MODELLING TONE AND INTONATION IN JAPANESE

by

Tadao Miyamoto
B.A., Momoyama Gakuin Daigaku, 1970
B.A., University of Victoria, 1987

A Thesis Submitted In Partial Fulfillment of the
Requirements of the Degree of

MASTER OF ARTS

in the Department of Linguistics

We accept this thesis as conforming
to the required standard

[redacted]

Dr. H. J. Warkentyne, Supervisor (Department of Linguistics)

[redacted]

Dr. T. E. Hukari, Departmental Member (Department of Linguistics)

[redacted]

Dr. B. M. Young, Outside Member (Department of Pacific and Asian Studies)

[redacted]

Dr. Hiroko Noro, External Examiner (Department of Pacific and Asian Studies)

University of Victoria

Supervisor: Dr. Henry J. Warkentyne

ABSTRACT

The purpose of this thesis is to investigate the acoustic characteristics of tones and the syntactic, semantic, and extra-linguistic nature of intonation in order to design a simple linear phrasing and pitch contour model for a text-to-speech program in Japanese.

In the model, sentences are, first, parsed semi-automatically into intermediate (major) phrases. In doing so, the interactions of syntactic, semantic, and extra-linguistic factors are framed as the algorithmic linear parsing model which creates at least three types of phrasing variations for given sentences.

Once the intermediate phrase boundaries are located, the accentual phrasing algorithm parses intermediate phrases into accentual (minor) phrases by means of examining accentuations and accentual features of words in a two-word-sized-window-cursor. Whenever there is at least one item in the cursor which is underlyingly (originally) accented, an accentual phrase boundary is inserted in between the items.

Based on Pierrehumbert and Beckman's (1989) mechanism, relevant tones are assigned to the accentual phrases. These tones are then scaled in transform spaces to have vertical spatial values. The scaled tones are linearly interpolated to draw a schematic pitch contours, completing the operations which provides a speech synthesis program with natural pitch contours.

Supervisor: Dr. Henry J. Warkentyne

ABSTRACT

Examiners:

[redacted]

Dr. H. J. Warkentyne, Supervisor (Department of Linguistics)

[redacted]

Dr. T. E. Hukari, Departmental Member (Department of Linguistics)

[redacted]

Dr. B. M. Young, Outside Member (Department of Pacific and Asian Studies)

[redacted]

Dr. Hiroko Noro, External Examiner (Department of Pacific and Asian Studies)

CONTENTS

Abstract ii

Contents iv

Tables viii

Figures ix

Acknowledgement xiv

Dedication xv

Chapter 1: Introduction 1

1.1 Aim 1
1.2 Approach 1
1.3 Topics 2

Chapter 2: Theoretical Background 6

2.1 Two Main Schools of Intonational Models 7
2.2 Pierrehumbert and Beckman's Hierarchical Autosegmental Tree Representation of Utterances 9
2.3 Comparison of Pierrehumbert and Beckman's Hierarchical Autosegmental Model with Other Models 12
2.3.1 Representations of Tones 12
2.3.1.1 Traditional Description of Tones 12
2.3.1.2 Kawakami's Tone Representation 14
2.3.1.3 Pierrehumbert and Beckman's Description of Tones 16
2.3.2 Description of Structural Components 18
2.3.2.1 McCawley's Model 20
2.3.2.2 Poser's Model 20
2.3.2.3 Pierrehumbert and Beckman's Model 21
2.4 Autosegmental Association 22
2.5 Tone Scaling 26
2.6 Concluding Remarks 30

Chapter 3: Accentual Phrasing **36**
3.1 Accentual Behaviour of Postpositions 36
3.2 Experiment on Accentual Phrasing 40
3.2.1 Aim of Experiment 40
3.2.2 Procedures 41
3.2.3 Summary Tables of Accentual Phrasing 45
3.2.4 Unmarked Phrasing 49
3.2.5 Marked Phrasing 51
3.2.6 A Psycholinguistic Implication 54
3.2.7 Algorithmic Rules of Accentual Phrasing and Tone Assignment 56
3.2.8 The Operation of Accentual Phrasing and Tone Assignment 59
3.2.9 Tone Scaling 63
3.2.10 Summary 66
3.3 Experiment on Narrow-focusing 77
3.3.1 Narrow-focus on the Modifiers 78
3.3.2 Narrow-focus on the Nouns 84
3.3.3 Narrow-focus on the Postpositions 86
3.3.4 Summary 90

Chapter 4: Catathesis **94**
4.1 What is Catathesis? 95
4.1.1 Kubozono's Claim 100
4.2 Experiment on the Correlation between Iterative Application of Catathesis and Syntactic Trees 102
4.2.1 Design of Stimuli 102
4.2.2 Catathesis Constants 108
4.2.3 Timing Differences in Triggering Catathesis 114
4.2.4 Right- and Left-branching Trees and Chains of Catathesis 115
4.2.5 Tree Overlays 116
4.2.6 Comparison with Kubozono's Results 121
4.3 Additional Experiment 124
4.3.1 Blocking? 124
4.3.2 Hypothesis Test 127

4.3.3 Results 128
4.4 Summary 132

Chapter 5: Intermediate Phrasing **135**
5.1 Previous Works 138
5.1.1 Selkirk's Intonational Phrasing (Semantic-based Approach) 138
5.1.2 Suzuki and Teranishi's Syntax-based Approach 141
5.1.3 Nespor and Vogel's Intonational Phrasing 143
5.1.4 Choice of Nespor and Vogel's Approach 147
5.2 Application of Nespor and Vogel's Approach to Japanese 148
5.2.1 Basic Framework 148
5.2.2 Performance Variables 152
5.2.2.1 Length as a Variable 153
5.2.2.2 Style of Speech as a Variable 155
5.2.3 The Syntactic and Semantic Factors in Obligatory Category ... 157
5.2.3.1 Extras 157
5.2.3.2 Subordinate Clauses 158
5.2.3.3 Lists 160
5.2.3.4 Embedded Structures 162
5.2.3.5 Coordination 163
5.2.3.6 Contrastive Prominence 164
5.2.4 The Syntactic and Semantic Factors in Variable Dependent Category 165
5.2.4.1 Theme 165
5.2.4.2 Argument Structures 167
5.2.4.3 Wide Scope 171
5.2.4.4 Every NP&ADP 172
5.3 Operations of the Intermediate Phrasing Model 173
5.3.1 Three Levels of Developing Phrasing Variations 174
5.3.1.1 The Phrasing at the Level of ([+Obligatory] [-Variable Dependent]) 174
5.3.1.2 The Phrasing at the Level of ([+Obligatory] [+Variable Dependent]([+Length] [Informal]) 175

5.3.1.3 The Phrasing at the Level of ([+Obligatory] [+Variable Dependent]([+Length] [Formal]) 175
5.3.2 Notations and Algorithmic Rules 176
5.3.2.1 Dictionary 176
5.3.2.2 Word Tagging 177
5.3.2.3 Algorithmic Rules (Pattern Matching) 177
5.3.2.4 Direction in Searching for Patterns in a Text 181
5.3.3 Demonstration 182
5.3.3.1 The Operations at the Level of ([+Obligatory] [-Variable Dependent]) 183
5.3.3.2 The Operations at the Level of ([+Obligatory] [+Variable Dependent]([+Length] [Informal]) 187
5.3.3.3 The Operations at the Level of ([+Obligatory] [+Variable Dependent]([+Length] [Formal]) 195
5.4 Concluding Remarks 199

Bibliography **205**

Appendix: Stimuli for Experiment: 4.3 ..... **210**

TABLES

2.1 Higher Level Prosodic Units and Downtrends Postulated in Three Japanese Models 19

3.1 Accentual Types of Non-monomoraic Postpositions 37

3.2 A List of Stimuli Used in the Experiment on Accentual Phrasing 42

3.3 The Results of Accentual Phrasing of the Phrases with *Oma'me* 45

3.4 The Results of Accentual Phrasing of the Phrases with *Nimame* 46

3.5 Insertion and Non-insertion of an Accentual L% Before the Narrow-focused postpositions 87

4.1 Stimuli for the Experiment Investigating the Correlation Between the Syntactic Tree Branchingness and the Iterative Application of Catathesis 103

4.2 Computations of Local and Accumulative Catathesis Constants 110

4.3 Comparison of the Accumulative Catathesis Constants Between Those of the VP Embedding the Left-branching NPs with Double *No*-phrase Modifiers and Those of the VP Embedding the Right-branching NPs 120

5.1 A List of Three Major Components in the Intermediate Phrasing Model 149

5.2 A list of Algorithmic Intermediate Phrasing Rules 178

5.3 A list of Notations Used in the Algorithmic Intermediate Phrasing Rules 181

FIGURES

2.1 Diagrammatic Representation of Pierrehumbert and Beckman's Prosodic Model 7

2.2 Hierarchical Tree Representation of an Utterance of *Ao'i Okama'de Ooisogide Ikima'suka* 10

2.3 Phoneme Association to the Mora and Syllable Nodes in the Prosodic Tree for a Part of the Utterance in Figure 2.2 10

2.4 Traditional Tonal Description of the Words: *Oba'asama* "Grandmother", *Ka'geboosi* "Silhouette", *Kyooikusha* "Educator", and *Omosirozuku* "the Fun of the Thing" 12

2.5 Pierrehumbert and Beckman's Rendition of Tone Patterns for the Words in Figure 2.4 14

2.6 Tonal Representation of *Arewa U'makatta* by Way of the Traditional Description and by Way of Kawakami's Kinetic (Dynamic) Description 15

2.7 Association of the Bitonal Accent Tone, HL, to an Accented Mora in *Ao'i* "Blue"; *Ma'de* "To"; and *Ikima'suka* "Do (You) Go?" 23

2.8 Association of the Phrasal H in the Accentual Phrases: *Okama'de* "To the Hill"; *Ooisogide* "Quickly"; and *Ikima'suka* "Do (You) Go?" 24

2.9 Association of the Accentual Phrase Final L% to the Initial Mora of the Following Accentual Phrase in *Ao'i Okama'de* "To the Blue Hill" and *Ooisogide Ikima'suka* "Do (You) Go Quickly?" 25

2.10 Association of the L% to the Initial Mora of the Initial Accentual Phrase in the Utterance of *Ao'i Okama'de Ooisogide Ikima'suka* "Do You Go Quickly to the Blue Hill?" 25

3.1 A Schematic Pitch Contour of (1), +/ + − : L% HL L% HL L% 68

3.2 A Schematic Pitch Contour of (2), +/ − + : L% HL L% H HL L% 69

3.3 A Schematic Pitch Contour of (3), +/ − − : L% HL L% H H L% 70

3.4 A Schematic Pitch Contour of (4), −/ + − ; −/ − + : L% H L% HL L% 71

3.5 A Schematic Pitch Contour of (5), [− − +] : L% H HL L% 72

3.6 A Schematic Pitch Contour of (6), −/ − − : L% H L% H L% 73

3.7 A Schematic Pitch Contour of (7), [− − −] : L% H H L% 74

3.8 A Schematic Pitch Contour of (8, Marked Phrasing), [+ − −] : L% HL L% 75

3.9 A Schematic Pitch Contour of (9, Marked Phrasing), [+ − +]: L% HL HL L% 76

3.10 Schematized Pitch Contours Which are Due to the Narrow-focusing the Phrase Initial Accented Modifier 79

3.11 Allophony of F0 Contours for the Phrases with At Least Two Underlying Accents 80

3.12 Schematized Pitch Contours Which Are Due to Narrow-focusing the Phrase Initial Unaccented Modifiers 81

3.13 Schematized Pitch Contours of the Phrases with the Narrow-focused Accented *No*-phrase Modifier, *A'ni-no* 82

3.14 A F0 Contour of *A'ni-no* L% *OMA'ME -Jyuu* 84

3.15 Schematic F0 Contours of *Ao'i-OMA'(ME----)* and *Ao'i-NIMA(ME----)* 85

3.16 A F0 Contour of *A'ni-no-OMA'ME-Gu'rai* 86

3.17 The Appearance of Accentual L% Before the Narrow-focused Postposition in *Ao'i* L% *Oma'me* L% *MA'DE* 87

3.18 A F0 Contour of *Ao'i* + *Oma'me* + *Ni* "To the Blue Beans" 89

3.19 A F0 Contour of *A'ni-no* + *Nimame* + *Ni* "To the Cooked-beans" 89

4.1 Mean Peak F0 Values of Adjective and Noun in the Phrases (1) *Ao'i* (Accented)- *U'mi* "Blue Sea" and (2) *Arai* (Unaccented)- *U'mi* "Rough Sea" 97

4.2 F0 Value of the Peak in the Noun (*U'mi*) Plotted Against the Peak in the Preceding Adjective (*Ao'i* or *Arai*) 98

4.3 Kubozono's Stimuli for the Left-branching NP and for the Right-branching NP As Well As the Corresponding Schematic Pitch Contours 100

4.4 A F0 Contour of an Utterance of *Ao'i- U'mi* "Blue Sea" 106

4.5 Accumulative Catathesis Constants and Local Catathesis Constants for the VP Embedding the Right-branching NPs 111

4.6 Accumulative Catathesis Constants and Local Catathesis Constants for the VP Embedding the Left-branching Symmetrical NPs 112

4.7 Accumulative Catathesis Constants and Local Catathesis Constants for the VP Embedding the Left-branching Relative Clause NPs 113

4.8 F0 Value of the Third Peak (P^3) Plotted Against That of the Second Peak (P^2) in Three Different Phrasal Types of the NPs All of Which are Right-branching 117

4.9 F0 Value of the Third Peak (P^3) Plotted Against That of the Second Peak (P^2) in Three Different Phrasal Types of the Left-branching Relative Clause NPs 118

4.10 F0 Value of the Third Peak (P^3) Plotted Against That of the Second Peak (P^2) in Three Different Phrasal Types of the Left-branching Symmetircal NPs 119

4.11 Schematic Catathesis Contours of the Left-branching NPs and Right-branching NPs Obtained from Kubozono's Experiment and from the Current Experiment 121

4.12 Different Prosodic Tree Renditions for the Two Utterances of *Ta'roo-no Waka'reta Okusan* "Taro's Divorced Wife" 123

4.13 Syntactic Tree Illustrations of a Working Hypothesis That the Right-node of a Mutually C-commanding Unit in the NPs Does Not Participate in the Chain of Catathesis 125

4.14 Hypothetical Applications of Catathesis and Final Lowering in the VP Embedding the Left-branching NPs and in the Right-branching NPs 126

4.15 F0 Values of P^4 Plotted Against F0 Values of P^2 in the VP Embedding the Right-branching NPs 128

4.16 F0 Values of P^3 Plotted Against F0 Values of P^1 for the Left-branching Symmetrical NPs 130

4.17 F0 Values of P^3 Plotted Against F0 Values of P^1 for the Left-branching Relative Clause NPs 131

ACKNOWLEDGEMENTS

I wish to express my thanks and appreciation to my supervisor, Dr. Henry J. Warkentyne for his support not only to this study but also throughout my undergraduate and graduate studies at the Department of Linguistics.

I wish to extend my thanks to my committee members, Dr. Thomas E. Hukari and Dr. B. Morgan Young.

I am very grateful also to Dr. John H. Esling and Dr. Stephen J. Eady who constantly encouraged me to keep work on this study.

Thanks also to Dr. Yong-Bum Kim, with whom friendship I had enjoyed till he returned to Korea recently, who helped me in the syntactic tree analyses in Japanese.

Thanks also to Kimiko Tsukada for her help in Tokyo accent and to Laura Proctor for her advice on computational notations.

My special thanks is for Andrea Giles who read carefully the whole thesis to correct the typographical as well as grammatical errors. Without her help, my thesis writing would have been a far more difficult one.

Lastly, I wish to extend my thanks and appreciation to British Columbia Science Council for providing me, a non-science student, with two years of generous award. Without the financial support, my M.A. program would have taken much longer to finish.

DEDICATION

To my parents, Takaji and Tomiko Miyamoto, and to my wife, Susan.

Chapter 1
Introduction

1.1 Aim

The purpose of this thesis is to investigate the acoustic characteristics of tones and the syntactic, semantic, and extra-linguistic nature of intonation in order to design a simple phrasing and pitch contour model in Japanese. The aim of designing such a model is to assist the development of a text-to-speech synthesis program which is small enough to be implemented on a personal computer. One meaningful application of the speech synthesis program is the development of a Japanese language instruction software. The speech synthesis program should be able to convert a fairly simple typed-in text to a human-like sound using such a technology as LPC-encoded[1] demisyllables (Browman, Fujimura and Ohira 1982; cf. Klatt 1987).

Although linguistic theories and notions are employed throughout the thesis, it must be made clear that it is not the aim of this study to develop a coherent linguistic theory of Japanese prosody.

1.2 Approach

In developing a speech synthesis program, there must be linguistic input which assists the synthesized speech to have a natural, non-machine-like pitch contour. The way to create such a pitch contour is, of course, to model a contour which is observed in natural speech. One way of doing so is to: (i) parse a sentence into different sizes of phrases; (ii) mark the phrases by tones which will be set at strategic points; and (iii) linearly interpolate these tones to draw a schematic pitch contour which is close to a natural contour. In general, this is what Pierrehumbert and Beckman's (1989) prosodic model does and it is the model on which I will mainly rely.

1.3 Topics

In Chapter 2, Pierrehumbert and Beckman's prosodic model is examined. The aim is to understand the nature of Japanese tones and to acquire tools to create a basic framework of the phrasing and pitch contour model suited for speech synthesis.

Pierrehumbert and Beckman's model has some unique features. The first unique feature is that in their representation of an utterance in Tokyo Japanese, tones are sparsely specified on the surface representation of an utterance. The second unique feature is that an utterance is represented as a hierarchical prosodic tree which incorporates three major components of an utterance; i.e., the segmental component, the tonal component, and the structural component. The third unique feature is that the model is not merely a phonological model but has a device which transforms the tree representation of an utterance into a schematic pitch contour suited for speech synthesis. All these features of Pierrehumbert and Beckman's model are examined. Also, the model is compared with some other phonological and phonetic models.

Chapter 3 investigates how a larger intonational phrase, which, in general, is marked by a pause, is parsed into smaller, accentual phrases which, very roughly speaking, correspond in size to that of a content word or two. An acoustic experiment is conducted to discover the factor which determines the accentual phrasing. That is, is the conditioning factor of the accentual phrasing syntactic or phonological? This finding is, then, implemented in an accentual phrasing and pitch contour model which can automatically parse an intonational phrase into accentual phrases and assign relevant tones to the phrases to create a schematic pitch contour of the intonational phrase. Prior to the discussion of the above main issue, there is a discussion of the accentual characteristics of Japanese postpositions (particles) because they play significant roles in the accentual phrasing. Following the main issue, then, there is another experiment to investigate the effect of narrow-focus on the accentual phrasing.

What is investigated in Chapter 4 is the correlation between the iterative application of catathesis (pitch compression caused by an accent) and syntactic trees. Despite the claims of Poser (1984) and of Pierrehumbert and Beckman (1989) that an accent conditions catathesis in any environment, Kubozono (1985) makes an interesting claim that catathesis is blocked in a right-branching phrase. An experiment is conducted to examine whether there is any difference between the iterative application of catathesis in right-branching and left-branching phrases. If the difference is found and the blocking of catathesis is observed, another experiment is conducted to examine whether catathesis is actually blocked or that the catathesis effect is merely obscured by a syntactic configuration of a phrase.

In Chapter 5, based on Nespor and Vogel (1986), a simple algorithmic intonational phrasing model is designed. The model is simple in the sense that it can parse simple sentences but cannot parse complex, embedded sentences into intonational phrases. This limited capability of phrasing is due to the linear parsing which regards a sentence merely as a string of lexical items which are tagged with such information as parts-of-speech classifications and syntactic and sematntic features. Sentences are phrased into intonational phrases with no recourse to hierarchical syntactic structures of the sentences. Although primitive, the model has the capability of predicting the locations of boundaries which may vary depending upon the interaction of the syntactic, semantic, as well as extra-linguistic factors, such as length of sentence or style of speech. To design thus such a phrasing model, the following two issues are examined: (i) the syntactic and semantic factors which condition or may condition an intonational phrase boundary; and (ii) the effect of the extra-linguistic factors on the syntactic and semantic factors in conditioning boundaries. Examples are used to demonstrate: (i) how these syntactic, semantic and extra-linguistic factors interact with each other to provide variations in the intonational phrasing; and (ii) how algorithmic rules identify these syntactic and semantic factors and insert intermediate phrase boundaries at appropriate locations in the text of a sentence.

In an application, the above algorithmic intonational phrasing model is attached to the accentual phrasing and pitch contour model developed in Chapter 3 so that a typed-in text will, if successful, create schematic pitch contours which resembles those observed in natural speech once the whole model is implemented in a speech synthesis program.

Endnotes

[1] Linear prediction of speech wave (cf. Atal and Hanauer 1971).

Chapter 2
Theoretical Background

The issue of this chapter is to examine Pierrehumbert and Beckman's (1989) prosodic model in comparison with other engineering and phonological models. The aims of doing so are:
(i) to understand the characteristics of Japanese prosody;
(ii) to obtain tools to investigate such issues as accentual phrasing and its conditioning factor (in Chapter 3) and the correlation between catathesis (pitch compression) and syntactic tree branchingness (in Chapter 4); and
(iii) to acquire a basic framework to design an algorithmic phrasing and pitch contour model for speech synthesis.

Pierrehumbert and Beckman's model is binate in that, as a speech synthesis model, it is a target level model and as a phonological model, it is a hierarchical autosegmental model. The hierarchical autosegmental model is a sophisticated prosodic model and is autonomous in that it does not depend on the target level model. Contrary, the target level model depends on the hierarchical autosegmental model to function as an application model to speech synthesis. While employing Ladd's (1983) classification of intonational models, I will introduce Pierrehumbert and Beckman's model.

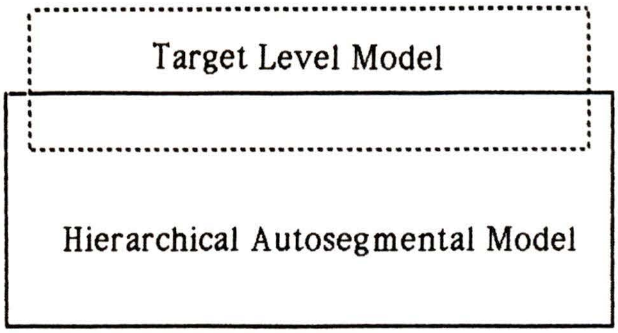


Figure 2.1
Diagrammatic representation of Pierrehumbert and Beckman's prosodic model.

2.1 Two Main Schools of Intonational Models

In spite of the common assumption that a continuously changing contour can be regarded as a realization of a series of discrete units, according to Ladd (ibid), in the study of intonational models (applied to speech synthesis), there are two main schools of thought, i.e., the Contour Interaction (CI) approach and the Tone Sequence (TS) approach. The CI approach (Cohen and t'Hart 1967; O'Shaughnessy 1979; O'Shaughnessy and Allen 1983; etc.), takes either a phrasal or sentential contour as a basic unit of intonation. Factors affecting fundamental frequency (F0)[1], such as focus or segmental effects, are superimposed on the basic contour.

In O'Shaughnessy's (1979) feature model, for example, a sentential F0 contour is regarded as a product of interactions between the two basic features, i.e., "F0 rise" and "F0 fall" at four different prosodic levels, i.e., the phoneme, the word, the phrase, and the sentence. At the phonemic level, a "F0 rise" represents a F0 effect of a voiced consonant whereas a "F0 fall" represents that of a voiceless consonant. At the word level, a "F0 rise" represents an emphatic stress contour before a nucleus whereas a "F0 fall" represents the one after the nucleus. At the phrasal level, syntactic boundaries are represented either by a non-stress "continuous rise" or by a stress "broad fall and rise". Finally, at the sentential level, the features represent a terminal F0 fall of a statement and a terminal F0 rise of a question.

Fujisaki and Hirose's (1980) sentential model for Tokyo Japanese is another type of CI model, evolved from Fujisaki's earlier word contour model (Fujisaki and Sudo 1971, Fujisaki and Sugito 1978) based on the framework laid out by Ohman and Lindqvist (1966). In the word model, a word contour is treated as the product of interactions between a baseline which arises at the beginning of a word and declines steadily towards its end and an accent contour which rises and declines rapidly at the location of accentuation. Basically, the same mechanism is applied to synthesizing a sentential F0 contour which is regarded as an exponential function of the control signals of two components, i.e., a phrasal component which is a phrase-sized declining contour and an accent component which is a local F0 hump of a pitch accent.

The second school of intonational models takes a different approach which is based on the inventory of phonological units, such as High and Low tones and analyzes a contour as a series of tones which are associated with accents and boundaries. Pierrehumbert and Beckman's target level model belongs to the TS (tone sequence) approach and originates, in many aspects, in Pierrehumbert (1980), Liberman and Pierrehumbert (1984), Anderson, Pierrehumbert and Liberman (1984), Beckman, Hertz and Fujimura (1983) and Beckman (1986). In Pierrehumbert and Beckman's target level model for Tokyo Japanese, tones which function as target levels are linearly interpolated to create a schematic pitch contour. Non-local F0 effects, i.e., declination and final lowering, are applied on the pitch contour which will then be filtered to gain a smooth contour close to the one observed in real speech. One of the aspects of Pierrehumbert and Beckman's framework is that their target level model is not a mere speech synthesis model but constitutes a part of an innovative phonological model, i.e., a hierarchical autosegmental model.

2.2 Pierrehumbert and Beckman's Hierarchical Autosegmental Tree Representation of Utterances

In the hierarchical autosegmental model, an utterance is represented as a prosodic tree, such as the one shown in Figure 2.2, which incorporates three major components of the utterance, i.e., the segmental component, the tonal component and the structural component.

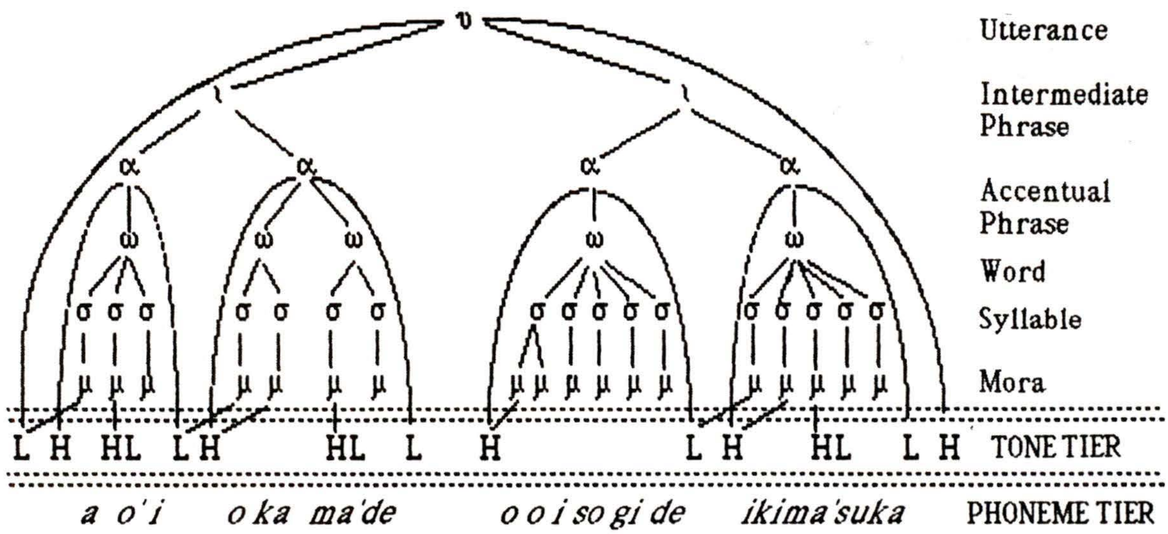


Figure 2.2
Hierarchical tree representation of an utterance of the phrase, *ao'i okama'de ooisogide ikima'suka* "Do you go quickly to the blue hill?" The prosodic tree consists of three major components: i.e., segmental component (phoneme tier), tonal component (tone tier), and structural component (six prosodic units).

It is, perhaps, unnecessary to say anything about the segmental component other than that in the prosodic tree, vowels and mora-nasals are associated with the lowest level of the structural component, i.e., the mora and consonants are associated with second lowest level of the component, the syllable, as shown in Figure 2.3.

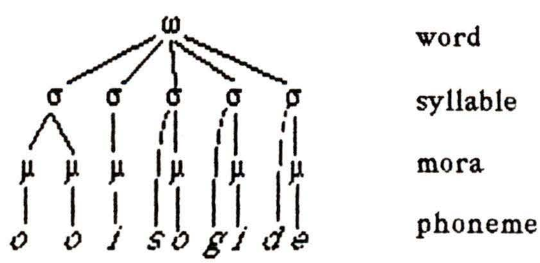


Figure 2.3
Segment, or phoneme, association to the mora and syllable nodes in the prosodic tree for a part of the utterance in Figure 2.2. In the tree, vowels are associated to the mora nodes whereas consonants are associated with syllable nodes.

The structural component consists of six prosodic units, i.e., the mora (μ), the syllable (σ), the word (ω), the accentual phrase (α), the intermediate phrase (ι), and the utterance (υ). Among these, the three higher levels of the prosodic units play major roles in contouring an utterance because each is a domain of specific F0 rises or F0 falls. That is, the accentual phrase is the domain of an initial delimitative F0 rise and a possible occurrence of an accent. The intermediate phrase, which may contain more than one accentual phrase, is the domain of catathesis which iteratively compresses the pitch range at each accent H.[2]. The utterance is, then, the domain of a declination as well as the domain of a final F0 lowering for a statement and a final F0 raising for a question.

The third, tonal, component consists of four types of tones with different structural origins, i.e., an accent H which is a property of an accented word; a phrasal H which originates at the level of an accentual phrase, an interrogative H boundary tone which is inserted at the level of an utterance, and a L tone which can be a boundary tone for an accentual phrase, an intermediate phrase, or an utterance.

Having provided a brief description of Pierrehumbert and Beckman's prosodic model, I would like to look in detail at the mechanisms of the model by postulating the following questions: (i) How does Pierrehumbert and Beckman's tonal description differ from the tonal description by traditional phonologists? (ii) What are the characteristics of the tones which constitute the tonal component? (iii) What are the characteristics of the higher level prosodic units of the structural component and how do they differ from the equivalent prosodic units proposed by other scholars? (iv) How are the tonal component and the structural component associated with each other in the prosodic tree? Lastly, (v) How can the information represented as a well-formed prosodic tree be passed onto the target level model for an application?

2.3 Comparison of Pierrehumbert and Beckman's Model with Other Models

2.3.1 Representations of Tones

2.3.1.1 Traditional Description of Tones

Regarding tonal description, Pierrehumbert and Beckman claim that in Tokyo Japanese, tones are sparsely specified even at the surface level. This underspecification of tones proposed by Pierrehumbert and Beckman presents quite a different picture in the tonal representation from the traditional description, such as by Kindaichi (1967), which assigns either a High or a Low tone to each tone-bearing unit exhaustively, as seen in Figure 2.4.

Accented	*o ba' a sa ma* L H L L L	
	ka' ge bo o si H L L L L	(initial accented)
Unaccented	*kyo o i ku sha* H H H H H	(heavy initial syllable)
	o mo si ro zu ku L H H H H H	

Figure 2.4
Traditional tonal description of the words: *oba'asama* "grandmother", *ka'geboosi* "silhouette", *kyooikusha* "educator", and *omosirozuku* "the fun of the thing".

In the traditional tonal description, the following basic characteristics are observed:

(i) Words are divided into two groups: accented words which have a pitch shift from H to L somewhere in their domain and unaccented words which do not have such a pitch shift.

(ii) Word initial syllables are classified into two categories: the heavy initial syllable, containing two sonorant morae (i.e., a long vowel, a diphthong, or a vowel + a mora nasal) and the light initial syllable (i.e., a short vowel or a short vowel followed by a geminate obstruent).

(iii) Once the above features are specified, the assignment of H and L tones to tone-bearing units is rule-governed. That is, in the case of an accented word, first, assign a L tone to every tone-bearing unit after the accented mora. Next, assign a H tone to the accented mora and to every tone-bearing unit preceding it except to an initial light syllable which is assigned a L tone. In the case of an unaccented word, assign a H tone to every mora except to a light initial syllable which is assigned a L tone.

Such an exhaustive tonal specification is respected even by autosegmental phonologists, such as Haraguchi. In his autosegmental framework of Japanese, Haraguchi (1977) exhaustively assigned a tone to every tone-bearing unit by one of the Well-Formedness Conventions: "[a]ll tones should be associated with at least one tone-bearing unit, and conversely, all tone-bearing units should be associated with at least one tone in the tone melody" (ibid: 55).

One of the important claims made by Pierrehumbert and Beckman in the study of Tokyo Japanese is that "the phonetic rules of Japanese operate on a surface string of tones that are very sparsely distributed relative to the number of tone-bearing units in the text of an utterance" (Pierrehumbert and Beckman 1989: 5). Pitch patterns of the areas with no specific tones are derived from the interpolation between the target values of two tones. According to Pierrehumbert and Beckman's tonal specification, the tone patterns of the words in Figure 2.4 are represented as in Figure 2.5.

```
   o  ba' a sa ma
  /   |
L%   HL         L%

   ka' ge bo o shi
    |
L%HL             L%

  kyo o i ku sha
     |
L% H             L%

  o mo si ro zu ku
 /     |
L%     H          L%
```

Figure 2.5
Pierrehumbert and Beckman's rendition of tone patterns for the words in Figure 2.4.

2.3.1.2 Kawakami's Tone Representation

In fact, it must have been Kawakami (1961, 1962, 1966) who first claimed, based on acoustic observations, that the traditional tonal description by way of exhaustively assigning a H or L tone to each tone-bearing unit is not an adequate way of representing tone patterns, especially, those of phrases. To remedy the inadequacy, Kawakami used kinetic (dynamic) tone markers: "┌" for an initial rise, "┐" for an accent fall, "↘" for a gentle fall or declination and " | " for an interphrasal boundary.

(a) *a re wa u' ma ka t ta*
LH H H L L LL

(a') *a┌rewa u┐makatta*

(a'') *a↘rewa | ┌u┐makatta*

Figure 2.6
Tonal representation of the phrase, *arewa u'makatta* "that was tasty" by way of the traditional description on the left and by way of Kawakami's kinetic (dynamic) description on the right.

As seen in Figure 2.6, Kawakami's description is capable of representing the allophonic phrasal contours conditioned by the presence and absence of an interphrasal boundary as well as representing the initial delimitative rise after the boundary in (a''). Obviously, the traditional tonal description is not endowed with such a representative power and fails to describe the tone pattern of (a'') which has a declination starting from the phrasal maxima at *re* to the boundary minima at *wa* in *arewa* and an acute initial delimitative rise at the very beginning of the following phrase, *umakatta*. Although there is no comparison with Pierrehumbert and Beckman's model in terms of sophistication, Kawakami's tone representation exhibits some similarities to that of Pierrehumbert and Beckman in that (i) the initial delimitative rise is properly represented; (ii) the tone patterns of the areas with no tonal specifications are derived by the linear interpolation between a phrasal maxima and a boundary minima; and (iii) phrasing plays a significant role in determining tone patterns.

2.3.1.3 Pierrehumbert and Beckman's Description of Tones

Based on the assumptions that in Tokyo Japanese, tones are sparsely distributed and that pitch contours of the areas without tones arise from the linear interpolation between the assigned tones, Pierrehumbert and Beckman claim that it is feasible to represent the pitch contours of any utterances with the inventory of just four types of tones, i.e., the phrasal H, the bitonal accent HL, the utterance final H boundary tone, and the L boundary tone.

2.2.1.3.1 H Tones

First, the phrasal H, which is inserted at the accentual phrase level and which represents the target for the end of an initial delimitative rise, is usually linked to the second mora of an accentual phrase[3], as in *o mo*-H*shirozuku* "fun of the thing". However, if an accentual phrase begins with a heavy syllable, the phrasal H is linked to the initial sonorant mora, as in *kyo*-H*oikusha*[4] "educator". The linking of the phrasal H applies not only to unaccented accentual phrases but also to accented accentual phrases. In the case of the accented accentual phrases, however, the F0 effect of the phrasal H becomes apparent only when it is well separated in distance from an accent in a word such as *a ma*-H *zup pa*-HL*i* "sweet and sour". Furthermore, the linking of the phrasal H does not take place if an accentual phrase has an accent at the initial mora, as in *ka*-HL*gebooshi* "silhouette" or at the second mora, as in *o ba*-HL*asama* "grandmother".

Second, in reference to the accent H, Pierrehumbert and Beckman say that there are four tonal aspects of accent whose representation must be accounted for, i.e., (i) the location of the accent; (ii) the shape of the accent; (iii) the height of the accent relative to the phrasal H; and (iv) the accent as a condition of catathesis. By treating the accent as a bitonal phonological unit, HL, rather than a sequence of a H tone and a L tone, Pierrehumbert and Beckman account for all the four aspects of the accent.

First, in regard to the location of the accent, Pierrehumbert and Beckman agree with Poser (1984) and Pulleyblank (1983) that the accent is a property of the lexicon so that its location is lexically determined. Second, in regard to the shape of the accent, in the bitonal treatment of the accent, because the L tone is an inseparable part of the HL accent, the duration of the L is invariant regardless of the length of the mora which immediately follows the accented mora. This durational characteristic of the accent is, in fact, implemented in Pierrehumbert and Beckman's synthesis model. Even though the H of the bitonal HL is assigned the durational value which corresponds to the length of an accented mora, the L is treated as a mere point in a pitch contour. Third, for the height of the accent, Pierrehumbert and Beckman create a rule which makes an accented H higher than a phrasal H in an accentual phrase. Lastly, the treatment of the accent as a bitonal unit makes it possible to differentiate the HL tone sequence which triggers catathesis, i.e., the accent, from a HL tone sequence which does not trigger catathesis, such as a sequence of a phrasal H tone and a boundary L tone. Obviously, the main advantage of the treating the accent as a bitonal unit is to be able to account for the last aspect of the accent.

The third type of the H tone is a H boundary tone which is inserted at the end of an interrogative utterance. The H boundary tone immediately follows the L boundary tone of the last accentual phrase in an interrogative utterance, causing a pitch contour at the end of a question to have a sharp rise starting from the L boundary tone to the H boundary tone.

2.3.1.3.2 L Tone

Although there is only one type of L tone, which is a boundary tone, it assumes different names (and, in fact, different strengths) depending on the locations. First, an "accentual phrase final L%"[5] is inserted at the end of each accentual phrase, providing the onset of the initial delimitative rise for the following accentual phrase. Second, an "utterance initial L%" is inserted at the beginning of an utterance, functioning as the onset of the utterance's initial F0 rise. No L% is inserted at the intermediate phrase level. Because the L% of the last accentual phrase in an intermediate phrase functions as an "intermediate phrase L%", it is not necessary to insert an additional L tone at the intermediate phrase boundary.[6]

2.3.2 Descriptions of Structural Components

The tones we have looked at in the previous section are different in the sense that they originate at the different levels of prosodic units. In this section, I would like to look at the three higher levels of prosodic units, i.e., the accentual phrase, the intermediate phrase, and the utterance, because they play the major role in contouring an utterance. For a better understanding, two other prosodic models will be examined, i.e., one by

McCawely (1968, 1977) and one by Poser (1984). Table 2.1 provides a list of the units and corresponding downtrends postulated in these models.

Table 2.1
Higher level prosodic units and downtrends postulated in three Japanese models; i,e., McCawley (1968, 1977), Poser (1984) and Pierrehumbert and Beckman (1989).

Models	Higher Prosodic Units	Downtrend Mechanisms
McCawley (1968, (1977)	Minor phrase (= contains at most one accented syllable)	Accent Reduction Rule (= accents a primary accent and reduces all other accents)
	Major phrase (= consists of one or more minor phrases)	Pitch Assignment Rules (= assign H, M or L tone to each tone-bearing unit)
	(Sentence)	
Poser (1984)	Minimal minor phrase (= a lexical item + particle(s))	Morphoaccentual processes = prosodic words
	Larger minor phrase (= consists of two or more minor phrases)	LPCFM [Local Phonologically Condition F0 Modification] ‖ Catathesis
	Major phrase (= domain of catathesis and phrase-final tone insertion)	Chaining of catathesis Phrase-final Tone Insertion
	Sentence	Declination (10Hz/sec)

Pierrehumbert Beckman (1989)	Accentual phrase (= contoured by phrasal H & L%; possible occurrence of HL)	Making phrasal H lower than accent H
	Intermediate phrase (= marked by low-F0 valued L%)	Catathesis
	Utterance	Declination Final lowering / Final raising

2.3.2.1 McCawley's Model

In McCawley's phonological model, a minor phrase is defined as a unit which has a L-toned initial mora (unless it is accented) and a unit which contains at most one accent. One or more minor phrases, then, form a major phrase. In the major phrase, it is only the leftmost accent which is assigned a H tone and all other accents in the phrase are reduced to a mid (M) tone. As far as mechanisms are concerned, the Accent Reduction Rule and Pitch Assignment Rules account for the downtrends in both minor and major phrases. That is, first, the Accent Reduction Rule accents a primary accent thereby reducing all other accents in the phrases. Then, it is the Pitch Assignment Rules which actually assign a H tone to the primary accent, a M tone to reduced accents and a L tone to every non-accented mora.

2.3.2.2 Poser's Model

In Poser's model, there are two types of minor phrases, i.e., a minimal minor phrase and a larger minor phrase. The minimal minor phrase is equivalent to a "phonological word" or a "clitic group", consisting of a lexical item plus particle(s), such as *fu'ne-o* "ship (accusative)". Morpho-accentual processes are responsible for resolving accent-conflicts, deleting all but one accent in the domain; e.g., *u'mi* + *ma'de* --> *u'mi-made* "to the sea"; *u'mi* +*gu'rai* --> *umi-gu'rai* "as much as the sea". Two or more minimal minor phrases form a larger minor phrase. One or more larger minor phrases, then, form a major phrase. Catathesis, or, in Poser's terminology, Local Phonologically Condition F0 Modification (LPCFM) accounts for the downtrend in the larger minor phrase. The iterative application of catathesis and a phrase-final tone insertion are responsible for the downtrends in the major phrase. A sentence, the highest level of the units, is defined as the domain of a declination whose ratio is said to be 10 Hz per second.

