [*kimaarukku*] *peṭṭiyai kuɽukkaatee*
Kumar.dat echo box.acc give.neg.imp
'Don't give the box to Kumar or any of his type.'

31b. [*kumaar*] [*kimaar*]*ukku peṭṭiyai kuɽukkaatee*
Kumar echo.dat box.acc give.neg.imp

32a. *nii* [*peecaatee*] [*kiicaatee*]
you speak.neg.imp echo
'Don't speak and so forth!'

32b. *nii* [*peeca*] [*kiica*]*atee*
you speak.infin echo.neg

33a. *iŋkee* [*paɽikka*] [*kiɽikka*] *kuuɽaatu*
here study.infin echo should.neg.3sn
'One should not study and so forth in this place!'

33b. *iŋkee* [*paɽi*][*kiɽi*]*kka kuuɽaatu*
here study echo.infin should.neg.3sn

34a. *avanooɽu* [*caappiɽale*] [*kiippiɽale*]
he.soc eat.infin.neg echo
'I don't eat and so forth with him.'

34b. *avanooɽu* [*caappiɽa*] [*kiippiɽa*]*le*
he.soc eat.infin echo.neg

35a. *kumaar* [*paɽikka*] [*kiɽikka*] *maaṭṭaan*
Kumar study.infin echo will-not.3sm
'Kumar will not study etc.'

35b. *kumaar* [*paɽikka maaṭṭaan*] [*kiɽikka maaṭṭaan*]
Kumar study.infin will-not.3sm echo

36a. *avanukku* [*viiɽu*] [*kiiɽu*] *illai*
he.dat house echo neg
'He has no house and family etc.'

36b. *avanukku* [*viiɽu illai*] [*kiiɽu illai*]
he.dat house not echo

37a. *kumaar* [*paaʈiʈʈu*] [*kiiʈiʈʈu*] *viiʈʈukku poonaan*
Kumar sing.asp.vbp echo house.dat go.past.3sm
'After singing and so forth, Kumar went home.'

37b. *kumaar* [*paaʈi*] [*kiiʈi*]*ʈʈu viiʈʈukku poonaan*
Kumar sing.vbp echo.asp.vbp house.dat go.past.3sm

38a. [*tuuŋkiʈʈu*] [*kiiŋkiʈʈu*] *caappiʈaatee*
sleep.asp.vbp echo eat.neg.imp
'Don't eat after sleeping etc.'

38b. [*tuuŋki*] [*kiiŋki*]*ʈʈu caappiʈaatee*
sleep.vbp echo.asp.vbp eat.neg.imp

39a. *enakku* [*teriyaatu -ŋŋu*] [*kiriyaatu-ŋŋu*] *collaatee*
I.obl.dat know.neg.3sn quote echo say.neg.imp
'Don't say "I don't know!".'

39b. *enakku* [*teriyaatu*] [*kiriyaatu*]*-ŋŋu collaatee*
I.obl.dat know.neg.3sn echo quote say.neg.imp

39c. [*enakku teriyaatu -ŋŋu*] [*kinakku teriyaatu-ŋŋu*] *collaatee*
I.obl.dat know.neg.3sn quote echo say.neg.imp

40a. [*kumaar paakkale*] [*kimaar paakkale*]*-ŋŋu poy collaatee*
Kumar see.infin.neg echo quote lie say.neg.imp
'Don't lie that Kumar didn't see or some such nonsense.'

40b. *kumaar* [*paakkale*] [*kiikkale*]*-ŋŋu poy collaatee*
Kumar see.infin.neg echo quote lie say.neg.imp

41. [*kumaar*] [*kimaar*] *paakkale -ŋŋu poy collaatee*
Kumar echo see.infin.neg quote lie say.neg.imp
'Don't lie that Kumar or someone like that didn't see.'

42. [*paakkale*] [*kiikkale*]*-ŋŋu poy collaatee*
see.infin.neg echo quote lie say.neg.imp
'Don't lie that you didn't see and so forth.'

43a. [*kumaarukku kuʈutteen -ŋŋu*] [*kimaarukku kuʈutteen-ŋŋu*] *poy collaatee*
Kumar.dat give.past.1s quote echo lie say.neg.imp
'Don't lie that you gave it to Kumar or some such nonsense.'

43b. [*kumaarukku kuʈutteen*] [*kimaarukku kuʈutteen*]*-ŋŋu poy collaatee*
Kumar.dat give.past.1s echo quote lie say.neg.imp

44. *kumaarukku* [*kuʈutteen*] [*kiʈutteen*]*-ŋŋu poy collaatee*
Kumar.dat give.past.1s echo quote lie say.neg.imp
'Don't lie that you gave it or something to Kumar.'

45. [*kumaarukku*] [*kimaarukku*] *kuʈutteen -ŋŋu poy collaatee*
Kumar.dat echo give.past.1s quote lie say.neg.imp
'Don't lie that you gave it to Kumar or anyone else!'

46a. [*kumaar onnum paattaan -ŋŋu*] [*kimaar onnum paattaan-ŋŋu*] *poy collaatee*
Kumar anything see.past.3sm quote echo lie say.neg.imp
'Don't lie that Kumar saw anything and so forth.'

46b. [*kumaar*] [*kimaar*] *onnum paattaan -ŋŋu poy collaatee*
Kumar echo anything see.past.3sm quote lie say.neg.imp
'Don't lie that Kumar or someone like that saw anything.'

46c. *kumaar onnum* [*paattaan*] [*kiittaan*]*-ŋŋu poy collaatee*
Kumar anything see.past.3sm echo quote lie say.neg.imp
'Don't lie that Kumar saw anything and so forth.'

47. *naaḷekki naan* [*vantu*] [*kintu*]*kiṯṯirukka maaṯṯeen*
tomorrow I come.vbp echo.dur.be.infin will-not.1s
'I won't be coming and so forth tomorrow.'

48a. *innikki* [*appaa varaar -ŋŋu*] [*kippaa varaar-ŋŋu*] *collaatee*
today father come.pres.3pl quote echo say.neg.imp
'Don't say "Father's coming today and so forth".'

48b. *innikki appaa* [*varaar*] [*kiraar*]*-ŋŋu collaatee*
today father come.pres.3pl echo quote say.neg.imp

48c. *innikki* [*appaa*] [*kippaa*] *varaar -ŋŋu collaatee*
today father echo come.pres.3pl quote say.neg.imp
'Don't say "Father and so on are coming today".'

48d. *appaa innikki varaar -ŋŋu kinnikki appaa varaar -ŋŋu collaatee*
father today come.pres.3pl quote echo father come.pres.3pl quote say.neg.imp
'Don't say "Father's coming today and so forth".'

49a. *patil* [*collaatee*] [*killaatee*]
answer say.neg.imp echo
'Don't answer and so forth.'

49b. [*patil*] [*kitil*] *collaatee*
answer echo say.neg.imp

49c. [*patil collaatee*] [*kitil collaatee*]
answer say.neg.imp echo

Appendix E: Responses to the questionnaire

Subject	1	2	3	4	5	6	7	8	9	10	11	12	Mean
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	3	4	1	4	1	2	1	1	1	4	4	4	2.5
3	4	2	4	1	1	4	4	1	4	4	2	4	2.9
4	4	1	1	4	1	4	4	1	4	4	2	3	2.8
5a	1	1	1	1	1	1	1	1	1	2	1	1	1.1
5b	4	4	2	3	4	4		3	4	2	4		3.4
6a	2	1	1	3	1	1	2	3	4	4	1	1	2
6b	3	1	2	1	1	1	1	1	1	2	1		1.4
7a		1	1	1	1	1	1	1	3				1.3
7b		2	1	3	4	1	3	3	1				2.3
8a	3	4	1	3	1	1	3	1	3	1	4	4	2.4
8b	3	1	1	1	1	1	1	4	1	1	1	1	1.4
8c	2	4	2	3	1	1	1	1	3	4	4	3	2.4
9a		1	1	1	1	1	1	1	1	1			1
9b		4	4	3	1	1	4	1	3	1			2.4
9c		4	4	3	4	4	4	4	3	1			3.4
10a	1	1	1	1	1	2	1	3	3	1	1	2	1.5
10b	3	4	1	3	1	2	3	1	1	1	4	3	2.3
11a	1	4	1	3	1	3	3	3	3	4	1	1	2.3
11b	4	1	1	1	4	3	1	1	1	1	1	2	1.8
12a	1	1	3	3	1	1	3	3	1				1.9
12b	3	3	3	3	3	2	4	3	3				3
12c	3	3	4	3	3		3	1	2				2.8
12d	3	3	4	3	1		3	3	3				2.9
13a	4	3	2	3	2	1	1	3	3	1			2.3
13b	3	1	1	1	4	3	3	1	1	1			1.9
14a	1	1		3	1	2	1	1	1	1	1	4	1.6
14b	3	3	4	1	4	2	3	3	3	4		4	3.1
15a	1	1	1	1	1	1	3	3	1	1	1	2	1.4
15b	4	4		3	4	1	1	1	3	1		1	2.3
16a	1	1	1	1	1	1	1	1	1	1	1	3	1.2
16b	4	4	4	3	4	4	3	3	3	4	4		3.6
17a	1	1	1	1	1	1	1	1	1	1		1	1
17b	4	4	4	4	4	4	3	3	3	4			3.7
18a	1	1	1	1	1	1	1	1	1	1	1	2	1.1
18b	4	4	4	3	3	4	3	3	3	4	4	4	3.6
19a		1	1	1	1	1	1	3	1	1		1	1.2
19b	1	1	1	3	4	1	1	3	1	1		1	1.6
19c	4	4	4	3		4	3	1	4	4	4	4	3.6
19d	3	4	3	3	1	4	4	3	4	4	4	3	3.3
20a	3	3	1								2		2.3
20b	3	3	3								4		3.3
20c				1	4	1	3	1	3	4			2.4
20d				3	1	1	1	3	1	1			1.6
21a	4	1	1	1	4	2	1	3	1	1	1	1	1.8
21b	1	4	4	4	1	2	3	1	3	4	4	2	2.8
22a	1	1	1	3	4	1	1	1	1	1	2	1	1.5
22b	2	1	1	1	1	1	3	3	3	1	1		1.6
22c	3	3	2	3	4	4	3	3	3	1	1		2.7
22d	3	3	3	3	1	1	3	3	3	1	3	4	2.6
22e	1	3	1	3	1	1	3	1	3	1	3		1.9

