

**An Algorithm for the Assignment of Lexical  
Stress in Converting Text to Speech**

Suzanne C. Urbanczyk  
Bachelor of Science, University of Victoria, 1985

A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of


Linguistics

We accept this thesis as conforming  
to the required standard

  
Henry J. Warkentyne, Ph.D.

  
Stephen J. Eady, Ph.D.

  
John H. Esling, Ph.D.

  
R. Lynn Kirilin, Ph.D.

  
John J. Tucker, Ph.D.

© Suzanne C. Urbanczyk, 1989

UNIVERSITY OF VICTORIA

1989

All rights reserved. This thesis may not be reproduced  
in whole or in part, by mimeograph or other means,  
without permission of the author.

ACCEPTED

FACULTY OF GRADUATE STUDIES

DATE

Sept 08, 1989

DEAN

Supervisor: Dr. Warkentyne and Dr. Eady

## ABSTRACT

This thesis describes the development and testing of an algorithm to assign main stress to English words. It has been developed for inclusion in a text-to-speech conversion system. Within the field of speech technology, the correct assignment of lexical stress is important in ensuring that the correct rules of prosody are applied to the speech output. Some of the problems associated with word stress in English have been addressed, including aspects of the morphological structure of English and a quantity-sensitive stress rule. The algorithm developed is as accurate or more accurate than other stress-assigning algorithms reported on in the literature.

During the evaluation of the algorithm, some observations have been made that may have future implications for theories of lexical stress assignment for English.

Examiners:

[REDACTED]

Henry J. Warkentyne, Ph.D.

[REDACTED]

Stephen J. Eady, Ph.D.

[REDACTED]

John H. Esling, Ph.D.

[REDACTED]

R. Lynn Kirlin, Ph.D.

[REDACTED]

John J. Tucker, Ph.D.

## CONTENTS

<b>Abstract</b> .....	<b>ii</b>
<b>Contents</b> .....	<b>iv</b>
<b>Figures</b> .....	<b>vi</b>
<b>Acknowledgements</b> .....	<b>vii</b>
<b>Introduction</b> .....	<b>1</b>
<b>The Problem: Automatic Stress Assignment</b> .....	<b>1</b>
Importance of Correct Stress Assignment .....	2
Problems of Lexical Stress Assignment in English .....	4
<b>Previous Approaches</b> .....	<b>6</b>
Hunnicuttt .....	6
Nessly .....	10
Church .....	12
<b>Problems with Previous Approaches</b> .....	<b>14</b>
Accuracy .....	14
Computational Efficiency .....	17
Dialect .....	17
<b>Proposed Solution: Morphological Emphasis</b> .....	<b>18</b>
Advantage to Morphological Decomposition .....	18
Influence of Affixes on Stress Patterns .....	20
Previous work with Morphological Approach .....	23
Proposed Approach .....	24
<b>The Algorithm</b> .....	<b>27</b>
Preliminaries to the Algorithm .....	27
Description of the Algorithm .....	28
Basic Stress Rule .....	29
<b>Affixes and Stress</b> .....	<b>31</b>
Stress-Neutral Suffixes .....	32
'Autostressed' Suffixes .....	32
'Pre-Stressed' Suffixes .....	33
'Mixed' Suffixes .....	34
Prefixes .....	35
<b>Affix Stress Rules</b> .....	<b>36</b>
<b>Examples</b> .....	<b>37</b>
Example 1: LITTLE .....	38
Example 2: SOVIET .....	38
Example 3: ALASKA .....	39
Example 4: HIMSELF .....	39
Example 5: POSSIBLE .....	40
Example 6: WITHOUT .....	40

Examples 7 and 8: DESIGN and DESIGNATION .....	41
Summary .....	42
<b>Evaluation .....</b>	<b>43</b>
Evaluation .....	43
Hunnicutt .....	44
Bernstein and Nessly .....	45
Church .....	46
McCallister .....	47
Current Methodology .....	47
Level of Accuracy .....	49
Error Analysis .....	50
Results .....	50
Disyllabic Words .....	50
Trisyllabic Words .....	54
Polysyllabic Words .....	56
Generalizations .....	57
Improvements to Algorithm .....	58
Implications for Future Theoretical Work .....	59
Summary .....	60
<b>Conclusion .....</b>	<b>61</b>
The Algorithm .....	61
Further Developments .....	62
Linguistic Insights .....	64
Summary .....	66
<b>Bibliography .....</b>	<b>67</b>
<b>Appendix A: Practical Orthography .....</b>	<b>70</b>
<b>Appendix B: Affixes .....</b>	<b>72</b>
Suffixes .....	72
Prefixes .....	81
<b>Appendix C: Corpus .....</b>	<b>83</b>
Brown Corpus .....	83
Running Text .....	93

## FIGURES

1. Flow-chart of Lexical Stress Component . . . . . 26

## ACKNOWLEDGEMENTS

Throughout the research and writing of this thesis Steve Eady has provided more encouragement and guidance than one could ever ask for. I should like to express my deepest gratitude to Steve for all the patience and help that he has provided me with. I would also like to thank the Science Council of British Columbia for providing the financial assistance that has supported me over the past two years. My other committee members, Dr. Warkentyne, Dr. Esling and Dr. Kirilin have been very helpful in providing comments and taking the time to read the thesis.

Among those at CSTR who I have worked with over the last four years I would like to thank Craig Dickson, Jocelyn Clayards, Alan Wynrib, Jim Woolsey and Tybring Hemphill.

And of course my fellow grad students, for their friendship, support and sharing of ideas and generally making the experience of studying Linguistics at UVic a pleasant and happy environment. Thank-you Andrea, Pannakitti, Jennifer, Erica, Evan, Lin Hua, Kimiko, Michael, Laura and Xue Ping.

I would like to thank my good friend David Merino and my family for understanding what I was doing without understanding. They are not too sure what exactly it is that I do, but support it because they know it is what I want to do.

Finally, I would like to thank Dawn Bates for introducing me to current phonological theory and encouraging me to continue, and Geoff O'Grady and Barry

Carlson for showing me the devotion and care that must be taken when studying a language first-hand. I would also like to thank Tom Hukari because he has ensured that many of us will be aware of the alternative theories in syntax.

And once again a huge thank-you to Steve Eady, for making the completion of this thesis a rather painless experience.

## INTRODUCTION

Text-to-speech conversion is the process involved in proceeding from unrestricted text as input to natural sounding synthetic speech as output. This complex task can be divided into two general stages. The first stage is to convert unrestricted text to some form of abstract linguistic representation. The second stage converts this representation to speech output. These two general stages can be further subdivided into various components. The focus of this work is to discuss one aspect of the first stage of text-to-speech conversion: the assignment of lexical stress to English words. The purpose of this research is to develop an algorithm that can be used for automatic stress assignment in a working text-to-speech system.

### The Problem: Automatic Stress Assignment

The linguistic feature of stress, in English, can be defined in a number of ways. It can be described most generally as a relationship between the syllables of a word. Specifically, stress is defined as the relative prominence of a syllable in a word. The term "stress" is sometimes used to refer to the relative prominence of other linguistic items (such as words in a sentence). This thesis will restrict its discussion to word-level, or lexical stress.

The domain of lexical stress is the word. While there can be several levels of stress within a word, (primary, secondary and unstressed) the major concern here

is to develop an algorithm for the location of the primary stressed syllable. For example, in the word elementary, the primary stress is the syllable which is most prominent perceptually (in this case, the third syllable ment). The first syllable also receives some emphasis, but this is not as great. This is the secondary stressed syllable. The remainder of the syllables, the second, fourth, and fifth are all unstressed.

Developing algorithms for the correct assignment of stress in English has posed several problems to both man and machine. A number of factors contribute to the complexity of the English stress system. However, recent developments in phonological theory have added knowledge to our understanding of English word-stress.

### **Importance of Correct Stress Assignment**

Determining the stressed syllable for words is important, because there are many acoustic correlates of stress in English. In English, stress is known to affect the fundamental frequency, the duration of the syllables, the intensity of syllables, as well as the vowel quality (Fry, 1955, 1958, 1965.)

In a series of experiments conducted between 1955 and 1965, Fry studied four phonetic and acoustic correlates of stress: intensity, duration, fundamental frequency and the formant frequencies of the vowel. Fry synthesized homographic words with alternations in stress, such as object, digest, and permit. The four acoustic parameters were adjusted and listeners were asked to make judgements on which form of the word they perceived. He found that a change in the fundamental frequency was the most significant cue to the perception of stress. Duration and intensity were less significant, and Fry proposed that the

formant structure of the vowels may be even less effective than the intensity cue in the perception of stressed syllables.