2.3.2.3 Pierrehumbert and Beckman's Model

The following passage summarizes concisely the characteristics of the higher prosodic units and corresponding downtrends posited in Pierrehumbert and Beckman's model.

> "The smallest level, the accentual phrase, is characterized by the occurrence of the phrasal H, the occurrence of the boundary L%, and the possible occurrence of an accent. It is the domain for the rule making a phrasal H lower than an accent H, and for the expansion of pitch range associated with narrow focus. The next larger level, the intermediate phrase, is the domain of catathesis. Catathesis is not carried over to the next H tone if an intermediate phrase boundary intervenes. In addition, the L% is lowered at an intermediate phrase boundary. The utterance is the level at which an initial L% is assigned, as well as a final H% in questions. --- It is also the level at which final raising or lowering and declination apply." (Pierrehumbert and Beckman 1989: 176).

One common feature among the three models is that they all postulate several prosodic units of different sizes. So, Pierrehumbert and Beckman are not entirely original in setting such hierarchical prosodic units as the accentual phrase, the intermediate phrase, and the utterance. However, there is a distinctively different and innovative feature of Pierrehumbert and Beckman model that the other two models do not have, i.e., the structural units are autosegmentally linked to the elements in the tonal component (and to those in the segmental component). The linking between the two components makes it possible for the tones in the tonal component to assume structural information. That is, in the realization of the tones, they can encode structural information and can reflect it in their realization values.

2.4 Autosegmental Association

In the following section, I will examine how the elements of the tonal and structural components are linked together to form a well-formed prosodic tree which represents an utterance of a sentence.

Before looking into the actual mechanisms of the linking, I must mention one unique feature of the prosodic tree; i.e., it is not a binary branching but *n*-ary branching tree, allowing a node to have more than two daughter nodes (cf. Beckman 1986, Nespor and Vogel 1986). The *n*-ary tree can be characterized as follows: (i) the tree is flatter than the binary tree which is commonly employed by metrical phonologists, such as Liberman and Prince (1977), Hayes (1980), or Selkirk (1984); and (ii) the depth of the tree is determined by the number of the prosodic units which a language may have and which are either phonetically or phonologically motivated. In other words, the depth of the n-ary prosodic tree cannot be infinite as can a binary tree. In the case of Tokyo Japanese, as seen in Figure 2.2, the depth of the prosodic tree is delimited by the number of six prosodic unit; $\{\mu, \sigma, \omega, \alpha, \iota, \upsilon\}$.

The truly innovative feature of Pierrehumbert and Beckman's representation of an utterance is autosegmental tone associations. Unlike the 1970's autosegmental phonology, such as Haraguchi's work (1977), whose main concern is the alignment between the segmental tier and the tonal tier in the course of time, in Pierrehumbert and Beckman's autosegmental tree representation, the tones in the tonal component, or in the "substantive tier", are linked not only to the lowest prosodic unit, i.e., mora , but also to any nodes in the structural component, or "structural tier". Taking the prosodic tree in Figure 2.2 as an example, let us see how tones are associated with structural nodes.

First, at the word-level, the bitonal accent tone, HL, which is a property of a lexical word, is linked to an accented mora, as seen in Figure 2.7.

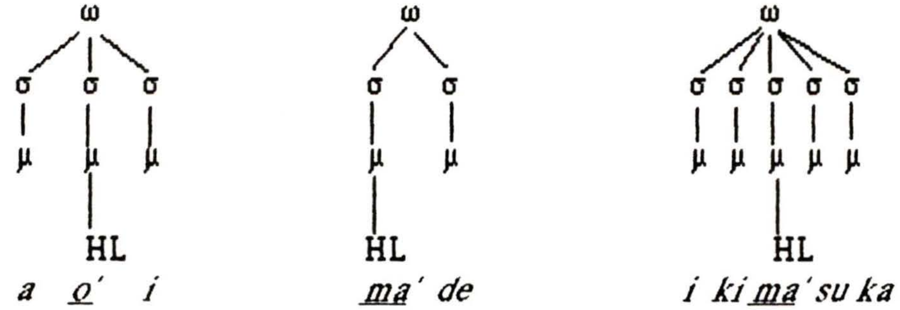


Figure 2.7
Association of the bitonal accent tone, HL, to an accented mora in *ao'i* "blue"; *ma'de* "to"; and *ikima'suka* "do (you) go?".

Next, the phrasal H and the phrase final L%, both of which are properties of the accentual phrase, are linked to an accentual phrase node. As seen in Figure 2.8, the phrasal H may be linked to the second mora of the following accentual phrase if it is toneless, as in *o ka*-H *ma'de*" to a hill" or in *i ki*-H *ma'suka* "do (you) go?". Or, the phrasal H may be linked to the first mora if the mora is the first sonorant mora of an initial heavy (long) syllable, as in *o*-H *oisogide* "quickly". The phrasal H is not, however, linked to the second mora in *a o*'-HL*i* "blue" where the second mora has already been linked to the accent HL.

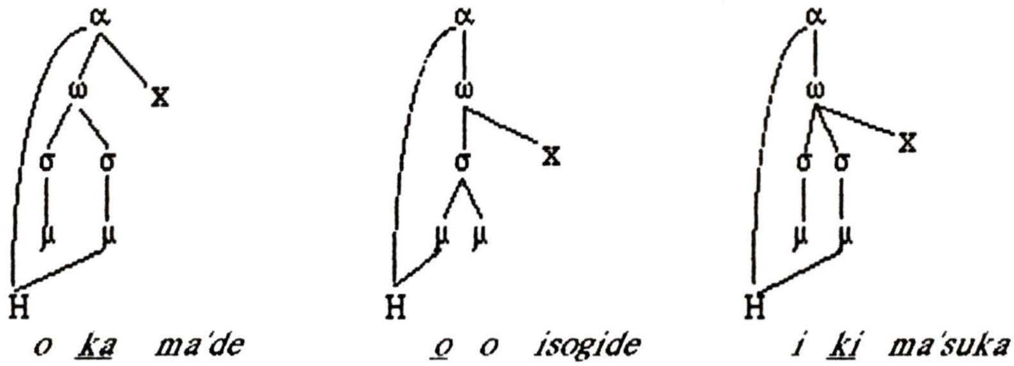


Figure 2.8
Association of the phrasal H in the accentual phrases: *okama'de*"to the hill"; *ooisogide* "quickly"; and *ikima'suka*"do (you) go?". The symbol, "x", indicates an abbreviation of nodes which are not relevant to the case in question.

Also, as seen in Figure 2.9, the accentual phrase final L boundary tone may be linked to the initial mora of the following accentual phrase if the mora is neither accented nor part of a heavy syllable.

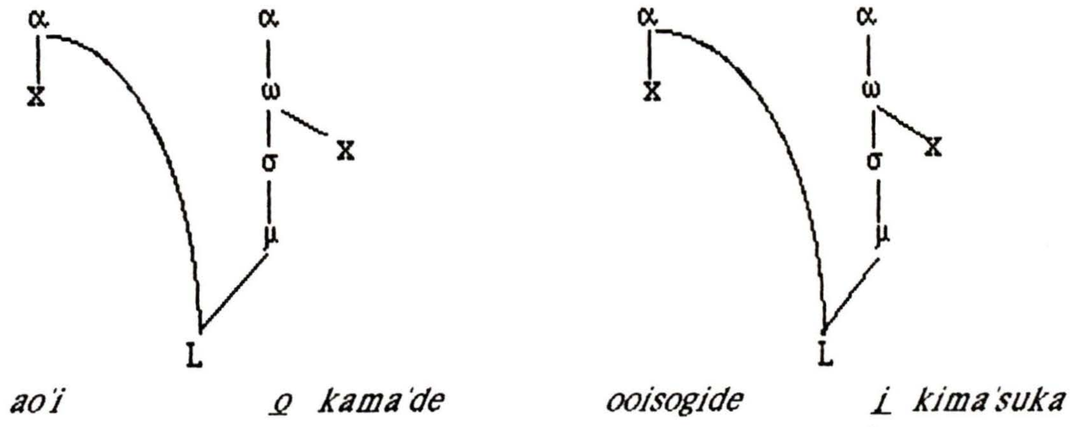


Figure 2.9
Association of the accentual phrase final L% to the initial mora of the following accentual phrase in *ao'i okama'de* "to the blue hill" and *ooisogide ikima'suka* "do (you) go quickly?"

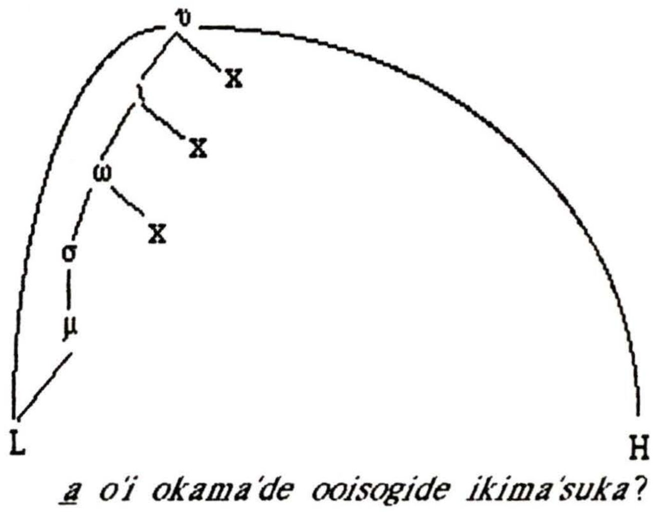


Figure 2.10
Association of the L% to the initial mora of the initial accentual phrase in the utterance of the sentence, *ao'i okama'de ooisogide ikima'suka* "Do you go quickly to the blue hill?" The interrogative H% of the utterance level is not associated with a mora.

Finally, two types of tones are associated with the utterance node, i.e., the utterance initial L% and the utterance final interrogative H%. In the case of the utterance of the sentence in Figure 2.2, because the initial mora of the utterance's initial accentual phrase, i.e. *a* of *aoi*, carries no tone, the utterance initial L% is linked to the mora. In contrast, the utterance's final H% is never associated with a mora.

One may wonder why some tones are associated with a mora and some tones are not. What is the difference? Pierrehumbert and Beckman say that if a tone is associated with a mora, the tone assumes the durational value of the mora, whereas a tone which is not associated with a mora is a mere point in a pitch contour. Except for the interrogative H%, all the other tones may be linked to a mora to assume a mora-length duration. That is, the accent H assumes the duration of an accented mora. The phrasal H assumes the duration of the first or second sonorant mora in an accentual phrase if the mora is not assigned to an accent H. In the case of the L%, it is assigned a durational value only when it is linked to a tone-less initial mora of a following accentual phrase.

2.5 Tone Scaling

We have looked at the components of the hierarchical autosegmental model and how the elements in the components are linked together in a prosodic tree which represents an utterance of a sentence. In this final section, I examine how the information represented as a prosodic tree is actually applied in the target level model. In other words, I will see how tones which lie flat at the bottom of the prosodic tree with no vertical spatial values can stand up and participate to create an actual pitch contour for a speech synthesis.

The basic mechanism of uplifting the tones is called "tone scaling" which, assisted by mathematical equations, provides spatial values to the tones. The idea of "tone scaling" originates in Pierrehumbert (1980) and Liberman and Pierrehumbert (1984) where F0 values for tones are computed by equations in a transform of the F0 domain. The transform is defined as the distance from the reference or base line to a tone. For example, the next equation provides a transform value for a peak (i.e., a H tone):

$T(P) = P - r$ (Liberman and Pierrehumbert 1984: 192).

In the formula, T(P) is a transform value in Hz; P is a F0 value of the peak and r is a F0 value of a reference line.

In Pierrehumbert and Beckman's (1989) target level model, not only a reference line but also a high-tone line (or top line) are employed to set a "transform space" and the tones between these two lines are expressed by normalized values ranging between 1 and 0. Thus, in the model, the above equation is changed to:

(1) $T(H) = (H - r) / (h - r)$ (ibid: 182)

In the formula, T(H) is a transform value for a H tone. H is a F0 value for the H tone and h is a F0 value of a high-tone line which is set "for a hypothetical H tone of maximal prominence in the phrase" (ibid: 182).

The main advantage of computing transform values for tones is that they can express the linguistic notion of relative prominence (cf. Liberman and Prince 1977). That is, the transform value can abstract away, for example, the prominence relationship between peaks in the context of "answer" and "background"; or the prominence relationship among peaks affected by an iterative application of pitch compression (catathesis); or the prominence relationship between a modifier and a noun head. In fact, these transform values ranging from 1 to 0 can not only express the prominence relationships among tones but also be used to uplift the tones and to scale them vertically. The vertically scaled tones are, then, used as targets for linear interpolation which then creates a schematic pitch contour which approximates to a pitch contour observed in real speech.

Pierrehumbert and Beckman postulate several equations for computing transform values and relative prominence constants of tones. One such equation, i.e., the equation (1) above, is for computing a transform value of the accent H. The following set of equations is for computing a catathesis constant (2), which is, in fact, a transform value (3) of the H tone affected by catathesis:

(2) $c = (h_{new} - r) / (h_{old} - r)$

(3) $T(h_{new}) = c$ (ibid: 184)

In the formulae, c is a catathesis constant; h_{old} is a F0 value of a high-tone line set prior to a catathesis, and h_{new} is a F0 value of a catathesized high-tone line set as an upper F0 limit of a new transform space for tones after the accent H. The catathesis constant, c, provides a value for the strength of the catathesis. The strength is directly reflected on the height of the new high-tone line because of the formula; $T(h_{new}) = c$, which indicates that the catathesis constant is equal to the transform value of a new high-tone line.

To capture the fact that an accent H is always higher than an unaccented H (i.e., phrasal H), the following set of equations are postulated:

(4) $T(Hu) = u \quad u < 1$

(5) $T(Ha) = 1$ (ibid: 185)

In the formulae, T(Hu) is a transform value of an unaccented H; T(Ha) is a transform value of an accented H; and u is an unaccented phrase constant which specifies the degree of lowering of the phrasal H, whose transform value is always smaller than 1, in relation to the accent H, whose transform value is set at 1.

For computing the prominence relationship between the accent H and the unaccented H which occur in the same accentual phrase, the next equation is postulated: (ibid: 185)

(6) $T(Hu) = u \cdot T(Ha)$

The following equation is, then, used to obtain the value of the constant, u.

(7) $u = T(Hu) / T(Ha) = ((Hu - r) / (h - r)) / ((Ha - r) / (h - r))$
$= (Hu - r) / (Ha - r)$

Next, the main characteristic of the boundary tone is that the lower the F0 value, the stronger the boundary. This F0 characteristic of the boundary tone is expressed by the following equation:

(8) $T(L) = 1 - (L - r) / (h - r)$ (ibid: 188)

This equation is similar to the one for H tones, i.e., the equation (1), except that in the case of L%, the lower the F0 value of L, the closer the transform value is to 1, rather than the other way around. That is, in extreme cases, when L%'s transform value is 1 (i.e., when the L% is strongest), the L% is on the reference or base line and when L%'s transform value is 0 (i.e., when the L% is least strong), the L% is right on the high-tone line.

In Pierrehumbert and Beckman's speech synthesis program, the transform values and constant values obtained by the above computations are used to scale the tones which will then become target levels for linear interpolations. Declination as well as final lowering are applied to the above schematic contour which will then be filtered to be smoothed. At the last stage, segmental effects are imposed on the contours.

2.6 Concluding Remarks

In this chapter, Pierrehumbert and Beckman's prosodic model has been examined to understand its major components and the interaction of the components. The main purpose of examining the model was to acquire tools to design an algorithmic phrasing and pitch contour model for speech synthesis. The following three main features are adopted from Pierrehumbert and Beckman's model and will become basic components of the algorithmic phrasing and pitch contour model:

(i) Inventory of tones, i.e., four types of tones which are sparsely distributed in a sentence, i.e., accent H, phrasal H, boundary L, and boundary H.

(ii) Hierarchical prosodic units; in particular, the accentual phrase and the intermediate phrase are of prime importance as far as modelling pitch contours of utterances is concerned.
(iii) Tone scaling, i.e., a computational device to draw schematic pitch contours by way of assigning transform values to tones.

Relying mainly on Pierrehumbert and Beckman's model,[7] and using these tools, this thesis investigates, to a major extent, the nature of phrasing, i.e., the phonological nature of accentual phrasing and the syntactic, semantic, and extra-linguistic nature of intermediate phrasing due to the following reasons:
(i) Pierrehumbert and Beckman (1989) do not discuss the issue of phrasing.
(ii) An example used to demonstrate Pierrehumbert and Beckman's speech synthesis program suggests that their phrasing is not automatic in that different levels of phrase boundaries must be manually inserted in the text of a sentence.
(iii) To develop a text-to-speech program, it is essential to be able to predict the locations of different types of boundaries because once the locations of these boundaries are predicted, tones can be automatically placed at the appropriate locations and a pitch contour can be automatically drawn by interpolating these tones.

In theoretical terms, the above proposal of designing an automatic[8] phrasing and pitch contour model can be seen as the investigation of the interfaces between phonological component (including phonetic realization processes) and the other components of the grammar.

As a phonological theory, Pierrehumbert and Beckman's model is termed a "domain theory". That is, it is a theory which postulates prosodic units each of which is defined as a domain of specific phonological rules and phonetic processes (i.e., downtrends). In other words, to limit the domain to the three higher levels of the prosodic units, the accentual phrase (α) is the domain of delimitative raise and possible occurrence of an accent; the intermediate phrase (ι) is the domain of catathesis; and the utterance (υ) is the domain of declination and final lowering.

Although these prosodic constituents are domains of specific phonological rules and/or phonetic processes, it is implausible to regard these downtrends as being solely the products of the phonological rules and/or phonetic processes. It would be more realistic and correct, if we regard these downtrends as the products of "mapping rules" between the phonological component and the other components of the grammar, such as morphological, syntactic, and semantic components.

The issue of the mapping rules or interfaces between the phonological component and the other components of the grammar is not within the scope of Pierrehumbert and Beckman (1989) and they do not discuss such an important issue as how the demarcation of the higher prosodic constituents (more specifically, the accentual phrase and the intermediate phrase) make reference to nonphonological notions. Inevitably, it becomes the task of any study which depends on Pierrehumbert and Beckman's model as a theoretical basis to investigate the issue of interfaces between the phonological component and the other components.

As will be seen in Chapter 3, at the level of the accentual phrase, there is an interface between the phonological component and the morphological component in determining accentual configurations of the phrases (if postpositions are treated as morphemes). Idiosyncratic accentual peculiarities of (non-monomoraic) postpositions play roles in accentuation of the phrases. However, besides the interface with the morphological component, at this level of the accentual phrase, the phonological component has a minimal interface with the syntactic component. The only syntactic information required in the demarcation of the accentual phrase (i.e., accentual phrasing) is the information whether the items are content words or postpositions. Once the syntactic information is provided, accentual phrasing is determined by phonological rules. That is, the accentual phrasing is determined by the underlying accent of words.

Once however, we move up in the hierarchy to the next level of the prosodic unit, i.e., the intermediate phrase, syntactic information plays a visible role in determining phonetic processes of the phrase. As will be seen in Chapter 4, the difference in syntactic tree configurations result in different pitch contours of the phrases in phonetic realizations of intermediate phrases. The issue at hand is the correlation between syntactic tree configurations and the iterative application of catathesis. (Such a correlation between syntax and phonetic processes does not, however, support the existence of isomorphism between the two components because of a simple fact that the domain of the intermediate phrase does not correspond to any specific syntactic unit.)

Furthermore, in the case of the demarcation of the intermediate phrase (i.e., intermediate phrasing), it is not only the syntactic component but also the semantic component that interfaces with the phonological component. In other words, not only syntactic information but also semantic information, such as theme, argument, contrastive prominence, or scope of modifiers, are required to phrase a sentence into intermediate phrases.

Thus, it becomes clear that the higher the prosodic units are in the hierarchical prosodic tree, the more involved are the notions of nonphonological components.

Although the above is a brief description of the issues of this thesis from a theoretical perspective, it must be made very clear that arguments in this thesis are not meant for developing a theory, but rather meant for empirical feasibility, i.e., designing linguistic input to a speech synthesis program. And as will be noticed, occasionally, the simplicity required by an application overrides theoretical adequacy.

Endnotes

[1] Fundamental frequency or F0 refers to the lowest frequency component of the sound wave. It is important in the study of intonation due to its close correlation to the sensation of pitch (Crystal 1985).

[2] H is an abbreviation of high tone whereas L is an abbreviation of low tone. These are two tonal primitives assumed by Pierrehumbert and Beckman.

[3] In the following examples, there is no difference between an accentual phrase and a word. In other words, each word is realized as an accentual phrase.

[4] Due to type-setting, tone is assigned to a relevant tone-bearing unit horizontally.

[5] "%" is Pierrehumbert's rendition of a boundary tone.

[6] Non-insertion of L% at the intermediate phrase level does not mean that there is no intermediate phrase boundary. It simply means that no L%L% at the end of an intermediate phrase, one L% being for an accentual phrase and the other being for an intermediate phrase.

[7] As will be seen in Chapter 5, there is another prosodic model which this thesis depends on.

[8] As will be seen in Chapter 5, the locations of some of the intermediate phrase boundaries are manually specified so that the phrasing is not absolutely automatic.

Chapter 3
Accentual Phrasing

The aims of this chapter are to examine the main factor which determine the accentual phrasing and how intermediate phrases (which end with a postposition) can be parsed into accentual phrases in order to develop algorithmic rules of accentual phrasing and tone assignment. Prior to the above issues, the accentual behaviour of Japanese postpositions (particles) will be discussed. Also, the effect of narrow-focus on accentual phrasing will be examined to understand variability in accentual phrasing. To assist in achieving the above objectives, two series of acoustic experiments are conducted.

In the following section, the accentual behaviour of Japanese postpositions is examined because they play significant roles in the accentual phrasing and because Pierrehumbert and Beckman (1989) are silent about the issue of the postpositions which do not, in many cases, obey their "left-win rule" in solving accent-conflicts.

3.1 Accentual Behaviour of Postpositions

In combinations of postpositions and their hosts, many of the postpositions exhibit peculiarities in accentual behaviour. These accentual behaviour of postpositions are well documented (e.g. Hirayama 1969; NHK

1966; McCawley 1968, 1977; Higurashi 1983; Poser 1984). Table 3.1 provides a convenient, though not exhaustive, summary of accounts on the accentual behaviour of non-monomoraic postpositions.

Table 3.1
Accentual types of non-monomoraic postpositions shown in the forms with the accented host, *i'noti* "life" and the unaccented host, *miyako* "capital".

(1) *ma'de* -type [+Left-winning]: an unmarked type which obeys the left - win rule; e.g. *de'su* "copula", *yo'ri* "from", *ba'kari* "only".

i'noti + *ma'de* --> *i'noti-made*

miyako + *ma'de* --> *miyako-ma'de*

(2) *kara* -type [+Anonymity]: an unaccented counterpart of the type (1); all the monomoraic postpositions should also be included in this type.

i'noti + *kara* --> *i'noti-kara*

miyako + *kara* --> *miyako-kara*

(3) (a) *gu'rai* -type [+Deaccenting]: a marked type; some speakers treat the postpositions of this type as the type (1).

i'noti + *gu'rai* --> *inoti-gu'rai*

miyako + *gu'rai* --> *miyako-gu'rai*

(b) *jyuu* -type [+Deaccenting]; an unaccented counterpart of the *gu'rai* type postpositions.

i'noti + *jyuu* --> *inoti-jyuu*

miyako + *jyuu* --> *miyako-jyuu*

(4) (a) *'sika* -type [+Preaccenting (partial)]: a marked type of postposition.

i'noti + *'sika* --> *i'noti - sika* (obeying the left-win rule)

miyako + *'sika* --> *miyako'- sika*

(b) *jyuu*-type [+Preaccenting (total)]: a marked type which has not been documented, but produced by two of the five subjects in my experiments.

i'noti + *'jyuu* --> *inoti' - jyuu*

miyako + *'jyuu* --> *miyako'- jyuu*

As listed in Table 3.1, postpositions may be categorized into four major types; (1) [+Left-winning] postpositions; (2) [+Anonymity] postpositions; (3) [+Deaccenting] postpositions; and (4) [+Preaccenting] postpositions. The first type, [+Left-winning] postposition is an unmarked case. Non-monomoraic postpositions, such as *ma'de* "to", *de'su* "copula", or *ba'kari* "only" are classified in this type. If a [+Left-winning] postposition has any accent-conflict; i.e., when both the host and the postposition are accented, it is the host's accent which is realized, and the accented postposition loses its accent, as in *i'noti* + *ma'de* --> *i'noti- made.* If there is no accent-conflict, an available accent is realized as the accent of the accentual phrase (noun + postposition), as in *miyako* + *ma'de* -> *miyako-ma'de*.

The second type of postposition marked by [+Anonymity] is the unaccented counterpart of [+Left-winning] postpositions, and being a part of a host noun, they are never independent in accentuation and never cause any accent-conflicts. All the monomoraic postpositions, such as *o* "accusative", *ni* "dative", or *wa* "topic marker" should also be included in this type.

The third type of postposition is marked by the feature [+Deaccenting] and postpositions, such as, *gu'rai* "as much as", *da'ke* "only", or *jyuu* "throughout" are classified in this type. In the case of a [+Deaccenting] postposition, the accent of the host will not be realized because of the predominant power associated with the [+Deaccenting] postposition which deaccents the accent on its left, as in *i'noti* + *gu'rai* --> *inoti -gu'rai* The unaccented [+Deaccenting] postpositions, *jyuu* and *dake*, create an unaccented accentual phrase regardless of the accentuation of the host, as in *i'noti* + *jyuu* --> *inoti-jyuu*; *miyako* + *jyuu* --> *miyako-jyuu*.

The fourth type of postposition is marked by the feature [+Preaccenting] because the postposition of this type places an accent on the last syllable of the preceding host. In the literature, only *'sika* "only" is listed as a postposition of this type. However, *jyuu* which is usually classified as a [+Deaccenting] postposition may also be included in this type because two of the subjects who participated in one of the experiments conducted in this chapter treated *jyuu* as a [+Preaccenting] postposition. There must be speaker variation in the accentuation; *jyuu*$^{[+\text{Deaccenting}]}$ ~ *jyuu*$^{[+\text{Preaccenting}]}$. When *jyuu* is treated as a [+Preaccenting] postposition, it accents the last syllable[1] of the host regardless of the accentuation of the host, as in *i'noti* + *'jyuu* --> *inoti'-jyuu* and *miyako* + *'jyuu* --> *miyako'-jyuu*. *'Sika*, the other [+Preaccenting] postposition, is different from *'jyuu* in that *'sika* places an accent on the last syllable of the host only if the host does not have an accent (i.e., unaccented) as in *miyako* + *'sika* --> *miyako'-*

sika. If the host is accented, *'sika* obeys the left-win rule as in *i'noti* + *'sika* --> *i'noti-sika.* Thus, in Table 3.1, *'jyuu* is specified as [+Preaccenting (total)] and *'sika* as [+Preaccenting (partial)].

A few words may be needed to describe the underlying accents of monomoraic postpositions. In Bloch (1950) and McCawley (1977), monomoraic postpositions are treated as having an underlying accent as *ni'* "to", *mo'* "also", or *wa'* "topic marker". The underlying accent is manifested only when two monomoraic postpositions are combined together to form a new dimoraic postposition in the context of contrast; e.g. *Yamada ni'wa okane-o agetaga, Honda ni'wa agenakatta* "(I) gave money to Yamada, but not to Honda". The postpositions created by combining two monomoraic postpositions obey the left-win rule.

These postpositions play important roles in accentual phrasing. The effects of postpositions on accentual phrasing and on pitch contours of accentual phrases will be observed in the experiment on accentaul phrasing (3.2) as well as in the experiment on narrow-focusing (3.3).

3.2 Experiment on Accentual Phrasing

3.2.1 Aim of Experiment

The main aim in conducting this acoustic experiment is to obtain a generalization about accentual phrasing. More precisely, I would like to know whether it is a syntactic configuration or an accentual configuration which determines how an intermediate phrase is parsed into accentual phrases. For example, given the phrase, *a'oi* + *oma'me* + *ma'de*, "to the blue beans", is it possible to predict how many accentual phrases are created from the phrase? Although unlikely, will the phrase be uttered with two interphrasal L%'s, creating three accentual phrases in the phrase because there are three underlying accents? Or, more likely, will the phrase be uttered with just one interphrasal L% which is inserted before the noun, creating only two accentual phrases, as *ao'i* L% *oma'me-made* because there are two surface accents ? Or, will the whole phrase be realized as just one accentual phrase, having a culminative accent at the leftmost unit, *ao'i* ? Or, will it be that accentual phrasing is not conditioned by the accentual configuration, but by the syntactic configuration: modifier + noun + postposition? Of course, there ought to be variations in phrasing but also there ought to be a general trend in accentual phrasing which must be determined either by a phonological or syntactic condition. It is the trends and conditions of the accentual phrasing which are what I would like to elicit from the experiment in order to develop algorithmic rules of accentual phrasing. Such algorithmic rules should parse intermediate phrases into accentual phrases as well as assigning relevant tones to the phrases.

3.2.2 Procedures

Table 3.2 is the list and the possible combinations (4 * 2 * 4 = 32) of lexical items used as stimuli in the experiments. The phrases made of the possible combinations of the lexical items are set in a carrier sentence; "...... *te-ga todokimasu*," (I can reach out my hand for ......) except for the possible combinations with *gu'rai*. The phrases with *gu'rai* are placed in a carrier sentence, "...... AJ-N *wa arimasen*" as "*ao'i omame-gu'rai ao'i oma'me-wa arimasen*," (there are no beans which are as blue as the blue beans). It is the meaning of *gu'rai* which demands the different carrier sentence.

Table 3.2
A list of stimuli used in the experiments examining accentual phrasing

Modifier		Noun		Postposition	
ao'i	"blue"			*ma'de*	"to"
omoi	"heavy"	*oma'me*	"beans"	*gu'rai*	"as much as"
a'ni-no	"brother's"	*nimame*	"cooked beans"	*jyuu*	"all over"
ane-no	"sister's"			*ni*	"to"

There are accentual reasons why these items are chosen. First, in the second noun slot, the accented noun, *oma'me*, is contrasted with the unaccented noun, *nimame*, in that both nouns have the same number of morae as well as similar phonemic configurations. Second, in the postposition slot, *ma'de* represents [+Left-winning] postpositions; *gu'rai* is an accented postposition marked by the feature [+Deaccenting]; and the postposition, *jyuu*, is treated as an unaccented [+Deaccenting] postposition. The [+Preaccenting] postposition, *'sika*, is not included in the list because its segments, /s/, devoiced /i/, and /k/ are all invisible in F0 analysis. *Ni* represents monomoraic postpositions. Third, in the modifier slot, there are two pairs of accented and unaccented items, i.e. *ao'i* versus *omoi* and *a'ni-no* versus *ane-no*, and it may look somewhat redundant to have the two pairs of the modifiers. The inclusion of the *no*-phrase modifiers is motivated by my contention that a *no*-phrase modifier has a different locality in deaccentuaion from an adjectival modifier: e.g.,

a) *oma'me* + *ma'de* --> *oma'me-made*

b) *ao'i* + *nimame* + *ma'de* --> *ao'i-nimame-made*

c) *a'ni-no* + *nimame* + *ma'de* --> *a'ni-no-nimame-ma'de*

As shown above, the [+Left-winning] postpositions, such as *ma'de* is deaccented not only by an immediately preceding accented *word* as in (a) but may also be deaccented by a preceding accented *phrase* as in (b). However, this may not be the case for the accented noun phrase with the *no* -phrase. That is, *a'ni-no* + *nimame* does not deaccent *ma'de*. In the case of the noun phrase with a *no*-phrase modifier, it is not the accentuation of the whole phrase, *a'ni-no* + *nimame* (accented), but the accentuation of the noun head of the NP, *nimame* (unaccented), which determines the accentuation of the following [+Left-winning] postposition, *ma'de*. To see whether any acoustic evidence can be found to support such a claim, the pair of the *no*-phrase modifiers was inserted in the modifier slot.[2]

The above mentioned stimuli embedded in the carrier sentences were written, in a random order, on sheets of paper in Japanese. Each sentence was paired with its echo question. The data for analyses were taken only from the answers because, being old information, none of the items in the phrases in the answers should have received any narrow-focus. The subjects were requested to utter the stimuli in a well articulated manner. The total of 160 (((4 * 2 * 4 *) * 5) = 160) utterances were recorded by five female subjects.

Measurements were taken by using MSL (Micro Speech Lab) and MSLPITCH which are IBMPC compatible speech analysis programs developed at the Centre for Speech Technology Research, associated with the Department of Linguistics at the University of Victoria. The recorded items were analyzed with 10 bits, 10k/sec sampling rate.

3.2.3 Summary Tables of Accentual Phrasing

Table 3.3
The results of accentual phrasing of the phrases whose head is the accented noun, *oma'me.*

OMA'ME - Set

	Unmarked Phrasing				Marked Phrasing	
ao'i						
A11	+/ + −	(*ma'de*)	(+ + +)	(1)	[+ − −] h/s	(8)
A 12	+/ − +	(*gu'rai*)	(+ + +)	(2)	[ ] h	(9)
A13	+/ − −	(*jyuu*)	(+ + −)	(3)	Ø	
A14	+/ + −	(*ni'*)	(+ + −)	(1)	[+ − −] s	(8)
omoi'						
A21	−/ + −	(*ma'de*)	(− + +)	4)	Ø	
A22	−/ − +	(*gu'rai*)	(− + +)	(4)	Ø	
A23	−/ − −	(*jyuu*)	(− + −)	(6)	Ø	
A24	−/ + −	(*ni'*)	(− + −)	(4)	[ ] h/s	(8)
a'ni-no						
A31	+/ + −	(*ma'de*)	(+ + +)	(1)	[+ − −] h	(8)
A32	+/ − +	(*gu'rai*)	(+ + +)	(2)	[+ − −] h	(8)
A33	+/ − −	(*jyuu*)	(+ + −)	(3)	Ø	
A34	+/ + −	(*ni'*)	(+ + −)	(1)	[+ − −] h	(8)
ane-no						
A41	−/ + −	(*ma'de*)	(− + +)	(4)	[ ] h	(5)
A42	−/ − +	(*gu'rai*)	(− + +)	(4)	[ ] h/s	5)
A43	−/ − −	(*jyuu*)	(− + −)	(6)	[ ] s	(7)
A44	−/ + −	(*ni'*)	(− + −)	(7)	[ ] h/s	(5)

Table 3.4
The results of accentual phrasing of the phrases whose head is the unaccented noun, *nímame*.

	Nímame - Set				
	Unmarked Phrasing			Marked Phrasing	
ao'i					
B11	+/ − +	(*ma'de*)	(2)	[ ] h/s	(9)
B 12	+/ − +	(*gu'rai*)	(2)	[ ] s/t	(9)
B13	+/− −	(*jyuu*)	(3)	[ ] h (8); +/−/− k	
B14	+/ − −	(*ni*)	(3)	[ ] s/h	(8)
omoi					
B21	[− − +]	(*ma'de*)	(5)	Ø	
B22	[− − +]	(*gu'rai*)	(5)	Ø	
B23	[− − −]	(*jyuu*)	(7)	−/−/−k	
B24	[− − −]	(*ni*)	(7)	Ø	
a'ni-no					
B31	+/ − +	(*ma'de*)	(2)	Ø	
B32	+/ − +	(*gu'rai*)	(2)	[+ − −] h	(8)
B33	+/ − −	(*jyuu*)	(3)	Ø	
B34	+/ − −	(*ni*)	(3)	[ ] h	(8)
ane-no					
B41	[− − +]	(*ma'de*)	(5)	Ø	
B42	[− − +]	(*gu'rai*)	(5)	−/ − +k	(4)
B43	[− − −]	(*jyuu*)	(6)	−/ − − k	(6)
B44	[− − −]	(*ni*)	(7)	Ø	

The results of the experiment on accentual phrasing are summarized as Table 3.3 and Table 3.4 . Table 3.3 is a summary of the phrasing of all the possible combinations with the accented noun, *oma'me* and Table 3.4 is a summary of the phrasing of those with the unaccented word, *nimame*. In both sets, i.e., *oma'me*-set and *nimame*-set, all the cases are divided into two groups, unmarked phrasing and marked phrasing. The markedness and unmarkedness are determined by the frequency of occurrences. In each table, there are four rows of phrase groups which differ in the modifier they take. In a group, each phrase is specified with its ending postposition. The + and – signs specify whether items in a phrase are accented (+) or unaccented (–). The reason why there are two series of + and – specifications in the unmarked phrasing case in the *oma'me*-set is that one on the left specifies a surface accentuation of a phrase and one on the right in a parenthesis specifies underlying (original) accentuation of the phrase, i.e., the accentuation prior to an application of a [+Feature] of a postposition. The *nimame*-set does not have two types of accentual specifications because surface and underlying accentual specifications are the same in a phrase in the set. A slash between symbols indicates the presence of an interphrasal accentual boundary. If a phrase is realized as a single phrase without an interphrasal boundary, such a phrase is marked by [ ]. If there are no symbols inside [ ], it shows that a phrase is realized without an interphrasal boundary and with the same accentuation as its unmarked phrasing. If a subscript is attached to the bracket, it identifies the subject who uttered the instance. The symbol, Ø, indicates the absence of an instance. Finally, the numeral at each case indicates a schematic F0 contour of the phrase presented in a later section (from Figure 3.1 to Figure 3.9) so that the reader can have visual understanding of the phrase in question.

For example, a part of the first, *ao'i*-group in the *oma'me*-set which is reproduced below can be read as follows:

OMA'ME - Set

		Unmarked Phrasing				Marked Phrasing	
ao'i							
	A11	+/ + −	(*ma'de*)	(+ + +)	(1)	[+ − −] h/s	(8)
	A13	+/ − −	(*jyuu*)	(+ + −)	(3)	Ø	

The case, A11, *ao'i* + *oma'me* + *ma'de* (+ + +) was realized, in the case of unmarked phrasing, as +/+− , i.e., *ao'i* L% *oma'me*-*made* with the insertion of an interphrasal accentual boundary. The schematic F0 contour of the phrase is (1) (which is listed in figure 3.1). The subjects H and S, however, uttered the same phrase as [+−−], i.e., *ao'i*-*omame*-*made* with no insertion of L% and with just one culminative accent on the left-most item, *ao'i*. The utterance is regarded as marked phrasing, and its schematic F0 contour is (8) (which can be seen in Figure 3.8).

Another case, A13, *ao'i* + *oma'me* + *jyuu* whose underlying accentuation is (+ + −) was realized as +/−−; *ao'i* L% *omame*-*jyuu*, i.e., an intermediate phrase consisting of two accentual phrases. The schematic F0 contour of the phrase is (3) (which is listed in Figure 3.3). All five subjects showed the same phrasing pattern because its marked case has Ø, a null-sign.

3.2.4 Unmarked Phrasing

First of all, let us look at unmarked phrasing in the *oma'me*-set. The phrases in the set have a consistent pattern of phrasing, i.e., the insertion of an interphrasal boundary between the modifier and the noun. The accentuation of the phrases seems to have no impact on the phrasing because there are the differences of all the possible combinations in accentuations. That is, if the accentuation of the postpositions are excluded from consideration, there are following accentual variations across the interphrasal boundary:

+/ + (*ao'i* L% *oma'me-*; *a'ni-no* L% *oma'me-*)
+/ − (*ao'i* L% *omame-*; *a'ni-no* L% *omame-*)
−/ + (*omoi* L% *oma'me-*; *ane-no* L% *oma'me-*)
−/ − (*omoi* L% *omame-*; *ane-no* L% *omame-*)

The above facts seem to suggest that a syntactic configuration rather than an accentual configuration determines accentual phrasing. That is, as unmarked phrasing, a phrase of "modifier + noun + postposition" is always uttered as an intermediate phrase consisting of two accentual phrases with L% inserted after the modifier. So, to account for the accentual phrasing, I can posit a very simple working hypothesis; i.e., "if a phrase has a syntactic configuration of modifier + noun + postposition, insert an interphrasal accentual boundary after a modifier".

Next, let us look at unmarked phrasing in *nimame*-set in Table 3.4 , and test whether the above hypothesis can account for all the phrasings. In the *nimame*-set, the working hypothesis based on the syntactic configuration is obviously denied because in *omoi*- and *ane-no*- groups, there is no instance which has an interphrasal L%. All the phrases were realized without an interphrasal L%. This discounts the syntax-based hypothesis. The question is, then, how to account for the fact that it is only the phrases in the unaccented modifier (*omoi* and *ane-no*) groups in the *nimame*-set that do not have an interphrasal L%. It looks as if the accentual configurations of the phrases, too, fails to condition accentual phrasing because in the *oma'me*-set, there are the cases where L% is inserted between an unaccented modifier (–) and an unaccented noun (–), i.e., "–/– " (cases: A22, A23, A42, and A43). On the other hand, in the *nimame*-set, there is no insertion of L% in the phrases which have exactly the same accentual configuration, i.e., [– –] (cases: all the phrases in *omoi*- and *ane-no*- groups). So, denying the previous syntax-based working hypothesis, it seems that accentual phrasing is arbitrary; i.e., the insertion of the interphrasal L% cannot be predicted either by a syntactic configuration or by an accentual configuration.