Subject	1	2	3	4	5	6	7	8	9	10	11	12	Mean
23a	1	1	1	1	1	2	1	3	3	1			1.5
23b	4	3	4	3	1	2	3	1	1	1			2.3
24a	1	1	1	1	4	2	3	1	3	1			1.8
24b	3	3	3	4	1	2	1	3	1	1			2.2
25a	1	1	1	1	1	2	1	3	1	1			1.3
25b	4	3	4	3	4	2	3	1	3	1			2.8
26a	1	1	1	3	4	4	1	3	1	2			2.1
26b	1	3	4	1	1	2	3	1	3	4			2.3
27a	1	1	1	1	1	1	1	3	1	1	1		1.2
27b	1	3	4	3	4	2	3	1	3	4			2.8
28a	1	1	4	1	1	2	1	3	1	1			1.6
28b	1	4	1	3	4	2	4	1	3	1			2.4
29	1	1	1	1	1	2	3		1	1	4		1.6
30a	1	3	4	3	4	1	3	3	1	1		1	2.3
30b	4	3	3	3	4	4		1	2		4	4	3.2
30c	1	1	1	1	1	1	1	1	3		1		1.2
31a	1	1	1	1	1	1	1	3	3	1			1.4
31b	1	3	1	4	4	1	3	1	1	4			2.3
32a	1	2	3	1	4	1	4	1	1	4		2	2.2
32b	4	4	4	3	1	1	4	3	3	4		4	3.2
33a	1	1	1	1	1	1	1	1	1	1	1	1	1
33b	4	4	4	3	4	4	3	3	3	4		4	3.6
34a	1	1	1	1	4	2	1	1	1	1			1.4
34b	4	4	1	3	1	4	3	3	3	4			3
35a	1	1	1	1	1	1	1	1	1	1	1		1
35b	3	4	4	3	4	1	3	3	3	3			3.1
36a	1	1	1	1	1	1	1	1	1	1	1	1	1
36b	4	4	4	3	4	4	3	3	3	3			3.5
37a	1	1	1	1	4	1	1	3	3	1		1	1.6
37b	4	3	4	3	1	4	3	1	1	2		1	2.5
38a	1	1	1	1	1	1	1	3	1	1			1.2
38b	4	4		3	1	4	3	1	3	1			2.7
39a	1	3	1	3	1	2	3	3	3	1			2.1
39b	1	1	1	1	1	1	1	1	1	2	1	1	1.1
39c	3	3	4	3	4	4	3	3	3	2	4		3.3
40a	1	3	1	3	1	1	3	3	3	1			2
40b	1	1	1	1	4	1	1	1	1	1			1.3
41	1	1	1	1	1	2	1	1	1	1			1.1
42	1	1	1	1	1	1	1	2	1	1			1.1
43a	1	1	1	3	1	2	3	3	3	1			1.9
43b	2	3	2	1	1	2	1	1	1	1		1	1.5
44	1	1	1	1	1	2	1	1	1	1		1	1.1
45	1	1	1	1	1	2	1	1	1	1		1	1.1
46a	2	3	2	3	1	2	4	3	2	1			2.3
46b	1	3	2	1	1	2	1	1	3	1			1.6
46c	1	1	2	3	4	2	3	2	2	1			2.1
47		3	3	3	3		3		3	4			3.1
48a	3	3	1	3	3	2	3	3	3	4		1	2.6
48b	1	2		3	3	2	1	3	3	1			2.1
48c	1	1	1	1	1	2	3	1	1	1			1.3
48d	4	3	4	3	1	4	3	4	3	1		4	3.1
49a						4	3			4			3.7
49b						1	1			1			1
49c						4	3			4			3

Appendix F: Stress sentences

1. *kumaar ippaɻi paɻiccaan*
   Kumar thus study.past.3sm
   'Kumar studied in this way.'

2. *naan oru kiɻi paatteen*
   I one parrot see.past.1s
   'I saw a parrot.'

3. *enakku paampu kiimpu piɻikkaatu*
   I.obl.dat snake echo like.neg.3sn
   'I don't like snakes and such creatures.'

4. *enakku un pai taa*
   I.obl.dat you.obl bag give
   'Give your bag to me.'

5. *kumaar vimmuraan*
   Kumar weep.pres.3sm
   'Kumar is weeping.'

6. *ilaiyai kilaiyai tinnaatee*
   leaf.acc echo eat.neg.imp
   'Don't eat leaves and such things.'

7. *kiɻi kiiccukiiccuŋŋu kattatu*
   parrot chirp cry.past.3sn
   'The parrot chirped.'

8. *kumaar kiirai caappiṯṯaan*
   Kumar greens eat.past.3sm
   'Kumar ate greens.'

9. *enakku civappu kivappu caṯṯai piɻikkaatu*
   I.obl.dat red echo shirt like.neg.3sn
   'I don't like shirts that are red and things like that.'

10. *kili nallatalla*
    fear good.3sn.not
    'Fear is not a good thing.'

11. *kumaarukku kuɻutteen kimaarukku kuɻutteen-ŋŋu poy collaatee*
    Kumar.dat give.past.1s echo quote lie say.neg.imp
    'Don't lie that I gave it to Kumar and so forth.'

12. *mookan kiɻavan*
    Mohan old.3sm
    'Mohan is an old man.'

13. *kumaar kiiraitaan caappiṯṯaan*
    Kumar greens.emph eat.past.3sm
    'Kumar ate *greens*.'

14. *enakku kiṯṯiyatu*
    I.obl.dat approach.past.3sn
    'I got it.'

15. *avan aɻakaana kiɻakaana paiyan-ŋŋu ninaikkaatee*
    he beautiful echo boy quote think.neg.imp
    'Don't think that he's a beautiful boy or anything like that.'

16. *iŋkee oru kiŋaru irukku*
    here one well be.pres.3sn
    'Here is a well.'

17. *anpu kinpu patti eŋkiʈʈa peecaatee*
    love echo about I.obl.near speak.neg.imp
    'Don't speak to me about love and so forth.'

18. *enakku naay piʈikkaatu*
    I.obl.dat dog like.neg.3sn
    'I don't like dogs.'

19. *paattirattai uʈaikka kuuʈaatu*
    cup.obl.acc break.infin should.neg.3sn
    'You must not break the cup.'

20. *en maattaantaay neettu vantaa*
    I.obl step-mother yesterday come.past.3sf
    'My stepmother came yesterday.'

21. *kumaarai naŋpan kiŋpan-ŋŋu ninaikkaatee*
    Kumar.acc friend echo quote think.neg.imp
    'Don't think Kumar is a friend or anything like that.'

22. *inta kukai ciriyatu*
    this cave small.3sn
    'This cave is small.'

23. *kumaarukku kuʈutteen kiʈutteen-ŋŋu poy collaatee*
    Kumar.dat give.past.1s echo quote lie say.neg.imp
    'Don't lie that I gave or donated or presented it to Kumar.'

24. *piccai kiccai keekkaatee*
    alms echo listen.neg.imp
    'Don't listen to begging and such like.'

25. *keŋʈi paɻapaɻakkum*
    spout shine.fut.3sn
    'The spout of the kettle will be shiny.'

26. *naan kirukiruŋŋu cuɻalureen*
    I giddiness whirl.pres.1s
    'I am feeling giddy.'

27. *talaivali kilaivali irukkutaa?*
    head-ache echo be.pres.3sn.qs
    'Do you have a headache or anything like that?'

28. *nalla paiyan killa paiyan-ŋŋu ninaikkaatee*
    good boy echo quote think.neg.imp
    'Don't think that he's a good boy and so forth.'

29. *uŋmai kiŋmai patti eŋkiʈʈa peecaatee*
    truth echo about I.obl.near speak.neg.imp
    'Don't speak about truth and such like to me.'

30. *enakku kari piɻikkaatu*
I.obl.dat curry like.neg.3sn
'I don't like curry.'