While the formant structure may not be an important perceptual cue for the detection of the stressed syllable, it is important when modeling speech production. When syllables are stressed, the target vowel is realized in the acoustic output. However, when a syllable is unstressed, the vowel quality is often reduced to that of a neutral vowel. For example in the words element and elementary there is similar phonological information. The difference lies in the addition of a suffix which causes a shift in the location of the primary stressed syllable in the word. In "element" the syllable represented as "ment" is unstressed and the vowel is realized as a reduced vowel. However when the primary stress is shifted to this syllable, the vowel is realized as the front mid lax vowel /ɛ/.

The study of the acoustic properties of stress has been of continuing concern in linguistic research. While much is known, much is still to be discovered about the complex relationship between the various acoustic and phonetic factors. It would seem that the location of stress in words would have a great importance when prosodic features such as intonation and duration are to be modeled in synthetic speech. It has been shown that the comprehensibility of synthetic speech is greatly increased when the prosodic components of a language are adequately modeled (Klatt, 1987).

## **Problems of Lexical Stress Assignment in English**

English has often been referred to as a sort of "mongrel" language. This label comes from its diverse ancestry. The lexicon and indeed some aspects concerning the structure of English are the result of millenia of borrowings from several languages, most notably the Romance and Germanic languages. Subsequently, the stress patterns of English are not as straight forward as those of other languages. Stress patterns in English are influenced by the phonological, morphological and etymological properties of the word (Church, 1985 and Woods, 1976.)

There can be two approaches to assigning lexical stress by computer. One is to have a large dictionary containing all the words necessary for the system. This poses problems when one is restricted in terms of storage space, as the English language contains over one hundred thousand words. A further drawback to this approach occurs when a word is encountered which is not in the dictionary. It has been suggested that the dictionary approach has an inherent limitation of 60% coverage when handling proper names (Church, 1985.)

Another approach is to generate the stress patterns based on rules. There are disadvantages with this method as well; for example, when an exceptional word is encountered (i.e., a word that does not follow the general pattern). Many loan words are included in this category. A rule-based approach is the more efficient approach as well as the linguistically more meaningful approach. The approach being adopted here, as in many other systems, will be to develop a rule for stress-assignment, combined with the use of a small exceptions dictionary.

The first problem that one encounters is in developing rules to convert the orthographic representation of English to a phonological representation

(letter-to-sound rules). In some instances the spelling of English reflects an earlier pronunciation of the language (e.g. ove in the words love, move, and clove). Other problems arise because there are features of English orthography that cannot be characterized by a single rule. That is to say that something represented orthographically may depict more than one item in the language.

An example of this is the case of homographs. Homographs are words which are spelled the same but belong to more than one part of speech. They are occasionally accompanied by differing pronunciations. This varying pronunciation may be realized in a stress alternation or the selection of a different vowel quality. Examples of these are the noun SUBject and the verb subJECT, and the past and present tense of read, /rɪd/ or /rɛd/. The treatment of homograph pairs has been of continuing interest in linguistic theory since the work of Chomsky and Halle (1968). With the amount of space devoted to this problem, one would assume that stress-differentiated homographs were quite productive in English.

In fact Sherman (1975) has provided an interesting account of this phenomenon, by taking a historical look at the development of noun-verb stress alternations. He examines noun-verb homograph pairs historically by tracking the occurrence of the stress-differentiated pairs from 1570 to 1934. The source of his information comes from a diverse selection of dictionaries that have preserved, or in some way indicated the stress patterns of words. The main points of interest that Sherman makes concern the distribution of homograph pairs which are differentiated by stress. There has been a gradual increase in the number of bisyllabic noun-verb stress-differentiated pairs from three in 1570 to 150 in 1934. When one considers that there are nearly 2,000 polysyllabic homograph pairs in the

English language, and of these only 220 are differentiated by stress (bisyllabic and polysyllabic words), this is not a very large percentage. In fact, Sherman (1975) calculated that only 15% of polysyllabic homograph words are differentiated by stress. If one were to calculate the percentage in terms of the entire number of words in the English language, or to restrict it to the 74,000 entries in the Shorter Oxford English Dictionary, one would end up with a paltry 0.3%. While this is indeed a trifling detail, it is important in the design of a text-to-speech conversion system. The trivial amount of homograph pairs differentiated by stress does not merit a rule system of its own, rather a slightly different method of analysis, namely a syntactic analysis.

### **Previous Approaches**

Information concerning the details and development of text-to-speech conversion systems which is available to the public is not abundant due to the proprietary nature of the research. There are three basic approaches that will be examined in this paper. These include the work of Hunnicutt(1976), Bernstein and Nessly(1981), and more recently Church(1985). In order to discuss the advantages and short-comings of these systems it is necessary to give a general description of each and how they are presented within a text-to-speech conversion system.

#### **Hunnicutt**

Possibly one of the most extensively documented systems developed was the MITalk system at the Massachusetts Institute of Technology (MIT). The approach taken by this group of researchers was to model or in some manner imitate the process of how humans read.

Jonathan Allen began work on the conversion of text-to-speech in the early 1970's. The first task was to analyze the textual representation and extract from this a suitably detailed linguistic representation that would be capable of driving the synthesis-by-rule mechanisms (Allen, 1973). This approach was undertaken in an attempt to model certain aspects of the human reading process (Allen, 1973). The work of Allen on the conversion of text-to-linguistic representation focussed on two domains: letter-to-sound rules and lexical stress assignment. The task of stress assignment was also divided into two components: prediction of stress based on a morphological decomposition, and the assignment of stress to monomorphemic words based on phonological principles.

Allen's discussion concentrates on rules developed to account for the appropriate morpheme decomposition. For example, strategies were used to account for the decomposition of the word formally into the morphemes formal+ly as opposed to form+all+y. Allen outlines a procedure for analysis that is given in the following brief description.

Text is first converted to a standardized orthographic representation (i.e. any numbers or abbreviations are spelled out). There is then a morphological decomposition which is based on a dictionary look-up combined with rules for grammatical morpheme combinations. If this is successful, then stress is assigned based on rules for polymorphemic words. If there can be no morpheme decomposition, the word is converted to a phonemic representation based on the letter-to-sound rules. The monomorphemic lexical stress rules are then applied. A final step, before the phonetic transcription is obtained, is the application of morphophonemic rules. Morphophonemics is "the study of the phonemic

differences between the allomorphs of the same morpheme, thus concerning the phonemic shape of the allomorphs (Carroll)" Mario Pei (1966.) Therefore the morphophonemic rules which Allen is referring to are those rules which would consider the different phonetically conditioned forms of a morpheme. An example of this would be the past tense morpheme, which is often realized as /əd/ following alveolar stops (e.g., baited /betəd/). It is realized as /d/ following other voiced consonants (e.g., begged /bɛgd/), and /t/ after voiceless consonants, except /t/, (e.g., baked /bekt/).

Allen incorporates two methods of lexical stress assignment. First, modern phonology at that time included the English stress rules proposed by Chomsky and Halle in their landmark work The Sound Pattern of English in 1968 (herein referred to as SPE). Allen discusses the adaptation of the SPE Main Stress Rule for the purpose of lexical stress assignment. Secondly, he makes note of certain suffixes which influence the stress of a word. In particular, Allen makes note of suffixes which attract main stress such as -ee, -eer, -ade, and -self, in the words referee, engineer, parade, and himself. He also makes note of suffixes which place the main stress on the previous vowel. Suffixes of this nature include -ic and -ial, as in the words sporadic, and confidential. The treatment of a morphological influence on the stress pattern of English is, as in SPE, kept to a minimal level, with most of the stress assignment performed by extensive rules.

Sharon Hunnicutt continued the work which Jonathan Allen had started on the MIT system. In Hunnicutt's system, most of the conversion of text-to-phonemes continued to be carried out by a morphological decomposition. Words were decomposed into their constituent 'morphs'. Morphs are defined as being "the

surface realization of underlying morphemes" (Allen et. al. 1987, p. 26). The system which Hunnicutt worked on required a morph lexicon with over 11,000 entries. This morph lexicon contains information reflecting the phonemic representation as well as part(s) of speech. The morph lexicon is quite extensive, consisting of free roots, bound roots, and affixes. It is claimed that a morph dictionary containing 10,000 entries is sufficient to account for or decompose virtually all the words of English (Klatt, 1987.)