Importantly, however, it becomes possible to obtain a generalization on accentual phrasing once the underlying (original) accentual configuration rather than the surface pattern is taken into account. That is, in all the underlying accentual forms (i.e., the accentuations of the phrases prior to the applications of the postpositional features) in the *oma'me*-set, there is at least one + either in the modifier slot or in the noun slot. It is, then, always the case that an interphrasal L% is inserted after a noun. Now, in the *nimame*-set, all the phrases in *ao'i*- and *a'ni-no*- groups have + specification in the modifier slot, and they all have an interphrasal L%. In the same *nimame*-set, however, all the phrases in *omoi*- and *ane-no*-groups which do not show any interphrasal L% have no + specification either in the modifier slot or in the noun slot. Thus, from these facts, I can deduce the following generalization: in the case of unmarked phrasing, a phrase of "modifier + noun + postposition" has an interphrasal accentual boundary after the modifier if at least one of the (content) words is underlyingly (originally) accented. This generalization accounts for the unmarked phrasing exhibited in all the data and may possibly be extended to account for accentual phrasing of any lengths of phrases with multiple modifiers by eliminating the proviso, "modifier + noun + postposition", from the statement. That is, in general, an intermediate phrase boundary is inserted between (content) words if at least one of the words is underlyingly (originally) accented.

3.2.5 Marked Phrasing

Roughly speaking, there are two types of marked phrasing; one caused by "culminative accentuation" and the other caused by "enunciative accentuation". Two of the subjects, S and, especially, H, constantly show the first type of marked phrasing, creating a single phrase with only one accent whereas the subject, K, shows, once in a while, the second, opposite type of phrasing, inserting L% at every possible location. Typical examples of the marked phrasing caused by culminative accentuation are found in the following cases: A11, A14, A31, A32, and A34. The unmarked phrasing in these cases have either +/+− or +/−+, whereas the marked phrasing shows only [+−−] which is characterized by (1) having just one culminative accent at the leftmost item and by (2) having no interphrasal L%, realizing the whole phrase as a single accentual phrase. A speculation is that such a phrasing is caused by a narrow-focus placed on the left-most item which deaccents any accents on its right. As you may recall, the stimuli of the experiment was designed in such a way that no narrow-focus was to be placed on any of the items in the phrases. In spite of such a design, it seems that two of the subjects, especially, H, constantly placed a narrow-focus on the left-most accented word in the phrases.

In the section on narrow-focusing (3.3.1), we will see whether such phrasing can be replicated when prominence is placed on the left-most item which is accented. I expect this type of phrasing occurs commonly in ordinary speech because it will not tax the speaker with accentuation as much as articulated speech does in that no rules are involved in solving an accent-conflict. That is, once this style of speech is chosen, the speaker can ignore accentual features of postpositions. Just keep the left-most accent and delete any accent until the speaker reaches the end of a phrase.

In addition, though not quite the same as the above [+−−] phrasing, there is a similar type of marked phrasing which does not insert an interphrasal accentual L% in a phrasal domain. The following are instances; A12, A24, A41-A44, B11-B14, and B34 (which are all specified by an empty [ ]). This type of phrasing has the same number of accents as in corresponding unmarked phrasings. The only the difference is that the marked phrasings create no interphrasal L%. According to Pierrehumbert and Beckman(1989), the prominence placed on a phrase initial item, which may be accented or unaccented, causes the disappearance of an interphrasal

L% after the item. In the section on narrow-focusing, I would like to see if this is the case or not.

The second type of marked phrasing, the phrasing caused by enunciation, can be found in subject K's utterances in B13, B23[3], B42, and B43. Typical examples are seen in (B13) *aoi-nimame-jyuu* and (B23) *omoi - nimame-jyuu* where L% is inserted at every possible location, creating three accentual phrases in the sequence of three items.

Next, there are two cases (B42 and B43) where the subject, K, inserts L% between an unaccented modifier (–) and an unaccented noun (–). This fact may provide counterevidence to the earlier generalization that in the case of unmarked phrasing, the phrase of "modifier + noun + postposition" has an interphrasal accentual L% after the modifier if at least one of the (content) words is underlyingly (originally) accented. Such a generalization does not, however, exclude the possibility of inserting L% between an unaccented modifier and an unaccented noun. What the generalization says is that the insertion of interphrasal L%, such as in B42 or in B43, is highly marked (enunciative) phrasing. As the insertion of L% between the noun and the postposition in B13 and B23 is highly marked, so is the insertion of L% between the unaccented modifier (–) and the unaccented noun (–). The fact that both types of phrasing were produced by the same speaker enables one to say that the subject K's utterances in B42 and B43 should be regarded as marked phrasing and cannot be construed a counterevidence for the generalization obtained above; i.e., in general, an accentual phrase boundary is inserted between (content) words if at least one of the words is underlyingly (originally) accented.

3.2.6 A Psycholinguistic Implication

One psycholinguistic implication, which comes to mind, based on the results on the accentual phrasing is that there must be some sort of look-ahead mechanism in accentuation and phrasing. More precisely, there must be a look-ahead-one-word mechanism in accentuation and phrasing. Such a mechanism can be represented by a two-word sized window cursor which moves from left to right one word at a time. It is only in a (current) window cursor, that any accent-conflict between two words is resolved. Also, in the (current) window cursor, a phrasing decision is made; i.e., an interphrasal L% will be inserted if at least one of the two (content) words in the cursor is underlyingly accented (+).

What are the reasons for postulating a look-ahead-one-word mechanism for accentuation and phrasing? First, if there were no look-ahead mechanism at all, how is it possible to account for the resolution of an accent-conflict triggered, for example, by the feature, [+Deaccenting]; e.g., *ao'i + oma'me + jyuu --> ao'i-omame-jyuu*? To deaccent correctly *oma'me* as *omame* in the phrase, the speaker has to see the feature [+Deaccenting] before the speaker reaches the second mora of the noun, or more reasonably before the speaker starts to utter the noun. Thus, there must be some sort of look-ahead mechanism in accentuation. If, however, the speaker were able to look ahead the accentual configurations of items up to the end of the phrase, in other words, if there were a *phrase-sized* window cursor, it would not be possible to account for the phrasing difference between, for example, (A43) *ane-no-omame-jyuu*; (− + −), −/− − and (B43) *ane-no - nimame-jyuu*; (− − −), [− − −]. If the speaker were able to see the feature [+Deaccenting] prior to uttering the phrases, both phrases would have the same phrasing, i.e., [– – –]. That is, A43 should not have the interphrasal L% because the speaker would be able to see the feature [+Deaccenting] of the postposition prior to uttering the initial word and, thus, would treat the accentuation of the whole phrase as [– – –]. If this were the case, (A43) *ane-no-omame-jyuu* [– – –] and (B43) *ane-no-nimame-jyuu* [– – –] should have had the same phrasing, i.e., [– – –] according with the earlier generalization which inhibits the insertion of an interphrasal L% between two unaccented (–) content words. However, the fact that A43 was realized as –/– – whereas B43 was realized as [– – –] denies the existence of the phrase-sized window cursor; i.e., the speaker cannot look ahead to all the accentual configurations of a phrase before starting to utter it.

A look-ahead-one-word mechanism or an implementation of a two-word-sized window cursor will explain things nicely. Because there is a two-word-sized window cursor, an interphrasal L% is inserted after the modifier in A43 but not in B43 due to the generalization that a phrase will have L% between a modifier and a noun if at least one of them is underlyingly accented:

A43	B43
[*ane-no*(–) *oma'me*(+)] *jyuu*[+Deac]	[*ane-no*(–) *nimame*(–)] *jyuu*[+Deac]
‖	‖
ane-no L% *oma'me*	*ane-no* - *nimame*

The next movement of the cursor enables the speaker to see the feature [+Deaccenting] and to deaccent the noun, *oma'me*, in A43:

ane-no / [*oma'me jyuu*[+Deac]]   *ane-no* [*nimame-jyuu*[+Deac]][4]

|| ||

omame-jyuu   *nimame-jyuu*

The results, *ane-no* L% *omame-jyuu* and *ane-no-nimame-jyuu* are exactly what we want as the unmarked phrasing for the phrases. As the reader will see shortly, this two-word-sized window cursor is implemented as a device in algorithmic rules for accentual phrasing which will parse intermediate phrases automatically into accentual phrases and will assign appropriate tones to the phrases.

3.2.7 Algorithmic Rules of Accentual Phrasing and Tone Assignment

In this section, I explore the development of simple algorithmic rules to draw schematic F0 contours of the phrases used as the stimuli in the experiment. Input will be texts of the phrases with some additional information, and output will be schematic F0 contours of the phrases. The aim of drawing such F0 contours of the phrases is to show that drawing the schematic contours is actually the main part of the process which creates natural pitch contours for a speech synthesis program.

In Pierrehumbert and Beckman's program, there are following steps to be taken in synthesizing speech. First, a schematic version of the F0 contour of what one wants to synthesize is constructed by scaling tones and linearly interpolating these tones which function as targets. The equations discussed in Chapter 2 are used to compute necessary constants and transform values for the scaling. Durational values are assigned to mora-associated tones. "The phonological attachment of the tones to the various nodes in the prosodic structure determines how the single-point and mora-length tone levels will align with segments in the speech signal" (Pierrehumbert and Beckman 1989: 176). Then, the contour is smoothed by a filtering to produce a gradual schematic pitch contour close to the one observed in real speech.

There are a few differences in arriving at a schematic version of a natural F0 contour between Pierrehumbert and Beckman's program and a simple algorithmic program presented here. First, in drawing schematic versions of F0 contours presented here, each tone is treated as a point with no assignment of mora-sized durational value. The treatment is for the sake of graphic representations. Treating a tone as a point will produce a clear graphic representation of a natural F0 contour and, in fact, can better approximate to the final version of synthesized F0 contour which has undergone a filtering in Pierrehumbert and Beckman's program.

The second difference is the treatment of the phrasal H in an accented accentual phrase. In Pierrehumbert and Beckman (1989), an accented accentual phrase is marked by both the phrasal H and the accent H. In my drawing of schematic F0 contours, the phrasal H in an accented accentual phrase is treated as optional because it often falls right on the line interpolated between the preceding L% and the following accent H, showing no elbow, inflection point, of an initial delimitative rise as seen, for example, in the case of the schematic F0 contour (1) in Figure 3.1 in comparison with (2) in Figure 3.2.

The third difference is the number of the phrasal H which the unaccented accentual phrase can have. As seen in the schematic F0 contours, (3) in Figure 3.3 and (7) in Figure 3.7, I allowed unaccented accentual phrase to have not just one but two phrasal H's, the second of which is associated with the last mora of the phrase, in order to better approximate the natural contours of the phrases.[5]

The last difference is the accentual phrasing algorithm. In Pierrehumbert and Beckman's synthesis model, phrasing is not automatic; i.e., phrasing information has to be specified manually. Here, in drawing schematic F0 contours, I develop simple algorithmic rules for accentual phrasing so that once an intermediate phrase boundary is specified (a topic of Chapter 5), the simple phrasing algorithm will automatically parse an intermediate phrase into accentual phrases as well as assign relevant tones to the phrases.

To make the simple algorithmic rule of accentual phrasing and tone assignment to work, the following information are required: item specification (whether items are content words or postpositions); the feature specification (the types of the features assumed by some of the postpositions), and accentual specification (whether items are accented or unaccented). Taking (A11) *ao'i + oma'me + ma'de* as an example, I show how simple algorithmic rules can be developed.

3.2.8 The Operation of Accentual Phrasing and Tone Assignment

1) Translate accented items (+), unaccented items (−), and features ([Feature]) into + − specifications; e.g.,

ao'i + *oma'me* + *ma'de* --→ + + +[+Left-win]

(If a specifier (+ or −) has [+Feature], it is a non-monomoraic postposition; if it is marked by [-Feature], it is a monomoraic postposition; if no [+/-Feature] is attached, it is a content word.)

2) Insert an initial boundary, /, and to the boundary assign a L% (which may be a property of an utterance or may be an intermediate phrase boundary[6]):

```
/ +   +   +[+Left-win]
|
L%
```

3) Set a two-specifier-sized window cursor at the initial position:

```
[/ +   +] +[+Left-win]
 |
 L%
```

4) Assign a HL to an initial specifier if it is +. If it is −, assign a H to the − iff it is preceded by /:

```
[/   +   +] +[+Left-win]
 |   |
 L%  HL
```

5) Insert an interphrasal boundary, /, and assign a L% to it iff at least one of the specifiers in the cursor is +:

```
[/   +   /  +]   +[+Left-win]
|    |   |
L%  HL  L%
```

6) Move the cursor to the next position:

```
/    +   [/  +  +[+Left-win]]
|    |    |
L%  HL   L%
```

7) If the second item in the cursor is a postposition which is marked by [+Feature], realize the feature. If it is not a postposition marked by [+/-Feature][7], go back to (4):

```
/    +   [/  +  -[+Left-win]]
|    |    |
L%  HL   L%
```

8) Assign a HL to a + in the cursor. If a + is not found, then, either assign a H to the first - iff it is preceded by / or assign a H to the second - iff the first - is not preceded by /:

```
/    +   [/     +   -[+Left-win]]
|    |    |     |
L%  HL    L%   HL
```

9) Remove the cursor and insert a boundary, /, and to it assign a L%:

```
/   +    /   +    -[+Left-win]  /
|   |    |   |                  |
L%  HL   L%  HL                 L%
```

The above nine rules should adequately handle the unmarked phrasing and tone assignment for all the phrases discussed in the previous sections

Also, due to the loop created between Rule (4) and the Rule (7), the rules can handle phrases with multiple modifiers. The only problem with the set of the rules is that the operation of the tone assignment stops when the cursor reaches a postposition. However, as we will see in Chapter 5, there are many possible cases that (long) intermediate phrases have more than one postposition in the phrases. To deal with such a situation, first, the proviso, "(content) word" in the above generalization should be changed to "(phonological) word"; i.e., in general, an accentual phrase boundary is inserted between two (phonological) words if at least one of the words is underlyingly (originally) accented. Second, let a phonological word be a content word which may or may not be accompanied by postposition(s), i.e., a phonological word = a content word (+ postposition(s)).[8] Third, some of the above rules must be changed as follows:

(i) the end of an intermediate phrase boundary (//) must be specified.

(ii) the following series of the rules should replace Rule (8) and Rule (9):

(8') Assign a HL to a + in the cursor. If a + is not found, then, assign a H to the first − iif it is preceded by /.

(9') If there is no // (intermediate phrase boundary marker) in the cursor, go back to (4). If there is //, proceeds to (10').

(10') Assign a H to the second – iff there are two –'s and there is no / in the cursor.

(11') Remove the cursor, and assign a L% to //.

In regard to tone, the following are the results of these rules:
(i) insertion of an accentual phrase boundary between two (phonological) words if one of them is underlyingly (originally) accented;
(ii) assignment of a HL to a (surface) accent;
(iii) assignment of a phrasal H to an unaccented word if it is preceded by a boundary;
(iv) assignment of a phrasal H to the intermediate phrase final item if the preceding unaccented word is not preceded by a boundary; and
(v) assignment of a L tone to a boundary.

Next, the tones have to be aligned with the segments in the phrase. In Pierrehumbert and Beckman's model, the alignments are made by the phonological attachment of the tones to the various nodes in the prosodic structure of a tree representation. Although not discussed in this paper, such alignment can easily be translated into a small set of algorithmic rules which specify whether or not a tone will be assigned to a mora to assume a mora-length durational value. In the example here, because tones are treated as points, alignments of the tones to the text is simple. From left to right, align the tones to each possible point as shown:

a	*o'*	*i*	*oma'me*	*made*
|	|	|	|	|
L%	HL	L%	HL%	L%

3.2.9 Tone Scaling

The next step is assigning transform values to the tones so that the tones which lie flat can be vertically scaled. To do so, in the present case, the averaged F0 values for the tones were obtained from 15 tokens of phrasal utterances:

a	*o'*	*i*	*oma'me*	*made*
|	|	|	|	|
L%	HL	L%	HL%	L%
222 Hz	294 Hz	189 Hz	242 Hz	155 Hz

Next to obtain transform values for the tones, a high-tone line is set at the same height as the highest peak in the phrase;[9], i.e., the first HL (294 Hz), and a reference line is set at the same height as the lowest point in the phrase, i.e., the final L% (155 Hz). In the transform space set by these two lines, transform values of the tones ranging between 1 and 0 are computed as follows:

$T(HL^1) = (HL^1 - r) / (h - r) = (294-155) / (294-155) = 1.000$

(Transform value of the initial accent HL; because a high tone line is set at the same height as the tone, its value is 1.000)

$c = (h_{new} - r) / ((h_{old} - r) = (242-155) / (294-155) = 0.626$

$T(h_{new}) = c = 0.626$ (Transform value of the second, catathesized accent HL)

$T(L^1) = 1 - (L^1 - r)/(h - r) = 1-(222-155)/(294-155) = 0.528$

(Transform value of the phrase initial L%)

$T(L^2) - 1 - (L^2 - r)/(h - r) - 1 - (189-155) / (294-155) - 0.755$

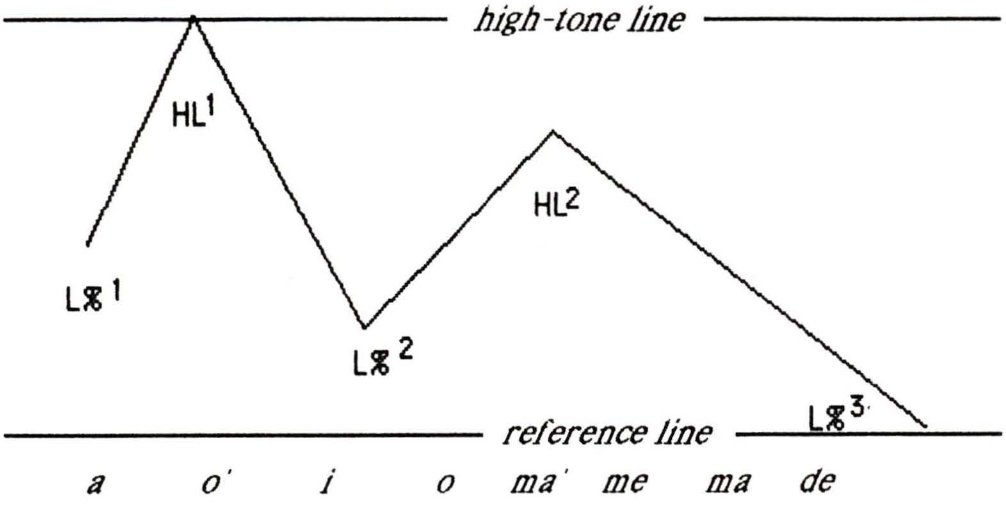
(Transform value of the interphrasal L%)

$T(L^3) = 1 - (L^2 - r) / (h - r) = 1 - (155 - 155) / (294 - 155) = 1.000$

(Transform value of the final L%: because a reference line is set at the same height as the tone, its value is 1.000)

a	*o'*	*i*	*oma'me*	*made*
|	|	|	|	|
L%	HL	L%	HL%	L%
0.518	1.000	0.755	0.626	1.000

Linearly interpolating the tones whose target values are encoded as the transform value , a schematic F0 contour for the phrase, *ao'i-oma'me-made* is drawn as follows:



One may wonder why it is necessary to compute transform values to scale the tones, instead of using actual F0 values to do so. The first reason, which I have already mentioned in Chapter 2, why transform values are used for tone scaling is that they can express the linguistic notion of relative prominence. The transform values enable us to compare a tone with any other tones in a transform space in terms of prominence. The prominence relationships among the tones can be encoded in the form of transform values regardless of changes of the whole pitch ranges. So, for example, in a speech synthesis program, we can assign any F0 values to a high-tone line and to a reference line without distorting the prominence relationships among the tones. The second reason why transform values should be used instead of actual F0 values is that accumulation of data of transform values will eventually enable us to develop rules. Right now, as we have seen, to compute transform values, we have to go back to F0 contours of real speech. Going back and forth between F0 contours of real speech and schematic F0 contours will eventually lead us to accumulate a fair size of data from which we can extract rules to assign transform values to any tones in any prosodic environments without referring to F0 contours of real speech.

Although it was with just one example, I have demonstrated how a text of a phrase, *aoí-oma'me-made*, can be converted into its schematic pitch contour. All schematic contours shown in the Figures from 3.1 to 3.9 should be understood that they are drawn in the same fashion.

3.2.10 Summary

(1) As a preliminary to investigating accentual phrasing, I have looked at accentual peculiarities of various postpositions because they play significant roles in the phrasing and phrasal contours. Depending on their accentual behaviour, the postpositions are classified into four groups: (i) [+Left-winning] postpositions whose accent is deleted by a preceding accented host ; (ii) [+Anonymity] postpositions which do not have any independence in accentuation; (iii) [+Deaccenting] postpositions which deaccent a preceding accented host ; and (iv) [+Preaccenting] postpositions which place accent on the last syllable of a preceding host.

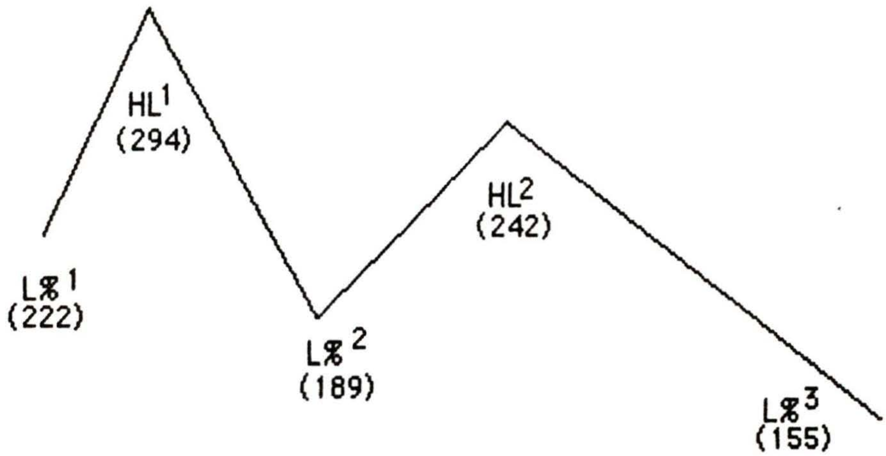
(2) Conducting an acoustic experiment on accentual phrasing, I have obtained some insight: (i) accentual phrasing is not determined by syntactic configurations but by accentual configurations; (ii) however, the accentual phrasing is conditioned, not by surface, but by underlying accentual configurations of phrases. The finding was generalized as that: in general, an interphrasal accentual boundary is inserted between two (content) words if at least one of the words is underlyingly (originally) accented.

(3) The generalization motivated the claim that there is a look-ahead mechanism in accentuation and accentual phrasing. That is, the speaker can look just one word ahead to determine the accentuation and phrasing of the current word. I have implemented such a mechanism as a two-word sized window cursor. Such a cursor was, then, used as a device in developing algorithmic rules of accentual phrasing and tone assignment.

(4) I have shown how such algorithmic rules can parse intermediate phrases into smaller, accentual phrases and can assign tones to the phrases to draw schematic pitch contours, which can be used for speech synthesis.

Figure 3.1

A schematic pitch contour of (1), +/ + − : L% HL L% HL L%. (F0 values in Hz are means of 15 tokens.)



Transform values:

$T(HL^1) = (HL^1 - r) / (h - r) = (294 - 155) / (294 - 155) = 1.000$

$c = (h_{new} - r) / (h_{old} - r) = (242 - 155) / (294 - 155) = 0.626$

$T(HL^2) = c = 0.626$

$T(L^1) = 1 - (L^1 - r) / (h - r) = 1 - (222 - 155) / (294 - 155) = 0.518$

$T(L^2) = 1 - (L^2 - r) / (h - r) = 1 - (189 - 155) / (294 - 155) = 0.755$

$T(L^3) = 1 - (L^3 - r) / (h - r) = 1 - (155 - 155) / (294 - 155) = 1.000$

Cases:

A11: *aoi'* L% *oma'me-made*

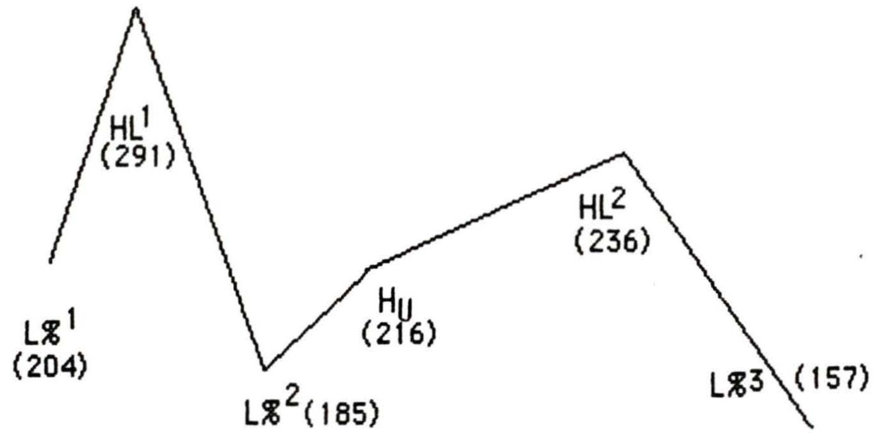
A14: *aoi'* L% *oma'me -ni*

A31: *a'ni-no* L% *oma'me-made*

A34: *a'ni-no* L% *oma'me -ni*

Figure 3.2

A schematic pitch contour of (2), + / – + : L% HL L% H HL L%. (F0 values in Hz are means of 22 tokens.)



Transform values:

$T(HL^1) = (HL^1 - r) / (h - r) = (291 - 157) / (291 - 157) = 1.000$

$c = (h_{new} - r) / (h_{old} - r) = (236 - 157) / (291 - 157) = 0.590$

$T(HL^2) = c = 0.590$

$T(HU) = u * T(HL^2) = 0.590\,u$

$u = T(HU) / T(HL^2) = ((216 - 157) / (291 - 157)) / ((236 - 157) / (291 - 157)) = 0.747$ (The constant, u, expresses the prominence relationship between the phrasal H and the accent HL^2.)

$T(HU) = 0.590 * 0.747 = 0.441$ (T (HU) is a transform value of the phrasal H)

$T(L^1) = 1 - (L^1 - r) / (h - r) = 0.649$

$T(L^2) = 1 - (L^2 - r) / (h - r) = 0.791$

$T(L^3) = 1 - (L^3 - r) / (h - r) = 1.000$

Cases:

A12: *ao'i* L% *omame-gu'rai*

A32: *a'ni-no* L% *omame-gu'rai*

B11: *ao'i* L% *nimame-ma'de*

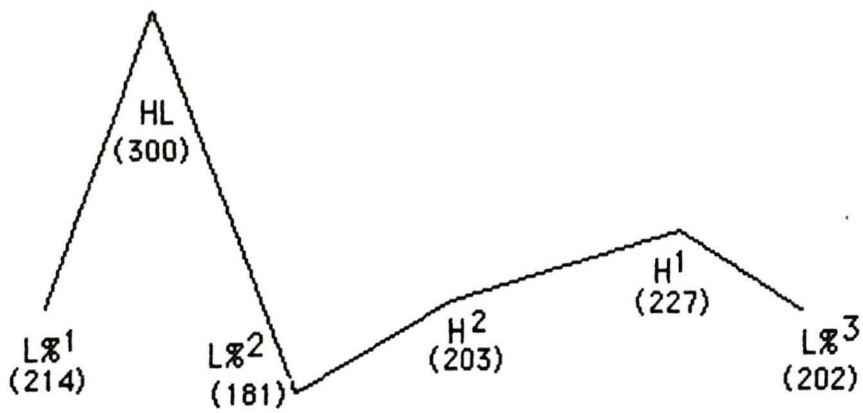
B12: *ao'i* L% *nimame-gu'rai*

B31: *a'ni-no* L% *nimame-ma'de*

B32: *a'ni-no* L% *nimame-gu'rai*

Figure 3.3

A schematic pitch contour of (3), + / − − : L% HL L% H H L%. (F0 values in Hz are means of 14 tokens.)



Transform values:

T (HL) = (HL - *r*) / (*h* - *r*) = (300 - 187) / (300 - 181) = 1.000

T (H^1) = *c* = (h_{new} - *r*) / (h_{old} - *r*) = (227 - 181) / (300 - 181) = 0.386

T(H^2) = *u* * T (H^1) = 0.386 *u*

u = (H^2 - *r*) / (H^1 - *r*) = (203 - 181) / (300 - 181) = 0.478

T (H^2) = 0.478 * 0.386 = 0.185

T(L^1) = 1 - (L^1 - *r*) / (*h* - *r*) = 1 - (214 - 181) / (300 - 181) = 0.723

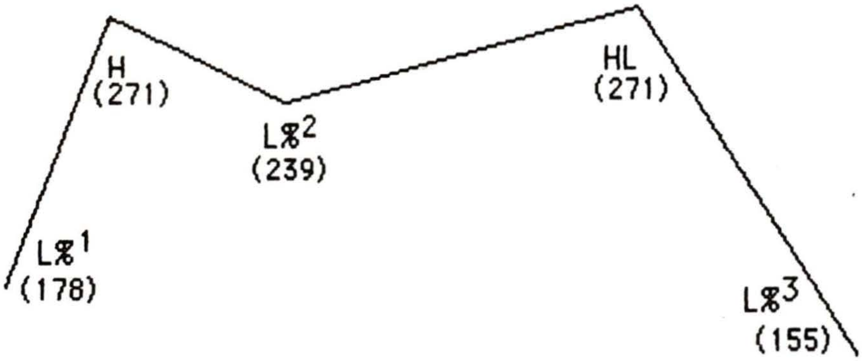
T(L^2) = 1 - (L^2 - *r*) / (*h* - *r*) = 1.000

T(L^3) = 1 - (L^3 - *r*) / (*h* - *r*) = 1 - (202 - 181) / (300 - 222) = 0.824

Cases:

A13: *aoi* L% *omame-jyuu*	B14: *aoi* L% *nimame-ni*
A33: *a'ni-no* L% *omame-jyuu*	B33: *a'ni-no* L% *nimame-jyuu*
B13: *aoi* L% *nimame-jyuu*	B32: *a'ni-no* L% *nimame-ni*

Figure 3.4
A schematic pitch contour of (4), –/ + – ; –/ – + : L% H L% HL L%. (F0 values in Hz are means of 23 tokens.)



Transform values:

T (H) = (H - *r*) / (*h* - *r*) = (271 - 155) / (271 - 155) = 1.000

T (HL) = (HL - *r*)/ (*h* - *r*) = (271 - 155) / (271 - 155) = 1.000

T (L^1) = 1 - (L^1 - *r*)/ (*h* - *r*) = 1 - (178 - 155) / (271 - 155) = 0.802

T (L^2) = 1 - (L^2 - *r*)/ (*h* - *r*) = 1 - (239 - 155) / (271 - 155) = 0.276

T (L^3) = 1 - (L^3 - *r*)/ (*h* - *r*) = 1 - (155- 155) / (2271 - 155) = 1.000

Cases:

A21: *omoi* L% *oma'me-made*

A41: *ane-no* L% *oma'me -made*

A22: *omoi* L% *omame-gu'rai*

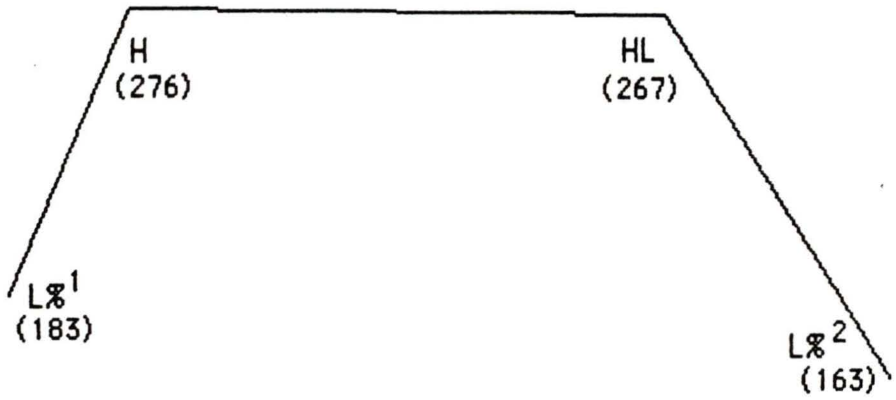
A24: *omoi* L% *oma'me-ni*

A42: *ane-no* L% *oma'me -gu'rai*

A42: *ane-no* L% *omame -ni*

Figure 3.5

A schematic pitch contour of (5), [− − +] : L% H HL L%. (F0 values in Hz are means of 19 tokens.)[10]



Transform values:

T (HL) = (HL - r) / (h - r) = (267 - 163) / (267 - 163) = 1.000

u = ((HU - r) / (h - r)) / ((HL - r) / (h - r)) = ((276 - 163) / (267 - 163)) / ((267 - 163) / (267 - 163)) = 1.086

T (HU) = u * T (HL) = 1.086

T(L^1) = 1 - (L^1 - r) / (h - r) = 1 - (183 - 163) / (267 - 163) = 0.808

T(L^2) = 1 - (L^2 - r) / (h- r) = 1.000

Cases:

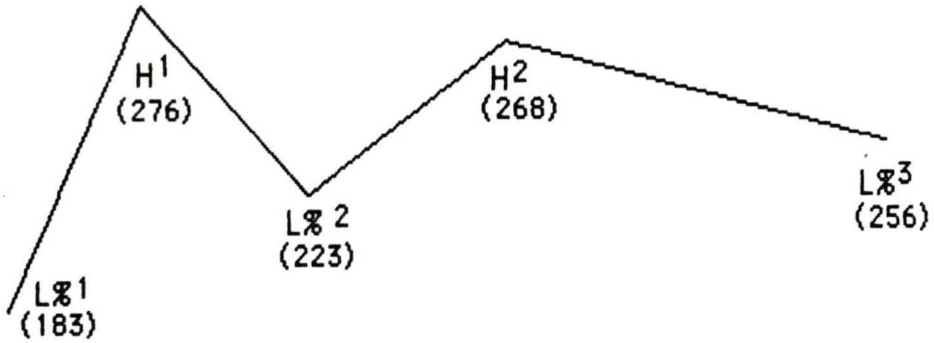
B21: *omoi-nimame-ma'de*

B22: *omoi-nimame-gu'rai*

B41: *ane-no-nimame-ma'de*

B42: *ane-no-nimame-gu'rai*

Figure 3.6
A schematic pitch contour of (6), −/ − − : L% H L% H L%. (F0 values in Hz are means of 3 tokens.)



Transform values:

$T(H^1) = (H^1 - r) / (h - r) = (276 - 183) / (276 - 183) = 1.000$

$T(H^2) = u$

$u = (H^2 - r) / (H^1 - r) = (268 - 183) - (276 - 183) = 0.913$

$T(H^2) = u = 0.913$

$T(L^1) = 1 - (L^1 - r) / (h - r) = 1.000$

$T(L^2) = 1 - (L^2 - r) / (h - r) = 1. - (223 - 183) / (276 - 183) = 0.570$

$T(L^3) = 1 - (L^3 - r) / (h - r) = 1 - (256 - 183) / (276 - 183) = 0.216$

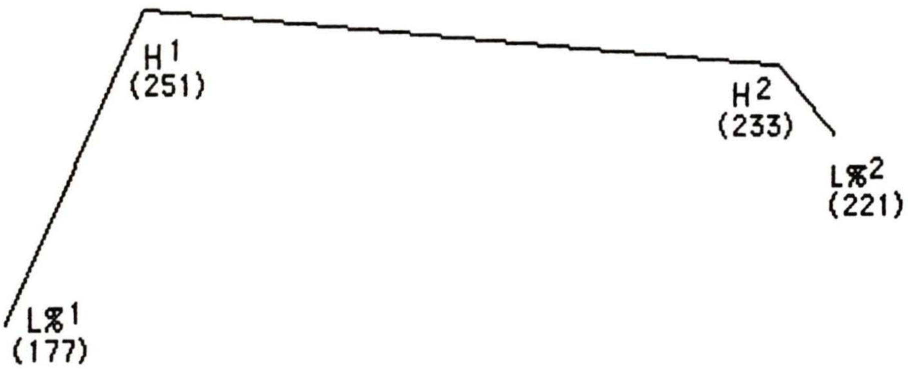
Cases:

A23: *ane-no* L% *omame-jyuu*

A43: *omoi* L% *omame-jyuu*

Figure 3.7

A schematic pitch contour of (7), [– – –] : L% H H L%. (F0 values in Hz are means of 12 tokens.)



Transform values:

$T(H^1) = (H^1 - r) / (h - r) = (251 - 177) / (251 - 177) = 1.000$

$T(H^2) = u$

$u = (H^2 - r) / (H^1 - r) = (233 - 177) / (251 - 177) = 0.750$

$T(H^2) = 0.750$

$T(L^1) = 1 - (L^1 - r) / (h - r) = 1.000$

$T(L^2) = 1 - (L^2 - r) / (h - r) = 1 - (221 - 177) / (251 - 177) = 0.406$

Cases:

B23: *omoi-nimame-jyuu*

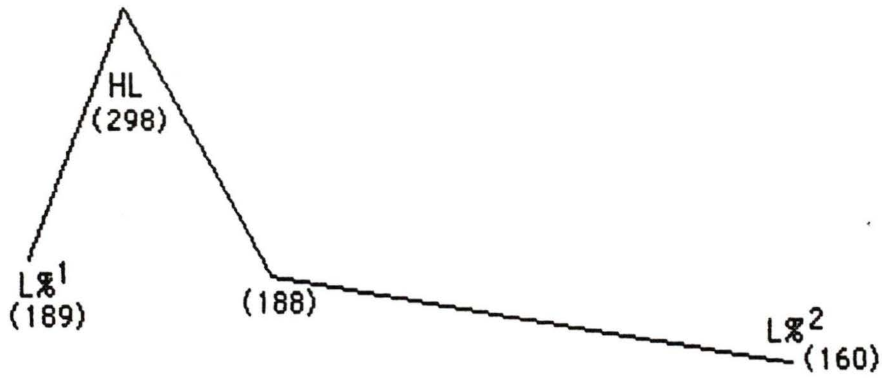
B24: *omoi-nimame-ni*

B43: *ane-no-nimame-jyuu*

B44: *ane-no-nimame-ni*

Figure 3.8

A schematic pitch contour of (8, marked phrasing), [+ − −] : L% HL L%. (F0 values in Hz are means of 10 tokens.)



Transform values:

$T(H_L) = (H_L - r) / (h - r) = (298 - 160) / (298 - 160) = 1.000$

$T(L^1) = 1 - (L^1 - r) / (h - r) = 1 - (188 - 160) / (298 - 160) = 0.790$

$T(^{H}L) = 1 - (^{H}L - r) / (h - r) = 1 - (189 - 160) / ((298 - 160) = 0.798$

$T(L^2) = 1 - (L^2 - r) / (h - r) = 1 - (160 - 160) / (298 - 160) = 1.000$

Cases:

A11: *aoʼi-omame-made*

A14: *aoʼi-omame-ni*

A31: *aʼni-no omame-made*

A32: *aʼni-no omame-gurai*

A34: *aʼni-no-omame-ni*

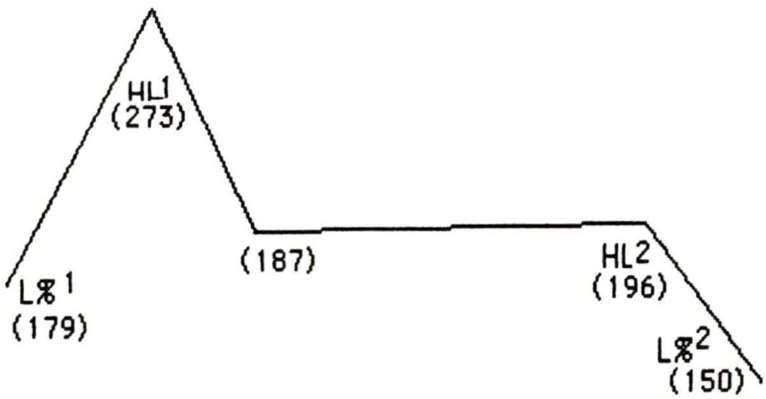
B13: *aoʼi-nimame-jyuu*

B14: *aoʼi-nimame-ni*

B34: *aʼni-no-nimame-ni*

Figure 3.9

A schematic pitch contour of (9, marked phrasing), [+ − +]: L% HL HL L%.
(F0 values in Hz are means of 4 tokens.)



Transform values:

$T(\mathrm{HL}^1) = (\mathrm{HL}^1 - r) / (h - r) = (273 - 150) / (273 - 150) = 1.000$

$c = (h_{new} - r) / (h_{old} - r) = (196 - 150) / (273 - 150) = 0.373$

$T(\mathrm{HL}^2) = c = 0.373$

$T(\mathrm{L}^1) = 1 - (\mathrm{L}^1 - r) / (h - r) = 1 - (179 - 150) / ((273 - 150) = 0.765$

$T(^{\mathrm{H}}\mathrm{L}) = 1 - (^{\mathrm{H}}\mathrm{L} - r) / (h - r) = 1 - (187 - 150) / ((273 - 150) = 0.700$

$T(\mathrm{L}^2) = 1 - (\mathrm{L}^2 - r) / (h - r) = 1.000$

Cases:

A12: *ao'i-omame-gu'rai*

B11: *ao'i-nimame-ma'de*

B12: *ao'i -nimame-gu'rai*

3.3 Experiment on Narrow-focusing

Using the same phrasal stimuli as in the previous experiment (in Table 3.2), an experiment is conducted to examine the effects of narrow-focus placed on each item of the phrases. To do so, one of the subjects who showed the clearest accentual phrasing (i.e., frequent insertion of interphrasal L% at possible locations) in the previous experiment participated in the experiment on narrow-focusing. In the experiment, each carrier sentence which embeds a stimulus (i.e., phrase) was elicited by three different interrogative sentences to which answers would place a narrow-focus on each item of the stimulus. A total of 96 utterances ((4 * 2 * 4) *3) = 96 were recorded and analysed.[11]

The following is a list of questions to be answered from the experiment:

(i) In the previous experiment, there were four types of the marked phrasing contours schematized as (5), (7), (8), and (9) (cf. Tables 3.3 and 3.4). Speculation was that these contours were the results of the narrow-focus placed on the initial modifiers. By placing prominence on the phrase initial modifiers, can these four types of the pitch contours be replicated?