31. *maaɻu marattai paakkumaa?*
cow tree.obl.acc see.fut.3sn.qs
'Will the cow see the tree?'

32. *mookanin tampi varaan*
Mohan.gen younger-brother come.pres.3sm
'Mohan's younger brother is coming.'

33. *enakku un uḷḷaŋkai kaaṭṭu*
I.obl.dat you.obl inner-hand show
'Show me the palm of your hand.'

34. *kumaar cennaikku pooraan*
Kumar Chennai.dat go.pres.3sm
'Kumar is going to Chennai.'

35. *enakku oonaay piɻikkaatu*
I.obl.dat wolf like.neg.3sn
'I don't like wolves.'

36. *enakku talaivaliyirukku*
I.obl.dat headache.be.pres.3sn
'I have a headache.'

37. *ammaa neettu vantaa*
mother yesterday come.past.3sf
'Mother came yesterday.'

38. *kumaar kiirai caappiṭṭaantaan*
Kumar greens eat.past.3sm.emph
'Kumar ate *greens.*'

39. *en paɻam illai*
I.obl fruit not
'It is not my fruit.'

40. *inta paɻattai eɻu*
this picture.obl.acc lift
'Lift up this picture.'

41. *enakku citamparam kitambaram piɻikkaatu*
I.obl.dat Chidambaram echo like.neg.3sn
'I don't like Chidambaram and such places.'

42. *appaa varaar -ŋŋu collaatee*
father come.pres.3sm quote say.neg.imp
'Don't say that father is coming.'

43. *eli oru puunaiyai paattatu*
rat one cat.acc see.past.3sn
'The rat saw a cat.'

44. *naaykkuṭṭi kiiykkuṭṭi viiṭṭukkuḷḷee koɳɻuvaraatee*
dog-calf echo house.dat.inside bring.neg.imp
'Don't bring puppies and so forth inside the house.'

45. *kumaar kiɻukiɻukkiraan*
Kumar tremble.pres.3sm
'Kumar is trembling.'

46. *kumaartaan kiirai caappiʈʈaan*
Kumar.emph greens eat.past.3sm
'*Kumar* ate greens.'

47. *iɻam kiɻam iŋkee illai*
place echo here not
'There is no place and so forth here.'

48. *kumaar kelippaay irukkaan*
Kumar successfully be.pres.3sm
'Kumar is successful.'

49. *avan aɻakaana paiyan kiɻakaana paiyan-ŋŋu ninaikkaatee*
he beautiful boy echo quote think.neg.imp
'Don't think that he's a beautiful boy and so forth.'

50. *naan koolam iɻuveen*
I pattern follow.fut.1s
'I will draw a pattern.'[1]

51. *kumaar enna tuukkiyirukkiraan?*
Kumar what lift-up.vbp.be.pres.3sm
'What has Kumar lifted up?'

52. *oru tiruɻan viiʈʈukkuḷḷee iruntaan*
one thief house.dat.inside be.past.3sm
'There was a thief inside the house.'

53. *uŋmai patti eŋkiʈʈa peecaatee*
truth about I.obl.near speak.neg.imp
'Don't speak to me about truth.'

54. *avan aɻakaana kiɻakaana paiyan-ŋŋu ninaikkaatee*
he beautiful echo boy quote think.neg.imp
'Don't think that he's a beautiful boy and so forth.'

55. *enakku un kai kaaʈʈu*
I.obl.dat you.obl hand show
'Show me your hand.'

56. *iɻam iŋkee illai*
place here not
'There's no place here.'

57. *avanukku paŋam kiŋam irukku*
he.dat money echo be.pres.3sn
'He has money etc.'

58. *kumaarukku kuɻutteen -ŋŋu poy collaatee*
Kumar.dat give.past.1s quote lie say.neg.imp
'Don't lie that I gave it to Kumar.'

---

[1] This refers to the patterns traditionally drawn in chalk or cow dung paste on the doorsteps of Tamil houses.

59. *enakku milaku kilaku pitikkaatu*
I.obl.dat pepper echo like.neg
'I don't like pepper and such spices.'

60. *mookan kilukiluppaa irukkiraan*
Mohan laughing loudly be.pres.3sm
'Mohan is laughing loudly.'

61. *appaa varaar kippaa varaar-nnu collaatee*
father come.pres.3pl echo quote say.neg.imp
'Don't say that father is coming and so forth.'

62. *paiyan kili paattaan*
boy parrot see.past.3sm
'The boy saw a parrot.'

63. *kumaarukku kimaarukku kututteen -nnu poy collaatee*
Kumar.dat echo give.past.1s quote lie say.neg.imp
'Don't lie that I gave it to Kumar or some such person.'

64. *enakku kaay pitikkaatu*
I.obl.dat unripe-fruit like.neg.3sn
'I don't like unripe fruit.'

65. *kili marattukku parantatu*
parrot tree.obl.dat fly.past.3sn
'The parrot flew to the tree.'

66. *enakku lantan kintan pitikkaatu*
I.obl.dat London echo like.neg3sn
'I don't like London and such places.'

67. *enakku rattam pitikkum*
I.obl.dat blood like.3sn
'I like blood.'

68. *ilaiyai tinnaatee*
leaf.acc eat.neg.imp
'Don't eat the leaf.'

69. *enakku anta patattai taa*
I.obl.dat that fruit.obl.acc give
'Give me that fruit.'

70. *naan viittule oru pillaiyai paatteen*
I house.loc one child.acc see.past.1s
'I saw a child in the house.'

71. *ennaal inta veelai ceyya mutiyum*
I.obl.instr this work do.infin can.3sn
'I can do this work.'

72. *ammaa aatai taikkiraa*
mother clothes sew.pres.3sf
'Mother is sewing clothes.'

73. *enakku maaŋkaay piɻikkum*
I.obl.dat mango-fruit like.3sn
'I like mangoes.'

74. *en taay neettu vantaa*
I.obl mother yesterday come.past.3sf
'My mother came yesterday.'

75. *anpu patti eŋkiɻɻa peecaatee*
love about I.obl.near speak.neg.imp
'Don't speak to me about love.'

76. *enakku kirukiruppu piɻikkaatu*
I.obl.dat giddiness like.neg.3sn
'I don't like giddiness.'

77. *avan naŋpan ena ninaikkaatee*
he friend say.infin think.neg.imp
'Don't consider him a friend.'

78. *enakku kaaykari piɻikkaatu*
I.obl.dat vegetable-curry like.neg.3sn
'I don't like vegetable-curry.'

79. *naan marattule oru kiɻippiɻɻai paatteen*
I tree.obl.loc one parrot-child see.past.1s
'I saw a parakeet in the tree.'

Appendix G: Gemination sentences

1. *kumaar iɻli* ***c****aappiʈʈaan*
   Kumar idli eat.past.3sm
   'Kumar ate idli(s).'

2. *kumaar ippaɻi****c*** *caappiʈʈaan*
   Kumar thus eat.past.3sm
   'Kumar ate in this way.'

3. *kumaar miʈʈaay* ***c****aappiʈʈaan*
   Kumar sweet eat.past.3sm
   'Kumar ate a sweet.'

4. *kumaar* ***c****aappiʈʈaan*
   Kumar eat.past.3sm
   'Kumar ate.'

5. *kumaar iɻli* ***k****iɻli* ***c****aappiʈʈaan*
   Kumar idli eat.past.3sm
   'Kumar ate idli(s) and so forth.'

6. *en uuraar* ***k****iiraar vara maaʈʈaaŋkaɭ*
   I.obl people-of-a-native-place echo come.infin will-not.3pl
   'People from my place and such like won't come.'

7. *kumaar* ***pa****ɻattai****k*** ***k****iirukiraan*
   Kumar fruit.obl.acc cut.pres.3sm
   'Kumar is cutting the fruit.'

8. *kumaar* ***k****iirai* ***c****aappiʈʈaan*
   Kumar greens eat.past.3sm
   'Kumar ate greens.'

9. *kumaar avanooʈu****p*** ***p****ooka maaʈʈaan*
   Kumar him.soc go.infin will-not.3sm
   'Kumar will not go with him.'

10. *kumaar* ***p****ooraan*
    Kumar go.pres.3sm
    'Kumar is going.'

11. *en aŋŋan (enakku) itai****k*** ***k****uʈikka****k*** ***k****uʈuttaan*
    I.obl elder brother I.obl.dat this.acc drink.infin give.past.3sm
    'My elder brother gave (me) this to drink.'

12. *en aŋŋan ena****kk****u****t*** ***t****aŋŋiir* ***k****uʈuttaan*
    I.obl elder brother I.obl.dat water give.past.3sm
    'My elder brother gave (me) water.'

13. *makanee,* ***pa****ɻattai****k*** ***k****uʈu*
    son.voc fruit.obl.acc give
    'Son, give (me) the fruit.'

14. *unnuʈaiya****p*** ***pa****ɻattai****k*** ***k****uʈu*
    you(sg).gen fruit.obl.acc give
    'Give (me) your fruit.'

15. *anta**p** pa.ɻattai**k** ku**ʈ**u*
that fruit.obl.acc give
'Give (me) that fruit.'

16. *inta**p** pa.ɻam cuvaiyu**ɭɭ**atu*
this fruit sweet.3sn
'This fruit is a sweet one.'