While there was a great deal of emphasis placed on the use of a morph dictionary, the main component of the lexical stress assignment was an extensive ordering of rules based on a modified SPE framework. These rules include the cyclic Main Stress Rule, Stressed Syllable Rule and Alternating Stress Rule. The non-cyclic rules consist of the Destressing Rule, Compound Stress Rule, Strong First Syllable Rule, Cursory Rule and Vowel Reduction Rules. Not all of these rules operate on a strictly phonological level, as some of them include conditions to check for suffixes that influence the stress of the word.

The morphological analysis developed by Hunnicutt is quite extensive, based on a large morph dictionary. One of the principle motivations for the large dictionary, and the extensive analysis is to derive the underlying root form of a word. Spelling recovery rules are also incorporated to restore the original orthographic form of the root. This is necessary in some instances when words such as coping have lost a letter due to orthographic conventions. If the morpheme boundary is inserted without reconstructing the root to cope the letter-to-sound rules in this system would operate on the form cop and thus produce an incorrect transcription. The lexical stress rules are then applied on this recovered root form.

The application of the stress rules proceeds in two phases (Hunnicut, 1976). The first phase is cyclic and involves the application of ordered rules to the root first, then to the root and closest suffix combined. This procedure is repeated, picking up suffixes until primary stresses have been located in the entire word. The second phase is non-reiterative and involves the reduction of all but one primary stress to secondary and zero stressed syllables. This is a necessary stage because the appropriate location of secondary stress and zero stressed syllables is an important step prior to the vowel selections (Hunnicut, 1976.)

Hunnicut provides a derivation of the stress determination for the word multinucleolated. The assignment of stress involves no less than ten steps, with the Main Stress Rule being applied more than once. This is clearly an attempt to generate the stress of a word based on its underlying representation, as opposed to its surface representation.

### Nessly

Bernstein and Nessly (1981) attempted to determine if computationally complex algorithms for the conversion of text-to-phonemes were more accurate in assigning lexical stress than less complicated algorithms. An experiment was designed to test the accuracy of a computationally simple algorithm against the more complex system developed by Hunnicutt. The algorithm that was used (Nessly's default rule) was proposed by Hill and Nessly (1973) in a review article of SPE, in which they attempt to supply a somewhat different account of the stress pattern of English. A key to the operation of the Main Stress Rule in SPE is whether a syllable is strong or weak. In SPE, strong syllables end in either a tense or long vowel (e.g., /i/, /o/, /u/) or a vowel followed by at least one consonant. On the other hand weak syllables typically end in open lax vowels. Chomsky and

Halle (1968) use the criteria of strong and weak syllables to determine if a syllable should receive primary stress. If the penultimate syllable is strong, it is stressed, if weak, the antepenultimate syllable is stressed. Contrary to the claims set forth in SPE, Hill and Nessly argue that, with a few scattered exceptions, one is able to assign stress accurately without reference to vowel tensity (Hill and Nessly, 1973.) Their claim is that what constitutes a strong syllable is based on the presence of a consonant on the coda of a syllable. The coda of a syllable is usually defined as the final portion of a syllable (Clements and Keyser, 1983).

Nessly's default rule is a very simplified version of the first approximation of the Main Stress Rule proposed in SPE. It differs in that no word class information is used and that Nessly's definition of a 'strong cluster' is indifferent to vowel length or tensity (Bernstein and Nessly, 1981).

Nessly's default stress rule can be stated in very simple terms:

1. Always stress monosyllabic words.
2. Stress the first syllable of disyllabic words.
3. If there are more than two syllables in the word, and if the penultimate syllable in the word contains a final consonant, (i.e. it is closed,) stress it.
4. Otherwise, stress the antepenultimate syllable.

This stress rule is quite simple and was found to perform at the same level of accuracy as the much more elaborate Hunnicutt rule system (Bernstein and Nessly, 1981). This was achieved without the analysis of words into morphemes and with reference only to the structure of the syllable.

## Church

The approach that Church takes differs from the previous two approaches in a number of ways. Advances made in phonological theory as well as computational techniques have helped to contribute to Church's quite eloquent solution to the problem of lexical stress assignment.

Church approaches the task of lexical stress assignment or "orthography -> stress," as he terms it, by first examining the notion of "syllable weight." This is a key to the system developed by Church. He uses weight as an intermediate level of representation in the stage between orthographic input and output stress (Church, 1985). He thus divides the conversion of orthography to stress into the two following sub-components: orthography to weight, and weight to stress.

The assignment of stress from syllable weight is similar to the approach developed by Hayes (1980) in regards to his treatment of the main stress rule and extrametricality. The weighting system was based on Chomsky and Halle's notion of strong and weak clusters (Church, 1985). Heavy syllables are closed (i.e., end in consonant clusters) or contain underlying long vowels. The final syllables of the following words are all heavy: obey, maintain, erase, torment, collapse, and exhaust. Light syllables end in open short vowels, such as the final syllables in the following words: consider, develop, astonish, edit, and promise. Each syllable is assigned a weight of either heavy or light. The treatment of word-final consonants is extrametrical, so each word is stripped of its final consonant prior to the determination of syllable weight.<sup>1</sup>

---

<sup>1</sup> The term extrametrical was first introduced by Liberman and Prince (1977) and later extended by Hayes (1982). This formal device is used to refer to a linguistic item that is "ignored by the stress rules" (Hayes, 1982).

Church developed a powerful technique for determining lexical stress. He utilized a table of possible stress assignments for words up to eight syllables in length, based on the weights of the syllables. The table that Church uses for stress assignment was developed by Sproat and, owing to its highly constrained nature, it offers a complete solution when going from weight to stress (Church, 1985.) The methods of going from orthography to weight are more complex, and Church devotes much time to this in his report.

Church's approach also involves a decomposition of the words into affixes. He distinguishes two types of affix. Level 1 affixes are of a Latinate etymology, including the following prefixes and suffixes: in+, ir+, ac+, +ity, +ion, +ive and +al. Level 2 affixes include Greek and Germanic affixes. The following are some examples: hetero#, un#, under#, #ness and #ly. Church details the motivations for dividing the affixes into these two categories.

Many generalizations about the morphological and phonological behavior of these two types of affixes are made. Level 1 affixes are found inside the scope of level 2 affixes, but not vice versa. Level 2 affixes must always attach to words or free roots, but level 1 affixes may attach to fragments or bound roots. These two points are of value because they also provide the system with generalizations that can be made on the nature of word-formations and valid morphological decompositions.

Of special interest to the application of stress-assignment are the following generalizations. Level 1 affixes accompany alternations in the stress pattern, while level 2 affixes are typically stress neutral. Further phonological rules apply with the addition of level 1 but not level 2 affixation. That is to say that the

domain of some phonological rules is at level 1 affixation but not level 2 affixation. Level 2 affixes behave as compounds in the prediction of stress. Since Church takes a metrical phonology framework as the basis of his English stress rules, he also makes the assumption that the treatment of some affixes as extrametrical accounts for the homographic words which are differentiated by stress, such as SUBject (noun) and subJECT (verb).

Church claims that his system is important because it addresses the following factors that figure in the assignment of stress: syllable weight, part of speech, morphology and etymology.

### Problems with Previous Approaches

As yet, no system for predicting English stress has been developed that is correct 100% of the time, or that can operate with the same efficiency as a human processor. Typically the shortcomings of previous systems have been in the areas of accuracy and complexity of the algorithms. These two areas will be addressed in this section.

### **Accuracy**

The accuracy of text-to-speech systems can be determined by a number of methods. One can test for the intelligibility of the speech output. The accuracy of the assignment of lexical stress is then confounded with the quality of speech, and other components of the text-to-speech conversion, especially the rules of prosody. For instance, if the duration and intonation rules are not well developed, then it is not an easy task to determine which are the stressed and unstressed syllables. It is therefore important to examine the output of the text-to-phoneme conversion rules before further linguistic rules are applied to this representation.

Often there are two components that operate together in the conversion of text to phonemes: the letter-to-sound rules and the lexical stress assignment rules. The letter-to-sound rules are important in influencing the lexical stress assignment rules because they provide information about syllable boundaries which must be known prior to the assignment of stress.