(ii) Does an intermediate L% always manifest right before a narrow-focused item, as Pierrehumbert and Beckman (1989) claim? For example, will the phrases of "modifier + noun + postposition" be split into two intermediate phrases if the noun is narrow-focused?

(iii) Does such an intermediate L% appear even before narrow-focused postpositions?

(iv) Can any evidence be found, in narrow-focus situations, which supports my contention that the accented *no*-phrase modifier behaves differently from the accented adjectival modifier in the locality of deaccentuation?

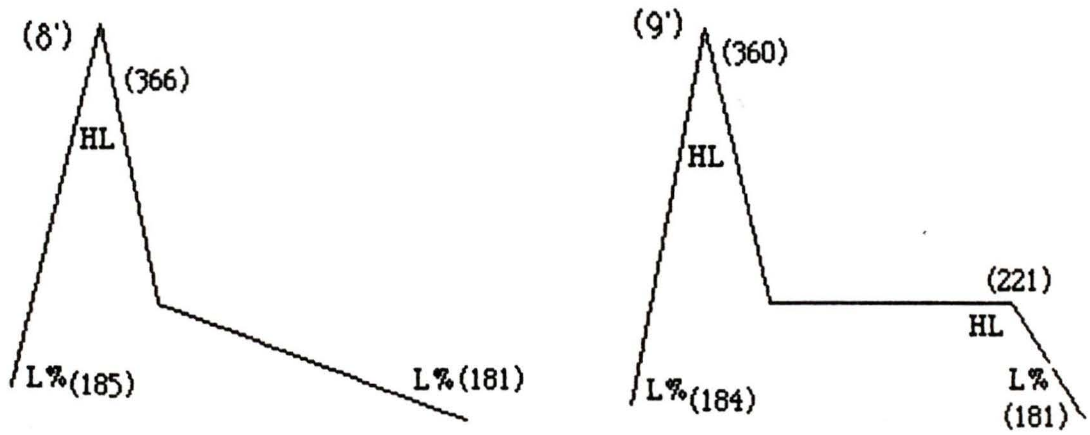
(v) In the phrases, such as *ao'i* + *oma'me* + *gu'rai*[+Deaccenting], does narrow-focus condition the accented noun to retain its accent which is, in a non-focus situation, deaccented by the feature [+Deaccenting]? Similarly, in the phrases, such as *ao'i* + *oma'me* + *ma'de*[+Left-winning], does narrow-focus condition the accented postposition marked by the feature [+Left-winning] to retain its accent, which is, in a non-focus situation, deaccented by the preceding accented noun?

(vi) Can monomoraic postpositions, such as *ni* "dative", which do not have any prosodic autonomy show any autonomy once they are narrow-focused?

3.3.1 Narrow-focus on the Modifiers

I will first discuss the effects of narrow-focus on the modifiers, grouping the modifiers into two groups: i.e., the accented modifiers and the unaccented modifiers. First, the noticeable effect of placing prominence on the accented modifiers is the disappearance of the interphrasal L% observed in the corresponding non-focused phrasings (cf. Figures 3.3 and 3.4), creating pitch contours schematized as (8') and (9') in Figure 3.10.

As is easily noticed, these pitch contours have the same shapes as the two of the four marked phrasing pitch contours in the previous experiment (3.2), i.e., the pitch contours schematized as (8) in Figure 3.8 and as (9) in Figure 3.9. Thus, it must be correct to say that the marked phrasing which was conditioned by culminative accentuation and whose pitch contours were schematized as (8) and (9) were, in fact, caused by narrow-focus placed on the initial accented modifiers.



Cases:
(A11) *ao'i* + *oma'me* + *ma'de.*
(A13) *ao'i* + *oma'me* + *jyuu.*
(A14) *ao'i* + *oma'me* + *ni.*
(B11) *ao'i* + *nimame* + *ma'de.*
(B13) *ao'i* + *nimame* + *jyuu.*
(B14) *ao'i* + *nimame* + *ni.*

Cases:
(A12) *ao'i* + *omame* + *gu'rai.*
(B12) *ao'i* + *nimame* + *gu'rai.*

Figure 3.10
Schematized pitch contours which are due to the narrow-focusing the phrase initial accented modifier. The F0 values in Hz in the left contour are means of 6 tokens and those in the right contour are means of 2 tokens.

It seems to be the case that for the phrases with at least two underlying accents, there is allophony of F0 contours, depending, perhaps, on the degrees of prominence.[12] That is, if no focus is involved, the phrases are realized as the pitch contour schematized as (1/2) in Figure 3.11. If the accented modifier of the phrases is narrow-focused, the phrases will have the binate pitch contour of (9/9'). However, if the prominence on the accented modifier is strong, the phrases will result in having the pitch contour of (8/8') where the left-most accent deaccents any accents on the right in the phrasal domain.

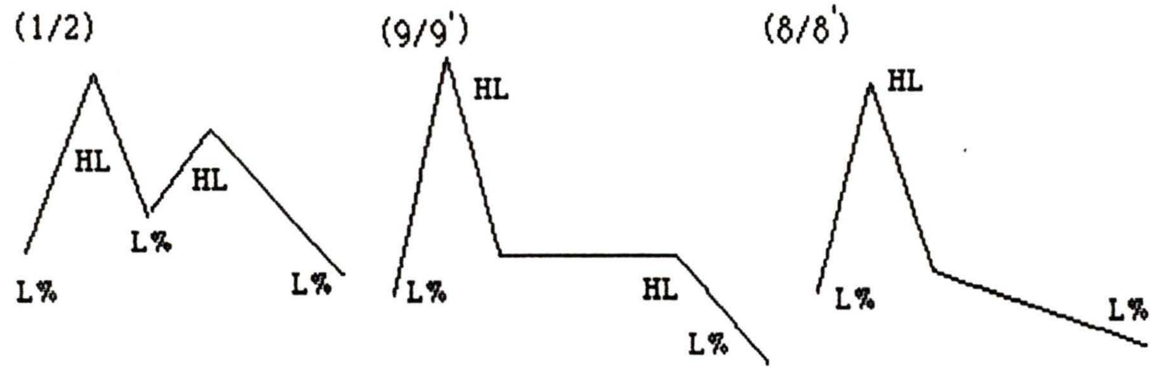


Figure 3.11
Allophony of F0 contours for the phrases with at least two underlying accents.

Such an allophony of the F0 contours can clearly be observed in (B11) *ao'i* + *nimame* + *ma'de*. As Table 3.4 tells us, the phrase is realized as the contour (2): *ao'i* L% *nimame-ma'de* as unmarked, i.e., non-focused, phrasing. It can also be realized as (9): [*AO'I-nimame-ma'de*], as the subject H and S did (in Table 3.4 /marked phrasing). Furthermore, it is possible to utter the phrase with a single accent, just as the subjects K did (in Figure 3.10), and to create the contour (8'): [*AO'I-nimame-made*]. It is quite likely that in conversational speech, the F0 contours schematized as (8/8') and (9/9') occur more regularly than that of (1/2).

In regard to the prominence placed on unaccented modifiers, Pierrehumbert and Beckman (1989: 105-106) demonstrate that in a phrase consisting of an unaccented adjective and a noun, prominence placed on the unaccented adjective erases the interphrasal L% observed in non-focus situations. This is exactly what happened in the experiment. Regardless of the accentuation of a following noun, the narrow-focus on the unaccented modifiers created plateau-like contours schematized as (5') and (7') in Figure 3.12. These schematic contours had, in fact, the same shapes as the contour (5) (in Figure 3.5) and the contour (7) (in Figure 3.7) which were classified as marked phrasing contours for the phrases in *ane-no* group in *oma'me*-set in Table 3.3 .

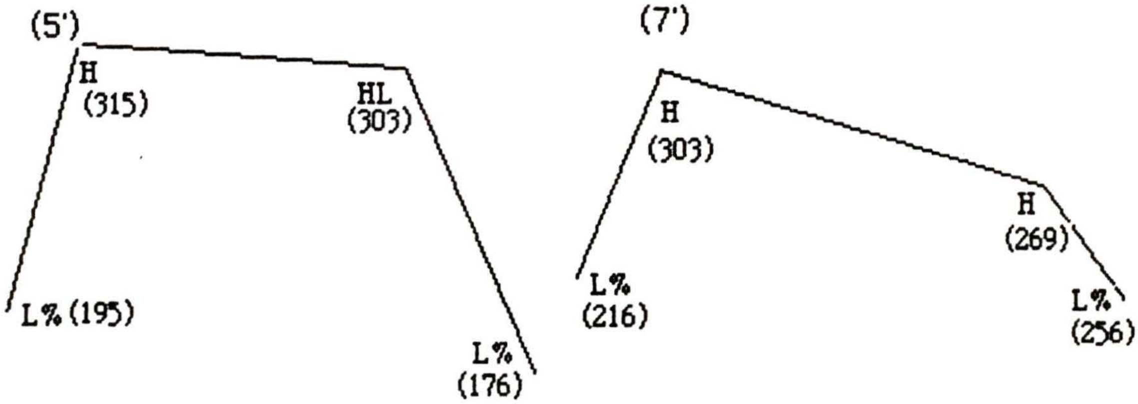
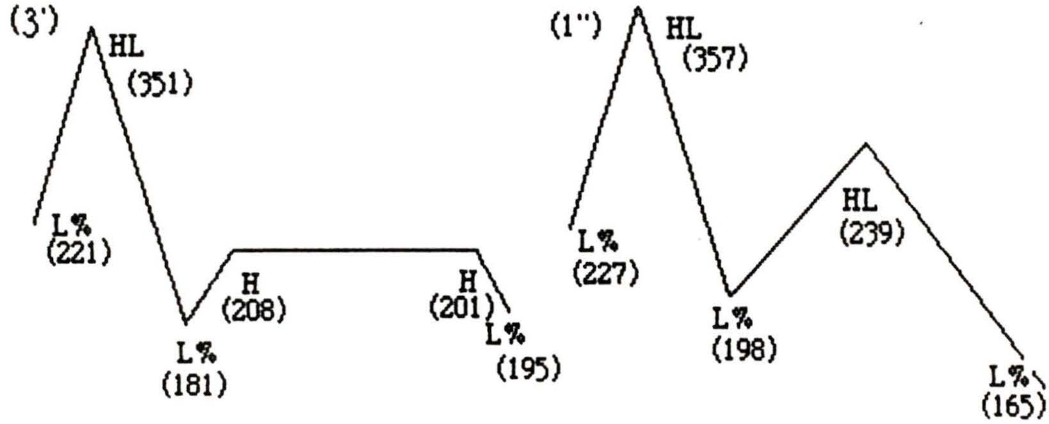


Figure 3.12
Schematized pitch contours which are due to narrow-focusing the phrase initial unaccented modifiers. The F0 values in Hz in the left contour are means of 8 tokens and those in the right contour are means of 3 tokens.

Thus, it can be concluded that these marked phrasing contours schematized as (5) and (7) as well as the marked phrasing contours schematized as (8) and (9) in the previous experiment were, surely, the results of the narrow-focus placed on the initial modifiers.

Before moving to the next section, something must be said about the pitch contours of the phrases whose modifier is the accented *no*-phrase, *a'ni*

-*no*, because the narrow-focus on the modifier created different pitch contours from the narrow-focus on the accented adjectival modifier, *ao'i*. That is, unlike the narrow-focus on the adjectival modifier, the one on the *no* -phrase modifier did not erase the interphrasal L% observed in the corresponding non-focused phrasings, as seen in Figure 3.13.



Cases:
(A33) *a'ni-no + oma'me + jyuu.*
(B33) *a'ni-no + nimame + jyuu.*
(B34) *a'ni-no + nimame + ni.*

Cases:
(A31) *a'ni-no + oma'me + ma'de.*
(A32) *a'ni-no + oma'me + gu'rai.*
(A34) *a'ni-no + oma'me + ni.*
(B31) *a'ni-no + nimame + ma'de*
(B32) *a'ni-no + nimame + gu'rai.*

Figure 3.13
Schematic pitch contours of the phrases with the narrow-focused accented *no*-phrase modifier, *a'ni-no*. The F0 values in Hz in the left contour are means of 3 tokens and those in the right contour are means of 5 tokens.

The contours in Figure 3.13, in fact, support my contention discussed in 3.2.2 that an accented adjectival modifier and an accented *no*-phrase modifier have different localities in deaccenting a phrase final [+Left-winning] postposition. As far as the subject K's utterances are concerned, unlike the narrow-focused accented adjectival modifier, for example, in (B11)*AO'I-nimame-made* (in Figure 3.10), the narrow-focused accented *no* -phrase modifier cannot deaccent the postposition. As seen in Figure 3.13, (B31) *A'NI-NO + nimame + ma'de* is realized as *A'NI-NO* L% *nimame-ma'de* , where the *no*-phrase modifier is unable to deaccent the postposition due to the interphrasal L%. That is, in (B11), the accented phrase, *AO'I-nimame*, deaccents the postposition, *ma'de*, while in (B31), being blocked by the L%, the *no*-phrase modifier cannot form an (accented) accentual phrase with the noun head, *nimame*, and fails to deaccent the postposition.

However, such phrasing and contour differences between the two modifiers cannot be observed in other speakers' utterances. For example, for the subject H, there is no difference in deaccentuation between (A11) [*AO'I-omame-made*] and (A31) [*A'NI-NO-omame-made*] both of which have the same schematized contour (8) (in Figure 3.8) with the deaccented postposition (cf. Table 3.3). Furthermore, if my contention were correct, (B11) *ao'i+nimame+ma'de* in Table 3.4 should be realized, even in the case of unmarked phrasing, as *ao'i-nimame-made* instead of *ao'i-nimame-ma'de*. Thus, although one subject, K, shows some difference in the locality of deaccentuation between the *no*-phrase modifier and the adjectival modifier in the narrow-focus situation, the evidence is too weak to be generalized. Because there is no other evidence to support my contention, I will withdraw the claim that *no*-phrase modifier and the adjectival modifier have different localities in deaccentuation.

3.3.2 Narrow-focus on the Nouns

Pierrehumbert and Beckman (1989) are correct in saying that an intermediate L% is manifested before a narrow-focused item because in the experiment, the insertion of the intermediate L% is clearly observed when prominence is placed on the nouns, *oma'me* and *nimame*. For example, an utterance of (A33) *a'ni-no + oma'me + jyuu* shows a strong, i.e., low F0-valued, intermediate phrase boundary immediately before the narrow-focused noun, as seen in Figure 3.14.

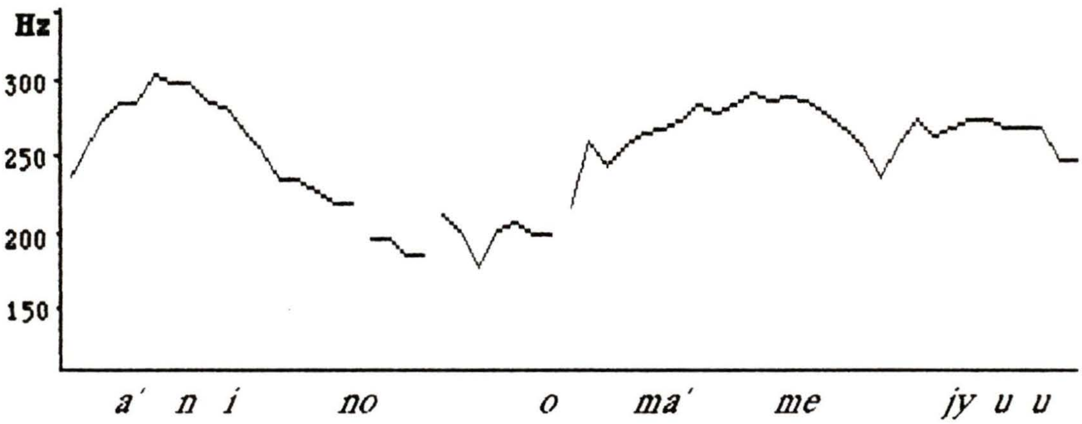


Figure 3.14
A F0 contour of *a'ni-no* L% *OMA'ME -jyuu* showing a strong, i.e., low F0-valued intermediate L% before the narrow-focused noun provides a piece of evidence which supports a claim that an intermediate L% appears before a narrow-focused item.

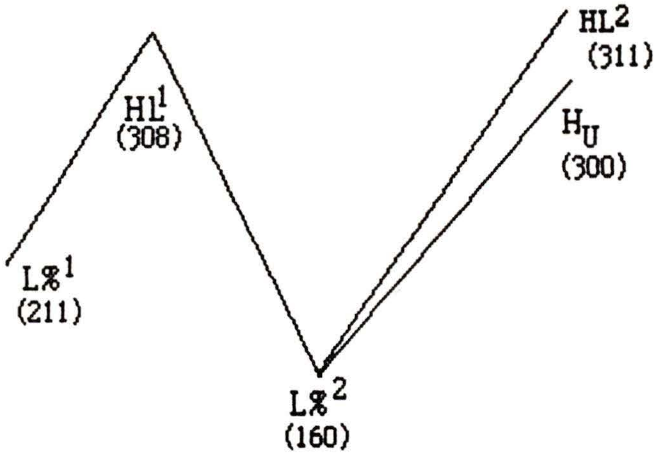


Figure 3.15
Schematic F0 contours of *ao'i-OMA'(ME----)* and *ao'i-NIMA(ME----)* which exhibit (i) the insertion of intermediate L% before the narrow-focused nouns, *oma'(me)* and *nima(me)*; and (ii) the non-propagation of the catathesis effect across the second L%. F0 values are means of 16 tokens.

The schematized pitch contour in Figure 3.15 also supports Pierrehumbert and Beckman's claim because it shows not only the low F0-valued L%, which must be an intermediate L%, but also the high F0-valued second HL (HL^2) and phrasal H (H_U) which are unaffected by catathesis. The catathesis effect of the initial HL does not affect the F0 value of the second HL or the phrasal H because the propagation of the catathesis is blocked by the intermediate L%.

Another rather expected effect of the narrow-focus on the nouns is that the accented noun retains its accent even if it is followed by the postposition marked by the feature [+Deaccenting], i.e., *jyuu* "distributive" or *gu'rai* "binary determiner": i.e.,

······ *omame-jyuu*	versus	······ *<u>OMA'ME</u>-jyuu*
······ *omame-gu'rai*	versus	······ *<u>OMA'ME</u>-gu'rai*

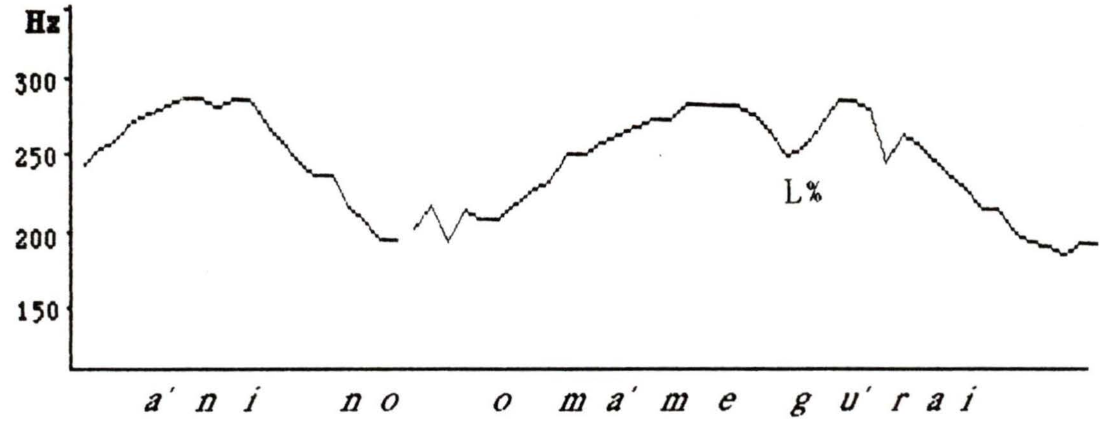


Figure 3.16
A F0 contour of *a'ni-no-OMA'ME-gu'rai* shows that the narrow-focused noun retains its accent despite of the fact that the following postposition is marked by the feature [+Deaccenting].

As seen in Figure 3.16 (and also in Figure 3.14), when the accent H of the noun is retained due to the prominence, an accentual L% appears before the following [+Deaccenting] postposition. The insertion of the accentual L% between the narrow-focused noun and the following postposition may well be to protect the accent H of the noun from being deaccented by the feature [+Deaccenting] as well as to protect the postposition from being focus-subordinated to the noun.

3.3.3 Narrow-focus on the Postpositions

Different from the placement of the narrow-focus on the content words, the placement of narrow-focus on the postpositions does not develop an intermediate L% before the focused items. Instead, as *ao'i-oma'me-MA'DE* in Figure 3.17 demonstrates, an accentual L% is inserted except for the two cases; i.e., when *ma'de* and *ni* are preceded by the unaccented noun

(−). Table 3.5 shows the insertion and non-insertion of the accentual L% before the narrow-focused postpositions.

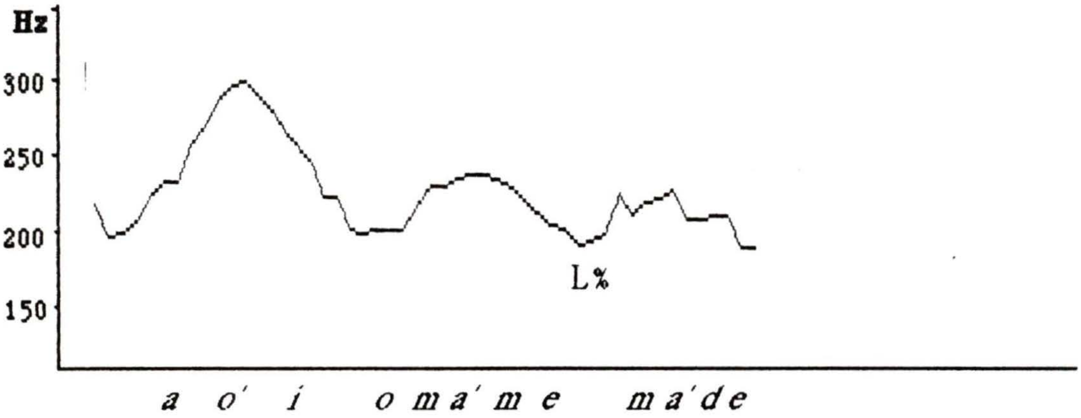


Figure 3.17
The appearance of accentual L% before the narrow-focused postposition in an utterance of *ao'i* L% *oma'me* L% *MA'DE* is displayed. The L% preceding the postposition cannot be an intermediate L% because of its high F0 value and because of its incapability of blocking the chain of catathesis. The L% functions to protect the accent of the postposition from being deaccented by the preceding accented noun, *oma'me*.

Table 3.5
The insertion and the non-insertion of the accentual L% before the narrow-focused postpositions is displayed. The accentual L% is not inserted between the unaccented noun and the postposition marked either by the feature [+Left-winning] or by the feature [+Anonymity]. The +'s and −'s specify the accentuations of the nouns.

Insertion of Accentual L% between Noun and POSTPOSITION			Non-insertion of Accentual L% between Noun and POSTPOSITION	
+,−	L%	*GU'RAI* [+Deaccenting]		
+,−	L%	*JYUU* [+Deaccenting]		
+	L%	*MA'DE* [+Left-winning]	−	*MA'DE* [+Left-winning]
+	L%	*NI* [+Anonymity]	−	*NI* [+Anonymity]

Next, in regard to the accent of [+Left-winning] postpositions, such as in *ao'i+ oma'me+ ma'de*, the narrow-focus conditions the postposition to retain its accent and realizes the phrase as *ao'i* L% *oma'me* L% *MA'DE*. However, the retention of the accent, which is, in a non-focus situation, deleted by the preceding accented noun may not be a direct result of a high F0 value due to the prominence but rather be a result of accentual phrasing. That is, the narrow-focus placed on the [+Left-winning] postposition protects the postposition's accent from being deaccented by the preceding accented noun by way of inserting an accentual L% before the postposition so that the left-win rule, whose domain is an accentual phrase, cannot apply across the accentual L%. The contour in Figure 3.17 illustrates the point.

Lastly, the effects of the narrow-focus on the monomoraic postpositions, such as *ni* "dative", can be seen in two manners. First, if the monomoraic postpositions are preceded by an accented noun, the narrow-focus on the postpositions conditions the insertion of an accentual L% before the postpositions, creating an independent accentual phrase contour for the postpositions, as seen in Figure 3.18. However, if the monomoraic postpositions are preceded by an unaccented noun, the effect of the narrow-focus is manifested merely as a final rising contour, as seen in Figure 3.19.

Figure 3.18
A F0 contour of an utterance of (A11) *ao'i + oma'me + ni* "to the blue beans" is realized as *ao'i* L% *oma'me* L% *NI* where the narrow-focused monomoraic postposition, which usually does not have any prosodic independence, shows an individual accentual phrase contour.

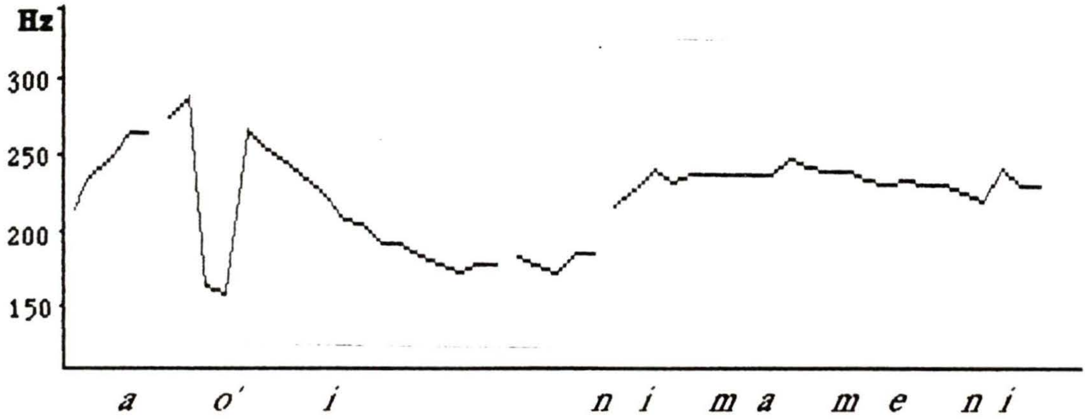


Figure 3.19
F0 contour of an utterance of (B34) *a'ni-no + nimame + ni* "to the brother's cooked beans" is realized as *a'ni-no* L% *nimame-NI* where the narrow-focused postposition shows merely a final rising contour when it is preceded by the unaccented noun, *nimame*.

3.3.4 Summary

By way of summarizing the results of the experiment on the effects of narrow-focusing, I will provide answers to the questions posited at the beginning of the section.

(i) By placing prominence on the modifiers in the phrases of "modifier + noun + postposition", I could replicate all the four pitch contours which were regarded as due to marked phrasing in the previous experiment. In other words, I could resolve my speculation that all the marked phrasing contours in the previous experiment were results of the narrow-focusing the initial modifiers.

(i') Furthermore, depending primarily on the degrees of prominence placed on the initial modifiers, there was allophony of pitch contours for the phrases with at least two underlying accents. That is, when no prominence was involved, the phrases were realized as a contour with two HL's separated by an interphrasal accentual L%. When the initial modifiers were narrow-focused, the phrases had either a binate contour characterized by two HL's with no intervention of an accentual L% or a one-accent-a-phrase contour which retained only the left-most accent in the phrases.

(ii) Pierrehumbert and Beckman (1989) were correct in saying that an intermediate phrase L% manifests right before a narrow-focused item. In the experiment, the insertion of the intermediate L% before a narrow-focused item was clearly observed when a noun in the phrases was narrow-focused, creating a low F0-valued (strong) L% which blocked the propagation of catathesis across the L%.

(iii) However, if the narrow-focused item was a postposition, instead of an intermediate L%, an accentual L% was manifested before the postpostion.

(iv) As far as the utterances of the participant to the experiment were concerned, there was some evidence which supported my contention that the *no*-phrase modifier and the adjectival modifier behave differently in the locality of deaccentuation. The evidence was, however, too weak to be generalized.

(v) The accents which were, in non-focus situations, deleted either by an immediately preceding accent (i.e., left-win rule) or by the feature [+Deaccenting] of a postposition were retained once the units containing the accent were narrow-focused. However, the retaining of the accent was not to do with high F0 values caused by the narrow-focus but rather to do with the phrasing which inserted an interphrasal L% to protect the accent from being deaccented by the left-win rule or by the feature [+Deaccenting].

(iv) Even the monomoraic postpositions, such as *ni* "dative", developed their own accentual phrase once they were narrow-focused. However, such prosodic autonomy was observed only when the postpositions were preceded by an accented item. When the monomoraic postpositions were preceded by an unaccented item, the narrow-focus effect was observed merely as a rising contour at the end of the postposition.

Although no tone scaling and accentual phrasing algorithm are provided for intermediate phrases some of whose tones are affected by narrow-focus, it is possible to translate all the above findings into tone scaling and accentual phrasing algorithm. Basically, what have to done are

(i) compute transform values of tones in intermediate phrases in whose domains narrow-focus is placed; (ii) create a narrow-focus parameter whose values are the ratio of transform values of narrow-focused tones to transform values of non-narrow-focused tones; (iii) develop a few sets of algorithmic accentual phrasing rules which can account for variability in accentual phrasing due to strengths of narrow-focus; and (iv) depending on the parameter values, i.e., the degrees of prominence, a different set of algorithmic rules is activated to develop variations in the accentual phrasing. In regards to the insertion of an intermediate phrase boundary which is conditioned by narrow-focus, there is a device in the intermediate phrasing algorithm which will be discussed in Chapter 5.

Endnotes

1 If the last syllable is long, the accent should be on the first mora of the syllable; e.g., *gakkoo* + *'sika* --> *gakko'o-sika*.

2 It looks like that instead of *a'ni-no*, I should have found a NP[+possessive]-item which has an accent not on the first but on the second mora so that it could have been contrasted with *ao'i*. But it is not possible to find such a NP[+Possessive]-item because of the well-known accentual behaviour of the *no*-postposition which deaccents an immediately preceding syllable (McCawley 1986, Poser 1984): e.g. *ie'* + *no* --> *ie-no* (house-of); *inu'* + *no* --> *inu-no* (dog's). Also, it was not possible for me to find, instead of *ao'i*, an adjective which has a similar phonemic configuration as *a'ni-no* and still has an accent on the first mora.

3 No schematic pitch contours are provided for Subject K's utterances of B13 and B23 because there is only one token for each case, making it impossible to provide generalized contours.

4 Because one of the items, i.e., postposition, in the cursor of the second move is not a content word, the generalization does not apply at this stage.

[5] Such a treatment may support the existence of "tone spreading" which Pierrehumbert and Beckman (1989) claim dose not exist in Tokyo Japanese. I have no intention to argue about whether or not tone spreading exists in Tokyo Japanese. (Cf. Kori's (1987) claim for Osaka Japanese).

[6] If it is an intermediate phrase boundary, in theory, it would be the last accentual phrase boundary of the preceding intermediate phrase which functions as an intermediate phrase boundary.

[7] That is "if it is a content word".

[8] I have no intention of postulating an additional prosodic unit, the "phonological word", in Pierrehumbert and Beckman's prosodic model. Postulating the phonological word is redundant because it can be represented either by a "word" or by an "accentual phrase".

[9] In Pierrehumbert and Beckman (1989), a high-tone line is set at a possible F0 value which a tone with maximum prominence will have.

[10] Although Pierrehumbert and Beckman (1989) say that an accent HL is always higher than a phrasal H in an accentual phrase, the schematic contour indicates that if an accent HL is at a postposition, the HL can be lower than a phrasal H.

[11] Only one subject was requested to participate in the experiment due to the size of the data. That is, having more than one subject would have made the size of the data too large to analyse in terms of time and labour involved.

[12] "Prosodic weight" of postpositions may also influence the allophony. For example, *jyuu*, an accented postposition with the feature [+Deaccenting] tends not to be focus-subordinated.

Chapter 4
Catathesis

Catathesis, a downtrend in an intermediate phrase, is the topic of this chapter. I will investigate relationships between iterative applications of catathesis and syntactic tree configurations.[1] Despite the claims of Poser (1984) and of Pierrehumbert and Beckman (1989) that catathesis takes place and chains regardless of the position and the number of its triggers, (i.e., accents), Kubozono (1985) makes an interesting argument that catathesis propagates only in a left-branching noun phrase. In a right-branching noun phrase, Kubozono says, even if a phonological condition is satisfied, catathesis does not chain beyond an innermost noun phrase, having the right-branching syntactic node as a constraint. I would like to examine, first, whether there is any difference between the iterative application of catathesis in the left-branching NPs and in the right-branching NPs. If there is any difference, then, I would like to investigate whether such branchingness can be regarded as a syntactic constraint on the application of catathesis.

An acoustic experiment which examines catathesis contours of various types of left- and right-branching NPs will show that there is a difference between left- and right-branching NPs in catatheis contours although the shapes of the catathesis contours in the experiment and those in Kubozono's study will not be the same. Furthermore, an additional experiment which examines blocking of catathesis will indicate that there is no such phenomenon as the *blocking* of catathesis.

4.1 What is Catathesis?

Catathesis, which is defined as an iterative pitch compression at each accent, is a tonal assimilation originally reported in the study of African tone languages (Schachter and Fromkin 1968, McCawley 1970). Depending on the tonal transparency, there are two types of catathesis; automatic catathesis and non-automatic catathesis. Automatic catathesis or "downdrift" is a tonal phenomenon in which any high tone separated from a preceding high tone by a low tone is realized with a lower pitch value than the preceding high tone. In other words, in a language with downdrift, high tones are lowered after each low tone, realizing the first high tone in a sentence as the highest tone in pitch and the last high tone in the sentence as the lowest high tone in pitch (Kenstowicz and Kisseberth 1979). Non-automatic catathesis or "downstep" is the same type of tonal assimilation as automatic catathesis except that the condition of the lowering the high tone, i.e., the low tone preceding the high tone, is not visible on the surface. Thus, in a language which has downstep, there is a difference between the sequence of H H where the second H has the same pitch value as the first H and the sequence of H !H[2] where the second !H has a lower pitch value than the first H due to the tonal assimilation conditioned by an invisible low tone which is floating underlyingly between the H and the !H (Clements and Ford 1979).

Catathesis was first proposed for English by Pierrehumbert (1980).

Pierrehumbert argues that English intonation can be described with the inventory of just two tones, High and Low, if a catathesis realization rule is postulated. That is, a catathesis realization rule can account for the tonal height which may motivate the postulation of Mid tone as, instead, a catathesized H tone and keep the inventory of tones as small as possible with no inclusion of the M tone. Furthermore, the catathesis realization rule can account for a down-stepping pitch contour as a result of iterative applications of catathesis. Without such a rule, for example, a contour of uttering (without a pause) a series of berry names, "blueberries !bayberries !raspberries !mulberries !brambleberries" will demand several discrete tonal levels for describing it. A problem with postulating so many different levels of tones is that by doing so it will not only make the tonal inventory cluttered but also fail to explain why so limited possible combinations of the tones are used only in certain environments, such as above (Beckman and Pierrehumbert 1986a: 272).

It is Poser (1984) who first proposed catathesis for Japanese, saying that much of the downtrend in the domain of a major (- intermediate) phrase can be attributed to catathesis triggered by an accent. Unlike in African languages (or in English), in Japanese, catathesis is not triggered by a mere sequence of H and L (such as a phrasal H and a boundary L in an unaccented accentual phrase) but is specifically triggered by an accent, HL. To show that the absolute pitch value of a H tone is lower after an accented word than after an unaccented word, Poser compared a H tone in minimally contrasting phrases: *uma'i mirin* "tasty cooking wine" and *amai mirin* "sweet cooking wine". Poser argues that the fact that F0 fall from the accented adjective to the noun is greater than F0 fall from the unaccented adjective to the noun is not to do with the intrinsic F0 value difference between accented *uma'i* and unaccented *amai*. If it were the case, the F0 values of the noun should be the same. However, the fact that the noun is lower in F0 value when preceded by the accented adjective than when preceded by the unaccented adjective proves that the accent in *uma'i* lowers a pitch value of *mirin*. The point can be illustrated from the data in Figure 4.1 where the F0 value of the second peak after the accented peak shows significant F0 reduction.

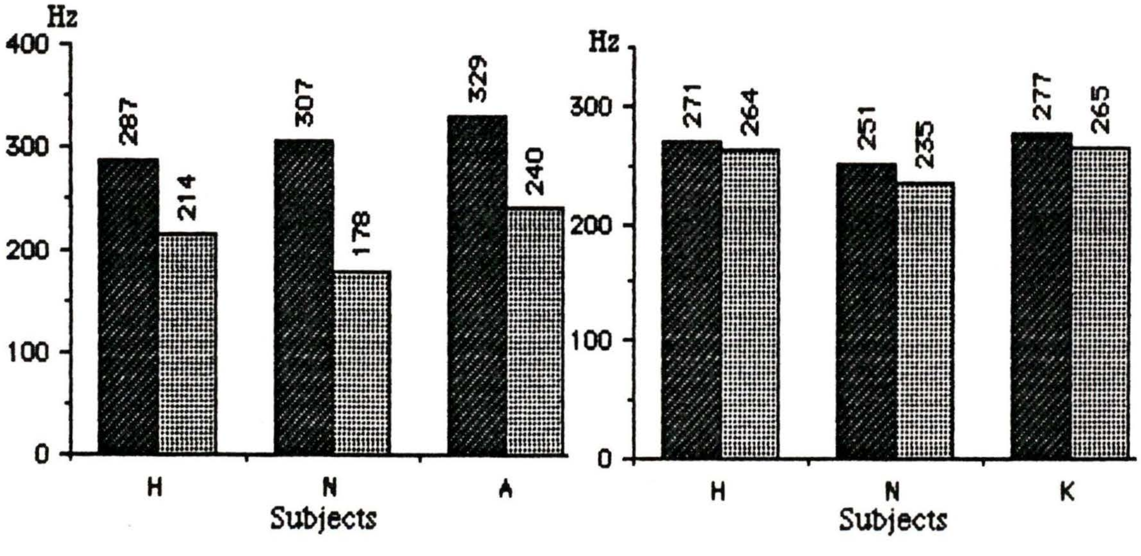


Figure 4.1
Mean peak F0 values of adjective and noun in the phrase; (1)*ao'i* (accented)- *u'mi* "blue sea" uttered by three subjects on the left-hand side and mean F0 values of adjective and noun in the phrase; (2)*arai* (unaccented)- *u'mi* "rough sea" uttered by the same subjects on the right-hand side.

Such a F0 lowering effect of an accent can also be plotted as a peak-to-peak relationship, as in Figure 4.2. If an accent were not to lower the following H, all the data points in the plot would be on the same line. The fact that the white dots representing *ao'i- u'mi* occupy a lower y-value region than the black dots representing *arai- u'mi* indicates that an accent does, indeed, lower the following H.

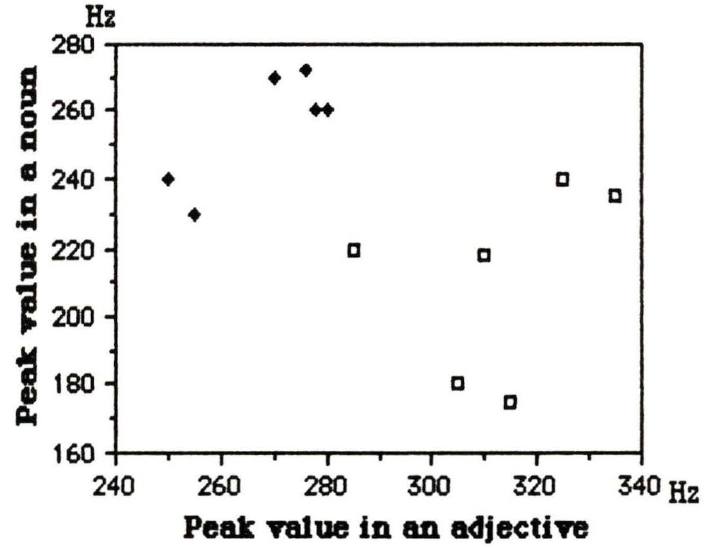


Figure 4.2
F0 value of the peak in the noun (*u'mi*) plotted against the peak in the preceding adjective (*ao'i* or *arai*). The Black diamonds are for the phrase, *arai* (unaccented)- *u'mi* "rough sea", and the white squares are for the phrase, *ao'i* (accented)- *u'mi* "blue sea".

It is Beckman and Pierrehumbert (1985, 1986a&b, 1989) who rigorously and extensively investigated the nature of catathesis in Tokyo Japanese. Their analysis is characterized as follows:

(i) Catathesis is a downtrend which takes place strictly in the domain of an intermediate phrase. An accent, bitonal HL, is the condition of catathesis which compresses the whole pitch range wherein following tones are realized. A chain of catathesis is realized if there are more than one accent in an intermediate phrase. The effect of catathesis, however, does not cross over an intermediate phrase boundary which is generally characterized by being strong, i.e., low F0-valued.[3]

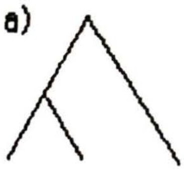
(ii) In reference to the timing of triggering catathesis, Beckman and Pierrehumbert agree with Poser that, judging from the fact that the L of an accent, HL, is lower in F0 than the boundary L which precedes the accent, the compression of F0 range starts within an accent, affecting the L tone of the accent itself.

(iii) Beckman and Pierrehumbert also agree with Poser about the treatment of catathesis as a register-shift, in which account what is lowered is not a tone but the whole F0 region in which tones are realized. Another more local way of representing catathesis proposed by Pierrehumbert (1980) obtains the transform value of each tone in relation to the preceding tone. In this treatment, the value of tone is computed relative not to the F0 region but to the preceding tone. The computation will be carried out strictly in a one-tone-to-one-tone fashion from left to right. The clumsiness of this treatment arises, in the case of Tokyo Japanese, when, for example, a value of a phrasal H is computed from a preceding tone, i.e., a boundary L which is subject to weak (high F0 valued)-strong (low F0 valued) allophony.[4] Because the value of the phrasal H is independent of the allophony, it is required to have two separate rules to compute the transform value of the phrasal H. Such clumsiness does not occur with the register-shift account of catathesis (Pierrehumbert and Beckman 1989: 140).