17. *enta**p** pa.ɻattai**p** **p**aattaay?*
which fruit.obl.acc see.past.2s
'Which fruit did you see?'

18. *yaaru**ʈ**aiya**p** pa.ɻam (itu)?*
who.gen fruit this
'Whose fruit is this?'

19. *raajaavoo **k**umaaroo **p**oonaa**ŋ**ka**ɭ***
Raja.dis Kumar.dis go.past.3pl
'Raja or Kumar went.'

20. *raajaavum **k**umaarum **p**oonaa**ŋ**ka**ɭ***
Raja.co Kumar.co go.past.3pl
'(Both) Raja and Kumar went.'

21. *kumaar raajaavaiyaa **p**aattaan?*
Kumar Raja.acc.qs see.past.3sm
'Was it Raja that Kumar saw?'

22. *kumaar **p**aattaanaa?*
Kumar see.past.3sm.qs
'Did Kumar see?'

23. *raajaa **p**aattaanaa?*
Raja see.past.3sm.qs
'Did Raja see?'

24. *nii oru **k**aa**ɭ**ai vaiccirukkaay*
you one bull possess.pres.2s
'You have a bull.'

25. *kumaar (oru) **k**aa**ɭ**aiyai**p** **p**aattaan*
Kumar one bull.acc see.past.3sm
'Kumar saw a bull.'

26. *kumaar atai**p** **p**atti **p**eecinaan*
Kumar that.acc about speak.past.3sm
'Kumar spoke about that.'

27. *pi**ɭɭ**ai **p**a**ʈ**ikkatu*
child study.pres.3sn
'The child studies.'

28. *kumaar **p**a**ʈ**ikkiraan*
Kumar study.pres.3sm
'Kumar studies.'

29. *kumaar appa**ʈ**i**p** **p**eecinaan*
Kumar in-that-way speak.past.3sm
'Kumar spoke in that way.'

30. *kumaar **p**eecinaan*
Kumar speak.past.3sm
'Kumar spoke.'

31. *kumaar (avan) uuriliruntu**p** **p**ooraan*
Kumar he native-place.abl go.pres.3sm
'Kumar is going from his native place.'

32. *kumaar **c**eennaikku**p** **p**ooraan*
Kumar Chennai.dat go.pres.3sm
'Kumar is going to Chennai.'

33. *kumaar eppaɻi**t** **t**uuŋkinaan?*
Kumar how sleep.past.3sm
'How did Kumar sleep?'

34. *kumaar **t**uuŋkinaanaa*
Kumar sleep.past.3sm.qs
'Did Kumar sleep?'

35. *itu **t**ooʈʈattu**k** **k**atavu*
this garden.obl door
'This is the garden-door.'

36. *kumaar (oru) **k**atavai**p** **p**aattaan*
Kumar one door.acc see.past.3sm
'Kumar saw a door.'

37. *unakku **t**ooʈʈattu**p**-**p**uu **p**iɻikkum*
you.obl.dat garden.obl-flower like.3sn
'You like garden flowers.'

38. *unakku **p**uukkaɭ piɻikkum*
you.obl.dat flower.pl like.3sn
'You like flowers.'

39. *kumaar **p**uukkaɭai**p** **p**aattaan(aa)*
Kumar flower.pl.acc see.past.3sm.(qs)
'Kumar saw the flowers.'

40. *kumaar **c**ivappu**p** **p**uukkaɭai**p** **p**aattaan(aa)*
Kumar red flower.pl.acc see.past.3sm(qs)
'Kumar saw the red flowers.'

41. *kumaar (enakku) oru**k** **k**atti kuɻuttaan*
Kumar I.obl.dat one knife give.past.3sm
'Kumar gave (me) a knife.'

42. *iŋkee oru (marra**k**) **k**araɳʈi irukkatu*
here one other spoon be.pres.3sn
OR *marra**k** **k**araɳʈi aŋkee irukkatu*
'There is another/the other spoon.'

43. *kumaar oru **k**araɳʈiyai**p** **p**aattaan*
Kumar one spoon.acc see.past.3sm
'Kumar saw a spoon.'

44. *kumaar avan tampiyai**k*** ***k**oopamaa**p**' **p**aattaan*
Kumar he younger-brother.acc anger.adv see.past.3sm
'Kumar looked angrily at his younger brother.'

45. *inta**p** **p**aattirattai**k** **k**uṭu*
this cup.obl.acc give
'Give (me) this cup.'

46. *kumaar **c**eennaikku**p** **p**ooy**p** **p**aṭiccaan*
Kumar Chennai.dat go.vbp study.past.3sm
'Having gone to Chennai, Kumar studied.'

47. *naay **p**aɻattai**p** **p**aattatu*
dog fruit.obl.acc see.past.3sn
'The dog saw the fruit.'

48. *nii **p**ooy**k** **k**umaarai teeṭaṇum*
you(sg) go.vbp Kumar.acc seek.infin.must
'You have to go and look for Kumar.'

49. *kumaar **p**ooy varuvaan*
Kumar go.vbp come.fut.3sm
'Kumar will go and come.'

50. *inta**p** **p**usattai vaaŋki**p** **p**aṭikkireen*
this book.obl.acc buy.vbp study.pres.1s
'Having bought this book I am studying (it).'

51. *pe**ṭṭ**iyai**t** **t**irntu**k** **k**umaar **p**aɻattai**p** **p**aappaan*
box.acc open.vbp Kumar fruit.obl.acc see.fut.3sm
'Having opened the box Kumar will see the fruit.'

52. *nii **p**aṭicca**p** **p**usattai marakkaatee*
you read.rel book.obl.acc forget.neg.imp
'Don't forget the book which you have read!'

53. *kumaar **p**usattai**p** **p**aṭiccaan*
Kumar book.obl.acc read.past.3sm
'Kumar read the book.'

54. *aŋkee oru **p**aampu irukku**p** **p**oolum*
there one snake be.pres.3sn seem.3sn
'There seems to be a snake there.'

55. *enakku atu **p**ootum*
I.obl.dat that enough.3sn
'That's enough for me.'

56. *inta**t** **t**ami.ɻ**p** **p**usattai**p** **p**aṭi*
this Tamil book.obl.acc read
'Read this Tamil book!'

57. *kumaar **t**ami.ɻ **p**eecinaan*
Kumar Tamil speak.past.3sm
'Kumar spoke Tamil.'

58. *iŋkee oru niir**k**-**k**umi.ɻi irukkatu*
here one water bubble be.pres.3sn
'Here is a water bubble.'

59. *kumaar enakku**k** ku**ʈ**ikka* *taṉṉ**i**ir ku**ʈ**uttaaṉ*
Kumar I.obl.dat drink.infin water give.past.3sm
'Kumar gave me water to drink.'

References

Abbi, Anvita. (1992a). *Reduplication in South Asian languages: an areal, typological and historical study*. New Delhi: Allied Publishers Ltd.

Abbi, Anvita. (1992b). Contact, conflict and compromise: the genesis of reduplicated structures in South Asian languages. In Dimock, Edward C., Kachru, Braj B., & Krishnamurti, Bh. (eds.), 131–148.

Abercrombie, David. (1965). *Studies in phonetics and linguistics*. London: Oxford University Press.

Alderete, John, Beckman, Jill, Benua, Laura, Gnanadesikan, Amalia, McCarthy, John J. & Urbanczyk, Suzanne. (1997). *Reduplication with fixed segmentism*. ROA 226. [Published in 1999 in a shortened version and with revisions in *Linguistic Inquiry* 30. 327–364.]

Anderson, Earl R. (1998). *A grammar of iconism*. Madison, NJ: Fairleigh Dickinson University Press.

Anderson, Stephen R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Andronov, M.S. (1973). Observations on accent in Tamil. *International Journal of Dravidian Linguistics* 2. 111–118.

Annamalai, E. (1968). Onomatopoeic resistance to sound change in Dravidian. In *Studies in Indian Linguistics: M.B. Emeneau Ṣaṣṭipūrti Volume*. Poona & Annamalainagar: Deccan College & Annamalai University.

Annamalai, E. & Steever, Sanford B. (1998). Modern Tamil. In Steever, Sanford B. (ed.), 100–128.

Anton, H. & Hellmann, D. (1975). *Tamil usage in mass media*. Madras: University of Hamburg.

Applegate, Richard B. (1976). Reduplication in Chumash. In Langdon, Margaret & Silver, Shirley (eds.), *Hokan Studies*. The Hague: Mouton. 271–283.

Apte, Mahadev L. (1968). *Reduplication, echo formation and onomatopoeia in Marathi*. Poona: Deccan College.

Arden, A. H. (1934). *A progressive grammar of common Tamil* (4th edn.). Madras: Christian Literature Society. [1st edn. 1891]

Arokianathan, S. (1981). *Tamil clitics*. Trivandrum: Dravidian Linguistics Association.

Aronoff, Mark. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark, Azhar, Arsyad, Basri, Hasan & Broselow, Ellen. (1987). Tier configuration in Makassarese reduplication. In Bosch, Anna, Need, Barbara & Schiller, Eric (eds.), *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society: Parasession on Autosegmental and Metrical Phonology*. Chicago, IL: Chicago Linguistic Society. 1–15.