The interaction between the letter-to-sound rules and the stress assignment rules is quite apparent and there is often a close relationship between the two sets of rules. For example there has been some debate as to which component should have priority. It is clear that the stress-assignment must wait until the syllable boundaries have been determined. As there have been no substantial rules for the determination of syllable boundaries from simple orthographic representation, the letter-to-sound rules must be ordered first. However, when selecting between a target vowel and a reduced vowel, (e.g., the third syllable of eleMENTary versus that of eleMENT), one must have information on the stress of the syllable to know if the vowel is to be adjusted in any way. Some ways in which a vowel may be adjusted are: change in vowel quality, (i.e., changes in the formant structure of the vowel,) reduction of the vowel to schwa, and vowel elision. (i.e., vowel deletion). An example of this phenomenon is the change that occurs in the second syllable in the words explain, explanatory and explanation. This rule is known as Trisyllabic Shortening (TSS) because in certain affixation processes, the antepenultimate vowel (third to the last) is reduced in quality. Some letter-to-sound rules are capable of handling this alteration of vowel quality without reference to syllable stress. The present version of letter-to-sound rules in place at the Centre for Speech Technology Research is capable of handling

alternations in vowel quality such as in the words supreme and supremacy, without reference to lexical stress.<sup>2</sup>

Often the accuracy of a lexical stress assignment system is measured assuming 100% accuracy of the letter-to-sound rules. Klatt (1987) discusses the accuracy of several systems in terms of different methods of evaluation. Accuracy can be discussed in terms of overall accuracy, (i.e. the number of errors in phoneme selection or stress assignment per one hundred words). It may also be discussed in terms of the performance of the stress assignment rules as they are applied to a corpus of words. The typical method of testing is to apply the rules for stress assignment to a corpus of words. This corpus of words is most often taken from the list of frequent words of English as compiled by Kucera and Francis (1967).

At the present time there is no comprehensive or standard test that can be applied to any system in order to ascertain a degree of accuracy that can be used for comparison against other systems.

In the work documented by Hunnicutt (1976), she reports a value of 67% correct on a random sampling of 250 six-letter words from the Brown Corpus. The errors in this system include inconsistencies in both phoneme selection and stress assignment. Bernstein and Nessly (1981) report a success rate of approximately 75% on a lexical sample consisting of 430 words of five or more letters from the Brown Corpus with frequencies between 40 occurrences per million and 34 occurrences per million. Church (1985) reports 82% correct primary stress assignment on words contained in a dictionary which holds the entire Brown

---

<sup>2</sup> This is done by not reconstructing the final e of supreme and is not necessarily related to TSS.

corpus.

### **Computational Efficiency**

Computational efficiency is a desirable attribute of any system. However, this is a very difficult aspect of an algorithm to measure as several factors are involved. This discussion of computational efficiency will be limited to complexity of the algorithm.

Bernstein and Nessly (1981) report that the lexical stress assignment rules developed by Hunnicutt require 3000 bytes of memory, while their simplified version only requires 150 bytes of code to be operational. The cyclic nature of the Hunnicutt rules would lead one to believe that they are not as computationally efficient as a system which would apply primary stress to only one syllable on the surface structure.

Church does not discuss the efficiency of his system in detail. However, one can surmise from his use of a table with built-in constraints that the assignment of stress patterns would entail a simple binary search through the table, and would therefore be quite efficient.

### **Dialect**

It is interesting to note that reports on the accuracy of text-to-speech systems make no mention of the dialect of English that the system is attempting to emulate. For example, if a speaker of Canadian English were to examine what Sharon Hunnicutt reports to be accurate transcriptions, there would be an error in certain words based on such factors as vowel selection and perhaps notably British stress patterns.

### **Proposed Solution: Morphological Emphasis**

It seems clear that any algorithm for automatic stress assignment in English must involve morphological decomposition to some extent. Examination of the first 500 words of the Brown Corpus shows that many of the errors made by Nessly's default rule could be alleviated if some morphological decomposition were performed. For example, the word government is incorrectly assigned stress on the second syllable, because it is a closed syllable. If the word were analyzed into its constituent morphemes govern+ment, one would find that, due to the stress-neutral nature of the suffix ment, Nessly's default rule would only apply to the portion of the word preceding the suffix (i.e. govern). Hence, the primary stress would correctly fall on the first syllable of the word. An explanation of further advantages to a morphological decomposition will be given below.

### **Advantage to Morphological Decomposition**

The construction of words in any language is of interest to both phonology and morphology. As the representation of every morpheme must be realized phonologically it seems natural that the phonology of a language must in some way be influenced by the morphology of that language. The advantages to taking a morphological approach can be discussed in terms of three areas that it affects or provides information for:

1. the letter-to-sound rules
2. stress prediction and
3. part-of-speech determination.

The morphological decomposition of a word influences the letter-to-sound rules in one very important way. It has been noted that many letter-to-sound

rules can not apply across morpheme boundaries (Allen, 1973). Therefore it is important to know where these boundaries exist before the letter-to-sound rules are applied. Often the morpheme boundary functions as a word boundary. For example, word boundary rules remain in place even when a suffix or prefix is concatenated onto a free root (e.g., the word unknown retains the silent k in word initial position and the word manageable retains the silent e for word final position in the word manage). Also, certain language specific phonotactic principles may be ignored at morpheme boundaries. For example, the word obnoxious contains the phonological sequence /bn/ which would never occur word initially, or in any environment other than at a morpheme boundary. Furthermore, rules of letter-to-sound conversion do not apply across morpheme boundaries.

In English it is known that the stress pattern of a word is often influenced by the type of affix that is attached to it. Prior to SPE, Kingdon (1958) set out the 'Groundwork of English Stress' in which he categorized many of the word-stress patterns of English in terms of a morphological analysis. He enumerated many of the affixes of English and discussed how they influence the placement of the primary stressed syllable in words. More recently, Fudge (1984) has continued along this line. Fudge has also combined some valuable features from the work of Chomsky and Halle (1968) on word stress. By combining the work of Kingdon, SPE and Paul Garde (1968), Fudge has created a useful technique of word stress prediction based on a morphological analysis, combined with a phonological analysis of the undecomposable portion of English words. In particular, Fudge along with others (Church, 1985) makes a distinction between the various types of affixes. There are two types of affixes: those influencing stress, and

stress-neutral affixes. The stress-neutral affixes play an important role in the prediction of primary stress as they must be stripped off or ignored before locating the primary-stressed syllable.

In addition to influencing the location of the primary-stressed syllable in a word, the morphological structure of a word is important in conveying grammatical relationships. While morphemes may be divided into inflectional morphemes and derivational suffixes, both in some way indicate the grammatical role of the word in a sentence. For example, the phonological unit of /s/ in word-final position can indicate either a plural noun, or the singular form of a regular verb paradigm. This is an example of inflectional morphology. Similarly, derivational morphemes often cause a shift in meaning of the root along with a change of the part-of-speech of a word. For example the morpheme -al often causes a noun to become an adjective, as in the words function and functional. This is important in text-to-speech systems when a large dictionary is not feasible for determining the part-of-speech of a word.

A final point on the importance of morphological information is that often affixes reflect the etymology of words. For example the suffix -ette is of French origin and reflects a French stress pattern of the final syllable receiving primary stress, as in the words cigar and cigarette.

### **Influence of Affixes on Stress Patterns**

It is known that certain affixes influence the stress patterns of words in English. Kingdon (1958) shows how suffixes affect the stress-placement in words consisting of a root plus affixes (Fudge, 1984). Kingdon makes a number of generalizations about the stress of English words based on the morphological

make-up of the words. He distinguishes three types of word formations, or compounds. These include: Romanic (root plus affixes), Greek (root plus root, where roots are bound morphemes), and English (root plus root, where roots are free words.) Romanic words include humanity and boundary, Greek include satisfy and magnification, while examples of English-type words include lambswool and spokesman. The work and generalizations made by Kingdon are indeed valuable in providing us with early efforts at developing rules for word-stress in English based on the morphological make-up of words.

Since Kingdon, a number of other researchers have continued in this area of research. Work in this area can be divided into two fields: theories about word-syntax (Aronoff, 1978; Selkirk, 1982; Mohanan, 1985), and methodology for teaching English as a second language (Woods, 1979; Fudge, 1984). While theories of word-syntax focus on describing well-formed structures for English words, much of the data on the effect of suffixes on word stress comes from researchers specializing in teaching English as a second language. Fudge (1984) provides a good summary of the work of Kingdon, while at the same time proposing a comprehensive phonological solution for monomorphemic words and roots which do not have the stress determined by suffixes. Fudge (1984) notes the effect of affixes on the location of primary stress in words. He makes a distinction between prefixes and suffixes, the majority of the data devoted to suffixes.

Suffixes may be classified in several ways according to the prediction of stress. Fudge (1984) makes the following four distinctions:

1. inflectional and other stress-neutral suffixes when attached to a free form, leave the stress-pattern unchanged (Fudge, 1984).