In short, catathesis in Tokyo Japanese is defined as an iterative pitch compression after each accent, operating strictly in the domain of an intermediate phrase.

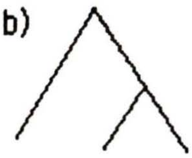
4.1.1 Kubozono's Claim

It is Kubozono (1985) who claims that there is a syntactic constraint on the iterative application of catathesis, saying that "[catathesis] is dependent on the syntactic structure of the whole phrase and does not apply over any right-branching syntactic node" (Kubozono 1985: 77).

a) 

Ta'roono ito'kono o'kusan

"Taro's cousin's wife."

b) 

Ta'roono waka'reta o'kusan

"Taro's divorced wife."

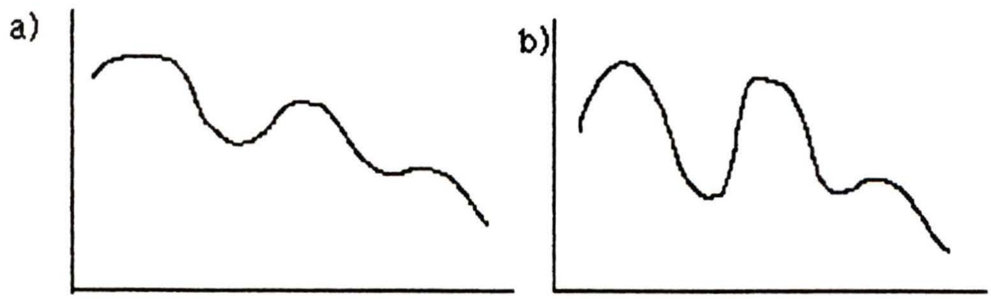


Figure 4.3
Kubozono's stimuli for the left-branching NP (a) and for the right-branching NP (b) as well as the corresponding schematic pitch contours.

Showing schematic contours somewhat like the ones in Figure 4.3, Kubozono says that in the contour (a) of the left-branching NP, F0 peaks for the second and the third words are both considerably lowered; however, in the contour (b) of the right-branching NP, although the third peak is substantially lower than the other two peaks, the second one is hardly lowered. Despite that both phrases have the same accentual patterns, they differ in the F0 contours because of the syntactic configurations. That is,

> "the left-branching NP undergoes the process of downdrift as a whole, while its right-branching counterpart shows a clear effect of the process only within its innermost NP ..... the right-branching syntactic node in right-branching NPs prevents the process of downdrift or, at least, it weakens the effect of the prosodic process" (ibid; 77-78).

Obviously, Kubozono made his claim depending on a very limited size of data, i.e., one set of contrasting phrases uttered ten times by just one subject. The size of the data is definitely not large enough to substantiate the claim. In the ensuing section, relying on a larger size of data, I will examine whether syntactic tree configurations have anything to do with the iterative application of catathesis.

4.2 Experiment on the Correlation between Iterative Application of Catathesis and Syntactic Trees

4.2.1 Design of the Stimuli

First, I describe how the stimuli are designed in the experiment to examine the issue: "Do left- and right-branching tree configurations have any correlations with the iterative application of catathesis?" Because there are only five subjects available for the experiment and because the principle of one subject per case (i.e., non-repetition of a case by the same subject) is respected, it is necessary to prepare quite a number of phrases in order to amass a fair size of data. Table 4.1 lists the noun phrases of "modifier + modifier + noun" used as stimuli.

Table 4.1
Stimuli for the experiment investigating the correlation between the syntactic tree branchingness and the iterative application of catathesis.

I. Right-branching Noun Phrases:

A2 *a'wa-no ya'bona a'ni-o* "clumsy brother in Awa".

A3 *mi'nami-no a'warena mi'nari*-o "Minami's pitiful appearance".

B2 *a'wa-no a'ji-no na'be-o* "fish pot of Awa".

B3 *a'rima-no a'wabi-no o'iru-o* "abalone oil from Arima".

C2 *ya'bona a'wa-no a'ni-o* "clumsy brother in Awa".

C3 *a'warena mi'nami-no mi'nari*-o " pitiful Minami's appearance".

II. Left-branching Symmetrical Noun Phrases:

A'2 *ya'bona a'ni-no o'no-o* "<u>clumsy brother's</u> ax".

A'3 *a'warena mi'nami-no mi'nari*-o "<u>pitiful Minami's</u> appearance".

B'2 *a'wa-no a'ni-no o'no-o* "brother in Awa's ax".

B'3 *a'ruje-no a'wabi-no one'dan-o* "Algerian abalone's price".

C'2 *ya'ya moro'i ama'do-o* "somewhat fragile shutter".

C'3 *iya'ni manuru'i yama'uri*-o " awkwardly warm wild-melon".

III. Left-branching Relative Clause Noun Phrases:

D2 *e'mi-no nu'ida bu'ra-o* "brassiere which Emi took off".

D3 *mi'nami-no no'nda nomimo'no*-o "drink which Minami had".

E2 *e'ngi-no ya'bona a'ni-o* "brother whose play is clumsy".

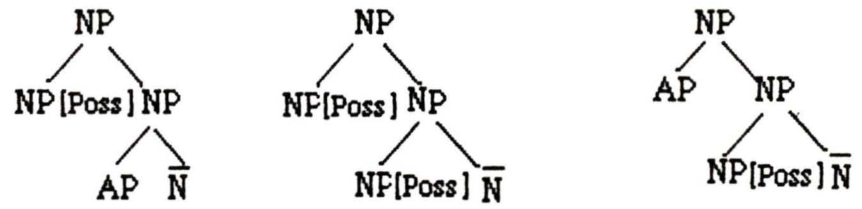
E3 *mi'nari-no a'warena mi'nami-o* "Minami whose appearance is pitiful".

F2 *ya'bona e'ngi-no a'ni-o* "brother whose play is clumsy".

F3 *a'warena mi'nari-no mi'nami-o* "Minami whose appearance is pitiful".

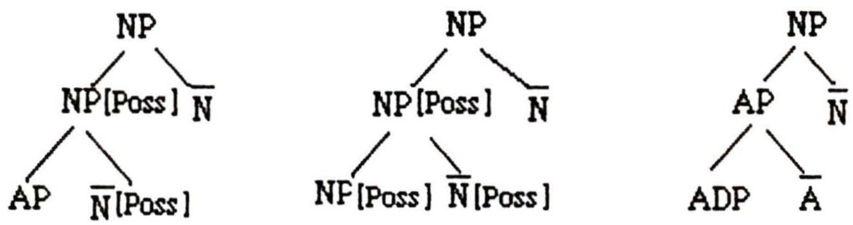
First, as right-branching NPs, three types of NPs with slightly different phrasal configurations are prepared. The differences are due to the types of modifiers they take:

(A) (B) (C)



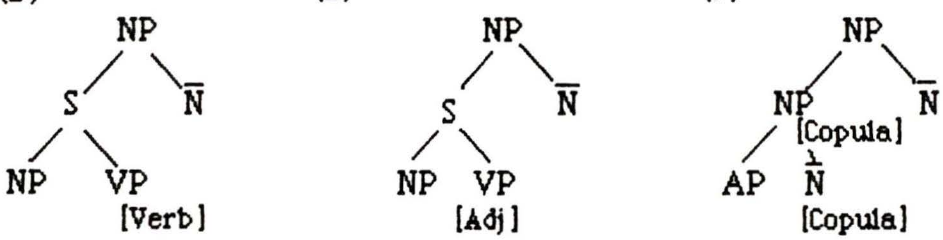
Next, for left-branching NPs, two different groups are prepared. One group has the phrases whose syntactic trees are somewhat symmetrical to the right-branching trees of the above phrases[5] in the sense that the innermost NPs in the right-branching phrases are moved to the left.

(A') (B') (C')



The other group is for left-branching relative clause NPs:

(D) (E) (F)



One may wonder why it is necessary to have noun phrases with so many different phrasal configurations. My answer is that if, as Kubozono says, the difference between the left- and right-branchingness is the factor affecting F0 contours of noun phrases in the iterative application of catathesis, the slight differences among the phrases should not matter because they are, basically, either left-branching or right-branching noun phrases. There should be a clear difference in contour patterns between the two syntactic types of phrases if the branchingness is the real factor.

In Table 4.1, as will have been noticed, each tree has a pair of noun phrases which differ in length. In the first member in a pair, each accent is separated by two morae; e.g., (A2) *a'wa-no ya'bona a'ni-o* and in the second member in the pair, each accent is separated by three morae;[6]: e.g., (A3) *mi'nami-no a'warena mi'nari-o*. The aim of having for each tree a pair of phrases which differ in the distance between accents is to examine whether the timing differences in triggering catathesis play any role in lowering a pitch range. For example, does a shorter distance between accents condition a higher pitch compression ratio than a longer distance does? If it is the case, then, it will be indicated that, besides catathesis, there must be a downtrend which is sensitive to proximity of triggers. For example, a F0 contour in Figure 4.4 may suggest that, besides catathesis, there might be an unreported downtrend by which the first accented item severely compresses the F0 contour of the following item when two accents are as closely located as just one mora apart.

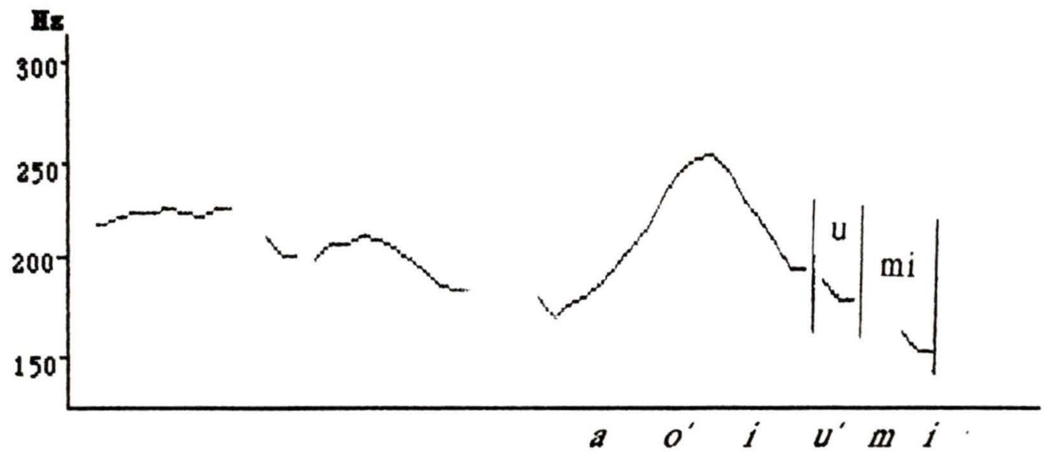


Figure 4.4
A F0 contour of an utterance of *aoi'-u'mi* "blue sea", implying that there might be, besides catathesis, a downtrend which is sensitive to the proximity of the accent H's.

If there exists such a downtrend sensitive to the proximity of accents, catathesis cannot account for the downtrend because catathesis is simply an iterative process which compresses a F0 range after each accent regardless of its position so that a catathesis constant (compression ratio of F0 range) should not be affected by the proximity of triggers. If there are significant F0 lowering differences due to the differences in the distance between accents, there must be an unreported downtrend. If, on the other hand, the differences in the distance between accents has no significance in F0 lowering, we will be assured that it is only catathesis which plays a significant role in downtrend in the domain of the intermediate phrase.

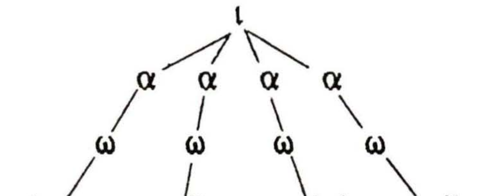
The stimuli are set in a carrier sentence: S- *wa* NP-*o mi'ru*, "S sees NP". As an attempt to break isochrony created by either two or three mora apart accentual distributions, the items in the subject slot are constantly changed with unaccented last names of different mora lengths; e.g., *Ono*, *Yamada*, *Miyamoto*, etc. These sentences are, then, treated as answers to the questions which paraphrase the noun phrases in the answers: e.g.,

Honda-wa arima-me'ibutsu-no awabi-o'iru-o mima'suka?
"Does Honda see the abalone oil, a speciality of Arima?"

(B31) *Honda-wa a'rima-no a'wabi-no o'iru-o mi'ru.*
Honda sees the abalone('s) oil from Arima."

The device has two aims. First, as being old information, any item in the NPs (which is underlined in the example) would not be narrow-focused. Second, the question which paraphrase the NPs in the answers should help subjects understand branchingness of the NPs clearly.

Lastly, I must mention that because the aim of the experiment is to investigate the correlation between the iterative application of catathesis which takes place in one intermediate phrase and the left- and right-branchingness of syntactic trees, the subjects were requested to utter the predicates (NP + V) of the answers with no insertion of a pause. Although the request did not guarantee that the predicates would be uttered as a unit without an intermediate boundary, any data which showed the insertion of the boundary was eliminated from analyses. In other words, all the predicates used for the analyses had the following prosodic tree of a single intemediate phrase:



E.g.: (A21) *a'wano ya'bona a'nio mi'ru*

"(I) see the clumsy brother in Awa."

Having the above prosodic tree configuration for the utterances of the predicates in the answers does not mean that they cannot have other prosodic trees. However, there are two reasons why such long chains of catathesis are generated. First, because there is no report of such long catathesis chains in the literature, I would like to see whether it is possible to produce, relatively naturally, the chains of four catatheses. Secondly, because the catathesized P4 values from the current stimuli will be needed in the following experiment (4.3), not just the NPs, but the whole predicates (NP + V) are realized as an intermediate phrase.

4.2.2 Catathesis Constants

To examine iterative applications of catathesis in noun phrases, two types of catathesis constants are computed: a local catathesis constant and an accumulative catathesis constant. A local catathesis (c) constant indicates how much each peak (actually, F0 range where a peak is realized) is compressed relative to the preceding peak. The following formula is used for the computation:

$$c = (h_{new} - r) / (h_{old} - r)$$

This type of c-constants specifies the ratio of peak compression in relation to the preceding peak.

Another type of c-constant is an accumulative c-constant which indicates a compression ratio of a peak globally. That is, the compression ratio of a peak is described in terms of an entire transform space so that an accumulative c-constant makes it possible to see a relative prominence relationship of a peak not only with the preceding peak but also with any other peaks in the space. An accumulative c-constant of a peak can be obtained by multiplying a local c-constant of a peak by a local c-constant of the preceding peak. Or, it may be obtained by the following formula:

$$c_{accumulative} = (h_{new} - r) / (h_{global} - r)$$

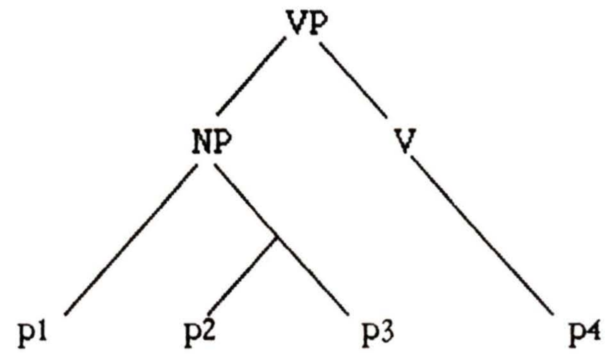
In the current analyses, h_{global}, which is a F0 value of a high-tone line for an entire transform space, is set at 320 Hz and r (reference line) is set at 120 Hz.

As an example, Table 4.2 demonstrates how local c-constants and accumulative c-constants are computed for the peaks in the right-branching NPs whose peaks (P^1, P^2, P^3) as well as the V peak (P^4) are two morae apart.

Table 4.2
Computations of local and accumulative catathesis constants for the peaks in the right-branching NPs whose peaks (P^1, P^2, P^3) as well as the V peak (P^4) are two morae apart. F0 values assigned to the peaks are means of 15 tokens.

Peaks	F0 in Hz	$(h_{new} - r)$ / $(h_{old} - r)$	Local c-constant	$(h_{new} - r)$ / $(h_{global} - r)$	Accumulative c-constant
P^1	303 Hz	303 - 120 / 320 -120 =	0.915	303 - 120 / 320 - 120 =	0.915
P^2	258 Hz	258 - 120 / 303 - 120 =	0.754	258 - 120 / 320 - 120 =	0.690
P^3	236 Hz	236 - 120 / 258 - 120 =	0.840	236 - 120 / 320 - 120 =	0.580
P^4	191 Hz	191 - 120 / 236 - 120 =	0.612	191 - 120 / 320 - 120 =	0.355

The results of computing the local and accumulative constants are summarized in Figure 4.5 for the VP embedding the right-branching NPs; in Figure 4.6 for the VP embedding the left-branching symmetrical NPs; and in Figure 4.7 for the VP embedding the left-branching relative clause NPs. In each figure, a table of the c-constants is accompanied by two schematic pitch contours: one for the phrases in which each accent is two mora apart and the other for the phrases whose accents are three morae apart. The schematic pitch contours are obtained by linearly interpolating F0 peaks and valleys. Also, in the pitch contours, accumulative c-constants are schematically approximated by numerals ranging from 1 to 4.



		p1	p2	p3	p4
Accumulative	(2)	0.915	0.690	0.580	0.355
c-constants	(3)	0.940	0.725	0.605	0.370
Differences		(0.025)	(0.035)	(0.025)	(0.015)
Local	(2)	0.915	0.754	0.840	0.612
c-constants	(3)	0.940	0.771	0.834	0.612

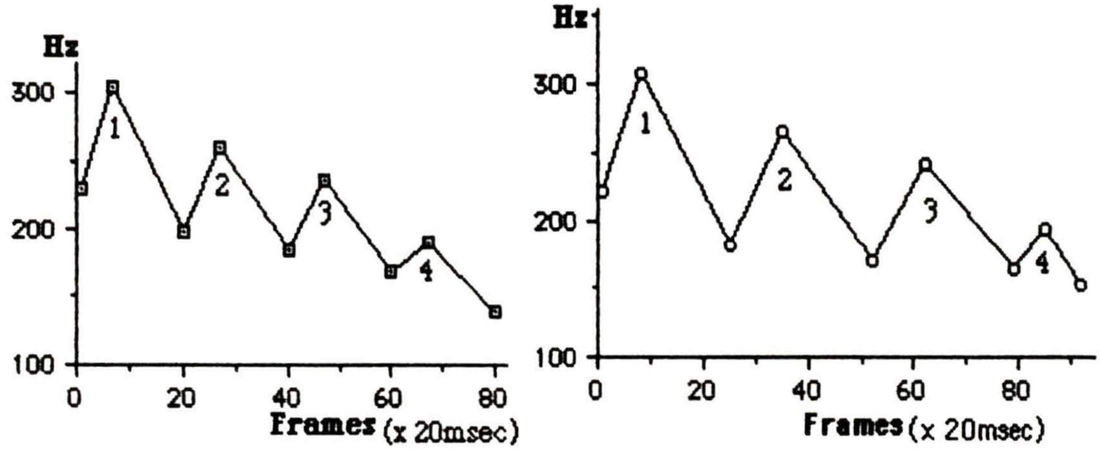
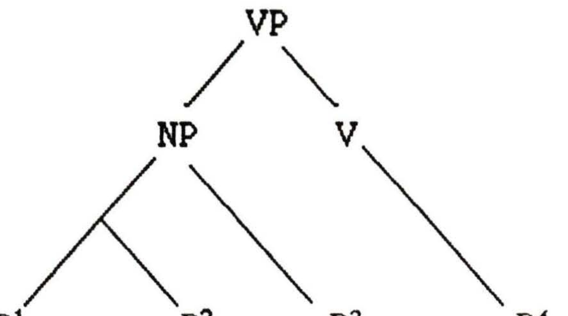


Figure 4.5
A list of accumulative catathesis constants and local catathesis constants for the **VP embedding the right-branching NPs** whose accents are two mora apart in one member (2) and three mora apart in the other member (3). The value differences in the accumulative *c*-constants between the two members are specified in the parentheses. Accompanied schematic pitch contours are drawn by linearly interpolating F0 maximas and minimas whose F0 values are means of 15 tokens. The pitch contours on the left is for the VP embedding the NPs in which each accent is two morae apart and the one on the right is for the VP embedding the NPs in which each accent is three morae apart.



		p1	p2	p3	p4
Accumulative	(2)	0.860	0.655	0.635	0.350
c-constants	(3)	0.895	0.625	0.625	0.350
Differnces		(0.025)	(0.030)	(0.010)	(0.000)
Local	(2)	0.860	0.762	0.970	0.551
c-constants	(3)	0.895	0.698	1.000	0.560

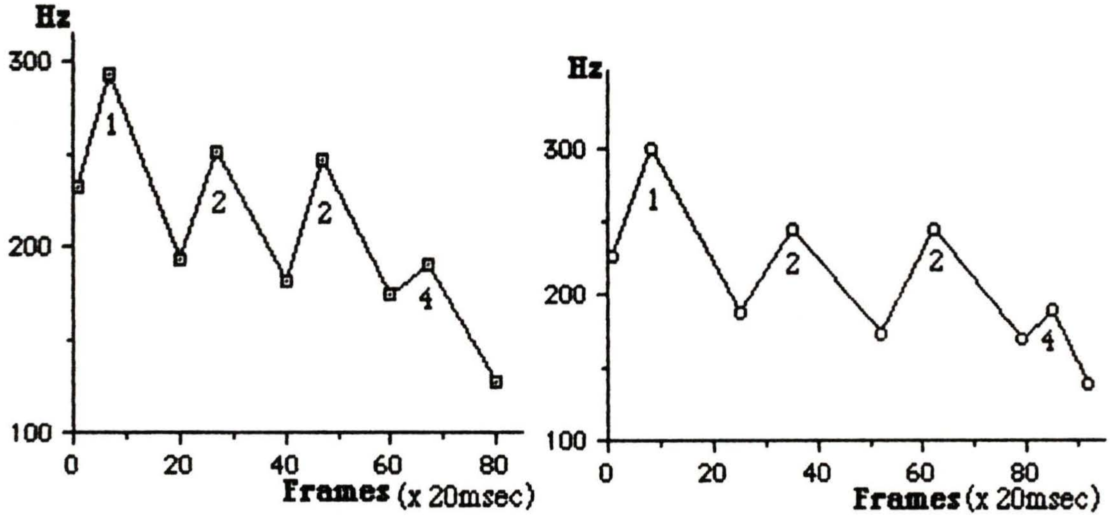
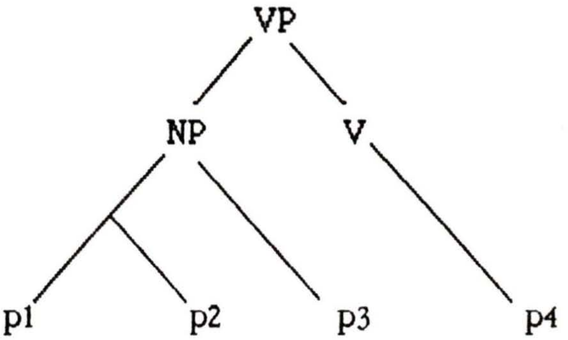


Figure 4.6
A list of accumulative catathesis constants and local catathesis constants for the **VP embedding the left-branching symmetrical NPs** whose accents are two mora apart in one member (2) and three mora apart in the other member (3). The value differences in the accumulative *c*-constants between the two members are specified in the parentheses. Accompanied schematic pitch contours are drawn by linearly interpolating F0 maximas and minimas whose F0 values are means of 15 tokens. The pitch contours on the left is for the VP embedding the NPs in which each accent is two morae apart and the one on the right is for the VP embedding the NPs in which each accent is three morae apart.



		p1	p2	p3	p4
Accumulative	(2)	0.930	0.615	0.610	0.355
c-constants	(3)	0.940	0.635	0.650	0.355
Differnces		(0.010)	(0.020)	(0.040)	(0.000)
Local	(2)	0.930	0.611	0.999	0.591
c-constants	(3)	0.940	0.675	1.023	0.542

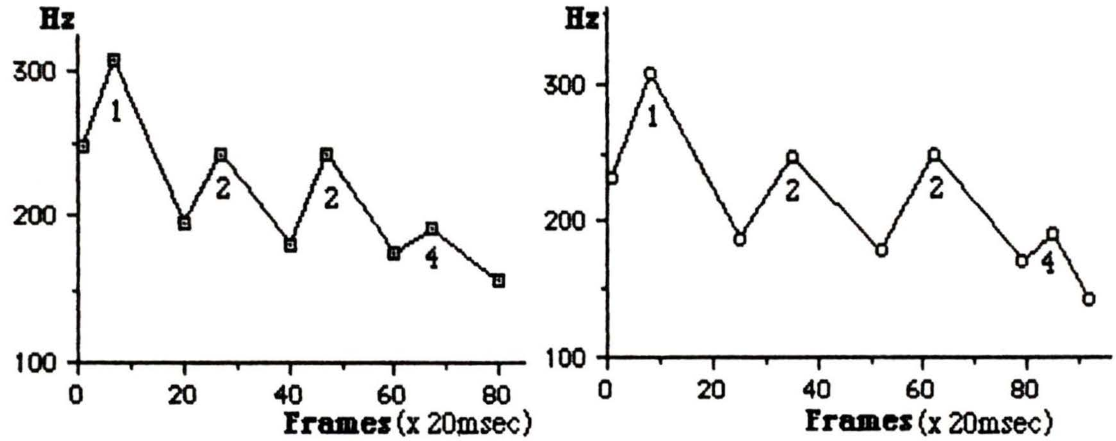


Figure 4.7
A list of accumulative catathesis constants and local catathesis constants for the **VP embedding the left-branching relative clause NPs** whose accents are two mora apart in one member (2) and three mora apart in the other member (3). The value differences in the accumulative *c*-constants between the two members are specified in the parentheses. Accompanied schematic pitch contours are drawn by linearly interpolating F0 maximas and minimas whose F0 values are means of 15 tokens. The pitch contours on the left is for the VP embedding the NPs in which each accent is two morae apart and the one on the right is for the VP embedding the NPs in which each accent is three morae apart.

4.2.3 Timing Differences in Triggering Catathesis

First, taking the accumulative *c*-constants in Figures 4.5, 4.6, and 4.7, I will compare the phrases whose accents are two mora apart with the phrases whose accents are three mora apart in order to examine whether there are any systematic differences in the *c*-constant values due to the distance differences between accents. The value differences in the accumulative *c*-constants between the two members are indicated in parentheses in each figure.

In each tree group, there are somewhat noticeable value differences in the second and third peaks. However, except for the third peak (P^3) of the left-branching relative clause NPs (in Figure 4.7) which has the difference of 0.040, the value differences of the accumulative *c*-constants between the two members are only around 0.030 or less, which can be translated to, merely, 6 Hz or less. The differences should be regarded as negligible.

There also seem to be no fixed patterns of the value differences except that the *c*-constant values for the last V peak are almost identical in both members. That is, the differences are 0.000, or null, in the VPs embedding the left-branching symmetrical and left-branching relative clause NPs and 0.015 in the VP embedding the right-branching NPs. The fact that the accumulative *c*-constant value difference for P^4 between the two members is almost identical across the three tree groups may imply that at the end of utterances, there may be a fixed F0 value at which the other downtrends, i.e., declination and final lowering, aim to drop. Such overall downtrends may affect the uniformity of the *c*-constants for the V peaks.

The fact that, in each tree group, there is no significant difference in accumulative *c*-constants between two members of the NPs which differ in the distance between accents makes it possible to say that the distance differences between accents do not contribute to the effect of catathesis. In other words, there is no downtrend which is sensitive to proximity of accents. If there is any severe pitch compression which looks like being conditioned by the proximity of accents, as seen in Figure 4.4, it must be due to other factor(s), possibly, narrow-focus. In ensuing analyses, I will consolidate the data from both members of the NPs into one set of the data.

4.2.4 Right- and Left- branching Trees and Chains of Catathesis

Now, I would like to examine the iterative application of catathesis in the right- and the left-branching NPs. First, the results of the experiment encoded in the form of the local catathesis constants in Figures 4.5 show that in the right-branching NPs, there is an iterative application of catathesis. That is, in the right-branching NPs, P^2 is compressed 75 - 77 % relative to P^1, and P^3 is compressed 83 - 84 % relative to P^2. Even P^4 at the V node which is the outside the domain of the NPs is affected by the chain of catathesis and lowered 61 % relative to the preceding peak. Thus, among the three peaks at the terminal nodes in the right-branching NPs as well as the peak at the V node, there is a chain of four catatheses. That is, none of the syntactic nodes functions as a constraint on the iterative application of catathesis.

The situation is different in the left-branching symmetrical NPs in Figure 4.6 and the left-branching relative clause NPs in Figure 4.7 because their third peak, P^3, does not show any effect of catathesis. All the local *c*-constant values for the peak are around 1.000, i.e., 0.970 and 1.000 for the left-branching symmetrical NPs and 0.999 and 1.023 for the left-branching relative clause NPs. The c-constants clearly say that the third peak is not compressed at all. Thus, it seems to be that, unlike in the right-branching NPs, in the left-branching NPs, catathesis applies only within an innermost NP, i.e., between P^1 and P_2, having a left-branching node (i.e., the NP node immediately dominated by the VP node) as a constraint on the iterative application of catathesis.

Thus, my results seem to be polar opposite to Kubozono's results. That is, in Kubozono's study, there is an iterative application of catathesis in the left-branching NP whereas in the right-branching NP, catathesis effect is seen only in the innermost NP. In my study, it is in right-branching NPs where an iterative application of catathesis is seen and in left-branching NPs, catathesis effect is seen in the innermost NPs

4.2.5 Tree Overlays

Before comparing Kubozono's results with my results in detail, there is one issue at which I would like to look briefly. As may be recalled, in the setting of the stimuli (4.2.1), each syntactic tree group was provided with three slightly different types of NPs. As the grouping of the NPs was syntactic, there would be no guarantee that all three different phrasal types in each group would not behave differently in terms of catathesis applications. To examine whether all three types of the NPs in each group have uniformity in catathesis contours, F0 values of the third peak are plotted against F0 values of the second peak as a scatter graph. The reason why, among the peaks, P^2 and P^3 are chosen is that, as far as my data are concerned, allophony of the catathesis contours can be judged as the peak-to-peak relationship between P^2 and P^3. If all three phrasal types in a tree group show the same catathesis contour, all the dots representing the correlation between P^2 and P^3 should be well blended. If, however, one phrasal type behave differently from other two types in catathesis contours, the dots assigned to the phrasal type will become separated from the cluster of the dots assigned to the other two phrasal types.

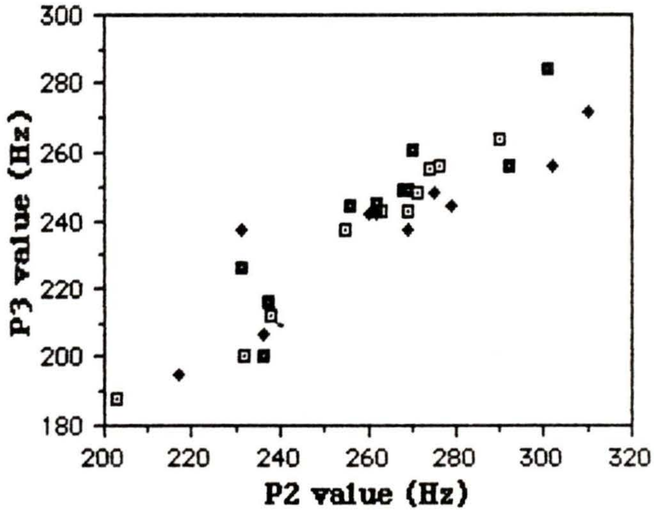


Figure 4.8
F0 value of the third peak (P^3) plotted against that of the second peak (P^2) in three different phrasal types of the NPs all of which are right-branching. The white squares are for the NPs of (A) "*no*-phrase + adjective + noun". The black diamonds are for the NPs of (B) "*no*-phrase + *no*-phrase + noun". The black squares are for the NPs of (C) "adjective + *no*-phrase + noun".

First, in the case of the right-branching NPs, as Figure 4.8 shows, all the dots are well blended, having no cluster of dots for a specific phrasal type. The scatter graph assures us that all three phrasal types in the right-branching NP group have a single catathesis contour. In other words, any right-branching NPs have just one type of catathesis contour regardless of the kinds of the modifiers they take.

Next, let us look at the scatter graph in Figure 4.9 which is for the left-branching relative clause NPs. Although the dots in the graph are somewhat wide-spread, it does not show any cluster of the dots which can be attributed to a specific phrasal type, showing that in this tree group, too, none of the phrasal type behave differently from each other as far as the catathesis contour is concerned.

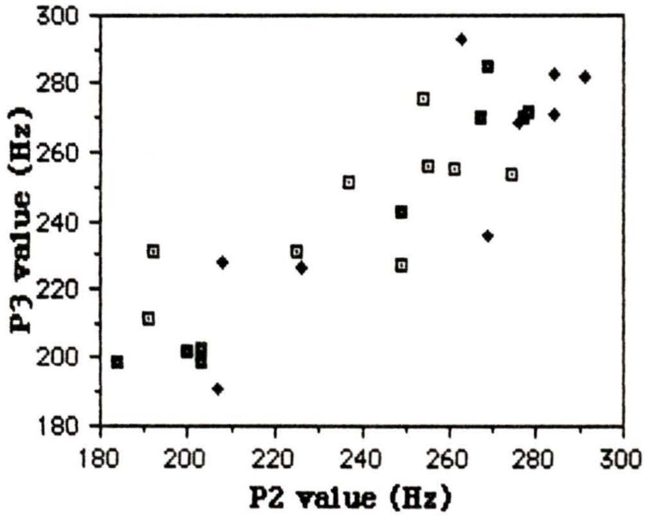


Figure 4.9
F0 value of the third peak (P^3) plotted against that of the second peak (P^2) in three different phrasal types of the left-branching relative clause NPs. The white squares are for the NPs of (D) "*no*-phrase + verb phrase + noun". The black diamonds are for the NPs of (E) "*no*-phrase + adjective + noun". The black squares are for the NPs of (F) "adjective + *no*-copula + noun".

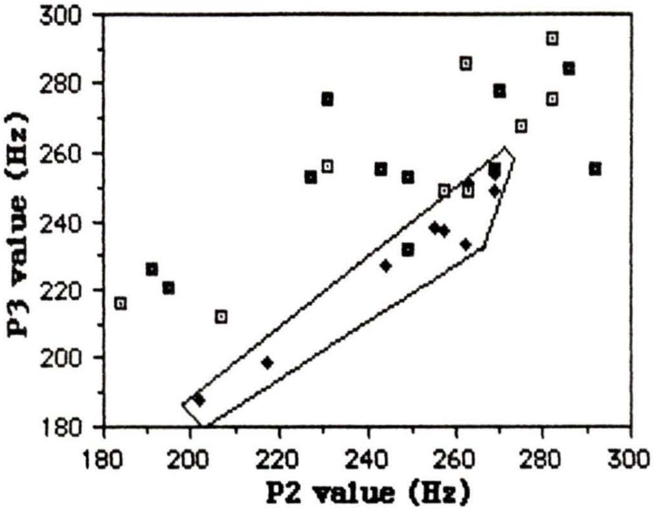


Figure 4.10
F0 value of the third peak (P^3) plotted against that of the second peak (P^2) in three different phrasal types of the left-branching symmetircal NPs. The white squares are for the NPs of (A') "adjective + *no*-phrase + noun". The black diamonds are for the NPs of (B') "*no*-phrase + *no*-phrase + noun". The black squares are for the NPs of (C') "adverb + adjective + noun".

Of the left-branching symmetrical NP group, however, the scatter graph in Figure 4.10 shows an isolated cluster of dots (identified by a circle) which belong to the NPs with double *no*-phrase modifiers. The P^3 in the NPs has much lower F0 values than those of the P^3 of the other two phrasal types. That is, in the NPs with double *no*-phrase modifiers, P^3 is catathesized. Comparing the accumulative *c*-constants of this type of NPs with those of the right-branching NPs in Table 4.3 makes it clear that, despite being left-branching phrases in which the chain of catathesis seems to be blocked between P^2 and P^3 (as far as my data are concerned, cf. Figure 4.6), the left-branching NPs with double *no*-phrase modifiers show the same chain of catathesis pattern as in the right-branching NPs. Thus, regardless of its branchingness, in the case of the NPs with double *no*-phrase modifiers, there is only one catathesis pattern which is the same as the one for the right-branching NPs.

Table 4.3
Comparison of the accumulative catathesis constants between those of the VP embedding the left-branching NPs with double *no*-modifiers and those of the VP embedding the right-branching NPs.

Types	Right-branching NPs		Left-branching NPs with double *no*-phrases	
Mora Distance between HL	Two Mora	Three Mora	Two Mora	Three Mora
P^1	0.915	0.940	0.885	0.895
P^2	0.690	0.725	0.650	0.655
P^3	0.580	0.605	0.565	0.565
P^4	0.355	0.370	0.350	0.340

4.2.6 Comparison with Kubozono's Results

Now, I would like to compare Kubozono's results with my results both of which are schematically represented in Figure 4.11.

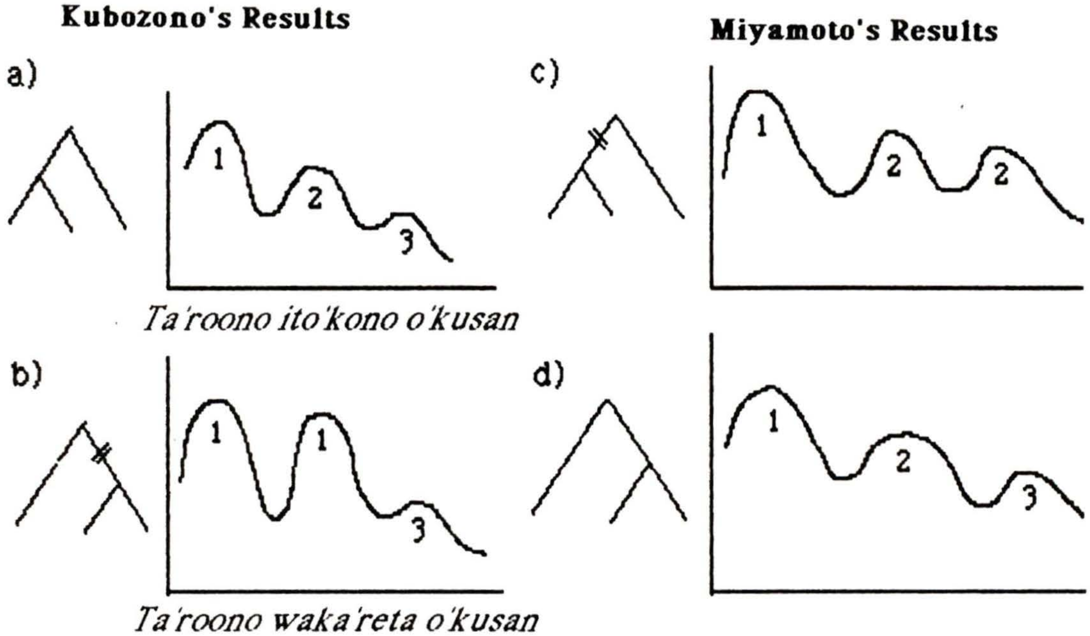


Figure 4.11
Schematic catathesis contours of the left-branching NPs and right-branching NPs obtained from Kubozono's experiment[7] and from my experiment. Numerals assigned to the peaks are schematic representations of the pitch compression ratios due to catathesis. The bouble-slash indicates the non-propagation of the catathesis chain.

As the schematic representation of the contours and syntactic trees in Figure 4.11 show, Kubozono' results and my results are not the same. In regard to the pitch contours, the first question to ask is, "Why does Kubozono's left-branching NP: (a) *Ta'roo-no ito'ko-no o'kusan* have a down-stepping 1-2-3 contour and why, in my case, is it not the left-branching NPs, (c), but the right-branching NPs, (d), which have the same 1-2-3 contour?" Why is there such a polar opposite result? To provide an answer, I must depend on the earlier account on the NPs with double *no*-phrase modifiers. As we have seen in the previous section, the NPs with double *no*-phrase modifiers have a down-stepping 1-2-3 contour regardless of their syntactic tree branching. That is, both in the left-branching NPs, such as *Ta'roo-no ito'ko-no o'kusan* "Taroo's cousin's wife" and in the right-branching NPs, such as (B3)*A'rima-no a'wabi-no o'iru* "abalone oil from Arima", there will be a chain of catathesis having the 1-2-3 down-stepping contour. Such NPs, thus, cannot solely be used to examine the correlation between syntactic branchingness and catathesis contours. It is, however, what Kubozono did. Therefore, I would dismiss Kubozono's claim that in left-branching NPs, there is a down-stepping contour, and adhere rather to my result which shows that it is in the right-branching NPs, (d), where such a down-stepping 1-2-3 contour occurs due to a chain of catathesis.