Aronoff, Mark & Sridhar, S.N. (1983). Morphological levels in English and Kannada or Atarizing Reagan. In Richardson, John F., Marks, Mitchell & Chukerman, Amy (eds.), *Papers from the 19th Regional Meeting of the Chicago Linguistic Society: Parasession on The Interplay of Phonology, Morphology and Syntax*. Chicago, IL: Chicago Linguistic Society. 3–16.

Arunachalam, M. (1977). Echo-words in Tamil. *Journal of Tamil Studies* 11. 1–10.

Asher, R.E. (1982). *Tamil*. Amsterdam: North-Holland.

Austin, Peter. (1981). *A grammar of Diyari, south Australia*. Cambridge: Cambridge University Press.

Baker, Mark C. (1988). *Incorporation: a theory of grammatical function changing*. Chicago: University of Chicago Press.

Balasubramanian, T. (1972). *The phonetics of colloquial Tamil*. Ph.D. dissertation, University of Edinburgh.

Balasubramanian, T. (1980). Timing in Tamil. *Journal of Phonetics* 8. 449–467.

Balasubramanian, T. (1982). The two r's and the two n's in Tamil. *Journal of Phonetics* 10. 89–97.

Balasubramanian, T. & Asher, R.E. (1984). Intervocalic plosives in Tamil. In Higgs, Jo-Ann W. & Thelwall, Robin (eds.), *Topics in Linguistic Phonetics in honour of E.T. Uldall: Occasional Papers in Linguistics & Language Learning* 9. The New University of Ulster: Coleraine. 49–63.

Benua, Laura. (1995). Identity effects in morphological truncation. *Papers in Optimality Theory: University of Massachusetts Occasional Papers* 18. 77–136.

Beschi, Joseph. (1831). *A grammar of the common dialect of the Tamulian language*. Translated by Horst, Christopher Henry. Madras: Vepery Mission Press.

Beythan, Hermann. (1943). *Praktische Grammatik der Tamilsprache*. Leipzig: Otto Harrassowitz.

Bhaskararao, Peri. (1977). *Reduplication and onomatopoeia in Telugu*. Poona: Centre of Advanced Study in Linguistics.

Bird, Steven. (1995). *Computational phonology*. Cambridge: Cambridge University Press.

Blevins, Juliette. (1996). Mokilese reduplication. *Linguistic Inquiry* 27. 523–530.

Bloch, Jules. (1934). *L'indo-aryen du Veda aux temps modernes*. Paris: Adrien-Maisonneuve.

Bloomfield, Leonard. (1933). *Language*. New York: Holt, Rinehart & Winston.

Booij, Geert & Rubach, Jerzy. (1987). Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry* 18, 1–44.

Borer, Hagit. (1988). On the parallelism between compounds and constructs. *Year of Morphology* 1. 45–66.

Bosch, Anna R.K. (1991). *Phonotactics at the level of the phonological word*. Ph.D. dissertation, University of Chicago.

Bosch, Anna R.K. & Wiltshire, Caroline R. (1993). The licensing of prosodic prominence in Tamil. In Stvan, Laurel Smith (ed.), *Papers from the 3rd Annual Meeting of the Formal Linguistics Society of Midamerica*. Bloomington, IN: Indiana University Linguistics Club. 1–15.

Britto, Francis. (1986). *Diglossia: a study of the theory with application to Tamil*. Washington, DC: Georgetown University Press.

Broe, Michael. (1993). *Specification theory: the treatment of redundancy in generative phonology*. Ph.D. dissertation, University of Edinburgh.

Broselow, Ellen. & McCarthy, John J. (1983). A theory of internal reduplication. *The Linguistic Review* 3. 25–88.

Bruening, Benjamin. (1997). Abkhaz Mabkhaz: M-reduplication in Abkhaz, weightless syllables, and base-reduplicant correspondence. In *PF: Papers at the Interface* (*MIT Working Papers in Linguistics*) 30. 291–329.

Butt, Miriam. (1995). *The structure of complex predicates in Urdu*. Stanford: CSLI.

Byrd, Dani, Kaun, Abigail, Naryanan, Shrikanth & Saltzman, Elliott. (2000). Phrasal signatures in articulation. In Broe, Michael B. & Pierrehumbert, Janet B. (eds.), *Papers in Laboratory Phonology* V. Cambridge: Cambridge University Press. 70–87.

Caldwell, R. (1856). *A comparative grammar of the Dravidian or South-Indian family of languages*. London: Harrison & Sons.

Carrier-Duncan, Jill. (1984). Some problems with prosodic accounts of reduplication. In Aronoff, Mark & Oehrle, Richard T. (eds.), *Language sound structure.* Cambridge, MA: MIT Press. 260–86.

Cataacivam, M. (1966). *Olikkurippakaraati* (Dictionary of Onomatopoeic Expressions). Madras: Paari Nilayam.

Chomsky, Noam. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Christdas, Prathima. (1987). On constraining the power of lexical phonology: evidence from Tamil. In McDonough, Joyce & Plunkett, Bernadette (eds.), *Proceedings of the North East Linguistic Society* 17. Amherst, MA: GLSA Publications. 122–146.

Christdas, Prathima. (1988). *The phonology and morphology of Tamil.* Ph.D. dissertation, Cornell University.

Clements, George N. (1985). The problem of transfer in nonlinear morphology. *Cornell Working Papers in Linguistics* 7. 38–73.

Clements, George N. & Sonaiya, Remi. (1990). Underlying feature representation in Yoruba. *Studies in the Linguistic Sciences* 20. 89–103.

Cohn, Abigail C. (1989). Stress in Indonesian and bracketing paradoxes. *Natural Language and Linguistic Theory* 7. 167–216.

Cohn, Abigail C., Ham, William H, & Podesva, Robert J. (1999). The phonetic realization of singleton-geminate contrasts in three languages of Indonesia. In Ohala, John J., Hasegawa, Yoko, Ohala, Manjari, Granville, Daniel & Bailey, Ashlee C. (eds.), 587–590.

Cohn, Abigail C. & McCarthy, John J. (1998). Alignment and parallelism in Indonesian phonology. *Working Papers of the Cornell Phonetics Laboratory* 12. 53–137.

Cole, Jennifer S. & Coleman, John. (1992). No need for cyclicity in generative phonology. In Denton, Jeannette Marshall, Chan, Grace P. & Canakis, Costas P. (eds.), *Papers from the 28th Regional Meeting of the Chicago Linguistic Society: Parasession on the Cycle in Linguistic Theory*. Chicago, IL: Chicago Linguistic Society. 36–50.

Cooper, William E. & Paccia-Cooper, Jeanne. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Corder, S. Pit. (1973). *Introducing applied linguistics*. Harmondsworth, Middlesex: Penguin.

Corré, Alan David. (1962). *The structure of Tamil*. Ph.D. dissertation, University of Pennsylvania.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11. 51–62.

Davis, John F. (1976). Some notes on Luiseño phonology. *International Journal of American Linguistics* 42. 192–216.

Diffloth, Gérard. (1994). i: big, a: small. In Hinton, Leanne, Nichols, Johanna & Ohala, John J. (eds.), *Sound symbolism*. Cambridge: Cambridge University Press. 107–114.

Dimock, Edward C., Kachru, Braj B., & Krishnamurti, Bh. (eds.). (1992). *Dimensions of sociolinguistics in South Asia*. New Delhi: Oxford & IBH Publishing Company.

Di Sciullo, Anna Maria & Williams, Edwin. (1987). *On the definition of word*. Cambridge, MA: MIT Press.

Dixon, Robert. (1977). *A grammar of Yidiɲ*. Cambridge: Cambridge University Press.

Dongare, V.S. (1975). Synonymic compounds in Hindi and Marathi. *Indian Linguistics* 36. 247–255.

Downing, Laura J. (1994). SiSwati verbal reduplication and the theory of generalized alignment. In Gonzàlez, Mercè (ed.), *Proceedings of the North East Linguistic Society* 24. Amherst, MA: GLSA Publications. 81–95.

Downing, Laura J. (1999). Verbal reduplication in three Bantu languages. In Kager, René, van der Hulst, Harry & Zonneveld, Wim (eds.), 62–89.

Dudas, Karen. (1976). *The phonology and morphology of modern Javanese*. Ph.D. dissertation, University of Illinois.

Elfenbein, Josef. (1998). Brahui. In Steever, Sanford B. (ed.), 388–414.

Emeneau, Murray Barnson. (1938). Echo-Words in Toda. *New Indian Antiquary* 1, 109–117. [Reprinted in Emeneau, Murray Barnson. (1967). *Dravidian linguistics, ethnology and folktales: collected papers*. Annamalainagar: Annamalai University. 37–45.]

Emeneau, Murray Barnson. (1955). *Kolami: a Dravidian language*. London: Cambridge University Press.

Emeneau, Murray Barnson. (1956). India as a linguistic area. *Language* 32. 1–17.

Emeneau, Murray Barnson. (1969). Onomatopoeics in the Indian linguistic area. *Language* 45. 274–299.

Emeneau, Murray Barnson. (1980). *Language and linguistic area*. Stanford: Stanford University Press.