2. 'autostressed' suffixes attract stress to themselves. They may be attached to free or bound forms. There are also a number of endings, which while not suffixes by nature, act like suffixes that attract stress to themselves, and are treated as such.

3. 'pre-stressed' suffixes are those which attract stress to a previous syllable.

4. 'mixed' suffixes may influence the placement of stress in a number of ways depending on the etymology of the root.

The 'pre-stressed' suffixes are the most common type of suffix and may influence the placement of the primary-stressed syllable in a number of ways. The primary-stressed syllable may be the immediately preceding one, or it may be located two syllables prior to the suffix. This classification of suffixes may be further divided, depending on which syllable it attracts stress to. Fudge refers to these suffixes as pre-stressed 1, pre-stressed 2, and pre-stressed 1/2. These will be elaborated on later as the need to clarify their function arises.

Mixed suffixes are perhaps the most difficult to model in a system. The major factor which determines how the mixed suffixes influence the location of the primary stress is if the suffix is attached to a free or bound root. Free roots are those which may stand alone as words, (e.g., change in changeable), while bound roots must have an affix attached, (e.g., -tain in detain, attain, and retain). In order to capture the distinction between bound and free roots, an extensive morph dictionary is required that would contain the relevant information.

Fudge also discusses the prediction of secondary stress by suffix. As this thesis is not focussing on the prediction of secondary stress, this will not be dealt with in detail.

While it is not known how extensively Fudge's analysis of English word stress has been incorporated into text-to-speech systems, McAllister (1987) has modeled the lexical-stress assignment component of a text-to-speech system quite closely after Fudge.

### **Previous work with Morphological Approach**

Most systems of text-to-speech conversion have utilized a morphological analysis to some extent. Of the systems that utilize a morphological analysis, some systems rely on the morphology more extensively than others, but all employ an analysis of words into their subsequent parts in the prediction of lexical stress and/or the determination of part of speech (O'Shaughnessy, 1987). While the decision to adopt a morphologically based approach to the assignment of lexical stress is not based on a desire to mimic other systems, it is clear that there is empirical as well as utilitarian evidence to support the validity of this move.

The extent to which other systems rely on the morphological analysis for converting text to some abstract linguistic representation can be viewed in terms of the basic approach that each takes. Some systems attempt to reconstruct the root in its entirety (Hunnicut, 1976). This may be called the generative approach. The phonological rules which operate in this system function on the underlying form of the root. Many of the rules are cyclic, with the primary stress being located first on the root, and then moving to another syllable if a stress-affecting suffix is encountered. The system developed by Bernstein and Nessly (1981) utilizes no morphological analysis at all in the assignment of stress, and it has achieved some reasonable measure of accuracy. Other systems operate on the surface structure of the word, relying on only the rules of the suffixes and the

letter-to-sound rules for the appropriate stress pattern and vowel selection (Church, 1985 and McAllister, 1987). The approach that will be adopted here will be to analyze only the surface form of the words and to apply rules only to the surface.

### **Proposed Approach**

The method proposed here will attempt to combine the features of several systems. In order to take advantage of the information about word-stress patterns in English that are associated with certain affixes, it will involve a morphological analysis. As well, to obtain a computationally efficient algorithm, it will incorporate a rule similar to Nessly's default rule to apply on portions of a word that are morphologically unanalyzable. It is hoped that by combining two simplified systems such as this, one will be able to obtain a reasonable degree of accuracy for the assignment of stress to words of English. The major part of the stress assignment rules will be carried out by a morphological decomposition. The remaining component will be a simple rule that approximates the SPE Main Stress Rule of English, and is phonological in nature.

A large portion of the letter-to-sound rules are already in place at the Centre for Speech Technology Research (CSTR) at the University of Victoria. Contained within these rules is a preliminary step to check for a limited number of suffixes. It has been suggested that this suffix list be augmented with information about the stress assigning properties associated with each suffix, and that it be increased in size to provide a more comprehensive analysis for the conversion of letters to sounds. The suffixes, as well as containing the phonological representation will also include diacritics for the marking of the stressed syllable. The bulk of this

information will be taken from the extensive Appendices provided by Fudge (1984) on the stress properties of over 200 suffixes and prefixes. Further, the classification and specifications of the affixes will follow the terminology set out in Fudge (1984).

After the suffixes have been converted to a phonemic representation, with diacritics to indicate the syllable to be stressed, the remainder of the word will be analyzed according to the letter-to-sound rules. A 'diacritic interpreter' will be used to ascertain whether Nessly's rule should apply to the base or root of the word. Stress will then be assigned according to this rule and will be sent to the prosodic rules of the system. This output will be detailed enough to provide an abstract linguistic representation of a dialect of Canadian English. A flow-chart of the proposed system is given in Figure 1.

A more detailed explanation of the various components of this flow-chart will be given in subsequent chapters.

It is felt that this approach is pragmatic enough to be used by text-to-speech conversion systems while still retaining some valuable linguistic insights. That is to say that it will rely on both phonological and morphological information about the word. To conclude this introduction to lexical stress assignment algorithms, it should be pointed out that while it has often been noted that the letter-to-sound rules and stress assignment rules are closely linked, the fusion of the two in this manner is unique and computationally efficient. It is hoped that, by combining a brief morphological analysis with a phonological approach to unanalyzeable portions of words, a reasonable measure of accuracy can be obtained within a computationally simple framework.

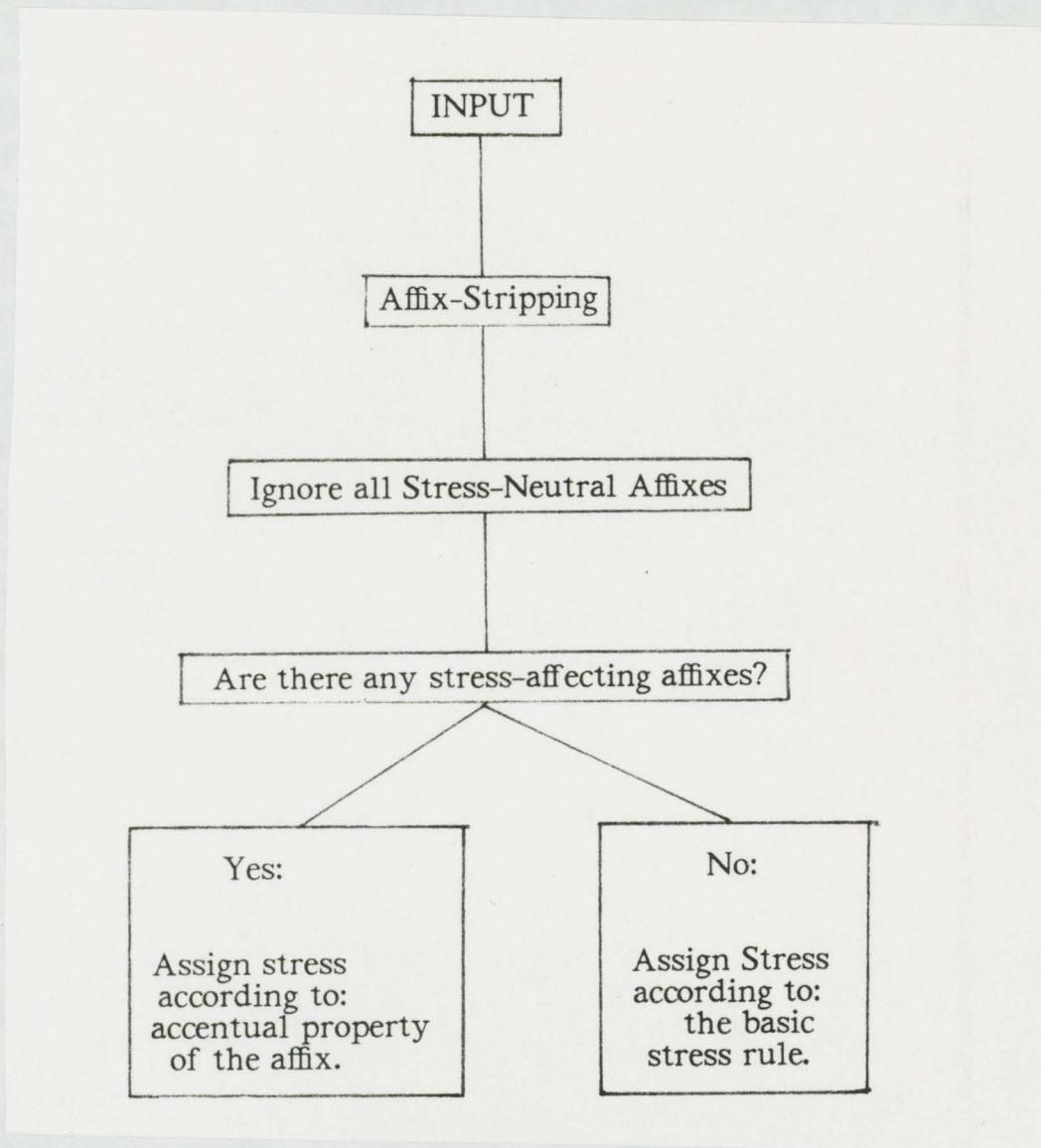


Figure 1: Flow-chart of Lexical Stress Component

## THE ALGORITHM

The purpose of this chapter is to outline the development and design of an algorithm for the assignment of stress to English words. The algorithm involves two components: the application of a basic stress rule and the location of stress-affecting affixes. The discussion of the algorithm is organized in four parts. The first of these will be to supply some preliminary remarks regarding the algorithm. Then the algorithm itself will be described. The basic stress rule will be described, followed by a detailed description of the affixes utilized, as compiled by Fudge (1984). The final area of discussion will be a detailed description and derivation of some relevant examples.