Next, it is in Kubozono's right-branching NP: (b) *Ta'roo-no waka'reta o'kusan* "Taroo's divorced wife", where a chain of catathesis seems to be blocked between P^1 and P^2, creating a 1-1-3 contour. In my case, however, such blocking is seen not in the right-branching NPs (d), but in the left-branching NPs, (c), and it is between P^2 and P^3, creating a 1-2-2- contour. Even if, as mentioned in the endnote (7), Kubozono's right-branching NP; *Ta'roo-no waka'reta o'kusan*, is treated as a left-branching NP, still it does not provide the same contour as my left-branching NPs whose contour is not 1-1-3 but 1-2-2. Admittedly, I do not have sufficient data to dispute the point. A reason, I can think of, why the phrase *Ta'roo-no waka'reta o'kusan* has a 1-1-3 contour is that Kubozono's subject constantly inserted an intermediate phrase boundary after the initial modifier, blocking an iterative application of catathesis between P1 and P2. In other words, Kubozono's phrase may have a prosodic tree representation on the left below, instead of the prosodic tree on the right below:

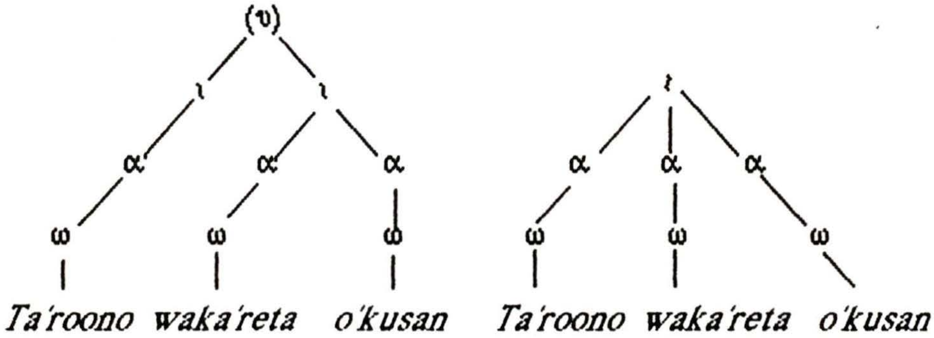


Figure 4.12
Different prosodic tree renditions for the two utterances of the phrase, *Ta'roo -no waka'reta o'kusan* "Taro's divorced wife". In the left prosodic tree, the utterance has two intermediate phrases whereas in the right prosodic tree, the utterance has only one intermediate phrase.

If this is the case, I am comparing different pieces of information. The issue here is the correlation between syntactic tree configurations and an iterative application of catathesis which takes place strictly in a single domain of an intermediate phrase. But Kubozono's contour of *Ta'roo-no waka'reta o'kusan* seems to be a result of *intermediate phrasing* involving two intermediate phrases and seem not to be a result of iterative applications of catathesis in an intermediate phrase. As briefly mentioned earlier, there is no reason that one phrase has to have just one prosodic tree. A phrase may have two or three or more prosodic trees, depending on how it is uttered or phrased. Because catathesis occurs exclusively in the domain of one intermediate phrase and an intermediate phrase boundary prevents its propagation, to discuss the issue of the relationship between a chain of catathesis and syntactic tree configurations, any data which show the insertion of an intermediate phrase boundary must be excluded from the data analyses. If my speculation that Kubozono's pitch contour of the right-branching NP is due to the insertion of intermediate L% between P^1 and P^2 is correct, his claim that "the right-branching syntactic node in right-branching NPs prevents the process of downdrift" (Kubozono 1985: 78) loses ground. Considering the size of my data in comparison to Kubozono's and the effort I took to eliminate the data which had intermediate L%, I think, my results are more convincing.

4.3 Additional Experiment

4.3.1 Blocking?

The fact that, in the previous experiment, the left-branching NPs have a 1-2-2 contour where a chain of catathesis seems to be blocked between the second accent H and the third accent H raises a question (cf. Figure 4.11). That is, can a left-branching tree configuration really be regarded as a constraint on the application of catathesis? There are two possibilities of accounting for the 1-2-2 contour in the left-branching NPs. The first argument is that an iterative application of catathesis is blocked between P^2 and P^3 due to the syntactic tree configuration. The second argument is that an iterative application of catathesis takes place after each peak (accent H), however, the effect of F0 lowering after the P^2 is obscured by other factor(s), primarily, the left-branching tree configuration.

In order to test which of these arguments is the more plausible one, let us assume the following hypothesis tentatively:

The accent at the right node of a mutually c-commanding unit in an NP does not participate in a chain of catathesis.

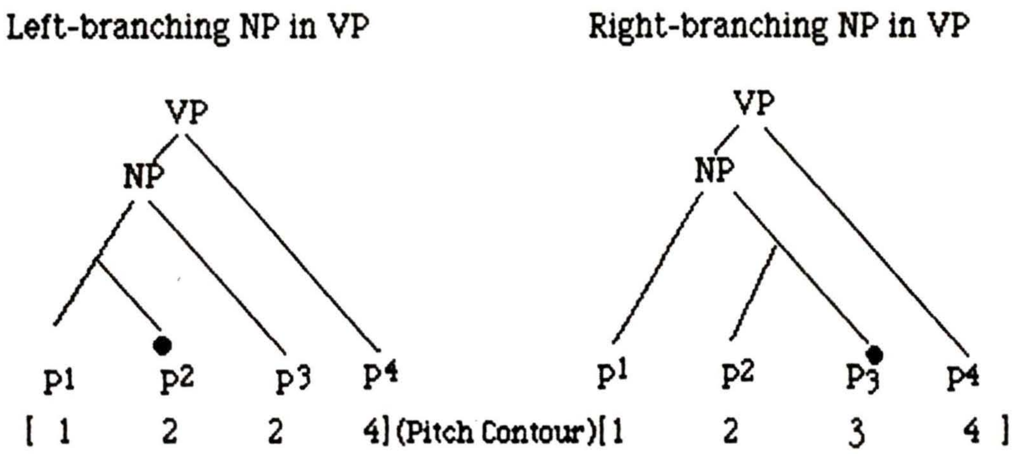


Figure 4.13
Syntactic tree illustrations of a working hypothesis that the right-node (marked by a small black circle) of a mutually c-commanding unit (i.e., the innermost phrase), in the NPs does not participate in the chain of catathesis. The VP embedding the left-branching NP has a 1-2-2-4 pitch contour whereas the VP embedding the right-branching NP has a 1-2-3-4 pitch contour.

What is the reason of such a working hypothesis? We have already looked at the fact that the left-branching NPs have a 1-2-2 pitch contour of which P^3 seems to be unaffected by catathesis. The fact motivates a speculation that the right-node (marked by a small black circle in the above tree) of a mutually c-commanding unit, i.e., the innermost phrase, in the NPs does not participate in the chain of catathesis. Next, how can the working hypothesis be posited even for the right-branching NPs whose peaks as well as the last V peak seem to be affected by the iterative application of catathesis, as seen in Figure 4.5. My reasoning can be expressed in the following manner.

	Left-branching NP in VP				Right-branching NP in VP			
	p1	p2	p3	p4	p1	p2	p3	p4
Catathesis at P2	1	2	2	2	1	2	2	2
Catathesis at P3	--------------------				1	2	3	3
Catathesis at P4	1	2	2	3	----------------------			
Final Lowering at P4	1	2	2	4	1	2	3	4
Output	1	2	2	4	1	2	3	4

Figure 4.14
Hypothetical applications of catathesis and final lowering in the VP embedding the left-branching NPs and in the right-branching NPs.

Judging from the accumulative catathesis constants listed in Figure 4.5, the VP embedding the right-branching NPs has 1-2-3-4 contour. We must realized that it is not only catathesis but also other downtrends, i.e., declination and, especially, final lowering, play significant roles in compressing the pitch contour near the end of the utterances. Now, consider in a schematic manner, as shown in Figure 4.14, that catathesis is applied only on the P^4 with the left-branching NPs but it is not applied on the P^4 with the right-branching NPs. Then, the final lowering affects P^4 values and it lowers the P^4 values by 1 in both VPs embedding the left- and right-branching NPs. If catathesis were to be applied on P^4 with the right-branching NPs, the contour would be 1-2-3-5, instead of 1-2-3-4. This line of reasoning makes it possible to posit the above working hypothesis even for the VP embedding the right-branching NPs.

4.3.2 Hypothesis Test

In order to test the working hypothesis, what must be done is simple; replace accented words at the test node[8] with unaccented words and plot the following peak[9] against the preceding peak[10]. If the replacement of the accented words, which are supposed to trigger catathesis, with the unaccented words, which are not supposed to trigger catathesis, does not affect the F0 values of the following peak, there will be a scatter graph where dots of both groups (i.e., a group of NPs whose test node has accented words and a group of NPs whose test node has unaccented words) are mingled together and somewhat on the same line, showing no effect of the word replacements. Such a scatter graph tells us that the working hypothesis is correct; that is, the accent at the test nodes is not participating in the chain of catathesis. If, however, such replacements condition the F0 value differences in the following peak, that is, if the following peak is higher when immediately preceded by an unaccented word than when immediately preceded by an accented word, there will be a scatter graph with two clusters of dots, one of which occupies a higher y-value region.

Such a graph will, then, dismiss the working hypothesis and support the alternative argument that a chain of catathesis takes place at each accent regardless of syntactic configurations even thought the lowering effect of the accent at the test node may be obscured.

The stimuli used for the experiment are listed in the Appendix for comparison with those used in the previous experiment. Measurements are taken from the utterances of the same five subjects, following exactly the same procedure as in the previous experiment. The following are the results.

4.3.3 Results

The Figure 4.15 shows a scatter graph for the VP embedding the right-branching NPs. In the graph, F0 values of P^4 are plotted[11] against F0 values of P^2.

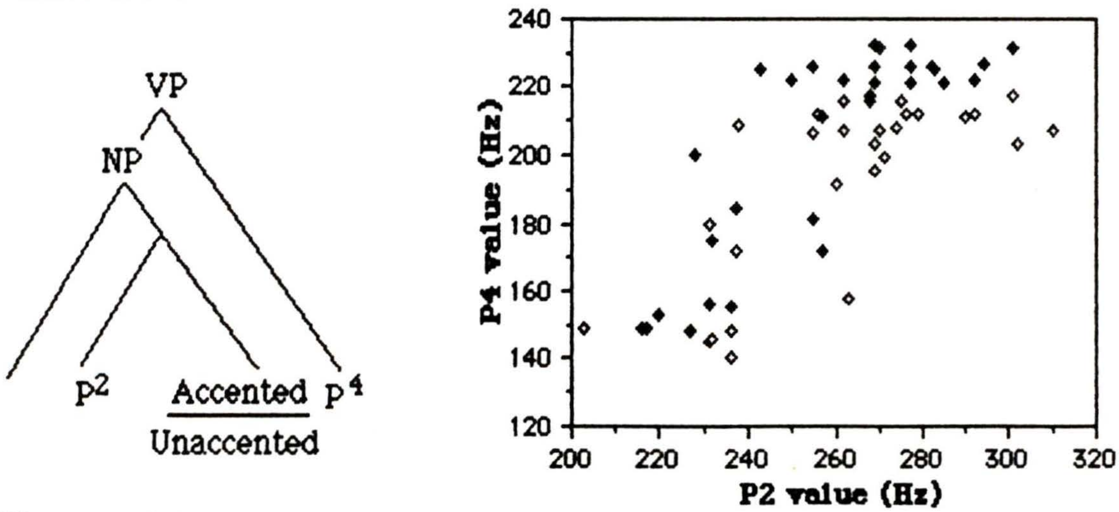


Figure 4. 15
F0 values of P^4 plotted against F0 values of P^2 in the VP embedding the right-branching NPs. White diamonds are those for the VPs whose accented P^3 is the right-node of the mutually c-commanding unit in the NPs. The black diamonds are for the VP whose unaccented P^3 is the right-node of the mutually c-commanding unit in the NPs.

An obvious point-occurring in the graph is that we are not getting clear separation of two types of legends, as we have seen in Figure 4.2 because in the current case, other factors, especially, final lowering is obscuring catathesis effect by P3. Though it must be the case, it is still possible to see that the black diamonds for the VPs with the unaccented P^3 occupy, in general, a higher y-value region than the white diamonds for the VPs whose P^3 is accented. The fact indicates that though the catathesis effect by the P^3 is obscured, the accent at the P^3 does catathesize the P^4. Thus, the working hypothesis must be rejected by saying that in the case of the right-branching NPs (immediatley dominated by VP), the accent at the right-node of a mutually c-commanding unit participates in the chain of catathesis.

Next, the scatter graph in Figure 4.16 is for the left-branching symmetrical NPs to test whether the accent at the right node of the innermost phrase participates in the chain of catathesis. Accented words at the test node, P^2, are replaced by unaccented words. The white diamonds are for the NPs whose P^2 is accented and the black diamonds are those for the NPs whose P^2 is unaccented. If the accent at P^2 does not catathesize P^3, we expect that the replacement of accented words with unaccented words at P^2 will not make any difference in F0 values of P^3. The scatter graph, however, seems to suggest otherwise. In the graph, although less clear than the plots in Figure 4.15 , the black diamonds occupy a slightly higher y-value region than the white ones, indicating that the replacement of accented P^2 with unaccented P^2 raised the F0 values of P^3. Or conversely, accented P^2 lowers P^3 even though the lowering effect is severely attenuated by other factor(s), mainly, the syntactic tree configuration of the NPs. Thus, in the case of the left-branching symmetrical NPs, as well, the working hypothesis must be rejected.

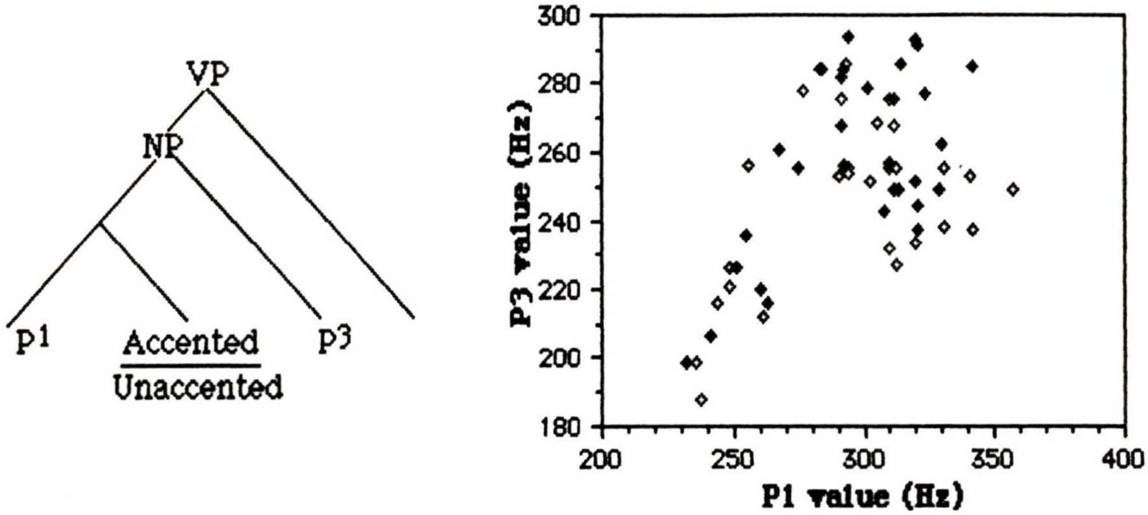


Figure 4.16
F0 values of P^3 are plotted against F0 values of P^1 for the left-branching symmetrical NPs. The white diamonds are for the left-branching NPs whose P^2 is accented and the black diamonds are for the left-branching NPs whose P^2 is unaccented.

Lastly the same procedure as above is taken to test the working hypothesis on the left-branching relative clause NPs. The result is shown in Figure 4.17.

Here, it is difficult to tell that the black diamonds for the NPs whose P^2 is unaccented occupy a higher y-value region than the white diamonds which are for the NPs with accented P^2. Only the regression lines help discern the slight difference of the P^3 versus P^1 relationships between the two groups because the regression line for the NPs whose P^2 is unaccented runs above the one for the NPs whose P^2 is accented. Although the F0 lowering effect by the accented P2 is almost nullified by the syntactic tree configuration of the NPs to the extent that catathesis appeared to be totally blocked, the plot still supports the argument that there is no blocking of catathesis between the accented P2 and P3.

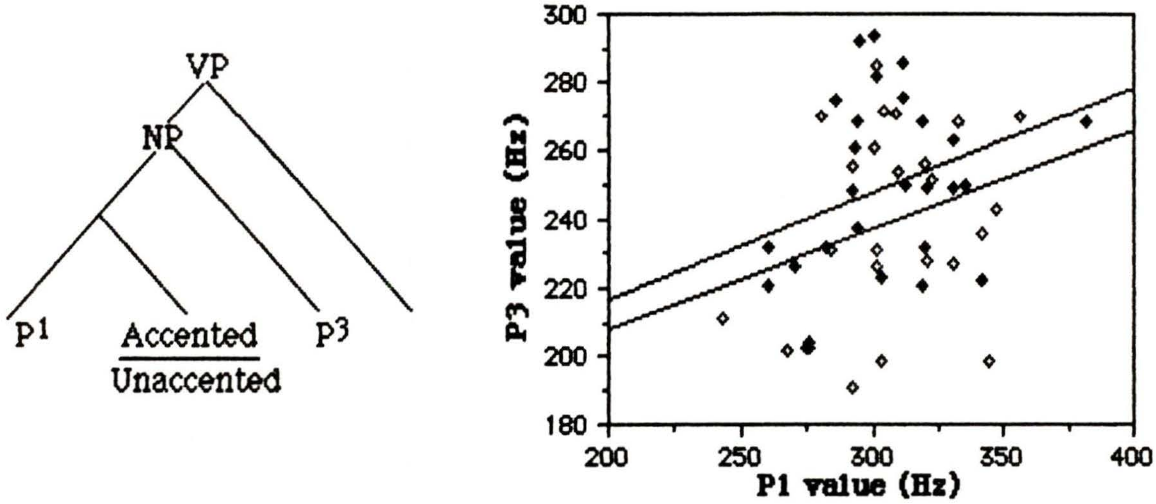


Figure 4.17
F0 values of P3 plotted against F0 values of P1 for the left-branching relative clause NPs. The white dots are for the NPs whose P2 is accented and the black dots are for the NPs whose P2 is unaccented.

Although I feel that further tests will be needed in the case of relative clause NPs to reject the working hypothesis, judging from the results from two previous scatter plots and from the slight negation obtained from the current plot, I have no choice other than to withdraw the working hypothesis. I would, then, say that in the domain of an intermediate phrase, whenever there is an accent, catathesis must take place even though there are cases where the catathesis effect is severely attenuated by other factors to the extent that the attenuation gives an impression that catathesis is blocked.

4.4 Summary

The following is a summary of the results from the first experiment:

(i) With a series of accented items, the chains of up to four catatheses were created to demonstrate that iterative applications of catathesis take place as long as there is no insertion of an intermediate boundary.

(ii) Two sets of stimuli are set up to test whether or not distance differences between accents contribute to pitch compression differences. It was proved that the distance differences between accents do not contribute to the F0 lowering effect. That is, the pitch compression ratio due to catathesis is not affected by the timing differences in triggering catathesis.

(iii) In regard to testing the correlation between left- and right-branching syntactic trees and catathesis, except for the case of the NPs with double *no*-phrase modifiers, there is a clear difference in catathesis contours between the right-branching NPs and the left-branching NPs of "modifier + modifier + noun". That is, the right-branching NPs have a downstepping 1-2-3 contour whereas the left-branching NPs have a 1-2-2 contour where a chain of catathesis seems to be blocked between the second accent H and the third accent H.

(iv) If, however, both modifiers are *no*-phrases in the NPs of "modifier + modifier + noun", the NPs will have a down-stepping 1-2-3 contour regardless of being left-branching or right-branching in their syntactic configurations.

The second experiment, then, proved that there is no such phenomenon called blocking of catathesis. As long as there is no intervention of an intermediate phrase boundary, catathesis effect propagates. That is, whenever there is an accent, catathesis must take place even though there are cases where the catathesis effect is severely attenuated by other factors to the extent that the attenuation gives an impression that catathesis is blocked.

Although the issue in this chapter had a strong flavour of a basic study and seemed to have little involvement in the development of the algorithmic phrasing and pitch contour model for speech synthesis, the findings made in the chapter are not only theoretical but practical in that they can easily be implemented on such a model to better approximate to natural pitch contours.

Endnotes

[1] The issue in this chapter has a strong flavour of a basic study and has no direct involvement in the development of the algorithmic phrasing and pitch contour model. However, findings made in this chapter can easily be implemented onto such a model to approximate to natural pitch contours.

[2] The symbol, " ! " indicates lowering effect of tone due to a tonal assimilation.

[3] Pierrehumbert and Beckman (1989) report allophony of intermediate phrase boundary. That is, if the boundary is preceded by an accent, HL, its F0 value is far lower than when it is preceded by a phrasal H.

[4] The F0 value of a boundary L is low when it is preceded by an accent and it is high when it is not peceded by an acent.

[5] There is no well-formed left-branching tree which is symmetrical to the tree of the phrase, (C). So, to retain some symmetricality, the innermost AP's configuration is changed to the AP dominating ADP and A-bar.

[6] In fact, phrases were prepared in a manner where there was only one mora separating each accent. However, the data from these phrases were not used for

analyses because all the subjects had difficulties in uttering the phrases with such an accentual configuration.

7 There are two technical problems with Kubozono's stimuli. The first problem is that voiceless consonants are incorporated into the strings of segments. Because voiceless consonants raise F0 values of a following vowel (cf. Lehiste 1970), they make it impossible to take accurate measurements. Another problem is Kubozono's right-branching NP, *Ta'roo-no waka'reta oku'san*, which can be treated either as a right-branching NP or a left-branching NP: i.e.,

8 That is, P2 in the left-branching NPs and P3 in the right-branching NPs.

9 That is, P3 in the left-branching NPs and P4 in the VP embedding the right-branching NPs.

10 That is, P1 in the left-branching NPs and P2 in the right-branching NPs.

11 Now, it becomes apparent why in the previous experiment, the subjects were requested to create four series of catathesis, including the one on P4 and why measurements were taken not just for NP peaks but for all the four peaks. P4 values were needed for conducting the current experiment.

Chapter 5

Intermediate Phrasing

The aim in this chapter is to design a simple algorithmic intermediate phrasing model which can predict the locations of intermediate phrase boundaries in sentences. In an application, the intermediate phrasing model is attached to the accentual phrasing and pitch contour model developed in Chapter 3 so that once the locations of intermediate phrase boundaries in a given sentence are predicted, the accentual phrasing algorithm can parse a given intermediate phrase into accentual phrases and assign tones to the phrases. These tones will, then, be quantized by means of tone scaling to create a schematic pitch contour for speech synthesis.

A problem in developing such an intermediate phrasing model is that at the level of the intermediate phrase, there is neither a one-to-one relationship between accentuation and phrasing, such as the one observed at the level of accentual phrasing, nor is there a one-to-one mapping relationship between the syntactic structure and the phonetic process, such as the one observed in the iterative application of catathesis. At the level of the intermediate phrase, accentuation has no role in phrasing. Also, it is not possible to rely on the syntactic structure to define the intermediate phrasing because the relationship between syntax and intermediate phrasing becomes a one-to-many mapping due to the obvious fact that one sentence can have more than one intonational realization. The higher a prosodic unit is in the prosodic hierarchical tree, the more general the definition of the unit and the nature of phrasing become.

Unfortunately, the study of the nature of intermediate phrasing in Japanese is very fragmental (cf. McCawley (1968, 1977), Poser (1984)) and the guides I have relied on so far; i.e., Pierrehumbert and Beckman (1989), are not helpful. A new guide is needed. Candidates for a guide in exploring the issue of intermediate phrasing are Selkirk (1984), Suzuki and Teranishi (1988), and Nespor and Vogel (1986). Following, then, Nespor and Vogel's approach, which I think to be the most theoretically sound[1] and, above all, the most practical in application, I will investigate factors which condition intermediate phrase boundaries in Japanese and develop an algorithmic intermediate phrasing model.

The basic assumption of the intermediate phrasing model, which is mainly based on Nespor and Vogel (1986), is that the intermediate phrasing is *loosely* constrained by the syntactic structure of a sentence and that it is not only syntactic factors but also semantic factors, such as theme, argument structures, contrastive prominence and scope of modifiers which play roles in the phrasing. Furthermore, there are extra-linguistic factors, such as length of sentence or style of speech, which affect the syntactic and semantic factors in developing phrasing variations. In this chapter, all these factors will be examined and the interaction of these factors will be moulded into an intermediate phrasing model.

The actual algorithmic intermediate phrasing model consists of a dictionary and algorithmic rules. The dictionary stores lexical information;

i.e., (i) syntactic categories (parts-of-speech classifications) and (ii) syntactic, semantic and phonological features. The algorithmic rules identify syntactic and semantic factors which condition intermediate phrase boundaries by way of "pattern-matching". Whenever there is a pattern which matches with a pattern in an algorithmic rule in a string of lexical items which are tagged with lexical information, the matched pattern is replaced with another pattern in the rule which specifies the insertion of intermediate phrase boundaries.[2]

Although the model is capable of parsing a sentence into intermediate phrases by way of predicting the location of an intermediate phrase boundary through "pattern-matching", there is a limit in the parsing capability due to its simplicity. Because of its linear parsing which regards a sentence as a mere string of lexical items tagged with syntactic categories and additional features, the model is not suited for parsing embedded complex sentences. In other words, the model is not able to predict the locations of intermediate phrase boundaries which are conditioned by such syntactic factors as deeply embedded relative clauses or subordinate clauses because the linear parsing is done with no recourse to the hierarchical syntactic structures of sentences.

In spite of this shortcoming, the linear parsing is designed because it is simple and suited to the purpose, i.e., the development of a phrasing and pitch contour model for a microcomputer-oriented speech synthesis program. Because the main application of the speech synthesis program is the development of a software program which assists beginning students to learn Japanese, the simple phrasing algorithm is adequate enough to accomplish its task.[3]

5.1 Previous Works

5.1.1 Selkirk's Intonational Phrasing (Semantic-based Approach)

Selkirk (1984) says that in any treatment of sentence intonation, it is necessary to account for the fact that a given sentence may not have just one but (several) different intonational realizations. Selkirk emphasizes the fact that there is no one-to-one relationship between the syntactic structure and the intonational structure of a sentence in such a sense that the syntactic structure *determines* its intonational structure. That is, the relationship between the syntactic structure and the intonational structure should be regarded as a one-to-many mapping. Thus, according to Selkirk, any syntax-based approaches, such as Cooper and Sorensen (1981), where a sentential pitch contour is computed based on its syntactic structure with no regard to its phonological representation, is misguiding. Opposed to syntax-based approaches, Selkirk postulates a semantic-based approach which has the following mechanisms for parsing intonational phrases whose end is marked by a grammatical pause.

First, a given sentence is exhaustively parsed into intonational phrases by a "Syntactic-Prosodic Correspondence Rule for Intonational Phrase":

> "A matrix sentence must be exhaustively parsed into a sequence of (one or more) intonational phrases." (ibid: 286).

A semantic-based constraint named "Sense Unit Condition" is, then, imposed on the intonational phrasings and filters out ungrammatical phrasings:

" *The Sense Unit Condition on Intonational Phrasing*
The immediate constituents of an intonational phrase must together form a sense unit."(ibid: 286).

The condition is substantiated by the following definition:

"Two constituents C_i, C_j form a sense unit if (a) or (b) is true of the semantic interpretation of the sentence:
a. C_i modifies C_j (a head)
b. C_i is an argument of C_j (a head)." (ibid: 291).

In other words, to be a well-formed intonational phrase, any elements in an intonational phrase must have either a modifier-head or argument-head relationship among the elements. If there is no such relationship among the elements, then, they cannot constitute an intonational phrase. To make the point, Selkirk provides the sentence: "Jane gave a book to Mary" as an example. Among the eight possible phrasings[4] for the sentence, the following two intonational phrasings are ungrammatical due to the fact that the sequence of the units, *the book to Mary*, has neither a modifier-head relationship nor an argument-head relationship, thus violating the Sense Unit Condition:

(1) *(Jane)$_{IP}$ (gave)$_{IP}$ (the book to Mary)$_{IP}$

(2) *(Jane gave)$_{IP}$ (the book to Mary)$_{IP}$

My objection to Selkirk's semantic-based approach is that the semantic-based constraint is too strong in that it may rule out, at least in the case of Japanese, even well-formed phrasings as ill-formed. For example, it is possible to have intonational phrases in Japanese which violate Selkirk's Sense Unit Condition and still to be well-formed, such a sentence as : *kyo'o* *watasi- wa eiga - o mini iki ma'su* "I go to see a movie today" which can be intonationally realized as (*kyo'o watasi- wa*)IP (*eiga - o mini iki ma'su*)IP. According to the Sense Unit Condition, *kyo'o watasi- wa* cannot be an intonational phrase because there is no modifier-head or argument-head relationship between the two units, i.e., *kyo'o* (today) and *watasi- wa* (thematized "I"), despite the fact that such a phrasing is very common (i.e., unmarked) and well-formed. Even such a simple case indicates that the semantic-based constraint is too strong and may not work properly in Japanese.

The second problem with Selkirk's approach, which has not to do with the theory itself, is its applicability to an intonational phrasing algorithm. In Selkirk's intonational phrasing, it is, first, necessary to derive all the possible intonational phrasings for a given sentence; then, ill-formed intonational phrasings will be filtered out by virtue of the Sense Unit Condition. A problem with such an approach, as far as an application is concerned, is that it will develop all the possible intonational phrasings for a given sentence with no regard to markedness and unmarkedness because all it does is check whether or not certain types of intonational phrasings are well-formed. Not being able to generate phrasings step by step from most marked phrasing to the least marked phrasing, or vice versa, may cause trouble in the application because in many cases it is not an objective of an application model to generate all the possible phrasings for a given sentence with no regard to markedness.

5.1.2 Suzuki and Teranishi's Syntax-based Approach

Suzuki and Teranishi's (1988) syntax-based phrasing algorithm is a part of the apparatus of their text-to-speech synthesis model in Japanese. The basic idea of the phrasing algorithm originates in Hakoda and Sato (1980) which derives a pitch contour of a sentence by using the Hashimoto-*bunsetsu* as a primitive. In Suzuki and Teranishi's approach, intonational phrasings (as well as phrasings at lower prosodic levels) are derived by virtue of the syntactic structure of a sentence with no recourse to its phonological structure. Taking Hashimoto-*bunsetsu*, syntactic units which consist of a head and postposition(s) or affix(s), as primitives, Suzuki and Teranishi combine the *bunsetsu* into a larger unit, such as a phrase and a clause, until the process of the combination reaches to the level of a sentence. The cyclic process of the combination of *bunsetsu* assumes that there are two types of *bunsetsu*: one which is independent in the sense that it is not involved in a modifier-head relationship with another *bunsetsu*, such as expletives or numerals[5] and the other which is dependent in the sense that it functions either as a modifier or a head, such as adjectives, adverbs, or nouns. With the combination of two dependent units at a time, *bunsetsu* are combined into a phrase, as in *mura-no* (village; possessive) + *ie'-ga* (houses; nominative); depending on a modifier-head relationship[6], the phrase and another *bunsetsu* are, in turn, combined into a larger phrase or a clause, as in *mura-no-ie'-ga* + *miema'su* (be seen); the larger phrase or the clause and another unit are, then, in the same fashion, combined into a sentence, as in *atira kotira-ni* (here and there) + *mura-no ie'-ga miema'su.*

The basic assumption of the cyclic combination of *bunsetsu* is that among modifier-head relationships, there are different degrees of strength which are reflected on the different levels of syntactic boundaries. Because the different degrees of strength in syntactic boundaries are reflected upon the degrees of prosodic boundary strengths, it is possible, then, to predict the locations of different levels of prosodic boundaries by measuring the strengths of syntactic boundaries. To do so, Suzuki and Teranishi heuristically scale the combination rules in strength and compute the strength of a syntactic boundary by way of computing the number of the combination processes involved (the higher the number is, the stronger the boundary is) and the strength of the rules involved (the stronger the modifier-head relationship is, the lower the strength degree of the rule is). An intonational phrase boundary is, then, assigned to a location where a strong syntactic boundary is indicated.

Furthermore, it is assumed that to provide the naturalness of prosody, it is necessary to impose an extra-syntactic modification due to length on the the syntactically motivated intonational phrasing. That is, no intonational boundary is inserted if a series of *bunsetsu* does not exceed more than five morae even if syntactic information specifies an insertion of an intonational phrase boundary. On the other hand, even if syntactic information does not specify an insertion of an intonational boundary in the string of over fifteen morae, a less strong syntactic boundary is upgraded as a marker for an intonational phrase boundary.

There are some problems in this approach. The first problem is the basic assumption that there is a one-to-one relationship between the syntactic structure and the prosodic structure of a sentence so that it is possible to derive an intonational phrasing from the syntactic configuration of a sentence. In the current phonological theories, such an assumption is strongly rejected (cf. Selkirk 1984; Nespor and Vogel 1986; etc.) because such an approach cannot account for the variability in the prosodic realization of a sentence.

The second problem is that, as easily guessed, the whole operation of phrasing is extremely complicated. There are fifteen groups of rules which combines units within various types of *bunsetsu* and there are another fifteen groups of rules which combine *bunsetsu* with another *bunsetsu*. The whole operation may require at least a few hundred rules to derive a phrasing pattern of a sentence. Suzuki and Teranishi's phrasing algorithm is, thus, not economical or at least it is not suited to the aim of this study, i.e., the development of a small-sized intonational phrasing model.

5.1.3 Nespor and Vogel's Intonational Phrasing

Like Selkirk, Nespor and Vogel (1986) assume that there is no one-to-one mapping relationship (isomorphism) between the syntactic structure and the intonational structure of a sentence. However, different from Selkirk who does not rely on any syntactic information in deriving intonational phrasing, Nespor and Vogel depend on syntactic information, saying that intonational phrasing is *loosely* constrained by syntax. Furthermore, in deriving well-formed intonational phrasings, Nespor and Vogel take a polar opposite direction to that of Selkirk who first creates all the possible phrasings for a given sentence, then, filters out ungrammatical phrasings. In contrast, Nespor and Vogel, first, set a rigid intonational phrasing rule which defines just one type of an intonational phrasing with no variation. Then, in virtue of restructuring rules, all the well-formed phrasing variations are created for a given sentence.

As far as mechanisms for intonational phrasing are concerned, first of all, the Basic Intonational Phrase Formation rule defines a constituent which is external to a root sentence as an intonational phrase, and defines also the remaining(s) in the sentence as intonational phrases, as in "(Kennedy)$_{IP}$ (as you know)$_{IP}$ (is a great musician)$_{IP}$". There are six types of constituents which are external to a root sentence, i.e., parenthetical expressions, non-restrictive relative clauses, tag questions, vocatives, expletives, and certain moved elements, all of which are said to be linearly represented at the s-structure level but not attached to a hierarchical sentence tree.

Because the Basic Intonational Phrase Formation rule does not allow any variability, it is necessary to have a process of restructuring to account for the intonational variations which a sentence may have. There are, Nespor and Vogel say, three performance and one semantic factors which affect the degrees of probability of restructuring, i.e., the length, the rate of speech, the style of speech, and the contrastive prominence. These four factors, or variables, warrant that the intonational phrasing created by the Basic Intonational Phrase Formation rule undergo restructuring. That is, the longer a given basic phrasing is, the more frequently are intonational phrase boundaries inserted. The faster the speech is, the less often is a given basic phrasing parsed into intonational phrases. The more formal the style of speech is, the more likely is a given basic phrasing parsed into intonational phrases. Contrastive prominence, too, which is a semantic factor, increases the probability of restructuring.

There are, then, three general constraints which specify where, in a sentence, intonational boundaries ought to be and ought not to be, meaning that the location of an intonational phrase boundary is not totally free. That is, if an intonational phrase phrase boundary occurs, it ought to be either at the end of an NP, or at the beginning of a S-bar and it ought not be between an obligatory argument and a V.

Taking as an example the sentence: "The giant panda eats only one type of bamboo in its natural habitat" (Nespor and Vogel 1986; 197), Nespor and Vogel say that an intonational phrase boundary coincides only with the end of an NP . That is, an intonational phrase boundary which is marked by a pause does not appear in such a place as: *--- giant)$_{IP}$; *--- eats)$_{IP}$; or *--- type)$_{IP}$ but appears only at the end of an NP, i.e., --- panda)$_{IP}$ or --- bamboo)$_{IP}$.

In regard to the beginning of an S-bar constraint, Nespor and Vogel say that an intonational phrase boundary takes place at the beginning of an S-bar as long as the insertion of the boundary does not interrupt an NP. Thus, for example, the following example (3) is well-formed because a boundary is inserted at the beginning of the S-bar: "that I was going to Japan". However, even though the phrasing conforms to the beginning of a S-bar constraint, the example (4) is ill-formed because it inserts a boundary in the middle of the NP: "the people who were coming to the party", violating thus the end of an NP constraint which overrides "the beginning of a S-bar" constraint.

(3) (I thought you already knew)$_{IP}$ (that I was going to Japan)$_{IP}$

(4) *(I though you knew the people)$_{IP}$ (who were coming to the party) $_{IP}$

The third, argument structure constraint is the same as one of the provisions of Selkirk's Sense Unit Condition which defines that the sense units must have an argument-head relationship among the elements in the units. The argument structure constraint is designed not to insert an intonational phrase boundary in such a way that it will separate an obligatory argument from a verb. Thus, in the following examples, (5) has a well-formed phrasing whereas (6) is ill-formed because it separates the second obligatory argument from its verb: (Nespor and Vogel 1986; 198)

(5) [$_{I}$That kind old lady always buys fresh meat]$_{I}$ [$_{I}$for the stray cats that live in the park]$_{I}$

(6) ?*[$_{I}$That kind old lady always gives fresh meat]$_{I}$ [to the stray cats that live in the park]$_{I}$

Furthermore, there are two special cases which override the above three general constraints, i.e., lists and embedded structures. The following is an example of lists which, even though it violates the end of an NP constraint, is a well-formed phrasing:

(7) (That big)$_{IP}$ (fat)$_{IP}$ (ugly)$_{IP}$ (black beast)$_{IP}$ (frightened the villagers)$_{IP}$

Embedded structures is another case which overides the general constraints. The following phrasing of the classic example of a sentence with embedded relative clauses is well-formed despite the fact that it violates the

the end of an NP constraint by splitting the NP, "the cat that ate the rat that ate the cheese", into three intonational phrases:

(8) (This is the cat)$_{IP}$ (that ate the rat)$_{IP}$ (that ate the cheese)$_{IP}$

Thus, embedded structures as well as lists are regarded as two special cases which cannot be accounted for by the general constraints.

5.1.4 Choice of Nespor and Vogel's Approach

Having examined the three different approaches, I would like to support Nespor and Vogel's approach. The following are the reasons why I do so. First, unlike Suzuki and Teranishi, Nespor and Vogel do not assume that the syntactic structure of a sentence *determines* its intonational structure and they deny the existence of the isomorphism between the two structures, conforming thus to the consensus of the current prosodic theories. Second, unlike Selkirk's semantic-based approach whose constraint is too strong, Nespor and Vogel's approach is less constrained and may, at least where Japanese is concerned, accommodate well-formed intonational phrasings which may be ruled out by the semantic constraint. Third and most importantly among the three approaches, Nespor and Vogel's approach is the most practical for application in the sense that it neither generates unnecessary phrasings, as Selkirk's approach does, nor does it undergenerate phrasings, as does Suzuki and Teranishi's approach which cannot accommodate variability in phrasing.

Seeing these advantages in Nespor and Vogel's approach over the other two, I will rely on Nespor and Vogel to develop an algorithmic intermediate phrasing model in Japanese which can parse a sentence into intermediate phrases by way of predicting the locations of intermediate phrase boundaries. Before discussing the actual development of such an algorithmic model (later in the chapter, 5.3), however, it is necessary to review Nespor and Vogel's arguments in the context of Japanese in order to set a basic framework for the model.

5.2 Application of Nespor and Vogel's Approach to Japanese

5.2.1 Basic Framework

Table 5.1 is a summary of an attempt to apply Nespor and Vogel's arguments on the intonational phrasing to Japanese. The aim of this attempt is to design a model which can create phrasing variations and can predict the locations of intermediate phrase boundaries by way of identifying in a sentence the syntactic and semantic factors which condition boundaries.

Table 5.1 shows three main components of the intermediate phrasing model, i.e., (i) the syntactic and semantic factors which condition an intermediate phrase boundary (L%)[7] on the left-hand side column; (ii) performance variables which affect some of the factors in conditioning L% on the centre column; and (iii) the general markers used to identify the factors in the text of a sentence on the right-hand side column. The symbol "Sy" specifies that a factor is syntactic whereas the symbol "Sm" specifies that a factor is semantic. The * indicates a marker which has to be manually specified in the text of sentences.

Table 5.1
Three major components in an intermediate phrasing model.

Syntactic & Semantic Factors Which Condition a L%	Performance Variables	Markers
[Obligatory]		
1. Extras (Sy)		*Commas
2. Subordinate Clauses (Sy)		Subordinators *Quotation Marks
3. Lists (Sy)		Null Postposition
4. Embedded Structures (Sy)		*Commas
5. Coordination (Sy)		Postpositions; Inflection
6. Contrastive Prominence (Sm)		*Capital Letters
[Variable Dependent]		
	Length	
7. Theme (Sm)	[10 < morae]	*wa*-postposition
8. Argument Structures (Sm)	[15 ≤ morae]	Position(s); Postpositions
	Style of Speech	
9. Wide Scope (Sm)	[formal/informal]	*Marker: <>
10. Every NP&ADP (Sy)	[formal/informal]	Postpositions

In the table, syntactic and semantic factors which constrain where an intermediate phrase boundary ought to be are classified into two categories, depending on the probability of conditioning an intermediate phrase boundary. That is, unlike Nespor and Vogel who rank the three general constraints in terms of strength, I set two categories for all the syntactic and semantic factors depending on: (i) whether the factors attract a L% without fail (Obligatory); or (ii) they attract a L% if a certain condition is met (Variable Dependent).

The factors in the Obligatory category condition L% without a fail. These factors are similar to external constituents in Nespor and Vogel's Basic Intonational Phrase Formation rule in the sense that they develop, without creating any phrasing variations, just one basic intermediate phrasing for a given sentence because they are not affected by performance variables. Thus, no performance variables are assigned to the factors in the Obligatory category.

The factors in the Variable Dependent category participate, to use Nespor and Vogel's terminology, in the restructuring process. These factors will contribute to create variations on the basic phrasing for a given sentence. In order to condition an intermediate phrase boundary, the factors in the Variable Dependent category must be licensed by the performance variables, i.e., the length and the style of speech. If the parameters of the performance variables are satisfied, or if a particular parameter is chosen, these semantic and syntactic factors are triggered to condition an intermediate phrase boundary and to develop variations in intermediate phrasing. In the model, the category Variable Dependent is created to accommodate variations in intermediate phrasing so that the model will not fall into the pitfall of assuming a one-to-one mapping relationship between the syntactic structure and the intonational phrasing, as seen in Suzuki and Teranishi's model. By activating the performance variables or even by changing parameter values of the variables, it becomes possible to provide a sentence with varieties in intermediate phrasing.