Emeneau, Murray Barnson & Hart, Kausalya. (1993). Tamil expressives with initial voiced stops. *Bulletin of the School of Oriental and African Studies* 56. 75–86.

Ferguson, Charles A. (1959). Diglossia. *Word* 15. 325–340.

Ferro-Luzzi, Gabriella Eichinger. (1992). *The taste of laughter: aspects of Tamil humour*. Wiesbaden: Otto Harrassowitz.

Firth, J.R. (1934). A short outline of Tamil pronunciation. Appendix to Arden.

Fitzpatrick-Cole, Jennifer. (1994). *The prosodic domain hierarchy in reduplication*. Ph.D. dissertation, Stanford University.

Fitzpatrick-Cole, Jennifer. (1996). Reduplication meets the phonological phrase in Bengali. *Linguistic Review* 13. 305–356.

Futagi, Yoko. (1998). Root-reduplicant faithfulness. In Curtis, Emily, Lyle, James & Webster Gabriel (eds.), *The Proceedings of the 16th West Coast Conference on Formal Linguistics*. Stanford: CSLI. 207–221.

Gazdar, Gerald & Daelemans, Walter (eds.). (1992). *Computational Linguistics* 18(2&3). (Special issues on inheritance).

Gnanasundaram, V. (1972). On echo words in Tamil. In Subramonian, V.I. (ed.), 247–254.

Goldsmith, John. (1979). *Autosegmental phonology*. New York: Garland Press.

Golston, Chris. (1995). Syntax outranks phonology: evidence from ancient Greek. *Phonology* 12. 343–368.

Goswami, Upendra. (1957). Onomatopoetic and echo-words in Kamrupi. *Indian Linguistics* 17. 161–164.

Gottfried, Michael, Miller, James D. & Meyer, Donald J. (1993). Three approaches to the classification of American English diphthongs. *Journal of Phonetics* 21. 205–229.

Grabe, Esther & Low, Ee Ling. (To appear). Acoustic correlates of rhythm class. In Warner, Natasha & Gussenhoven, Carlos (eds.), *Papers in Laboratory Phonology* 7. Cambridge: Cambridge University Press.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (ed.), *Universals of language*. Cambridge, MA: MIT Press. 58–82.

Haegeman, Liliane. (1994). *Introduction to Government and Binding Theory* (2nd edn.). Oxford: Blackwell.

Halle, Morris & Marantz, Alec. (1993). Distributed morphology and the pieces of inflection. In Hale, Kenneth & Keyser, Samuel Jay (eds.), *The view from Building 20*. Cambridge, MA: MIT Press. 111–176.

Harris, Barbara P. & O'Grady, Geoffrey N. (1976). An analysis of the progressive morpheme in Umpila verbs: a revision of a former attempt. In Sutton, Peter (ed.), *Languages of Cape York*. Canberra: Australian Institute of Aboriginal Studies.

Hawkins, John A. & Gilligan, Gary. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua* 74. 219–259.

Hayes, Bruce. (1989). The prosodic hierarchy in meter. In Kiparsky, Paul & Youmans, G. (eds.), *Phonetics and phonology: rhythm and meter* 1. San Diego: Academic Press Inc. 201–60.

Hayes, Bruce. (1995). *Metrical stress theory*. Chicago: The University of Chicago Press.

Hayes, Bruce & Lahiri, Aditi. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9. 47–96.

Hettiaratchi, D. E. (1959). Echo-words in Sinhalese. *University of Ceylon Review* 17. 47–50.

Hollenbach, Barbara E. (1974). Reduplication and anomalous rule ordering in Copala Trique. *International Journal of American Linguistics* 40. 176–183.

Hyman, Larry M. (1985). *A theory of phonological weight*. Dordrecht: Foris.

Hyman, Larry M. & Mtenje, Al. (1999). Prosodic morphology and tone: the case of Chichewa. In Kager, René, van der Hulst, Harry & Zonneveld, Wim (eds.), 90–133.

Inkelas, Sharon. (1990). *Prosodic constituency in the lexicon*. New York: Garland Press.

Inkelas, Sharon. (1999). Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. In Kager, René, van der Hulst, Harry & Zonneveld, Wim (eds.), 134–187.

Itô, Junko. (1990). Prosodic minimality in Japanese. In Ziolkowski, Michael, Noske, Manuela & Deaton, Karen (eds.), *Papers from the 26th Regional Meeting of the Chicago Linguistic Society: The Parasession on the syllable in phonetics and phonology*. Chicago, IL: Chicago Linguistic Society. 213–239.

Itô, Junko, Kitagawa, Yoshihisa & Mester, Armin. (1996). Prosodic faithfulness and correspondence: evidence from a Japanese argot. *Journal of East Asian Studies* 5. 217–294.

Jenner, Philip N., Thompson, Laurence C. & Starosta, Stanley (eds.). (1976). *Austroasiatic Studies Part II (Oceanic Linguistics Special Publication no. 13)*. Hawaii: University Press of Hawaii.

Johnson, Keith. (1997). *Acoustic and auditory phonetics*. Cambridge, MA: Blackwell.

Jones, Daniel. (1967). *The phoneme: its nature and use*. Cambridge: Cambridge University Press.

Kager, René, van der Hulst, Harry & Zonneveld, Wim (eds.). (1999). *The prosody-morphology interface.* Cambridge: Cambridge University Press.

Karunakaran, K. & Jeya, V. (1995). Language modernisation in Tamil with reference to modern use. In Koul, Omkar N. (ed.), *Sociolinguistics: south Asian perspectives*. New Delhi: Creative Books. 138–144.

Kenstowicz, Michael & Banksira, Degif Petros. (1999). Reduplicative identity in Chaha. *Linguistic Inquiry* 30. 573–585.

Kent, Ray D., Dembowski, James & Lass, Norman J. (1996). The acoustic characteristics of American English. In Lass, Norman J. (ed.), *Principles of experimental phonetics*. St. Louis, MO: Mosby. 185–225.

Kiparsky, Paul. (1982). Lexical morphology and phonology. In Yang, I.-S. (ed.), *Linguistics in the Morning Calm*. Seoul: Hanshin. 3–91.

Kothandaraman, Pon. (1983). A note on the echo words in Tamil. *Journal of Asian Studies* 1. 165–169.

Krishan, Shree (ed.). (1990). *Linguistic traits across language boundaries (a report of All India Linguistic Traits Survey)*. Calcutta: Director General, Anthropological Survey of India.

Krishnamurti, Bh. & Benham, Brett A. (1998). Koṇḍa. In Steever, Sanford B. (ed.), 241–269.

Kuiper, F.B.J. (1974). The genesis of a linguistic area. In Southworth, Franklin C. & Apte, Mahadev L. (eds.), 135–153.

Ladd, Robert D. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3. 311–340.

Ladefoged, Peter & Maddieson, Ian. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lahiri, Aditi & Hankamer, J. (1988). The timing of geminate consonants. *Journal of Phonetics* 16. 327–338.

Leben, William. (1973). *Suprasegmental phonology*. Ph.D. dissertation, MIT.

Lehiste, Ilse. (1985). An Estonian word game and the phonematic status of long vowels. *Linguistic Inquiry* 16. 490–492.

Lehmann, Thomas. (1989). *A grammar of modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.

Lidz, Jeffrey. (2000). Echo reduplication in Kannada: implications for a theory of word-formation. *University of Pennsylvania Working Papers in Linguistics* 6(3). 145–166.

Lieber, Rochelle. (1992). *Deconstructing morphology*. Chicago: University of Chicago Press.

Lindau, Mona, Norlin, Kjell & Svantesson, Jan-Olof. (1985). Cross-linguistic differences in diphthongs. *UCLA Working Papers in Linguistics* 61. 40–44.

Lisker, Leigh. (1958). The Tamil occlusives: short vs. long or voiced vs. voiceless? *Indian Linguistics* 19. 294–301.

Lisker, Leigh. (1972). On stops and gemination in Tamil. In Subramonian V.I. (ed.), 174–179.

Local, John & Simpson, Adrian P. (1988). The domain of gemination in Malayalam. In Bradley, David, Henderson, Eugénie J.A. & Mazaudon, Martine (eds.), *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg*. Canberra: Australian National University. 33–42.

Local, John & Simpson, Adrian P. (1999). Phonetic implementation of geminates in Malayalam nouns. In Ohala, John J., Hasegawa, Yoko, Ohala, Manjari, Granville, Daniel & Bailey, Ashlee C. (eds.), 595–598.

Louali, Naima & Maddieson, Ian. (1999). Phonological contrast and phonetic realization: the case of Berber stops. In Ohala, John J., Hasegawa, Yoko, Ohala, Manjari, Granville, Daniel & Bailey, Ashlee C. (eds.), 603–606.

Mahapatra, K. (1976). Echo-Formation in Gta?. In Jenner, Philip N., Thompson, Laurence C. & Starosta, Stanley (eds.), 815–831.

Malten, Thomas. (1989). *Reduplizierte Verbstämme im Tamil*. Stuttgart: Steiner Verlag Wiesbaden.

Marantz, Alec. (1982). Re reduplication. *Linguistic Inquiry* 13. 435–82.