### Preliminaries to the Algorithm

As mentioned, some restrictions have been imposed on the development of the algorithm. The first constraint was to combine the stress-assignment algorithm with existing letter-to-sound rules (under development at CSTR). The second constraint was not having access to a large morpheme dictionary. The consequences of these two constraints will be discussed below, and have influenced the development of the algorithm.

The letter-to-sound rules plus the stress-assignment algorithm make up the entire component of text-to-phoneme conversion. An affix dictionary has been developed to contain a large number of prefixes and suffixes, to facilitate the

conversion of letters to sounds as well as function in the stress-assignment algorithm. This affix dictionary contains the orthographic representation of each affix, along with its phonemic representation. The phonemic representation contains a code, identifying the accentual (stress-affecting) properties of each affix. The accentual property for each affix is represented by a single diacritic, that does not interfere with the other representations within the program.

The use of this affix dictionary and the phonemic representation containing diacritics is a critical component of the algorithm. How the program makes use of this information will be described below.

### Description of the Algorithm

The stress-assignment algorithm is essentially as follows. Text is input to the program by either entering it directly to the program or by specifying an external text file. The input is standard orthographic representation and the output is an abstract linguistic representation containing both the phonemic representation and the location of primary stress for each word. The algorithm begins after all text has been formatted, (i.e. when numbers and abbreviations have been converted to words, and similar such operations). The first stage of the algorithm involves the detection of affixes. Because suffixes have more of an effect in the location of stress than prefixes, the algorithm checks for the presence of these first. If a suffix has been found, there are two possible options. If the suffix is stress-neutral, then the program checks again to see if there are any suffixes preceding it. It repeats this task until a suffix has been found that is not stress-neutral or until no additional suffix has been found. If a stress-affecting

suffix has been found, then stress is assigned based on its accentual property. If no stress-affecting suffix has been found, the algorithm checks for the presence of stress-affecting prefixes. If a stress-affecting prefix is located, then stress is assigned based on its accentual property. If no stress-affecting affixes have been found then the algorithm assigns stress based on a basic stress rule.

The algorithm can be briefly stated as follows:

1. strip off all stress-neutral suffixes.
2. are there any stress-affecting affixes?
  - a. if yes: assign stress according to the affix stress rules.
  - b. if no: assign stress according to the basic stress rule.

The basic stress rule and the affix stress rules will now be described in greater detail.

### **Basic Stress Rule**

As was stated previously, the matter of assigning primary stress to English words is not a simple issue. A number of factors bear on the accurate assignment of stress in English. One is that English is known as a quantity-sensitive (QS) language (Liberman and Prince, 1977; Hayes, 1982; Halle and Vergnaud, 1987). This means that the location of the stressed syllable is dependent on what is referred to as syllable weight. This can be seen in the following list of words.

(1)veRANda AlASka

(2)PAMela AMERica

(3)tomAto torPEdo

Note that in (1) and (3) it is the penultimate syllable that receives main stress, while in (2) it is the antepenultimate syllable. In (1) and (3) the penultimate syllable is described as heavy, while in (2) the penultimate syllable is described as light.<sup>3</sup> Heavy syllables are those that contain either a tense vowel (as in (3)) or end in a consonant cluster (as in (1)). Light syllables, on the other hand, do not contain a tense vowel or end in a consonant cluster (as in (2)). These descriptions are based on the weight of the syllable. The distinction between heavy and light syllables is what determines where main stress will fall. Because in English the heavy syllable receives the stress, it is described as a QS language.

A QS basic stress rule has been adopted, based on that incorporated by Bernstein and Nessly (1981). The basic stress rule functions by examining the syllable that the program is currently looking at and determining its weight. If it is heavy, then it is stressed. If it is light, then the preceding syllable is stressed. Syllable weight is determined based on whether the syllable in question ends in a consonant cluster or not. If it ends in a consonant cluster, then it is heavy. A further discussion of the motivation to exclude vowel tensity as a condition for syllable weight will be reserved until a later section in the thesis. Suffice it to say that it was found that the algorithm was more accurate when vowel tensity was not included as a condition for syllable weight. The basic stress rule functions as follows:

1. locate the penultimate syllable of the stressable portion of the word.
2. if this syllable ends in a consonant cluster, then stress it

---

<sup>3</sup> From this point on the terms heavy and light will be adopted to describe strong and weak respectively. There is reason to believe that these terms more accurately describe the phonological character of the syllables in question (Selkirk, 1984; Liberman and Prince, 1977).

3. if this syllable does not end in a consonant cluster, then stress the preceding syllable

This is the basic stress rule of the program. Combined with the affix stress rules, this comprises the stress assignment algorithm. An examination of the treatment of affixes by Fudge (1984) will precede the description of the diacritic interpreter.

### Affixes and Stress

One of the most valuable aspects of Fudge (1984) is the extensive analysis and description he provides on how affixes can be used to predict the location of stress in English. Fudge has compiled a comprehensive listing of suffixes and prefixes, categorized by how they affect stress. In fact the bulk of his book is dedicated to the listing and description of these affixes. It is valuable as a resource from which to compile a list of affixes, as well as how they interact with the phonology and morphology of English. The discussion will focus on how suffixes affect stress, followed by a brief explanation of the effect of prefixes on stress.

It has been noted that there is a type of hierarchy for affixes in the prediction of stress. If a stress-affecting suffix is attached to a word, it invariably causes stress to be located based on its accentual properties. This overrides the effect of prefixes on stress, which in turn overrides the stress as located by the main stress rule. More discussion of this will be provided in the description of the algorithm itself.

## Stress-Neutral Suffixes

The first of the suffixes that Fudge examines are the stress-neutral ones. Stress-neutral suffixes invariably attach to free words and do not affect the stress of the word. They are composed of all inflectional suffixes, and some derivational suffixes. Inflectional suffixes include those that indicate grammatical functions such as plural -s/-es, (e.g. verandas, tomatoes), past tense -ed, (e.g., whispered), and present participle/ gerund -ing, (e.g., singing, skiing). Derivational suffixes include those that when added to a root, derive or create new words and often change the part of speech. Examples of derivational suffixes include nominalizers -hood, (e.g., womanhood), -dom, (e.g., kingdom), and adjectivizers -ous, -able, (e.g., beautiful, respectable).<sup>4</sup>

## 'Autostressed' Suffixes

'Autostressed' suffixes are those which attract stress to themselves. Examples of words containing autostressed suffixes include: Japanese, cigarette, and millionaire. Within this category Fudge includes endings which, while not strictly suffixes, function as other autostressed suffixes. Some of these include: -ee, -eer, -enne, -ese, and -ette.<sup>5</sup>

While the majority of autostressed suffixes are monosyllabic, there are a few that are polysyllabic. For this reason it is necessary to indicate the syllable which receives the main stress. Therefore the terms autostressed 1 and autostressed 2 have been adopted, indicating whether it is the first or the second syllable of the

<sup>4</sup> Stress-neutral suffixes may also be described as extrametrical. They attach to the outer most part of a word and are ignored in the assignment of lexical stress.

<sup>5</sup> Note that the inclusion of these endings will account for the many French loan-words that exhibit word-final primary stress.

suffix which receives the primary stress. Autostressed 1 suffixes include suffixes like -ation as in the words imagination and infatuation. An example of an autostressed 2 suffix is -ability as in the word acceptability. These suffixes are usually created by combining a number of suffixes to create a new suffix. In the example -ability, the suffix consists of -able and -ity. Because -ity has an accentual property associated with it, it causes the preceding syllable to be stressed.