Table 5.1 also shows all the general[8] markers which are used to identify the syntactic and semantic factors which condition intermediate phrase boundaries. There are two types of markers: automatic and non-automatic. In the table, the markers without * are automatic in that the corresponding syntactic and semantic factors can be automatically identified through "pattern-matching" (which will be discussed later in 5.3). In contrast, the markers with * are non-automatic in that the corresponding syntactic and semantic factors must be manually specified in the text of a sentence by using such notations as commas.

Thus, there is a noticeable difference between the accentual phrasing which we have looked at in Chapter 3 and the intermediate phrasing. That is, the accentual phrasing can be derived automatically, relying primarily on the information about the accentuation of the words which constitute the intermediate phrase. In contrast, the intermediate phrasing is not quite as simple and automatic as the accentual phrasing due to the number and the nature of the syntactic and semantic factors and performance variables involved.

Before looking into the semantic and syntactic factors which condition an intermediate phrase boundary and how the markers are used to identify the factors in a sentence, let us examine performance variables.

5.2.2 Performance Variables

As mentioned before, Nespor and Vogel list four factors, which I call variables, which affect the probability of restructuring, i.e., the length, the rate of speech, the style of speech, and contrastive prominence. In my model, contrastive prominence and the rate of speech are eliminated from the list.

First, the contrastive prominence is not treated as a variable because it is different in nature from the other variables. That is, unlike the other performance variables which merely license the syntactic and semantic factors to condition a L%, the contrastive prominence can actually condition the insertion of a L% without appealing to the syntactic and semantic factors. As you may remember, in one of the experiments in Chapter 3, we have looked at the fact that the contrastive prominence, or narrow-focus, conditions the insertion of an intermediate phrase boundary immediately before a narrow-focused item. This fact indicates that the contrastive prominence is not a variable which affects the syntactic and semantic factors in conditioning a L% but is, in fact, one of the factors which obligatorily trigger the insertion of a L%. For this reason, the contrastive prominence is treated as a factor in the Obligatory category rather than a variable.

Second, the rate of speech is not included as one of the variables because the variable of length can account for the role played by the rate of speech. That is, if it is necessary to create a phrasing which is suited for fast speech, the parameter value of the variable of length can be increased so that intermediate phrase boundaries will not appear often. Or, if a phrasing suited for slow speech is required, then, the parameter value of the variable of length is lowered so that a given sentence will have a fair number of phrase boundaries. Such a manipulation is supported by Nespor and Vogel's own words that "the faster the rate of speech, the longer the *I*s [intonational phrases] of a given utterances tend to be; conversely, slower the rate of speech, the shorter the *I*s tend to be" (Nespor and Vogel 1986; 195). To keep the mechanisms of the model less complicated, the rate of speech whose function can be represented by the length is, thus, not included as a performance variable.

5.2.2.1 Length as a Variable

In regard to the effect of the length on intermediate phrasing, there is acoustic evidence which indicates that in Japanese a long utterance is easily split into intermediate phrases. Gallaiford (1983) reports that in the comparison of three-word, four-word, and five-word sentences, utterances of longer sentences show more regular insertion of a phrase boundary than those of shorter sentences. Gallaiford says that in the utterances of the three-word sentences, an intermediate phrase boundary or, in his terminology, a "resetting of F0 declination" is hardly seen; in the utterances of the four-word sentences, half of the cases show a boundary (where a declination line is reset); and in the utterances of the five-word sentences, almost all of the cases have a boundary. Hakoda and Sato (1980), too, report that longer the sequence of morae is in a sentence, the more frequently inserted an intermediate phrases boundary is. Suzuki and Teranishi (1988) adopt the finding in their text-to-speech synthesis model which, as mentioned earlier, inserts an intonational phrase boundary in a string of over fifteen morae if the string does not have a boundary. Thus, it is clear that the length plays a significant role in intermediate phrasing in Japanese.[9]

The parameter values set for theme, i.e., [10< morae], and for argument structures, i.e., [15 ≤ morae] in Table 5.1 should be interpreted that if there is a thematized phrase in a sequence of more than 10 morae with no L%, insert a L% immediately after the thematized phrase. Also, if there is an argument, in a sequence of over 15 morae with no L%, insert a L% before the argument.

There is acoustic evidence which indicates that theme conditions a L% in a shorter sequence of morae more than argument structures do. Warkentyne (1984) demonstrates that in minimally contrasting pairs of three-word sentences of "Subject (-*ga*) / Theme (- *wa*) + Argument + Verb", thematized *wa*-phrases develop regularly a L% after the phrases, as in *Zo'o-wa* L% *hana-o itamema'sita* "Speaking of the elephant, he injured his trunk", but not before the arguments in the contrasting sentences with the subject *ga*-phrases, as in *Zo'o-ga hana-o itamema'sita* "The elephant injured his trunk". The above facts indicate that theme attracts a L% in a shorter sequence of morae more than argument structures do. If it were the case that argument structures attract a L% in a shorter or, at least, in the same length of mora strings as theme, we would see a L% before the arguments, as in ?*Zo'o-ga* L% *hana-o itamema'sita*.[10]

In regard to the setting of the actual parameter values for the length-variable, Warkentyne provides a clue. In Warkentyne (ibid), there is only one sentence with a thematized *wa*-phrase which does not show a L%. The sentence has ten morae. All the other sentences with thematized *wa*-

phrases show a L%. The mean number of the morae in these sentences is just over thirteen. Among them, the shortest sentence consists of twelve morae. With consideration of these facts, a parameter is set at [10 < morae][11] for theme.

Next, in Warkentyne (ibid), none of the sentences with *ga*-phrases show a L% before arguments. Among these sentences, the longest sentence has fourteen morae. Also, as mentioned earlier, Suzuki and Teranishi (1988) inserts an intonational phrase boundary in a string of over fifteen morae if the string does not have a boundary. Considering these facts, I will set a parameter for argument structures at [15 ≤ morae] so that if there is an argument in a string of over fifteen morae with no L%, a L% must be inserted before the argument.

Here, I would like to emphasize the fact that these parameter values are not absolute in that they can freely be changed depending on what kind phrasing variations one would like to create. The only restriction required in setting parameter values is that the value for theme must be smaller than that for argument structures.

5.2.2.2 Style of Speech as a Variable

In regard to the style of speech as a variable, I have no piece of acoustic evidence in Japanese which supports Nespor and Vogel's claim that in formal speech, an utterance is phrased into more intonational phrases than in informal speech. However, I take their claim to be universal and apply it to Japanese.

Unlike length whose parameter is gradient, style of speech has a parameter which is binary, i.e., "formal / informal". If "formal" is chosen as the style of speech, the corresponding factors (i.e., "wide scope" and "every NP & ADP") will be activated and L%'s will be inserted at all the possible locations in a sentence. The result will then be a phrasing which may be observed in well articulated, enunciative, formal speech which may typically occur in such a situation where a Japanese gives a language lesson to non-native speakers. If, however, "informal" is chosen, the corresponding semantic (i.e., wide scope) and syntactic (i.e., every NP&ADP) factors will not be activated so that there will be no further insertion of L% to the phrasing which has been defined by the factors in the Obligatory category and by the two semantic factors which are sensitive to the length in the Variable Dependent category (i.e., theme and argument structures). The result will then be a phrasing which has a relatively few intermediate phrase boundaries and which may be similar to the one observed in conversational speech.

The reason why wide scope and every NP & ADP (which will be explained later in 5.2.4.3 and 5.2.4.4) are regarded to be sensitive not to the length but to the style of speech can be explained by a simple fact that intermediate phrase boundaries conditioned by these factors can occur in very short spans of mora strings. The reason is that their main function is to slice a unit into intermediate phrases for the sake of clarity. Thus, these two factors are not sensitive to the length but sensitive primarily to the the style of speech: e.g.,

(9) Wide Scope: *aoi'* (blue) L% *u'mi-no* (sea's) *ie'* (house) *da* (to be)
"(That) is the blue house by the sea."

(10) Every Np & ADP: *Kyo'o* ADP] L% *watasi wa* NP] L% *a'sa kara* ADP] L% *gakkoo e* NP][12] *ikima'su*
"Today I go to school from the morning."

5.2.3 The Syntactic and Semantic Factors in Obligatory Category

In this section, applying Nespor and Vogel's argument on the intonational phrasing to Japanese, I would like to investigate the syntactic and semantic factors which condition intermediate phrase boundaries. Also I would like to explore the issue of how these syntactic and semantic factors can be identified in the text of a sentence so that the ways to identify these factors can be translated into algorithmic phrasing rules (later in 5.3) which specify the insertion of an intermediate phrase boundary in a sentence.

The factors in the Obligatory category are examined first, i.e., extras, subordinate clauses, lists, embedded structures, coordination, and contrastive prominence.

5.2.3.1 Extras

Nespor and Vogel say that the elements which are external to root sentences, i.e., the elements which are linearly represented at the s-structure level but not attached to hierarchical sentence trees, always constitute individual intermediate phrases. In Japanese, too, except for non-relative clauses,[13] these elements, which I call "extras", obligatorily form independent intermediate phrases regardless of lengths or styles of speech: e.g.,

(i) Parenthetical expression:

(11) *E'mi san-ni* L% *gozo'nji no yoo ni* L% *okane-o agema'sita*
"I gave money, as you know, to Emi."

(ii) Tag question:

(12) *Kore-ga Ta'roo-no ie' dewa nai'ndesu ka* L% *so'o desune?*
"This isn't Taro's house, is it."

(iii) Vocative:

(13) *Ta'roo san* L% *kottie irasshai*
"Taro, come here!"

(iv) Expletive:

(14) *Na'nte koto deshoo* L% *okane-o nakusita*
"Oh my gosh, I lost my money."

(v) Moved element:

(15) *Totemo kawai'i* L% *sono pa'nda wa*.
"They are so cute, those pandas."

[Identification]:

Except for vocative, these extras cannot be identified automatically. Unless the beginning and the end of an extra are demarcated by, for example, commas, there is no way, at least my model is concerned, to know the location of the extra in the text of a sentence because of the complication that an extra can be a word, a phrase, a clause, or a sentence. If it is marked that a noun is a name (proper noun),[14] vocative can be identified from its configuration of a noun with no case-marking postposition

5.2.3.2 Subordinate Clauses

One of the three general constraints proposed by Nespor and Vogel is "the beginning of an S-bar"constraint which conditions an intonational phrase boundary immediately before an S-bar. In this model, the syntactic notion of a S-bar or a subordinate clause which is cyclic in nature is limited in the sense that it refers only to a subordinate clause in not deeply embedded structures, i.e., a subordinate clause with just one complimentizer or a subordinate clause with just one subordinator.

First, in regard to the subordinate clause with a complimentizer, Nespor and Vogel's "the beginning of an S-bar" constraint applies literally to Japanese because it is obligatory to insert a L% at the beginning of a S-bar: e.g.,

(16) *Ta'roo-ga* L% [*Ha'nako-ga pa'atii-ni konai-to* $_{\text{S-bar}}$] *sitta*
(17) *Ta'roo-ga* L% [*Ha'nako-ga pa'atii-ni konai-koto*$_{\text{S-bar}}$] *o sitta*
(18) *Ta'roo-ga* L% [*Ha'nako-ga pa'atii-ni konai-no* $_{\text{S-bar}}$]*o sitta*
"Taro knew that Hanako would not come to the party."

Second, in regard to the subordinate clause with a subordinator, the following examples are sufficient to demonstrate the fact that the subordinate clause with a subordinator obligatorily conditions an intermediate phrase boundary at the end of the clause: e.g., (cf. Kuno 1973),

(i) Before clause (V + *ma'e ni*; *uti ni*):
(19) *Okane no nakunarana'i ma'e ni* L% *ryokoo ni iko'o*
"Before we spend all our money, let's go on a trip."

(ii) After clause (V + *kara*; *a'to de / ni*):
(20) *Go'han o ta'beta ato de* L% *benkyoo sima'su*
"After finishing my meal, I will study."

(iii) When clause (V + *to*; *to'ki*):
(21) *Natu'ni na'ru to* L% *te'nisu o suru*
"When it becomes summer, I play tennis."

(iv) Because clause (V + *no'de*)

(22) *Onaka ga suita no'de* L% *udon o ta'beta*

"I had noodles because I was hungry."

(v) Perfective (V + *ta'ra / na'ra*):

(23) *Ka'nada ni itta'ra* L% *tegami o kudasa'i.*

"When you will have gone to Canada, please send me a letter."

[Identification]:

To identify a subordinate clause marked by a subordinator in a text, the subordinator can be used. An intermediate phrase boundary should be inserted immediately after the subordinator which is preceded by a verb.

To identify a subordinate clause marked by a complimentizer, I would like to adopt the simplest way, i.e., specifying manually the subordinate clause by using, for example, quotation marks. The reason is that the automatic operation of identifying the beginning of the subordinate clause, i.e., the site of a L%, becomes quite complicated because of the combinatory possibilities of a S-bar, subject and indirect object.[15]

5.2.3.3 Lists

Nespor and Vogel claim that the "lists" category is one of two special cases which override the general constraints, such as "the end of an NP" constraint. The same argument applies to Japanese. That is, regardless of the syntactic categories of list items, the "lists" category obligatorily develops an intermediate phrase boundary after each item except for the last one: e.g.,

(24) *ao'i* $_{AP}$] L% *ao'i* $_{AP}$] L% *ao'i u'mi da*

"(This) is the blue, blue , blue sea."

(25) *totemo ookina*$_{AP}$] L% *totemo kireina*$_{AP}$] L% *totemo ko'okana hooseki*
"A very big, very beautiful, and very expensive jewel."

(26) *Hitori-no otoko*$_{NP}$] L% *hitori-no onna*$_{NP}$] L% *futari-no kodomo-ga ku'ru*
"One man, one woman, and two children are coming."

The example (24) typically demonstrates the fact that in the lists, length has no effect on intermediate phrasing because if the length were to play any role in the phrasing, the two intermediate phrase boundaries would not take place in such a short sentence.

[Identification]:

The lists can be detected automatically. First, a list of APs can be detected by the absence of nouns immediately following adjectives or nominal adjectives. In the case of NPs in a list, a dictionary is needed to identify them. The reason is that there are three types of nouns which occur without a case-marking postposition, i.e., noun(s) in a list, noun(s) in the vocative and a noun used as a temporal adverbial, such as *asu* "tomorrow" or *kinoo* "yesterday" (the latter does not necessarily condition a L%). Once a dictionary tags proper nouns with a feature, "name" and tags nouns used as temporal adverbials with a feature, "(temporal) adverbial" (as will be seen in 5.3), then, any nouns without a case-marking postposition and without either of the above feature can be treated as those in a list.

5.2.3.4 Embedded Structures

Embedded structures is another special case which, Nespor and Vogel say, overrides the general constraints, such as "the end of an NP" constraint and obligatorily conditions the insertion of an intonational phrase boundary. The claim applies to Japanese. For example, in the following sentences, the largest NPs are split into a few intermediate phrases and a L% is inserted immediately after a repeated node, i.e., a S in (27) and a PP in (28):

(27) *Kore-ga* L% [NP *ti'izu-o ta'beta* S] L% *nezumi-o ta'beta* S] L% *ne'ko-o ta'beta oni'* NP] *desu*
"This is the demon that ate the cat that ate the rat that ate the cheese."

(28) *Kore-ga* L% [NP *heya-no na'ka -no* pp] L% *tukue-no ue-no* pp] L% *ho'n-no ka'baa* NP] *desu*
"This is the cover of the book on the desk in the room."

[Identification]:

Next, how can embedded structures be identified in a text? The answer is that, as far as my model is concerned, embedded structures cannot be identified. Because the model's parsing is linear in that it relies on a string of words, which are tagged with syntactic categories and additional syntactic, semantic, and phonological features (as will be seen in 5.3.2), and does not rely on the syntactic hierarchical structure of a sentence, it cannot locate embedded elements which are hierarchical in structure. Thus, in the model, the beginning of each embedded element, i.e., the location of a L%, must be manually specified by using such a notation as a comma.[16]

5.2.3.5 Coordination

Although it may be included in the category of "lists", as Nespor and Vogel do, "coordination" is regarded in my model as one of the factors which obligatorily condition an intermediate phrase boundary. In Japanese, there are two ways of constructing coordinated structures; one by using coordinating postpositions and the other by using a gerundive form of verbals. First, to connect (29) nouns, (30) verbs, and (31) nominal adjectives, coordinating postpositions are used, i.e., *to* (*ni*, *ya*) for nouns, *yara* for verbs, and *de* for nominal adjectives. An intermediate phrase boundary is, then, obligatorily inserted immediately after each coordinated item except for the last coordinated item if it is not marked by a coordinating postposition: e.g.,

(29) *Ta'roo to* (*ni* / *ya*) L% *Ji'roo to* (*ni* / *ya*) L% *Ha'nako ga kita*
"Taro and Jiro and Hanako came."

(30) *Kankyaku ga naku yara* L% *warau yara* L% *oosa'wagi*
"Crying and laughing, the audience is excited "

(31) *yuueki de* L% *kaiteki de* L% *titeki na tabi'*
"The meaningful, comfortable, and intellectual trip."

Second, to connect (32) sentences, (33) clauses, and (34) adjectives, a gerundive form of verbals is used. An intermediate phrase boundary is, then, obligatorily inserted after each coordinated item inflected in the gerundive: e.g.,

(32) *Oji'i san ga yama' e itte* L% *oba'a san ga kawa' e itta*
"An old man went to a mountain and an old woman went to a river."

(33) *Oba'a san ga momo o mitukete* L% *sore o ie'e moti ka'etta*
"An old woman found a peach, and brought it to home."

(34) *a'oku* L% *hi'roku* L% *fuka'i u'mi*
"The blue, big, deep sea."

[Identification]:

Coordinated elements in the text of a sentence may be identified by relying on the locations of the coordinating postpositions[17] and of gerundive forms of verbals.

5.2.3.6 Contrastive Prominence

Pierrehumbert and Beckman (1989) claim that an intermediate phrase boundary appears immediately before a narrow-focused item. Also, in one of the experiments in Chapter 3, we have looked at the fact that in the phrases of "modifier + noun + postposition", the prominence placed on the nouns splits the phrases into two intermediate phrases by inserting an intermediate phrase boundary before the noun. Thus, it is obvious that in Japanese, too, narrow-focus, or contrastive prominence, obligatorily conditions an intermediate phrase boundary.

[Identification]:

In order to identify an item of contrastive prominence in the text of a sentence, it must be manually specified because any item in a given sentence may possibly have contrastive prominence. And there is no way to predict, first, whether or not a sentence may have a contrastive prominence and, second, if it does, which item in the sentence is assigned contrastive prominence. To mark the contrastive prominence, capital letters may be used and a L% will be inserted immediately before the item in capital letters: e.g.,

	[INPUT]	Watasi ga jidoosha o KARIMASITA
(35)	[Output]	*Watasi ga jidoosha o* L% *KARIMA'SITA*
		"I rented a car (I didn't buy it)."

5.2.4 The Syntactic and Semantic Factors in the Variable Dependent Category

All the factors in the Obligatory category, as we have seen, condition an intermediate phrase boundary with no regard to the performance variables. That is, they condition a L% regardless the length of a sentence and the style of speech. Unlike these factors in the Obligatory category, the following semantic and syntactic factors are all sensitive to a performance variable. Some of the factors condition a L% only when they occur in certain lengths of mora strings which are uninterrupted by a L% and the other factors condition a L% only in one specific style of speech.

5.2.4.1 Theme

Theme is a semantic factor which may condition an intermediate phrase boundary and which is sensitive to the length.

In reference to theme or thematization in Japanese, there are two different ideas. Martin (1975) argues that thematization does not depend on the postposition, *wa*, and that the position and the juncture are the main features for marking thematization. That is, if an item is placed at the sentence-initial position and if a major juncture (which I interpret as an intermediate phrase boundary) is placed after the item, it can be thematized: e.g., (Martin 1975; 225),

(36) *Kore o* || *ka'rera wa tootei rikai suru ko'to ga dekinai*
"This (THEME) they simply can't understand."

Kuno, however, argues that thematization is marked by the postposition, *wa*, saying that "Japanese is one of the few languages of the world that have a built-in mechanism for specifying the theme or topic of the sentence" (Kuno 1973; 5-6). As long as it is generic (37) or anaphoric (38), a sentence initial noun phrase marked by the postposition, *wa*, is treated as a theme: e.g., (Kuno ibid; 44),

(37) *Kujira wa honyuu-do'obutu desu*
"Speaking of whales, they are mammals."

(38) *Jo'hn wa watasi no tomodati de'su*
"Speaking of John, he is my friend."

For a practical reason, I adopt Kuno's argument because it makes it possible to predict the location of an intermediate phrase boundary by using the postposition, *wa*, as a theme-marker. If, on the other hand, Martin's argument is adopted, the issue of thematization and the insertion of an intermediate phrase boundary will become a circular argument. That is, in order to insert an intermediate phrase boundary, a theme must be found. However, a theme must be marked by an intermediate phrase boundary which cannot be inserted until the theme is found. To avoid such a circular situation, I adopt Kuno's argument on thematization.

There are sufficient data that a theme attracts an intermediate phrase boundary. Poser (1984) says that it is usually the case that a thematized phrase marked by *wa* constitutes an independent major (= intermediate) phrase. Furthermore, we have already looked at acoustic evidence in Warkentyne (1984) that even in relatively short sentences, thematized phrases develop an intermediate phrase boundary immediately after the phrases.

[Identification]:

Relying on the aove works, in the model an intermediate phrase boundary will be inserted after the phrase marked by the postposition, *wa* if no L% is found in a string of more than ten morae in the vicinity of the *wa*.

A minor problem in identifying a theme by way of the *wa*-maker is that a sentence may have more than one *wa*-phrase. In such a case, it is the sentence initial *wa*-phrase which is a theme (Kuno 1973). Thus, even if there are more than one *wa*-phrases in a sentence, it is only after the initial *wa*-phrase where a L% should be inserted: e.g.,

(39) *Watakusi wa* L% *eigo wa wakarima'su*
"Speaking of myself, I understand English."

5.2.4.2 Argument Structures

In regard to argument structures, which are semantic in nature, in Japanese, there seems to be a tendency to insert an intermediate phrase boundary immediately before the argument which immediately precedes a verb. An intermediate phrase boundary is inserted with no regard to whether it is an optional argument (40), such as an adjunct, or an obligatory argument (41): e.g.,

L% Optional Argument + V:

(40) *Watasi no ha'ha-ga Honda san-ni* L% *kinoo no go'go aimasita*
"My mother met Mr Honda yesterday afternoon."

L% Obligatory Argument + V:

(41) *Watasi no ha'ha-ga kinoo no go'go* L% *Honda san-ni aimasita*
"My mother met Mr Honda yesterday afternoon."

The tendency to insert an intermediate phrase boundary immediately before the argument which immediately precedes a verb may have to do with the "focus structure"[18] in Japanese. That is, according to Warkentyne (1978), "[t]he Japanese clause is unmarked for focus if the focal element is the constituent immediately preceding the verbal phrase" (ibid; 220). We have already seen that prominence attracts an intermediate phrase boundary. Now, because the argument preceding a verb is a focal element of a sentence, the focal element may easily attract a L% before the element. The above reasoning, based on the focus structure, may explain why in Japanese, there is a tendency to insert an intermediate phrase boundary before the argument which immediately precedes a verb.

When a verb takes two obligatory arguments (i.e., ditransitive verb), the insertion of an intermediate phrase boundary before the initial obligatory argument is preferred to the insertion of a boundary before the second argument, avoiding thus the initial argument from being split from a verb: e.g., L% Obligatory Argument + Obligatory Argument + V:

(42) *Watasi no ha'ha-ga* L% <u>*omoti-o*</u> <u>*Honda san-ni*</u> <u>*agema'sita*</u>
(rice-cakes ACC. Mr Honda DAT. gave)
"My mother gave Mr Honda rice-cakes."

There is nothing wrong in inserting a L% between the two obligatory arguments. However, doing so may place a contrastive prominence on the second argument: e.g.,

(43) *Watasi-no ha'ha-ga* <u>*omoti-o*</u> L% <u>*Honda san-ni*</u> <u>*agema'sita*</u>
"My mother gave rice-cakes to Mr Honda (not to Mr Suzuki)."

The claim, which is originally made by Nespor and Vogel, that two obligatory arguments constitute a single intonational phrase with a verb is phonetically supported, in the case of Japanese, by Gallaiford (1983) which shows that the utterances of the five-word sentences with two obligatory arguments exhibit a tendency to refrain from inserting a L% between two obligatory arguments. Thus, in regard to the argument structures, the general rules are: (i) an intermediate phrase boundary is inserted before the argument which immediately precedes a verb with no regard to whether it is optional or obligatory; and (ii) if, however, a verb takes two obligatory arguments, an intermediate phrase boundary is inserted before the initial obligatory argument.[19]

[Identification]:

An argument structure may be identified by a postposition and the position it takes; i.e., any phrase which is placed immediately before a verb ought to be regarded as an argument as long as the verb is not a ditransitive verb. How can we, then, isolate an argument to insert an intermediate phrase boundary immediately before it? An argument can be demarcated either by locating the postposition (except for *no*-possessive) which immediately precedes the argument, as in (44) or by locating the content word which immediately precedes the argument, as in (45).[20]

(44) *Watasi no ha'ha-ga Honda san-ni* L% *<u>kinoo no go'go</u>* *aimasita*
(Optional Arg.)
"My mother met Mr Honda <u>yesterday afternoon</u>."

(45) *Watasi no ha'ha-ga kinoo no go'go* L% *<u>Honda san-ni</u>* *aimasita*
(Obligatory Arg.)
"My mother met <u>Mr Honda</u> yesterday afternoon."

Even if a verb takes two obligatory arguments (i.e., ditransitive verb), case-marking postpositions can be still used to identify the arguments. First, identify the verbs which takes two arguments. The arguments of a ditransitive verb are usually marked both by the postpositions, *o* "accusative" and *ni* "dative": e.g.,

(46) *Ki'no'o ha'ha-ga* L% *<u>Honda san-ni</u>* *<u>hon'-o</u>* *ageta.*
(Mr Honda DAT. a book ACC.)
(47) *Ha'ha-ga kino'o* L% *<u>ho'n-o</u>* *<u>Honda san-ni</u>* *ageta.*
(a book ACC. Mr Honda DAT.)
"Yesterday, (my) mother gave a book to Mr Honda."

To locate the onset of the initial obligatory argument at which a L% is inserted, first, locate a penultimate postposition which is either *ni* (dative) or *o* (accusative). Then, the same procedure can be taken as the above case.

That is, the initial argument can be demarcated either by locating the postposition (except for *no*-possessive) which immediately precedes the argument or by locating the content word which immediately precedes the argument.

5.2.4.3 Wide Scope

As mentioned time and again, Nespor and Vogel set the end of an NP constraint to prohibit an NP from being split into two or more intonational phrases except for the two special cases, i.e., lists and embedded structures. In Japanese, there is another case in which an NP is easily split into two intermediate phrases. The case is "wide scope", another semantic factor. In the phrases of "modifier + modifier + noun", if the initial modifier does not modify the immediately following modifier, that is., if the initial modifier has wide scope, instead of narrow scope, of modification, a L% may be inserted between the two modifiers. Such a case can typically be seen when ambiguity is involved. For example, if the sequence of the words: *aoi'* (blue) + *u'mi-no* (sea's) + *ie'* (house) means, instead of "the house by the blue sea", "the blue house by the sea" where *aoi'* (blue) is not a modifier of the following item, *u'mi* (sea), the phrase may be realized as *aoi'* L% *u'mi-no ie'*. The intermediate phrase boundary between the two modifiers is a sort of a prosodic flag which cautions the listener against the fact that the initial modifier does not modify the immediately following item (i.e., the negation of the narrow scope). Thus, in Japanese, if an NP has two modifiers and if the initial modifier has wide scope, an intermediate phrase boundary may take place after the initial modifier.

In the model, wide scope is treated as a semantic factor which is not sensitive to the length but to the style of speech due to the obvious reason that the length has no role in the intermediate phrasing. If it were the case, it would not explain why in the narrow scope situation, exactly the same phrase of "modifier + modifier + noun" would not be split into two intermediate phrases. The style of speech or how enunciatively the phrase ought to be uttered must be a condition of the insertion of a L%.

[Identification]:

Wide Scope must be manually specified because, for example, given the text of the phrase: /aoi umi no ie/ "blue sea POSS. house", there is no way to tell whether or not the initial modifier has wide or narrow scope. The marker for wide scope is "< >" which should be inserted after the modifier of wide scope: e.g.,

Wide Scope:	[Input]	aoi <> umi no ie
(48)	[Output]	*ao'i* L% *u'mi-no ie'*

5.2.4.4 Every NP & ADP

What Nespor and Vogel's end of an NP constraint can basically do is insert an intonational phrase boundary at the end of every NP in a sentence. In the model, the idea of the constraint is adopted and modified to create a syntactic factor, "every NP & ADP", which may condition a L% after every NP and ADP in a sentence except the NP or ADP which immediately precedes a verb. In Japanese, it is not a well-formed phrasing if a L% is inserted immediately before a verb unless it is required to do so by Obligatory factors: e.g.,

(49) * *Honda san ga atarasii kuruma o* L% *katta*
"Mr Honda bought a new car."

(50) *Honda san ga atarasii kuruma o* L% *KATTA*
"Mr Honda bought a new car (He didn't rent it)." (cf. Warkentyne 1978)

This factor is created to account for a type of phrasing which exhaustively parse a sentence into intermediate phrases (although it is quite unlikely that such a phrasing occurs in our daily conversational speech): e.g.,

(51) *Kino'o* $_{ADP}$] L% *Ka're-ga* $_{NP}$] L% *ka'nojo-to* $_{NP}$] L% *susi'ya-de* $_{NP}$] L% *susi'-o* $_{NP}$] *ta'beta*
"Yesterday, he had sushi with her at a sushi-restaurant ."

[Identification]:

The way to find a noun and an adverb of the "every NP&ADP" category is simple. In the case of a noun in the category, find a lexcal item marked by the syntactic category N which is followed by a postposition (except for *no*-possessive)[21] and to which a L% has not been assigned yet. In the case of an adverb in the category, find a lexical category marked by the syntactic category AD which may or may not be followed by a postposition. Furthermore, in both cases, a verb should not immediately follow a noun or an adverb.

5.3 Operations of the Intermediate Phrasing Model

In this section, first, I would like to discuss the following issues:

(i) Three levels of developing phrasing variations.

(ii) The characteristics of the dictionary required by the model.

(iii) The algorithmic rules which are the translations of the automatic and non-automatic identifications of the syntactic and semantic factors discussed in 5.2.3 and 5.2.4.

(iv) A direction in searching for the syntactic and semantic factors in a text.

Then, using examples, I will demonstrate how the algorithmic intermediate phrasing model works.

5.3.1 Three Levels of Developing Phrasing Variations

In the model, it is assumed that there are three levels in creating phrasing variations for a given sentence (cf. Table 5.1); i.e.,

(i) {[+Obligatory] [-Variable Dependent]}

(ii) {[+Obligatory] [+Variable Dependent {[+Length] [Informal]}]}

(iii) {[+Obligatory] [+Variable Dependent {[+Length] [Formal]}]}

These phrasing variations are developed, from a least parsed (basic) phrasing to an exhaustively parsed phrasing, by applying the syntactic and semantic factors category by category.

5.3.1.1 Phrasing at the Level of {[+Obligatory] [-Variable Dependent]}

The first stage of creating phrasing variations is to develop a basic phrasing which is conditioned only by the factors in the Obligatory category ([+Obligatory]). The factors in the Variable Dependent category, which are switched off ([-Variable Dependent]), have no contribution to the basic phrasing. The operation at this level of phrasing, which is equivalent to Nespor and Vogel's Basic Intonational Phrasing Formation rule, will develop a kind of a phrasing which may be observed in casual speech with very few intermediate phrase boundaries.

5.3.1.2 Phrasing at the Level of ([+Obligatory] [+Variable Dependent ([+Length] [Informal])])

The next stage of creating phrasing variations is to activate, in addition to the factors in the Obligatory category, the factors which are sensitive to length ([+Length]) in the Variable Dependent category ([+Variable Dependent]). The factors which are sensitive to [Formal] style of speech in the Variable Dependent category are switched off by way of selecting the parameter: [Informal]. The operation of this level will create a type of phrasing which may be observed in informal but not too casual speech, such as when a Japanese gives a direction to a stranger.

5.3.1.3 Phrasing at the Level of ([+Obligatory] [+Variable Dependent ([+Length] [Formal])])

The last stage of developing phrasing variations is to create a phrasing which exhaustively parses a sentence into intermediate phrases by activating all the syntactic and semantic factors. Such an unusual phrasing may occur when an instructor of Japanese reads a passage from a textbook for beginning students of the language.

5.3.2 Notations and Algorithmic Rules

5.3.2.1 Dictionary

To identify the syntactic and semantic factors which condition intermediate phrase boundaries in a text, the model must be accompanied by a dictionary which contains the following information about lexical items: i.e.,

(i) Syntactic categories: e.g.,

a. Noun --> [N]

b. Verb --> [V]

c. Postposition --> [POSTP]

d. Modifier --> [MOD] (i.e., Adjective or Adverb)

e. Any Category --> [X]

(ii) Syntactic and semantic features which specify the types of lexical items, such as cases or grammatical functions: e.g.,

a. Verb --> [V]
|| Gerund(ive)

b. Verb --> [V]
|| Ditran(sitive)

c. Postposition --> [POSTP]
|| Subord(inator)

d. Postposition --> [POSTP]
|| Theme

(iii) Phonological feature specifications; i.e., the number of morae (M) and the location of an accent (A) in the lexical items in the format of (M/A): e.g,

a. umi (2/1) - the lexical item, /umi/ "sea" which consists of two morae (2/) with the accent on the initial mora (/1).

b. kokoro (3/2) = the lexical item, /kokoro/ "mind" which consists of three morae (3/) with the accent on the second mora (/2).

c. byooin (4/0) = the lexical item, /byooin/ "hospital" which consists of four morae (4/) with no accent (/0).

5.3.2.2 Word Tagging

In the operations of the algorithmic phrasing model, it is assumed that words are tagged with the lexical information stored in the dictionary prior to the application of algorithmic rules. For example, a simple typed-in text: watasi wa umi o miru "I see the sea" is tagged as follows:

[watasi $_N$] (3/0) [wa $_{POSTP}$] (1/0) [umi $_N$] (2/1) [o $_{POSTP}$] (1/0) [miru $_V$] (2/1)
|| Theme

5.3.2.3 Algorithmic Rules (Pattern Matching)

The basic device to find syntactic and semantic factors which condition an intermediate phrase boundary is "pattern matching" encoded as algorithmic rules. When the model finds patterns in a text which match with the patterns set on the left-hand side of the algorithmic rules, these identified patterns are replaced (-->) by the patterns set on the right-hand side of the rule. Because each pattern on the right-hand side of the rules has a specification of L%, the "pattern matching" inevitably conditions the insertion of L%.

Table 5.2 below is a partial list of the algorithmic rules which are translations of the automatic and non-automatic identifications of the syntactic and semantic factors discussed in 5.2.3 and 5.2.4. A list of notations used in the algorithmic rules will appear as Table 5.3 immediately after Table 5.2.

Table 5.2
A partial list of algorithmic intermediate phrasing rules.

<u>Vocative (Extra)</u>:

(1) [[N]]$^+$ --> [[N] L%]$^+$
|| Name || Name

<u>Subordinate Clause with a subordinator</u>:

(2) [V] [POSTP] --> [V] [POSTP] L%
|| Subord || Subord

<u>List of Nouns</u>:

(3) [[N] [POSTP]]$^+$ --> [[N] [POSTP] L%]$^+$
|| Null || Null

<u>Coordination Marked by Gerundive</u>:

(4) [V] --> [V] L%
|| Gerund || Gerund

<u>Coordination of Verbs</u>:

(5) [[V] [POSTP]]$^+$ --> [[V] [POSTP] L%]$^+$
|| Coord || Coord

Contrastive Prominence:

(6) [X] --> L% [X]

Theme:

(7) [POSTP]
||Theme

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) > 10;\ L\%(\%) \lll W_1 \ll [\text{POSTP}]\,||\text{Theme} \ll W_n \lll L\%(\%)$$

--> [POSTP] L%
||Theme

Argument Structure:

(8) [[N], [POSTP]] [MOD]*[N] ([POSTP]) [V]

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15;\ L\%(\%) \lll W_n \ll W_1 = [V]$$

--> [[N], [POSTP]] L% [MOD]*[N] ([POSTP]) [V]

Argument Structure with Copula:

(9) [[N], [POSTP]] [MOD]*[N/AJ] ([POSTP]) [V]
||Copula

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15;\ L\%(\%) \lll W_n \ll W_1 = [V]$$

--> [[N], [POSTP]] L% [MOD]*[N/AJ] ([POSTP]) [V]
||Copula

Argument Structure with Two Obligatory Arguments:

(10) [[N] , [POSTP]] [MOD]*[N] [POSTP] [MOD]*[N] [POSTP] [V]
||Acc/Dat || Dat /Acc || Ditrans

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15;\ L\%(\%) \lll W_n \ll W_1 = [\ V]$$

--> [[N] , [POSTP]] L% [MOD]*[N] [POSTP] [MOD]*[N] [POSTP] [V]
||Acc/Dat || Dat /Acc || Ditrans

Wide Scope:

(11) [<>] --> [<>] L%

Every NP:

(12) [N] [POSTP] [X] --> [N] [POSTP] L% [X]
|| -Poss || -V || -Poss || -V

Every ADP:

(13) [AD] ([POSTP]) [X] --> [AD] ([POSTP]) L% [X]
|| -V || -V

Table 5.3
A list of notations used in the algorithmic rules in Table 5.2.

(i) [ ]* = zero occurrence or one occurrence or more; e.g., [MOD]* (= there may be no modifier, or may be one or more than one modifier)

(ii) [ ]+ = at least one occurrence or more; e.g., [MOD]+ (= there must be at least one or more occurrence of a modifier).

(iii) [,] = no order in precedence; e.g., [A , B] (= A precedes B or B precedes A].

(iv) W_i = i-th lexical item (which constitutes a part of a string of morae).

(v) <<< = immediate precedence; e.g., A <<< B (= A immediately precedes B).

(vi) >>> = immediate following; e.g., A >>> B (= A immediately follows B).

(vii) << = precedence at a distance; e.g., A << B (= A precedes B at a distance).

(viii) >> = following at a distance; e.g., A >> B (= A follows B at a distance).

(ix) - = except for; e.g., || - Poss (= except for the feature "possessive").

(x) L%% = sentence initial or final boundary.

(xi) ( ) = optional; e.g., ([N]) (= there may or may not be a noun).

5.3.2.4 Directions in Searching for Patterns in a Text.

In regard to the direction of searching for patterns in a sentence, there is no preference. That is, patterns can be searched from left-to-right or from right-to-left. Nevertheless it appears that for finding a theme, the left-to-right direction is preferred because the initial encounter of a *wa*-postposition can immediately be regarded as a theme-marker even if there are more than one *wa*-postposition in a sentence. However, because word-tagging is done prior to the operation of pattern-matching, a theme-feature can be tagged only to the initial *wa*-postposition at the level of word-tagging. Thus, theme does not set any preference on the direction of searching patterns.

Following the direction of natural speech, in this model, left-to-right direction is adopted. A minor irregularity is that the model will take the opposite direction in mora-counting for argument structures in order to examine whether or not a string of morae in the vicinity of an argument structure is long enough for the argument structure to condition a L%. The counting starts from a verb and precedes rightward until it encounters a boundary. In the case of theme, mora-counting is conducted from left-to-right.

5.3.3 Demonstrations

Using a few examples, I would like to demonstrate how the whole algorithmic intermediate phrasing model works.

The following are the two sentences used as examples for the demonstration of the operation of the algorithmic intermediate phrasing model:

(51) *Watakusi-no ha'ha-wa mi'kan ringo momo budoo-o su'upaa-de katte, byooki-no ti'ti-o byooin-ni mima'tta.*

"Speaking of my mother, she bought oranges, apples, peaches and grapes at a supermarket and visited my sick father in the hospital.

(52) *Ha'ha wa ookina byooin-no kangofu-ga ti'ti-ni totemo si'nsetu d atta no'de, ka'nojo-ni mi'kan-o ageta*

Speaking of the mother, she gave the oranges (narrow-focused) to the big nurse in the hospital (wide scope) because the nurse was very kind to the father."

The following is the textual input of the sentences (i.e., a text which a user types in):

(53) Watakusi no haha wa mikan ringo momo budoo o suupaa de katte byooki no titi o byooin ni mimatta.
(54) Haha wa ookina <> byooin no kangofu ga titi ni totemo sinsetu d atta node kanojo ni MIKAN o ageta.

5.3.3.1 The Operations at the level of {[+Obligatory] [-Variable Dependent]}

The first possible phrasings for the sentences are the basic phrasings at the level of {[+Obligatory] [-Variable Dependent]}. To develop the basic phrasings, the model should be able to pick up relevant patterns. The algorithmic intermediate phrasing rules will then insert intermediate phrase boundaries at the appropriate locations. (Hereafter, all the numbers in the texts of sentences correspond to the numbers assigned to the algorithmic rules in Table 5.2.)

(55) Watakusi no haha wa mikan (3) ringo (3) momo (3) budoo o suupaa de katte (4) byooki no titi o byooin ni mimatta.
(56) Haha wa ookina <> byooin no kangofu ga titi ni totemo sinsetu d atta node (2) kanojo ni MIKAN (6) o ageta.

Among the factors in the Obligatory category, the following are the relevant factors in the basic phrasings of the sentences:

(a) List (3) of the nouns marked by having no case-marking postpositions.