Marantz, Alec. (1997). No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2). 201–225.

Marchand, Hans. (1952). Alliteration, Ablaut und Reim in den türkischen Zwillingsformen. *Oriens* 5. 60–69.

Marthandan, C.R. (1983). *Phonetics of casual Tamil*. Ph.D. dissertation, University of London.

Masica, Colin P. (1976). *Defining a linguistic area: south Asia*. Chicago: University of Chicago Press.

Masica, Colin P. (1992). South Asia as a sociolinguistic area? In Dimock, Edward C., Kachru, Braj B., & Krishnamurti, Bh. (eds.), 37–46.

Mastroianni, Michael & Carpenter, Bob. (1994). Constraint-based morpho-phonology. In *Computational Phonology: First Meeting of the ACL Special Interest Group in Computational Phonology*. New Mexico: Association for Computational Linguistics. 13–24.

McCarthy John J. (1982). Prosodic templates, morphemic templates, and morphemic tiers. In van der Hulst, Harry & Smith, Norval (eds.), *The structure of phonological representations (Part 1)*. Dordrecht: Foris. 191–223.

McCarthy, John J. (1985). *Formal problems in Semitic phonology and morphology*. New York: Garland Press.

McCarthy, John J. (1999). Sympathy and phonological opacity. *Phonology* 16. 331–399.

McCarthy, John J. & Prince, Alan S. (1986). *Prosodic morphology 1986*. Technical Report 32, Rutgers University Center for Cognitive Science.

McCarthy, John J. & Prince, Alan S. (1988). Quantitative transfer in reduplicative and templatic morphology. *Linguistics in the Morning Calm* 2. Seoul: Hanshin. 3–35.

McCarthy, John J. & Prince, Alan S. (1990). Foot and word in prosodic morphology: the Arabic broken plural. *Natural Language and Linguistic Theory* 8. 209–283.

McCarthy, John J. & Prince, Alan S. (1993a). *Prosodic morphology I: constraint interaction and satisfaction.* Technical Report 3, Rutgers University Center for Cognitive Science.

McCarthy, John J. & Prince, Alan S. (1993b). Generalized alignment. *Yearbook of Morphology* 6. 79–153.

McCarthy, John J. & Prince, Alan S. (1994a). The emergence of the unmarked: optimality in prosodic morphology. In Gonzàlez, Mercè (ed.), *Proceedings of the North East Linguistic Society* 24. Amherst, MA: GLSA Publications. 333–379.

McCarthy, John J. & Prince, Alan S. (1994b). An overview of prosodic morphology part I: template form in reduplication and part II: template satisfaction. Ms. from OTS/HIL Workshop on Prosodic Morphology, University of Utrecht. ROA 59.

McCarthy, John J. & Prince, Alan S. (1995). Faithfulness and reduplicative identity. *Papers in Optimality Theory: University of Massachusetts Occasional Papers* 18. 249–384.

McCarthy, John J. & Prince, Alan S. (1998). Prosodic morphology. In Spencer, Andrew & Zwicky, Arnold M. (eds.), *The handbook of morphology.* Oxford: Blackwell. 283–305.

Mester, Ralf Armin. (1986). *Studies in tier structure.* Ph.D. dissertation, University of Massachusetts.

Milsark, Gary. (1972). Re: doubl-ing. *Linguistic Inquiry* 3. 542–549.

Milsark, Gary. (1988). Singl-ing. *Linguistic Inquiry* 19. 611–634.

Mohanan, K. P. (1986). *The theory of lexical phonology.* Dordrecht: Reidel.

Mohanan, K. P. & Mohanan, Tara. (1984). Lexical phonology of the consonant system in Malayalam. *Linguistic Inquiry* 15. 575–602.

Mohanan, Tara. (1989). *The dual nature of complex predicates in Hindi.* Stanford University ms.

Mohanan, Tara. (1994). *Argument structure in Hindi.* Stanford: CSLI.

Moravcsik, Edith. (1978). Reduplicative constructions. In Greenberg, Joseph (ed.), *Universals of human language (vol. 3).* Stanford: Stanford University Press. 297–334.

Morin, Yves Charles. (1972). The phonology of echo-words in French. *Language* 48. 97–108.

Munro, Pamela & Benson, Peter John. (1973). Reduplication and rule ordering in Luiseño. *International Journal of American Linguistics* 39. 15–21.

Murthy, M. Chidananda. (1975). Formation of echo-words in Kannada. In Reddy, G.N. & Somasekharan Nair, P. (eds.), *Proceedings of the 2nd All India Conference of Dravidian linguists.* Trivandrum: University of Kerala. 158–164.

Mutaka, Ngessimo & Hyman, Larry M. (1990). Syllables and morpheme integrity in Kinande reduplication. *Phonology* 7. 73–119.

Myers, Scott & Carleton, Troi. (1996). Tonal transfer in Chichewa. *Phonology* 13. 39–72.

Nagarajan, Hemalatha. (1994). *A theory of post-syntactic phonology.* Hyderabad: T.R. Publications.

Narayanan, Shrikanth, Byrd, Dani & Kaun, Abigail. (1999). Geometry, kinematics, and acoustics of Tamil liquid consonants. *Journal of the Acoustical Society of America* 106(4). 1993–2007.

Nash, David. (1980). *Topics in Warlpiri grammar.* Ph.D. dissertation, MIT.

Nayar, V.R. Prabodhachandran. (1972). Phonetic and phonological exponents of transitivity in Malayalam. In Agesthialingom, S. & Shanmugan, S.V. (eds.), *3rd Seminar on Dravidian Linguistics.* Annamalainagar: Annamalai University. 205–222.

Nespor, Marina & Vogel, Irene. (1986). *Prosodic phonology.* Dordrecht: Foris.

Noyer, Rolf. (1998). Vietnamese 'morphology' and the definition of word. *University of Pennsylvania Working Papers in Linguistics* 5(2). 65–89.

Odden, David. (1987a). Arguments against the vowel plane in Gtaʔ. *Linguistic Inquiry* 18. 523–529.

Odden, David. (1987b). Kimatuumbi phrasal phonology. *Phonology Yearbook* 4. 13–36.

Odden, David & Odden, Mary. (1985). Ordered reduplication in Kíhehe. *Linguistic Inquiry* 16. 497–503.

Ohala, John J., Hasegawa, Yoko, Ohala, Manjari, Granville, Daniel & Bailey, Ashlee C. (eds.). (1999). *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley, CA: University of California.

Ohala, Manjari. (1983). *Aspects of Hindi phonology*. Delhi: Motilal Banarsidass.

Onn, Farid M. (1976). *Aspects of Malay phonology and morphology: a generative approach*. Ph.D. dissertation, University of Illinois.

Palmer, F.R. (1957). Gemination in Tigrinya. In *Studies in linguistic analysis (special volume of the Philological Society)*. Oxford: Basil Blackwell. 139–148.

Pandey, Pramod Kumar. (1989). Word accentuation in Hindi. *Lingua* 77. 37–73.

Patterson, Trudi A. (1988). Some morphological and phonological interactions in Lakhota. *Studies in the Linguistic Sciences* 18(1). 135–149.

Peperkamp, Sharon. (1997). *Prosodic words*. The Netherlands: HIL Dissertations.

Pierrehumbert, Janet & Nair, Rami. (1996). Implications of Hindi prosodic structure. In Durand, J. & Laks, B. (eds.), *Current trends in phonology* 2. Great Britain: European Studies Research Institute. 549–84.

Po-Ching, Yip. (2000). *The Chinese lexicon*. London: Routledge.

Politzer, Robert L. (1961). Synonymic repetition in late Latin and Romance. *Language* 37. 484–487.

Pope, G.U. (1859). *Tamil handbook* (2nd edn.). Madras: American Mission Press. [1st edn. 1855]

Prince, Alan S. (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.

Prince, Alan S. & Smolensky, Paul. (1993). *Optimality theory: constraint interaction in generative grammar*. Technical Report 2, Rutgers Center for Cognitive Science. (To appear in LI Monograph Series, MIT Press.)

Pulleyblank, Douglas. (1988). Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19. 233–270.

Pulleyblank, Douglas. (1994). Underlying mora structure. *Linguistic Inquiry* 25. 344–353.

Pulleyblank, Douglas. (1998). Yoruba vowel patterns: deriving asymmetries by the tension between opposing constraints. ROA 270.

Pulleyblank, Douglas & Akinlabi, Akinbiyi. (1988). Phrasal morphology in Yoruba. *Lingua* 74. 141–166.

Raghunathan, Vaijayanthi. (1995). The serial verb construction in Tamil. *Pondicherry Institute of Linguistics and Culture Journal of Dravidic Studies* 5. 57–66.

Raimy, Eric. (2000). *The phonology and morphology of reduplication*. Berlin: Mouton.

Rajaram, S. (1980). *Tamil phonetic reader*. Mysore: Central Institute of Indian Languages.

Ramasubramanian, N. & Thosar, R.B. (1971). Synthesis by rule of some retroflex speech sounds. *Language and Speech* 14. 65–85.

Ramus, Franck, Nespor, Marina & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265–292.

Ravisankar, G. (1994). *Intonation patterns in Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.