### **'Pre-Stressed' Suffixes**

The group known as 'pre-stressed' suffixes is by far the largest and most complex group of suffixes that influence stress. Fudge has broken this group down into a number of subgroups according to their effect on stress placement.

The first of these subgroups that Fudge describes are known as 'pre-stressed 1' suffixes, because they cause the primary stress to be located on the syllable immediately preceding the suffix. These include -ic, -ion, and -ity/ -ety, as in the words 'spoRADic', 'susPICion' and 'imPURity'.

The second group of pre-stressed suffixes are called 'pre-stressed 2' suffixes. These cause stress to be located two syllables prior to the suffix. Examples of these include -gon, -tude, and -cide, as in the words PARagon, SOLitude and HOMicide.

The third group is a rather large subgroup and operates on principles similar to those of the conditions of the basic stress rule. That is to say that determination of which syllable receives main stress depends on whether the preceding syllable is weak or strong. These are referred to as 'pre-stressed 1/2' because the syllable that is stressed may be either one or two syllables prior to the suffix, depending

on the weight of the preceding syllable. Fudge has developed a rule which will predict the location of the syllable to be stressed preceding a prestressed 1/2 suffix. The rule is as follows:

1. If the syllable before the suffix is strong, then it is stressed. (e.g., suffix -al, as in 'homiCIDal').
2. If the syllable before the suffix is weak, then the syllable before that is stressed. (e.g., suffix -al, as in 'orIGinal').

### 'Mixed' Suffixes

As a final matter to complicate the accentual properties of suffixes, a large number operate in a variety of ways, depending on whether the suffix is attached to a free root or a bound root. Free roots are those that can stand alone as words, while bound roots must have some type of suffixal, or morphological ending before they are perceived as words that can stand alone. Suffixes that function in this manner are known as 'mixed' suffixes. An example is the suffix -able in the words deFINable and inTERminable. In the former example, the suffix is attached to a free root or word, (i.e., define) and functions as a stress-neutral suffix. In the latter example the suffix is attached to a bound root, (i.e., intermine) and functions as a prestressed 2 suffix.

Mixed suffixes also function differently depending on the number of syllables in a word or root. An example of this is the suffix -ate. When attached to disyllabic words or stems, it functions as an autostressed suffix as in locATE and rotATE (in British English). However, if it is attached to a bound root of three or more syllables it functions as a pre-stressed 2 suffix, as in 'DEMonstrate'.

The set of 'mixed' suffixes has not, at the time of writing, been incorporated into the algorithm. These suffixes are dependent on whether they are attached to a free or bound root, in determining the correct pattern of stress assignment. Because there was a restriction of not allowing a morpheme dictionary for the algorithm, access to the morphemic structure of a word was not available to the stress rules.<sup>6</sup>

### Prefixes

The only prefixes that have been used in the stress-assignment algorithm are what Fudge calls 'stress-repellent'. Of these, only a limited number have been incorporated into the algorithm. The term 'stress-repellent' refers to the fact that many prefixes cause stress to move away from them in the case that there is not a stress-affecting suffix on the word. For example the prefix ex- will cause the stress to shift to the following syllable, such as in the word exCELL. However, when it is combined with a stress-affecting suffix, such as ent in the word EXcellent, the suffix has more priority in predicting where the stress will be located. The prefixes are only used to predict stress when there is not a suffix, or when the suffix is stress-neutral. A list of the affixes, and their accentual properties is given in Appendix B.

---

<sup>6</sup> Subsequently, a list of bound roots has been developed and the inclusion of the mixed suffixes will follow shortly.

### Affix Stress Rules

The affix stress rules are those rules that direct the algorithm in locating the primary stressed syllable based on the accentual properties of any affixes located. As mentioned previously, each affix contains a representation of its accentual property in the form of a diacritic. Each diacritic represents one of the six categories of affixes: stress-neutral, pre-stressed 1/2, pre-stressed 1, pre-stressed 2, autostressed 1 and autostressed 2. The affix stress rules determine the syllable to receive main stress based on these accentual properties. The following represents the stages the algorithm goes through to determine the location of the primary stressed syllable in a word.

1. | stress-neutral
  - a. skip to the preceding syllable
  - b. assign stress according to the basic stress rule
2. ^ pre-stressed 1/2
  - a. assign stress according to the basic stress rule
3. † pre-stressed 1
  - a. stress is placed on the preceding syllable
4. < pre-stressed 2
  - a. stress is assigned to the second preceding syllable
5. / autostressed 1
  - a. stress is assigned to the present syllable
6. > autostressed 2
  - a. stress is assigned to the following syllable

This determines, in the majority of instances, which syllable will receive the primary stress. It may now prove useful to examine how the algorithm works with a variety of examples, illustrating the various components of the algorithm.

### Examples

The following examples will range in complexity of analysis, beginning with a simple, monomorphemic word, and continuing to more morphologically, and phonologically complex examples. Eight examples will be given that show the various aspects of how the algorithm works. In each case, the word will be given first. This will be followed by the output of the letter-to-sound rules, containing both the phonemic representation and the appropriate diacritic. Then the stage involved in the assignment of stress will be described. And finally, the output, including the location of the primary stressed syllable, will be provided. Following each example, a representation of the output at each stage of the algorithm will be provided. This will include:

Input: input to the algorithm

LTS: output of the letter-to-sound rules

SA: output of the stress assignment algorithm.

A brief description of the structure of the affix dictionary will be given to facilitate the explanation of the examples. A typical representation of a suffix in the suffix dictionary is as follows:

ER |EOR<sup>7</sup>

---

<sup>7</sup> A practical orthography has been developed to represent the inventory of English phonemes on a standard keyboard. This practical orthography can be found in Appendix A.

The first item in this entry, ER, is the suffix as it is represented in English orthography. The next item, |EOR, provides the stress diacritic (i.e., | - stress neutral), and the phonemic representation of this suffix, EOR. This representation is used by the algorithm to convert the suffix -er to a its phonemic representation as well as signal that the suffix is stress-neutral and should be ignored in the assignment of stress.

**Example 1: LITTLE**

The first example is the word LITTLE. The letter-to-sound rules provide the phonemic representation: LITTOEL. Because there are no affixes located, stress is assigned by the basic stress rule. The algorithm begins by looking at the last syllable. It skips to the preceding syllable, and places the primary stress on this syllable, because the word contains no affixes and is a disyllabic word. This produces the abstract phonemic representation of LIT@TOEL.<sup>8</sup>

Input:     LITTLE  
LTS:     LITTOEL  
SA: LIT@TOEL

**Example 2: SOVIET**

The second example involves the polysyllabic monomorphemic word SOVIET. The letter-to-sound rules produce SOUVIYOET. The basic stress rule then examines the penultimate syllable (IY) to determine if it ends in a consonant cluster. Because it does not, the syllable preceding it is stressed. This produces the representation SOUV@IYOET.<sup>9</sup>

---

<sup>8</sup> Primary stress is indicated by the symbol @ following the stressed syllable.

Input: SOVIET

LTS: SOUVIYOET

SA: SOUV@IYOET

### Example 3: ALASKA

The third example illustrates how stress is correctly assigned to the word ALASKA. This is also a polysyllabic monomorphemic word. The algorithm proceeds in much the same manner as in the above example (SOVIET). However, when the basic stress rule is examining the penultimate syllable to determine if it ends in a consonant cluster, it finds that it does, and therefore stresses this syllable. Thus, the output is AELAESK@OE.

Input: ALASKA

LTS: AELAESKOE

SA: AELAESK@OE

### Example 4: HIMSELF

The fourth example, HIMSELF, illustrates the effect of a suffix on the placement of lexical stress. The output of the letter-to-sound rules is the following abstract representation: HIM/SELF. SELF is found to be an autostressed suffix with the representation /SELF. When the diacritic interpreter encounters the symbol '/', it ensures that the current syllable should receive the stress, and the output is HIMSELF@.

---

<sup>9</sup> Note that if vowel tensity were a condition on whether a syllable should receive stress, stress would incorrectly be assigned on the penultimate syllable.

(x)Input: HIMSELF  
LTS: HIM/SELF  
SA: HIMSELF@

**Example 5: POSSIBLE**

As the fifth example, a word with a prestressed 1/2 suffix was chosen. This is the word POSSIBLE, containing the suffix -IBLE. While this may not be a traditional morpheme, for the purposes of the operation of the algorithm in locating main stress, it is a suffix. It is analyzed to produce the output PAS^OEBOEL, recognizing the orthographic IBLE as a suffix. When the diacritic interpreter encounters the symbol ^, it is told to go directly to the section of the algorithm entitled, 'Basic Stress Rule'. At this point there is a check to see if the syllable ends in a consonant cluster. Because it does not, it is told to stress the previous syllable. However, because there is no previous syllable, it stresses the syllable it is on and produces the correct output of: PAS@OEBOEL.