(b) Coordination (4) marked by the gerundive form of the verb; *kau* "buy".

(c) Subordinate Clause (2) marked by the subordinator, *no'de* "because".

(d) Contrastive Prominence (6) marked by the capital letters.

5.3.3.1.1 List (3) of nouns

Hereafter, in the operations of the model, it is assumed that each lexical item is tagged with the appropriate information from the dictionary prior to the application of the algorithmic rules (even though only relevant tags will be indicated in examples).

i) The first series of the tags which the model will pick up is the nouns ([N]): /mikan ringo momo/ "oranges apples peaches" which are unmarked (|| Null) by case-marking postpositions ([POSTP]) (cf. 5.2.3.3): i.e.,

```
[mikan N] [POSTP]  [ringo N] [POSTP]  [momo N] [POSTP]
          || Null            || Null           || Null
```

The above pattern is, in fact, an expansion of:

```
[[N] [POSTP]] +
     || Null
```

(ii) The above pattern is picked up because it matches with the left-hand side of Rule 3: i.e.,

```
(3)   [[ N] [ POSTP]]+    -->    [[ N] [ POSTP]  L% ]+
            || Null                    || Null
```

(iii) In Rule 3, the matching pattern is replaced with the right-hand side pattern which specifies the insertion of L% after each noun ([[N] L%]+) which is unmarked by a postposition ([POSTP] || Null)[22]: i.e.,

```
[mikan N] [POSTP]  [ringo N] [POSTP]  [momo N] [POSTP]    -->
          || Null            || Null            || Null

[mikan N] [POSTP] L%   [ringo N] [POSTP]  L%  [momo N] [POSTP]  L%
          || Null                || Null                || Null
```

(iv) The result is:

mi'kan L% *ringo* L% *momo* L% (*budoo-o*)[23]

5.3.3.1.2 Coordination (4)

(i) The next pattern which the model will pick up is the sentence-medial verb, /katte/ "bought and" (cf. 5.2.3.5). It is tagged because of its gerundive ending /te/ as a verb ([v]) in the gerundive form (|| Gerund): i.e.,

```
[kat-te v]
       || Gerund
```

(ii) The above pattern is picked up because it matches with the left-hand side of Rule 4:

```
(4)   [ v]          -->     [ v]        L%
       || Gerund             || Gerund
```

(iii) Rule 4 replaces the identified pattern with the right-hand side pattern which then specifies the insertion a L% after the gerundive verb, /katte/.

(iv) The result is:

katte L% (*byooki-no ti'ti-o*)

5.3.3.1.3 Subordinate Clause (2)

(i) The next pattern which the model will pick up is at /atta node/ "was because" (cf. 5.2.3.2) which is tagged as a postposition ([POSTP]) functioning as a subordinator (|| Subord) that is preceded by a verb ([V] || Copula): i.e.,

```
[atta V]      [node POSTP]
   || Copula      || Subord
```

(ii) The above pattern is picked up because it matches with the left-hand side of Rule 2: i.e.,

```
(2)  [ V] [ POSTP]       -->    [ V] [ POSTP]    L%
          || Subord                  || Subord
```

(iii) What Rule 2 does is replace the identified pattern with the right-hand side pattern which specifies the insertion a L% after the postposition: i.e.,

```
[atta V]   [ node POSTP]     -->    [atta V]   [ node POSTP]    L%
   || Copula    || Subord              || Copula    || Subord
```

(iv) The result is:

no'de L% (*ka'nojo-ni*)

5.3.3.1.4 Contrastive Prominence (6)

(i) The next pattern which the model will pick up is the capitalized item (cf. 5.2.3.6), i.e., /MIKAN/ "orange" which is tagged as [MIKAN N].

(ii) The model picks up the pattern: [MIKAN $_N$] because it can identify the pattern in the text and match it with the left-hand side of Rule 6: i.e.,

(6) [X][24] --> L% [X]

(iii) Rule 6 replaces the identified pattern with its right-hand side pattern which specifies the insertion a L% before a capitalized item: i.e.,

[MIKAN $_N$] --> L% [MIKAN $_N$]

(iv) The result is:

(*ka'nojo-ni*) L% *MI'KAN*

Consequently, the following is the output of the phrasing operations at the level of {[+Obligatory] [-Variable Dependent]}:

(57) *Watakusi-no ha'ha-wa mi'kan* **L%** *ringo* **L%** *momo* **L%** *budoo-o su'upaa-de katte* **L%** *byooki-no ti'ti-o byooin-ni mima'tta.*

(58) *Ha'ha wa ookina byooin-no kangofu-ga ti'ti-ni totemo si'nsetu d'atta no'de* **L%** *ka'nojo-ni* **L%** *MI'KAN-o ageta*

5.3.3.2 The Operations at the level of {[+Obligatory] [+Variable Dependent {[+Length] [Informal]}]}

The next stage of creating phrasing variations is to activate, in addition to the factors in the Obligatory category, the factors which are sensitive to length ([+Length]) in the Variable Dependent category ([+Variable Dependent]), i.e., the "theme" and the "argument structures". The model should be able to pick up theme and argument structure patterns in the text of example sentences:

(59) Watakusi no haha wa (7a) mikan L% ringo L% momo L% budoo o (8a) suupaa de katte L% byooki no titi o (8b) byooin ni mimatta.
(60) Haha wa (7b) ookina <> byooin no kangofu ga titi ni (9) totemo sinsetu d atta node L% kanojo ni L% MIKAN o ((10)) ageta.

5.3.3.2.1 Theme (7a)

(i) In the first sentence (59), the model will pick up the theme-marker postposition, /wa/ (cf. 5.2.4.1), which is tagged as:

[wa POSTP]
|| Theme

(ii) The model picks up the above pattern because it can identify the pattern in the text and match it with the left-hand side of Rule 7: i.e.,

(7) [POSTP]
|| Theme

$$/ \sum_{i=1}^{n} \mathrm{Morae}(W_i) > 10;\ L\%(\%) \lll W_1 \ll [\ \mathrm{POSTP}] \,||\, \mathrm{Theme} \ll Wn \lll L\%(\%)$$

--> [POSTP] L%
|| Theme

What the rule says is that for a theme-marker ([POSTP]|| Theme) to have a L% immediately after the marker, a condition on length must be satisfied. That is, there must be a string of, in total, more than 10 morae ($\sum \mathrm{Morae}(W) = \mathrm{Morae}(W_1) + \text{---}\ \mathrm{Morae}(W_n) > 10$) from the item ($W_1$) which immediately follows (<<<) an intermediate phrase boundary (L%) or a sentence initial boundary (L%%) to the lexical item W_n which immediately precedes (<<<) an intermediate phrase boundary (L%) or a sentence final boundary (L%%). Relative to W_1 and W_n, the location of the theme-marker is defined as (W_i << [POSTP] || Theme << Wn). That is, the marker follows at a distance (<<) W_1 and precedes at a distance (<<) W_n.

(iii) Because information about the number of morae is tagged to each lexical item, the model can easily count whether or not the mora number of the string satisfies the parameter set for the variable of length: i.e.,

L%% watakusi (4/0) no (1/0) haha (2/1) wa (1/0) mikan (3/1) L%
(W_1) ---------------------------------------> (W_n)

"I of mother speaking of orange."

$$\sum_{i=1}^{n} \text{Morae}\ (W_i) = 4\ (W_1) + 1\ (W_2) + 2\ (W_3) + 1\ (W_4) + 3\ (W_5) = 11$$

(iv) Because, in the vicinity of the theme-marking postposition, there is a string of 11 morae which is uninterrupted by a L% and because the length parameter of a theme is set as [>10] (cf. Table 5.1), the length condition is satisfied. The rule thus replaces the identified pattern with the the right-hand side pattern which specifies the insertion of a L% after a theme-marking postposition: i.e.,

[wa POSTP] || Theme --> [wa POSTP] || Theme L%

(v) The result is:

(*watasi no ha'ha*) *wa* L% (*mi'kan*)

5.3.3.2.2 Theme (7b)

(i) For the theme marker at /haha wa/ "speaking of the mother" in the second sentence (60), the same procedure should be taken as the above case (7a).

(ii) Because in the vicinity of the theme-marking postposition, there is a string of, in total, 32 morae with no intervening L%, Rule 7 conditions a L% immediately after the postpostion.

(iii) The result is:

(*Ha'ha*) *wa* L% (*ookina byooin no kangofu ga ti'ti ni*
"Mother THEME big hospital POSS. nurse NOM. father DAT.
totemo si'nsetu d atta no'de)
very kind was because."

5.3.3.2.3 Argument Structure (8a)

i) In the search for an argument structure (cf. 5.2.4.2), a pattern that the model picks up is at /o suupaa de katte/ "ACC. supermarket LOC. bought and" which is tagged as:

[o POSTP] [suupaa N] [de POSTP] [katte V]

(ii) The pattern is picked up because it matches with the left-hand side of Rule 8:

(8) [[N] , [<u>POSTP</u>]] [MOD]*[<u>N</u>] ([<u>POSTP</u>]) [<u>V</u>]

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15;\ \text{L\%}(\%) \lll W_n \ll W_1 = [\ \text{V}]$$

--> [[N] , [POSTP]] L% [MOD]*[N] ([POSTP]) [V]

(iii) Although the identified pattern matches with the left-hand side pattern of Rule 8, it does not satisfy the length condition set in the rule because there is a string of only 12 morae between the verb /katte/ ([V] - W_1, the first word in the rightward mora-counting) and the item (W_n, the last word in the mora-counting) which immediately follows a L% (L% <<<W_n): i.e.,

L% budoo (3/0) o (1/0) suupaa (4/o) de (1/o) katte (3/0)
(W_n) <--------------------------- (W_1)

"Grapes ACC. supermarket LOC. bought and."

$$\sum_{i=1}^{n} \text{Morae}(W_i) = 3\ (W_1) + 1\ (W_2) + 4\ (W_3) + 1\ (W_4) + 3\ (W_5) = 12$$

(iv) Because the length value, 12, does not satisfy the parameter set at [≥ 15], Rule 8 fails to condition a L% before the argument, /suupaa de/.

5.3.3.2.4 Argument Structure (8b)

(i) In the search for an argument structure, the next argument structure pattern the model will pick up is at /o byooin ni mimatta/ "ACC. hospital at visited" which is tagged as:

[o POSTP] [byooin N] [ni POSTP] [mimatta V]

(ii) The above pattern is picked up because it matches with the left-hand side of Rule 8 (where the pattern match is made clear by underlining matching items): i.e.,

(8) [[N] , [POSTP]] [MOD]*[N] ([POSTP]) [V]

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15;\ L\%(\%) <<< W_n << W_1 = [\ V]$$

--> [[N] , [POSTP]] L% [MOD]*[N] ([POSTP]) [V]

(iii) To examine whether or not the mora number of a string (($\sum$ Morae(W) = Morae(W_1) + --- Morae(W_n)) satisfies the parameter set for the variable of the length, a mora-counting is conducted rightwards from the V to the word (W_n) which is immediately preceded by a L% (L% <<<W_n): i.e.,

L% byookì(3/0) no(1/0) titi(2/1) o(1/0) byooin(4/0) ni(1/0) mimatta(4/0)
(W_n) <-- (W_1)

"Sickness POSS. father ACC. hospital LOC. visited."

$$\sum_{i=1}^{n} \text{Morae}(W_i) = 4\ (W_1) + 1\ (W_2) + 4\ (W_3) + 1\ (W_4) + 2\ (W_5) + 1\ (W_6) + 3\ (W_7) = 16$$

(iv) Because the pattern in the sentence matches with the pattern on the left-hand side of Rule 8 and because the length condition is satisfied, the rule replaces the identified pattern with the right-hand side pattern which specifies the insertion of a L% after the penultimate postposition.

(v) The result is:

(*byooki no ti'ti*) *o* L% *byooin ni mima'tta*

5.3.3.2.5 Argument Structure (9)

(i) The argument structure pattern that the model will pick up in the second sentence is at / ni totemo sinsetu d[25] atta/ "DAT. very kind was" which is tagged as:

[ni POSTP] [totemo AD] [sinsetu AJ] [d POSTP] [atta V]
|| Copula

(ii) The model picks up the above pattern because it can identify the pattern[26] in the text and match it with the left-hand side of Rule 9: i.e.,

(9) [[N] , [POSTP]] [MOD]*[N/AJ] ([POSTP]) [V]
|| Copula

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15; \; L\%(\%) \lll W_n \ll W_1 = [\,V]$$

--> [[N] , [POSTP]] L% [MOD]*[N/AJ] ([POSTP]) [V]
|| Copula

(iii) To examine whether or not the mora number of a string satisfies the parameter set for the variable of the length, a mora-counting is conducted rightwards from the V, which is the first word (W_1 = [V]) in the counting, to a word which is immediately preceded by a L% (L% $\lll W_n$): i.e.,

L% ookina (4/0) byooin (4/0) no (1/0) kangofu (4/0) ga (1/0) titi
(W_n) <---------------------------------

"Big hospital POSS. nurse NOM. father

(2/1) ni (1/0) totemo (3/0) sinsetu (4/0) d (0/0) atta(3/0)
<-- (W_1)

DAT. very kind was."

$$\sum_{i=1}^{n} \text{Morae}(W_i) = 3\,(W_1) + 0\,(W_2) + 4\,(W_3) + 3\,(W_4) + 1\,(W_5) + 2\,(W_6) + 1$$

$$(W_7) + 4\,(W_8) + 1\,(W_9) + 4\,(W_{10}) + 4\,(W_{11}) = 27$$

(iv) Because the number of morae, 27, satisfies the parameter set at [≥ 15], the identified pattern is replaced by the right-hand side pattern in Rule 9 which specifies the insertion of a L% after the penultimate postposition /ni/.

(v) The result is:

(*kangofu ga ti'ti*) *ni* L% *totemo si'nsetu d atta*

5.3.3.2.6 Argument Structure ((10))

(i) In the search for a pattern of an argument structure, it may appear that the model picks up the argument structure pattern at /node ka'nojo ni MIKAN o ageta/ "because she DAT. oranges ACC. gave" which is tagged as:

[node POSTP] L% [kanojo N] [ni POSTP] L% [MIKAN N] [o POSTP] [ageta V]
||Acc/Dat || Dat /Acc || Ditrans

(ii) It appears as if the above syntactic pattern matches with the left-hand side of Rule 10: i.e.,

(10) [[N] , [POSTP]] [MOD]*[N] [POSTP] [MOD]*[N] [POSTP] [V]
||Acc/Dat || Dat /Acc || Ditrans

$$/ \sum_{i=1}^{n} \text{Morae}(W_i) \geq 15;\ L\%(\%) \lll W_n \ll W_1 = [\ V]$$

--> [[N] , [POSTP]] L% [MOD]*[N] [POSTP] [MOD]*[N] [POSTP] [V]
||Acc/Dat || Dat /Acc || Ditrans

(iii) However, due to the L%'s before [kanojo N] and before [MIKAN N], the argument structure pattern at /node kanojo ni MIKAN o ageta/ and the left-hand side pattern in Rule 10, in fact, do not match. Thus, the model ignores the argument structure pattern.

Consequently, the following is the output of the operations at the level of {[+Obligatory] [+Variable Dependent {[+Length] [Informal]}]}. (The bold faced boundary markers are newly inserted at this level.)

(61) *Watakusi-no ha'ha-wa* **L%** *mi'kan* L% *ringo* L% *momo* L% *budoo-o su'upaa-de katte* L% *byooki-no ti'ti-o* **L%** *byooin-ni mima'tta.*
(62) *Ha'ha wa* **L%** *ookina byooin-no kangofu-ga ti'ti-ni* **L%** *totemo si'nsetu d atta no'de* L% *ka'nojo-ni* L% *MI"KAN-o ageta*

5.3.3.3 The Operation at the Level of {[+Obligatory] [+Variable Dependent] {[+Length] [Formal]}]}

The phrasings of the last level, i.e., the level of {[+Obligatory] [+Variable Dependent] {[+Length] [Formal]}]}, will be created by activating all the syntactic and semantic factors. That is, the model activates, in addition to the factors of the previous level, two factors which are sensitive to the [Formal] style of speech, i.e., "wide scope" and "every NP & ADP". The model should be able to pick up a wide scope marker and relevant patterns of NPs and ADPs and insert intermediate phrase boundaries at appropriate locations. In the following text, relevant items at this last level of phrasing are underlined and numbered:

(63) Watakusi -no haha -wa L% mikan L% ringo L% momo L% <u>budoo -o</u> (12a) suupaa -de katte L% byooki -no titi -o L% byooin -ni mimatta
(64) Haha wa L% <u>ookina <></u> (11) byooin -no <u>kangofu -ga</u> (12b) titi -ni L% <u>totemo</u> (13) sinsetu d atta node L% kanojo -ni L% MI"KAN -o ageta.

5.3.3.3.1 Wide Scope (11)

(i) In the search for "wide scope" (cf. 5.2.4.3), the model picks up the wide scope marker, <>, at /ookina <> byooin no kangofu/ "big hospital POSS. nurse".

(ii) The pattern is picked up because it matches with the left-hand side of Rule 11 which specifies the insertion of a L%: i.e.,

(11) [<>] --> [<>] L%

(iii) The result is:

(*ookina*) L% (*byooin-no kangofu*)

The insertion of a L% after *ookina* "big" ascertains that the phrase *ookina byooin-no kangofu* does not mean "the nurse in the big hospital" but means "a big nurse in the hospital".

5.3.3.3.2 Every NP (12a)

(i) In the search of "every NP" (cf. 5.2.4.4), the model search for NPs whose pattern matches with the left-hand side of Rule 12: i.e.,

(12)	[N]	[POSTP]	[X]	-->	[N]	[POSTP]	L%	[X]
		‖ -Poss	‖ -V			‖ -Poss		‖ -V

(ii) In scanning NP patterns in the first sentence (63), until it reaches to /budoo o/ "grapes ACC.", the model ignores all the following NPs because none of them matches with the left-hand side pattern of Rule 12. (Underlines show mismatches.):

(a) [watakusi N] [no POSTP] [haha N] "I POSS. mother."
|| Poss

(b) [haha N] [wa POSTP] L% [mikan N] "Mother THEME oranges."
|| Theme

(c) [mikan N] [POSTP] L% [ringo N] "Oranges apples"
|| Null

(d) [ringo N] [POSTP] L% [momo N] "Apples peaches"
|| Null

(e) [momo N] [POSTP] L% [budoo N] "Peaches grapes."
|| Null

(iii) In the first sentence (63), the NP pattern which the model picks up is at /budoo o suupaa/ "grapes ACC. supermarket" which is tagged as:

[budoo N] [o POSTP] [suupaa N]

(iv) Because the pattern (i.e., a noun followed by a non-possessive proposition ([POSTP] || -Poss) which is, in turn, followed by a non-verbal ([X] || -V)) matches with the left-hand side pattern in Rule 12 which then replaces the matched pattern with its right-hand side pattern: i.e.,

(12) [N] [POSTP] [X] --> [N] [POSTP] L% [X]
|| -Poss || -V || -Poss || -V

(v) The results is:

budoo o L% *su'upaa* (*de*)

5.3.3.3.3 Every NP (12b)

In the continuation of the search for the "every NP" pattern, the model ignores such a NP pattern as:

(a) [suupaa N] [de POSTP] [katte V] "Supermarket LOC. bought and."

(b) [byooin N] [ni POSTP] [mimatta V] "Hospital LOC. visited."

The reason is that Rule 12 specifies that the item which follows a postposition cannot be a verb (|| - Verb) (cf, 5.2.4.4): i.e.,

```
(12)  [ N] [ POSTP]              [ X]  -->  [ N] [ POSTP]    L%   [ X]
            || -Poss/-Theme      || -V            || -Poss/-Theme  || -V
```

In the second sentence (64), it is at / kangofu ga titi / "nurse NOM. father" that the model picks up another "every NP" pattern. By virtue of Rule 12, a L% is inserted after the postposition /ga/. The result is:

kangofu ga L% *ti'ti* (*ni*)

5.3.3.3.4 Every ADP (13)

(i) In the search for an ADP (cf. 5.2.4.4), the model picks up an ADP pattern at /totemo sinsetu/ "very kind" which is tagged as:

[totemo AD] [sinsetu AJ]

(ii) The model picks up the above pattern because this pattern matches with the left-hand side pattern in Rule 13: i.e.,

```
(13) [ AD] ([ POSTP]) [ X]  -->  [ AD] ([ POSTP])  L%  [ X]
                      || -V                            || -V
```

(iii) By virtue of Rule 13, a L% is inserted after the ADP.

(iv) The result is:

totemo L% *si'nsetu*

Consequently, the following is the output of the operation at the level of {[+Obligatory] [+Variable Dependent] {[+Length] [Formal]}]}. (The bold faced boundary markers are newly inserted at this level.)

(65) *Watakusi-no ha'ha-wa* L% *mi'kan* L% *ringo* L% *momo* L% *budoo-o* **L%** *su'upaa-de katte* L% *byooki-no ti'ti-o* L% *byooin-ni mima'tta*

(66) *Ha'ha wa* L% *ookina* **L%** *byooin-no kangofu-ga* **L%** *ti'ti-ni* L% *totemo* **L%** *si'nsetu d atta no'de* L% *ka'nojo-ni* L% *MI'KAN-o ageta*

5.4 Concluding Remarks

In this chapter, employing basic components of Nespor and Vogel's (1986) intonational phrasing model, an application model was designed for Japanese to parse a sentence into intermediate phrases. Although it is primitive, the model can provide at least three different types of phrasings from a least parsed phrasing to an exhaustively parsed phrasing by way of applying, category by category, the syntactic and semantic factors which condition intermediate phrase boundaries.

The model is, in fact, designed to be flexible in that it makes it possible to create not only the above mentioned three types of phrasings but also other types of phrasing variations by changing the parameter values of the performance variables or by adding more performance variables or even by including in the model more syntactic and semantic factors which may condition an intermediate phrase boundary.

Once a type of intermediate phrasing is chosen and locations of intermediate phrase boundaries are specified for a given sentence, the accentual phrasing algorithm, developed in Chapter 3, parses intermediate phrases into accentual phrases by means of examining accentuations and accentual features of items in a two-word sized window-cursor. The cursor moves from the beginning of an intermediate phrase to the end of the phrase one word at a time. Whenever at least one of the two content words in the cursor is underlyingly (originally) accented, an accentual phrase boundary is inserted in between the words. Such an operation is supported by the finding, in one of the experiments, that an accentual phrase boundary is inserted between two content words if at least one of the words is underlyingly (originally) accented.

Furthermore, based on Pierrehumbert and Beckman's (1989) mechanism, discussed in Chapter 2, relevant tones are assigned to the accentual phrases. These tones are then scaled in transform spaces to have vertical spatial values. The scaled tones are linearly interpolated to draw a schematic pitch contour. Roughly speaking, the above are the operations which provide a speech synthesis program with natural pitch contours.

It must be made clear that to develop an actual speech synthesis program, there are at least a few more processes to be undertaken (not to mention technological aspects of the program). Obviously the studies on these processes, mentioned below, are beyond the scope of this thesis:

(i) In order to create accurate schematic pitch contours, it may further be required to implement the following three types of downtrends on the schematic pitch contours created by the current model; i.e., declination at the level of a sentence; final lowering at the domain which may be larger than an intermediate phrase; and micro, segmental effects.

(ii) Also, as briefly mentioned in Chapter 3, there must be a set of rules which evaluates tonal values so that once tones are assigned to the accentual phrases, transform values will be automatically assigned to the tones, instead of manually assigning transform values to these tones.

(iii) Also, there must be a device which specifies whether noun phrases of multiple modifiers are left-branching or right-branching so that different catathesis constants can be assigned to tones in the phrases to create different pitch contours, such as those seen in Chapter 4.

(iv) Also, nothing was mentioned in the thesis about the durational aspects of the tones including those of boundary tones. It will be necessary to develop a set of rules which determine which tones receive mora-sized durational values and which tones do not. Also, it may be required to study the durational characteristics of boundary tones to help create accurate schematized pitch contours.

In summary, a phrasing and pitch contour model was designed in this thesis as linguistic input to a speech synthesis program . The model needs to be implemented on the program to test its feasibility. Also, further studies on the above issues should be undertaken to develop the text-to-speech program as well as to better understand the characteristics of Japanese tone and intonation.

Endnotes

[1] This does not mean that this chapter accords with Nespor and Vogel's theoretical motivation. The simplicity required by the application model (i.e., linear phrasing) may not be able to respect their theoretical motivation (i.e., development of a hierarchical prosodic model).

[2] Because of the non-acoustic nature of the description in this chapter, I would like to assume that an intermediate phrase boundary is characterized by an insertion of a pause. In terms of phonetics, however, an intermediate phrase boundary is primarily characterized by non-propagation of the catathesis effect. Thus, in reality, there may be an intermediate phrase boundary which blocks the catathesis effect even though it may not be accompanied by a pause.

[3] It is quite unlikely that a beginner of Japanese will type in a sentence which has a complex syntactic structure.

[4] a. (Jane gave the book to Mary)
b. (Jane) (gave the book to Mary)
c. (Jane gave the book) (to Mary)
d. (Jane gave) (the book) (to Mary)
e. *(Jane) (gave) (the book to Mary)
f. *((Jane gave) (the book to Mary)
g. (Jane) (gave the book) (to Mary)
h. (Jane) (gave) (the book) (to Mary) (Selkirk 1984: 293)

[5] It seems what is meant here are phrases marked by classifiers: e.g., *Otokoga hitori onnaga hitori tatteiru* "One man and one woman are standing."

[6] It ought to be an argument-head relationship.

[7] As seen in Figure 1.2, in Pierrehumbert and Beckman's prosodic model, there is no (additional) intermediate phrase boundary inserted at the level of the intermediate phrase because the last accentual phrase boundary in an intermediate phrase functions as an intermediate phrase boundary. Thus, whenever, an intermediate phrase boundary is mentioned, it ought to be understood that it refers to the last accentual phrase in an intermediate phrase.

[8] The markers are general in the sense that they are not specific "patterns" in the algorithmic phrasing rules which are discussed later in the chapter (5.3.2.3).

[9] Cooper and Sorensen (1980), too, suggest that in Japanese, an utterance of a sentence tends to be more easily broken into intonational phrases than in English.

[10] The insertion of a L% between *Zo'o-ga* and *hana-o* may place contrastive prominence on the object, as *Zo'o-ga* L% *HANA-o itamema'sita*.

[11] It can also be [11 < morae].

[12] As will be seen (in 5.2.4.4), a L% is not inserted immediately before a verb.

[13] In regard to non-restrictive relative clauses, first, there is no distinctions in Japanese between restrictive and non-restrictive relative clauses in the phonological, morphological or syntactic form. Furthermore, unlike in English, it is not obligatory to single out intonationally a relative clause as a phrase: e.g., (Kuno 1973; 235)

> *Korega* (L%) *honyuu-do'obutu de aru* (L%) *kuzira* (non-restrictive)
> "This is the whale, which is a mammal."
> *Korega* (L%) *niho'n-kai ni su'nde iru* (L%) *kuzira* (restrictive)
> "These are the whales that live in the Japan Sea."

[14] The feature specification of "name" is required because there is a noun without a case-marking postposition which does not necessarily condition a L%; i.e., a noun used as a temporal adverbial, such as kyoo "today'.

[15] The possibilities of the combinations:
- i Subject [S-bar] Verb
- ii [S-bar] Subject Verb
- iii [S-bar] Subject /Indirect Object Verb
- iv Indirect Object [S-bar] Subject Verb
- v Indirect Object/Subject [S-bar] Verb

[16] Although somewhat heuristic, it may be possible to predict the locations of L%'s conditioned by some of embedded structures even by the linear parsing. In the examples (27, 28) of the embedded structures, an intermediate phrase is inserted immediately after each cyclic node. However, it is also true to say that a boundary is inserted immediately before the noun which follows a cyclic node. Thus, once nouns which occur in a series are identified, we can insert a L% before each of the nouns. To do so, first, search for nouns which have exactly the same postposition: e.g., (27) *ti'izu-o*, *nezumi-o*, *ne'ko-o*, and (28) *heya-no*, *tsukue-no*, *ho'n-no*. Then, insert a L% immediately before each of the nouns except for the last one.

[17] One problem in identifying the coordinated items by the postpositions they take is that there must be a way to differentiate the *to* and *ni* of the coordinating postpositions from the *to* of instrument and *ni* of dative, etc.. If there are more than one *to* or *ni* in a sentence, they may be regarded as the coordinating postpositions.

[18] The discussion of the "focus structure" in Japanese is beyond the scope of this paper. To understand the issue of the "focus structure", please refer to Selkirk 1984.

[19] If obligatory arguments are separated from a ditransitive verb by an optional argument, a L% should be before the argument which precedes a verb: e.g.,

> a. *Honda san-ni kinoo-no go'go* L% *hon'-o ageta*

b. *Ho'n-o Honda san-ni* L% *kinoo-no gogo ageta.*
"Yesterday, I gave a book to Mr Honda."

20 As will be seen, locating of these elements is done by way of pattern-matching.

21 A noun which is not followed by a postposition can either be a noun in vocative, a noun in a list, or a noun used as a temporal adverbial. Thus, they do not belong to the every NP category. Also, a noun followed by a *no*-postposition constitutes an AP. Thus, it is not a noun in the every NP category either.

22 Due to type-setting, a feature is represented horizontally

23 It is assumed that accentuation is performed at the level of accentual phrasing (discussed in Chapter 3).

24 As defined in Table 5.3, the notation "X" refers to an item of any syntactic category.

25 Admittedly I do not know the meaning and the function of the postposition/d(e)/.

26 It may require a following rule to modify the tag at [totemo]:
[AD] --> [MOD] / [AD] <<< [AJ] (AD becomes MOD when it immediately precedes AJ).

BIBILIOGRAPHY

Anderson, M. D., J. B. Pierrehumbert, and M. Y. Liberman (1984). "Synthesis by Rule of English Intonation Patterns." In *Proceedings of the (1984) IEEE International Conference on Acoustic, Speech, and Signal Processing,* 2.8.1-4.

Atal, B. S., and Hanauer, S. L. (1971). "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave." *Journal of the Acoustical Society of America* 50, 637-655.

Beckman, M. E. (1986). *Stress and Non-stress Accent*. (Netherlands Phonetic Archives 7.) Dordrecht: Foris.

Beckman, M. E., S. R. Hertz, and O. Fujimura (1983). "SRS Pitch Rules for Japanese." In *Working Papers of the Cornell Phonetics Laboratory* 1, 1-16.

Beckman M. E., and J. B. Pierrehumbert (1985). "Synthesizing Japanese Using a Downstep Model." In *Journal of the Acoustical Society of America* 77. s38.

Beckman M. E., and J. B. Pierrehumbert (1986a). "Intonational Structure in Japanese and English." *Phonology Yearbook* 3, 255-310.

Beckman M. E., and J. B. Pierrehumbert (1986b). "Japanese Prosodic Phrasing and Intonation Synthesis." In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 173-180. Assocation for Computational Linguistics, Morristown, New Jersy.

Bloch, Bernard (1950). "Studies in Colloquial Japanese: IV. Phonemics," *Language* 26, 86-125.

Browman, C. P., O. Fujimura, and E. L. Ohira (1982). "Demisyllablic Synthesis by Rule Using LINGUA." In *The Proceedings of the XIII-th International Congress of Linguists, Tokyo*, 1168-1170.

Clements, G. N., and K. Ford (1979). "Kikuyu Tone Shift and its Synchronic Consequences." *Linguistic Inquiry* 10, 179-210.

Cohen, A. R., and J. t' Hart (1967). "On the Anatomy of Intonation." *Lingua*, 19, 177-192.

Cooper, W. E., and J. M. Sorensen (1977). *Fundamental Frequency in Sentence Production*. New York: Springer-Verlag.

Crystal, David. (1985). *A Dictionary of Linguistics and Phonetics*. Oxford: Basil Blackwell.

Fujisaki, H., and K. Hirose (1980). "Acoustical Features of Fundamental Frequency Contours of Japanese Sentences." *Research on Human Information Processing Annual Report* 1, 37-47, University of Tokyo.

Fujisaki, H., and H. Sudo (1971). "A Model for the Generation of Fundamental Frequency Contours of Japanese Word Accent." *The Journal of the Acoustical Society of Japan* 27, 445-452.

Fujisaki, H., and M. Sugito (1978). "Analysis and Perception of Two-Mora Word Accent Types in the Kinki Dialect." *The Journal of the Acoustical Society of Japan* 34, 167-176.

Gallaiford, Neil (1983). *Pitch Declination in Japanese: An Instrumental Study*. M.A. Thesis, University of Victoria.

Hakoda, K., and H. Sato (1980). "Bun Onsei Gosei ni okeru Onchoo Kisoku (Prosodic Rules in Connected Speech Synthesis.)" *Densi Tuusin Gakkai Ronbunsi* 63-D, 715-722.

Haraguchi, Shosuke (1977). *The Tone Pattern in Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

Hashimoto, Shinkichi (1959). *Nihon Bunpoo Taikeiron* (The Structure of Japanese Grammar). Tokyo: Iwanamishoten.

Hayes, Bruce (1981). *A Metrical Theory of Stress Rules*. Doctoral Dissertation, MIT.

Higurashi, Yoshiko (1983). *The Accent of Extended Word Structures in Tokyo Japanese*. Tokyo: EDUCA.

Hirayama, Teruo (1960) *Zenkoku Akusento Jiten* (All Japan Accent Dictionary). Tokyo: Tookyoodoo.

Kawakami, Shin (1961). " On the Relationship between Word-Tone in Japanese Language [sic]." *Onsei no Kenkyuu* 9, 169-177.

Kawakami, Shin (1962). "Pitchiguramu de Mita Nihongo no Akusento (Japanese Accent Observed by the Aid of the Pitchgraph)." *Onsei no Kenkyuu* 10, 115-129.

Kawakami, Shin (1966). "Taigen ni Tsuku Ippaku no Joshi no Akusento (The Accentuation of Monosyllablic Postnominal Particles)." *Onsei no Kenkyuu* 12, 239-253.

Kenstowitz, M., and C. Kisseberth (1979). *Generative Phonology*. New York: Academic Press.

Kindaichi, Haruhiko (1967). *Nihongo Onin no Kenkyuu* (Study of Japanese Phonemes). Tokyo: Tokyoodoo Shuppan.

Klatt, Dennis H. (1987). Review of Text-to-Speech Converstaion for English." *Journal of the Acoustical Society of America* 82. 737-793.

Kori, Shiro (1987). "The Tonal Behavior of Osaka Japanese: An Interim Report." In *Paper from the Linguistics Laboratory*, 31-61. (Working Paper 36.) Department of Linguistics, Ohio State University, Columbus.

Kubozono, Haruo (1985). "On the syntax and prosody of Japanese compounds." In *Work in progress, Department of Linguistics, University of Edinburgh* 18, 60-87.

Kuno, Susumu (1973). *The Structure of Japanese*. Cambridge: MIT Press.

Ladd, Robert D. (1983). "Phonological Feature of Intonational Peaks." *Language* 59, 721-759.

Lehiste, Ilse (1970). *Suprasegmentals*. Cambridge, Massachusetts: The MIT Press.

Liberman, M. , and J. B. Pierrehumbert (1984). "Intonational Invariance under Changes in Pitch Range and Length." In M. Aronoff and R.T. Oehrle, eds., *Language Sound Structure: Studies in Phonology Presented to Morris Halle*, 157-233. Cambridge, Massachusetts: The MIT Press.

Liberman, M., and A. S. Prince (1977). "On Stress and Linguistic Rythm." *Linguistic Inquiry* 8, 249-336.

Martin, Samuel E. (1975). *A Reference Grammar of Japanese*. New Haven: Yale University Press.

McCawley, James D. (1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

McCawley, James D. (1970). "A Note on Tone in Tiv Conjugation." *Studies in African Linguistics* 1, 123-129.

McCawley, James D. (1977). "Accent in Japanese". In L. Hyman, ed., *Studies in Stress and Accent*, 261-302. (Southern California Occasional Papers in Linguistics 4.) Department of Linguistics, University of Southern California, Los Angeles.

Nespor, M., and I. Vogel (1986). *Prosodic Phonology*. Dordrecht: Foris.

NHK (1966). *Nihongo Hatsuon Akusento Ziten* (Dictionary of Japanese Pronunciation and Accent). Tokyo: Nippon Hoosoo Kyookai Shuppan.

Ohman, S., and J. Lindqvist (1966). "Analysis-by-Synthesis of Prosodic Pitch Contours." In *Speech Transmission Laboratory Quartely Progress and Status Report* 4, 1965, Royal Institute of Technology, Stockholm.

O'Shaughnessy, Douglas (1979). "Linguistic Features in Fundamental Frequency Patterns." *Journal of Phonetics* 7, 119-145.

O'Shaughnessy, D., and J. Allen (1983). "Linguistic Modality Effects on Fundamental Frequency in Speech." *Journal of the Acoustical Society of America* 74, 1155-1171.

Pierrehumbert, Janet B. (1980). *The Phonology and Phonetics of English Intonation*. Doctoral Dissertation, MIT.

Pierrehumbert, J.B., and M. E. Beckman (1989). *Japanese Tone Structure*. Cambridge, Massachusetts: The MIT Press.

Poser, William J. (1984). *The Phonetic and Phonology of Tone and Intonation in Japanese*. Doctoral Dissertation, MIT.

Pulleyblank, Douglas (1983). *Tone in Lexical Phonology*. Doctoral Dissertation, MIT.

Schachter, P., and V. Fromkin (1968). A Phonology of Akan: Akuapem, Asante and Fante. In *Working Papers in Phonetics*, 9. University of California

Selkirk, Elisabeth O. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Massachusetts: The MIT Press.

Suzuki, K. and R. Teranishi (1988). "Nihongo Onsei Kisoku Goosei no tame no Koobun Kaiseki (Parsing Algorithm for Text-to-speech in Japanese)." *The Journal of the Acoustical Society of Japan* 44, 335-343.

Warkentyne, Henry J. (1978). "Information Focus and Primary Accent." *Onsei no Kenkyuu* 18, 217-222.

Warkentyne, Henry J. (1984). "The Effects of Subject Thematization on the Tonal Pattern of Japanese Sentence." *Working Papers of the Linguistics Circle* 4, 114-126. Department of Linguistics, University of Victoria.

APPENDIX
STIMULI FOR EXPERIMENT 4.3

The stimuli used for Experiment 4.3 are listed together with those used for Experiment 4.2. Only the replaced items in Experiment 4.3 are shown.

I. Right-branching Phrases:

A21	*a'wa-no ya'bona a'ni*	"clumsy brother in Awa."
A22	.... *ane*	"............ sister ....".
A31	*mi'nami-no a'warena mi'nari*	"Minami's pitiful appearance".
A32	.... *emono*	"....................................... catch".
B21	*a'wa-no a'ji-no na'be*	"fish pot of Awa".
B22	.... *mai*	"...... dance in ....".
B31	*a'rima-no a'wabi-no o'iru*	"abalone oil from Arima".
B32	.... *nimono*	"............ food ....".
C21	*ya'bona a'wa-no a'ni*	"clumsy brother in Awa".
C22	.... *ane*	... "............ sister ....".
C31	*a'warena mi'nami-no mi'nari*	" pitiful Minami's appearance".
C32	.... *emono*	"..................................... catch".

II. Left-branching Symmetrical Phrases:

A'21	*ya'bona a'ni-no o'no*	"clumsy brother's ax".
A'22	.... *ane-no*	"............... sister's ......".
A'31	*a'warena mi'nami-no mi'nari*	"pitiful Minami's appearance".
A'32	.... *minae-no*	" .... Minae's ...."
B'21	*a'wa-no a'ni-no o'no*	"brother in Awa's ax".
B'22	.... *ane-no*	"sister .... ".
B'31	*a'ruje-no a'wabi-no one'dan*	"Algerian abalone's price".
B'32	.... *anago-no*	"................. eel's ....".
C'21	*ya'ya moro'i ama'do*	"somewhat fragile shutter".
C'22	.... *omoi*	"................. heavy ....".
C'31	*iya'ni manuru'i yama'uri*	" awkwardly warm wild-melon".
C'32	.... *mearai*	"...................... coarse ....".

III. Left-branching Relative Clause Phrases:

D21	*e'mi-no nu'ida bu'ra*	"brassiere which Emi took off".
D22	.... *monda*	"........................... hand-washed".
D31	*mi'nami-no no'nda nomimo'no*	"drink which Minami had".
D32	.... *yonda*	"............................... ordered".
E21	*e'ngi-no ya'bona a'ni*	"brother who play is clumsy".
E22	.... *kurai*	"brother whose knowledge of the play is poor
E31	*mi'nari-no a'warena mi'nami*	"Minami whose appearance is pitiful".
E32	.... *ayasii*	".. dubious".

F21	*ya'bona* <u>*e'ngi-no*</u> *a'ni*	"brother whose <u>play</u> is odd".
F22	.... <u>*shoogi-no*</u>	"......................... <u>chess-play</u> ...".
F31	*a'warena* <u>*mi'nari-no*</u> *mi'nami*	"Minami whose <u>appearance</u> is pitiful".
F32	.... <u>*minoue-no*</u>	"........................ <u>circumstance</u> ....".

VITA

Surname: Miyamoto Given Names: Tadao

Place of Birth: Sakai, Osaka Date of Birth: February 9, 1948

Educational Institutions Attended:

Momoyama Gakuin Daigaku, Osaka	1966 to 1970
University of Victoria	1982 to 1987
University of Victoria	1987 to 1989

Degrees Awarded:

B.A.	1970	Momoyama Gakuin Daigaku, Osaka
B.A.	1987	University of Victoria

Honours and Awards:

British Columbia Science Council G.R.E.A.T (Graduate Research Engineering And Technology) Award, 1988-1989

Publications:

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying if this thesis for scholarly purpose may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis

Modelling Tone and Intonation in Japanese

Author [redacted]
(Signature)

TADAO MIYAMOTO
(Name)

September 13, 1989
(Date)