Regier, Terry. (1998). Reduplication and the arbitrariness of the sign. In Gernsbacher, M. & Derry, S. (eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. New Jersey: Lawrence Erlbaum Associates. 887–892.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence* 13. 81–132.

Reynolds, Elinor. (1998). *Echo words, with particular reference to Hindi*. M.Phil. dissertation, University of Oxford.

Rhenius, C.G.E. (1845). *Tamil grammar* (abridged 2nd edn.). Madras: American Mission Press. [1st edn. 1836]

Roach, Peter. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In Crystal, David (ed.), *Linguistic controversies*. London: Edward Arnold. 73–79.

Ross, John Robert. (1972). 'Doubl-ing.' *Linguistic Inquiry* 3. 61–86.

Saksena, Baburam. (1937). *Evolution of Awadhi*. Allahabad: Indian Press.

Sapir, Edward. (1921). *Language*. New York: Harcourt, Brace & World.

Sapir, Edward. (1930). Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65, nos. 1–3.

Saussure, Ferdinand de. (1987). *Cours de linguistique générale*. Bally, Charles & Sechehaye, Albert (eds.). Paris: Payot. [1st edn. 1907]

Schein, Barry & Steriade, Donca. (1986). On geminates. *Linguistic Inquiry* 17. 691–744.

Schiffman, Harold F. (1983). *A reference grammar of spoken Kannada*. Seattle: University of Washington Press.

Schiffman, Harold F. (1998). Standardization or restandardization: The case for "Standard" Spoken Tamil. *Language in Society* 27. 359–385.

Schiffman, Harold F. (1999). *A reference grammar of spoken Tamil*. Cambridge: Cambridge University Press.

Schiffman, Harold F. & Arokianathan, S. (1986). Diglossic variation in Tamil film and fiction. In Krishnamurti, Bh. (ed.), *South Asian languages: structure, convergence and diglossia*. Delhi: Motilal Banarsidass. 371–381.

Schlindwein, Deborah. (1986). Tier alignment in reduplication. In Berman, Stephen, Choe, Jae-Woong & McDonough, Joyce (eds.), *Proceedings of the North East Linguistic Society* 16. Amherst, MA: GLSA Publications. 419–433.

Selkirk, Elisabeth O. (1980). *On prosodic structure and its relation to syntactic structure*. Bloomington, IN: Indiana University Linguistics Club.

Selkirk, Elisabeth O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, Elisabeth O. (1986). On derived domains in sentence phonology. *Phonology* 3. 371–405.

Selkirk, Elisabeth O. (1995). The prosodic structure of function words. *Papers in Optimality Theory: University of Massachussetts Occasional Papers* 18. 439–469.

Shaw, Patricia A. (1980). *Dakota phonology and morphology*. New York: Garland Press.

Shibatani, Masayoshi & Kageyama, Taro. (1988). Word formation in a modular theory of grammar: postsyntactic compounds in Japanese. *Language* 64. 451–484.

Sietsema, Brian M. (1988). Reduplication in Dakota. In MacLeod, Lynn, Larson, Gary & Brentari, Diane (eds.), *Papers from the 24th Annual Regional Meeting of the Chicago Linguistic Society*. Chicago, IL: Chicago Linguistic Society. 337–352.

Singh, Amar Bahadur. (1969). On echo words in Hindi. *Indian Linguistics* 30. 185–95.

Singh, Rajendra. (1995). *Linguistic theory, language contact and modern Hindustani*. New York: Peter Lang Publishing.

Southworth, Franklin C. (1974). Linguistic stratigraphy of north India. In Southworth, Franklin C. & Apte, Mahadev L. (eds.), 201–223.

Southworth, Franklin C. & Apte, Mahadev L. (eds.) (1974). *Contact and convergence in South Asian languages*. Trivandrum: University of Kerala.

Spencer, Andrew. (1991). *Morphological theory*. Oxford: Blackwell.

Spring, Cari Louise. (1990). *Implications of Axininca Campa for prosodic morphology and reduplication*. Ph.D. dissertation, University of Arizona.

Sridhar, S. N. (1990). *Kannada*. London: Routledge.

Steever, Sanford B. (1987). Tamil and the Dravidian languages. In Comrie, Bernard (ed.), *The world's major languages*. London: Croom Helm. 725–746.

Steever, Sanford B. (ed.) (1998). *The Dravidian languages*. London: Routledge.

Steever, Sanford B. (1998). Introduction to the Dravidian languages. In Steever, Sanford B. (ed.), 1–39.

Stemberger, Joseph P. (1981). Morphological haplology. *Language* 57. 791–817.

Stemberger, Joseph Paul. (1996). The scope of the theory: where does beyond lie? In McNair, Lisa, Singer, Kora, Dobrin, Lise M. & Aucoin, Michelle M. (eds.), *Papers from the 32nd Regional Meeting of the Chicago Linguistic Society: Parasession on Theory and Data in Linguistics*. Chicago, IL: Chicago Linguistic Society. 139–164.

Steriade, Donca. (1988). Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.

Stevens, Kenneth N. & Blumstein, Sheila E. (1975). Quantal aspects of consonant production and perception: a study of retroflex consonants. *Journal of Phonetics* 3. 215–233.

Subramonian, V.I. (ed.). (1972). *Proceedings of the 1st Conference of Dravidian linguists*. Trivandrum: University of Kerala.

Trivedi, G. M. (1990). Echo formation. In Krishan, S. (ed.), 51–82.

Trubetzkoy, N.S. (1969). *Principles of phonology*. Translated by Baltaxe, Christiane A. M. Berkeley, CA: University of California Press. [1st edn. 1939]

Uhrbach, Amy J. (1987). *A formal analysis of reduplication and its interaction with phonological and morphological processes*. Ph.D. dissertation, University of Texas.

Urbanczyk, Suzanne C. (1996). *Patterns of reduplication in Lushootseed*. Ph.D. dissertation, University of Massachusetts.

Vacek, Jaroslav. (1994). On semantic reduplication in Tamil. *Pondicherry Institute of Linguistics and Culture Journal of Dravidic Studies* 4. 145–150.

Vasanthakumari, T. (1989). *Generative phonology of Tamil*. Delhi: Mittal Publications.

Vennemann, Theo. (1971). The phonology of Gothic vowels. *Language* 47. 90–132.

Walldén, Ruth. (1973). Contributions to the study of the pronunciation of i- and e-sounds in Tamil, especially before retroflex consonants. *Orientalia Suecana* 22. 127–181.

Walldén, Ruth. (1974–1975). Contributions to the study of the pronunciation of i- and e-sounds in Tamil, especially before retroflex consonants V. *Orientalia Suecana* 23–24. 200–209.

Walsh, Laura. (1992). Tone in reduplication. In Canakis, Costas P., Chan, Grace P., & Denton, Jeannette Marshall (eds.), *Papers from the 28th Regional Meeting of the Chicago Linguistic Society*. Chicago, IL: Chicago Linguistic Society. 543–553.

Wenk, Brian J. & Wioland, François. (1982). Is French really syllable-timed? *Journal of Phonetics* 10. 193–216.

Wilbur, Ronnie Bring. (1973). *The phonology of reduplication*. Bloomington, IN: Indiana University Linguistics Club.

Wilkinson, Karina. (1988). Prosodic structure and Lardil phonology. *Linguistic Inquiry* 19. 325–333.

Williams, Jeffrey P. (1991). A note on echo word morphology in Thai and the languages of South and South-East Asia. *Australian Journal of Linguistics* 11. 107–11.

Wiltshire, Caroline R. (1992). *Syllabification and rule application in harmonic phonology*. Ph.D. dissertation, University of Chicago.

Wiltshire, Caroline R. (1999). Expressives in Tamil: evidence for a word class. In Singh, Rajendra (ed.), *The Yearbook of South Asian Languages and Linguistics*. New Delhi: Sage Publications. 1191–37.

Winslow, Miron. (1862). *A comprehensive Tamil and English dictionary of high and low Tamil*. Madras: P.R. Hunt.

Yip, Moira. (1982). Reduplication and C-V skeleta in Chinese secret languages. *Linguistic Inquiry* 13. 637–661.

Yip, Moira. (1992). Reduplication with fixed melodic material. *Proceedings of the North East Linguistic Society* 22. Amherst, MA: GLSA Publications. 459–474.

Yip, Moira. (1998a). Identity avoidance in phonology and morphology. In Lapointe, Steven G., Brentari, Diane K., & Farrell, Patrick M. (eds.), *Morphology and its relation to phonology and syntax*. Stanford: CSLI. 216–246.

Yip, Moira. (1998b). Multiple systems of reduplication in a single language: a markedness account. Ms. from talk delivered at Oxford University.

Zide, Norman H. (1976). A note on Gta? echo forms. In Jenner, Philip N., Thompson, Laurence C. & Starosta, Stanley (eds.), 1335–1343.

Zvelebil, Kamil V. (1964). Spoken language of Tamilnad. *Archiv Orientální* 32. 237–264.

Zvelebil, Kamil V. (1970). *Comparative Dravidian phonology*. The Hague: Mouton.

Zvelebil, Kamil V. (1990). *Dravidian linguistics: an introduction*. Pondicherry: Pondicherry Institute of Linguistics and Culture.