Input: POSSIBLE  
LTS: PAS^OEBOEL  
SA: PAS@OEBOEL

**Example 6: WITHOUT**

The sixth example illustrates the effect of stress-repellent prefixes in the algorithm. WITHOUT is analyzed in the letter-to-sound rules as >WITHAUT, because the prefix WITH is specified as a stress-repellent prefix in the suffix dictionary. The symbol > indicates that stress falls on the following syllable, and the output is WITHAUT@.

Input:       WITHOUT  
LTS:         >WITHAUT  
SA: WITHAUT@

### Examples 7 and 8: DESIGN and DESIGNATION

The final two examples will be discussed in tandem as they illustrate first the effect of a stress-repellent prefix, and second the effect of a stress-repellent prefix in combination with a stress affecting suffix. The words used are DESIGN and DESIGNATION. The first word, DESIGN, is given the phonemic analysis of >DOEZAIN. The diacritic interpreter encounters the symbol >, and assigns stress to the following syllable. Thus, the representation DOEZAIN@.

Input:       DESIGN  
LTS:         >DOEZAIN  
SA: DOEZAIN@

However, in the analysis of the word DESIGNATION, a word containing both a stress-affecting prefix and suffix, the suffix takes priority over the prefix in determining the location of the main stress. In particular, the portion of the algorithm that locates affixes finds both the prefix DE and the suffix ATION. Because the suffix ATION has an accentual property associated with it (it is represented as /EYSHOEN in the dictionary), this affix has priority over the prefix DE. Therefore the word is analyzed as DEZIGN/EYSHOEN. When the diacritic interpreter encounters the symbol /, it will cause the primary stress to be located on the present syllable. This will produce the abstract linguistic representation of DEZIGNEY@SHOEN.

Input: DESIGNATION

LTS: DEZIGN/EYSHOEN

SA: DEZIGNEY@SHOEN

This concludes the section on examples which illustrate the various components of the stress-assignment algorithm. It is hoped that a clear picture has been obtained on the various components of the algorithm, how they function, and how they interact.

### Summary

In summary, this algorithm represents a computationally efficient approach to the problem of stress assignment in English. It utilizes two properties that determine the placement of the primary stressed syllable: a quantity sensitive stress rule and a morphological analysis. As will be shown in the following section detailing the evaluation of the algorithm, it is as accurate as, or more accurate than other more complex algorithms.

## EVALUATION

The goal of this chapter is to evaluate the algorithm previously described. A method was chosen that is similar to that adopted by other researchers in evaluating their systems, in order to compare its accuracy on a predefined corpus of English words. The method will be described in the first section. This will be followed by a report of the results and an indepth discussion of the errors made by the algorithm.

### Evaluation

There were two methods of evaluation of this system. The first was to test its accuracy against other systems. The second was to test the accuracy of the system on a sample of running text.

Previous algorithms of stress assignment for English have been evaluated using a variety of procedures. While some methodology has been developed that could be said to be representative of the accuracy of the systems, testing has been inconsistent between research groups. This is problematic when one wishes to compare the accuracy of a system against those previously developed. In some instances it is difficult to assess whether a system has been evaluated objectively. A discussion will be provided of the methods of evaluation of previous systems and some apparent problems with these methods. This will be followed by a discussion of the current methodology.

## Hunnicutt

Hunnicutt (1976) evaluated both the letter-to-sound rules and the stress-assignment algorithm developed for MITalk on a number of corpii. She used as one of her corpii of words, 250 randomly sampled six-letter words taken from the Brown Corpus. The Brown Corpus is a listing of word frequencies for American English. The corpus was compiled by examining an amalgamation of texts totalling over one million words. The corpus contains 43,777 entries, rated from the most frequent to the least frequent in terms of numbers of occurrences within the larger corpus.

The corpus selected by Hunnicutt contained a random sampling of six-letter words from the Brown Corpus. Justification for choosing six-letter words as opposed to others was not given by the author. However, one can assume that the criteria of six letters was adopted in order to test the system on larger English words. The system functioned at a level of accuracy, varying between 62 to 69 percent, depending on the number of rules used in the algorithm.

One problem with this testing procedure is that no mention was made of whether the six-letter words were monosyllabic or polysyllabic. While it may be safe to assume that the majority of six-letter words are polysyllabic, there were a number of examples cited in the work that were monosyllabic. These include: should, theirs, streak, and realm. One would assume that these words are assigned stress correctly no matter what the algorithm is because they do not represent a choice of syllables.

A further problem with the testing of the MITalk system is that the letter-to-sound rules are not tested separately from the stress rules. Many of the

polysyllabic words which were assigned stress correctly were considered errors because of incorrect phoneme selection. Examples cited by Hunnicutt include exhaust, forfeit, and strategically.

Therefore there are two confounding effects. The first is that, due to the presence of monosyllabic words, the accuracy achieved is artificially high. That is to say that including monosyllabic words in the testing of the stress-assignment algorithm would cause the accuracy to be higher than if monosyllabic words were excluded. The second problem is that, due to errors in phoneme selection, the accuracy of the stress-assigning portion of the algorithm is artificially low. Words assigned stress correctly, but converted to phonemes incorrectly were considered errors. This makes the accuracy of the algorithm seem lower than it actually is. Therefore this testing procedure does not represent a valid methodology for the testing of stress-assigning algorithms since it is unclear how accurate the component for assigning stress actually is. Subsequently the MITalk system has been evaluated by Bernstein and Nessly (1981) against a more simple algorithm. This will be the next area of discussion.

### **Bernstein and Nessly**

Bernstein and Nessly (1981) discuss the difficulties of comparing the accuracy of text-to-phoneme conversion systems as due to problems associated with measurement techniques. They confirm the points previously made that the confounding of the letter-to-sound rules with the stress-assignment rules is problematic. They describe the comparison of two algorithms for the automatic assignment of lexical stress in English. The two algorithms they compare are the MITalk system, as described by Hunnicutt (1976), and a simplified version of a

stress rule proposed by Hill and Nessly (1973), referred to as a first approximation of the Main Stress Rule (MSR) of SPE. The testing was performed on a sample of 430 words selected from the Brown Corpus. The words were five or more characters in length and occurred between 34 to 40 parts per million (i.e., occurred 34 to 40 times in the large sample of text). A number of words were rejected from the results due to either being monosyllabic or having an incorrect syllabification by the letter-to-sound rules. Both algorithms, MITalk and the first approximation of the MSR, achieved a 75 percent level of accuracy.

This procedure of testing seems valid because it ensures that only polysyllabic words are included, and the results are not confounded with the letter-to-sound rules.

### **Church**

Church (1985) tested his rules on the entire Brown Corpus (43,777 words) and achieved 82 percent accuracy. Because the Brown Corpus contains a large number of monosyllabic words this figure may be artificially high. While it is not certain that the number of monosyllabic words contained within the Brown Corpus is evenly distributed (there would seemingly be a greater percentage among the high frequency words than the low frequency words), a brief examination was made to see if there was a substantial portion of monosyllabic words included in the sample. In examining the first 5000 words of the Brown Corpus, 1350 were found to be monosyllabic. When subtracted from the percentage correct (82% of 43,777 = 35,897) ( $35,897 - 1350 = 34,547$ ) and divided by the total number of words in Church's dictionary, this yields an accuracy of 79 percent ( $34,547$  divided by  $43,777 = 79\%$ ). This is a substantial difference in the accuracy achieved, even in

the initial portion of the corpus. Therefore it may be safe to say that the rules developed by Church are not as accurate as reported by the figure of 82%.

### **McCallister**

McAllister (1986) tested his algorithm on a "small, randomly-selected corpus of polysyllabic words which the system is known to decompose correctly into constituent morphemes" (McCallister, 1986), and achieved a 94 percent level of accuracy.

The method of testing as reported by McAllister (1986) is suspect for several reasons. He does not explain how the "randomly-selected" corpus is obtained, i.e., which words were tested. Nor does he make note of how many words are contained in the "small" corpus. Further, the fact that the program decomposes the words correctly is not random, and implies an intervention in the testing procedure. As the algorithm developed by McAllister depends on an accurate morphemic analysis, and the correct decomposition of morphemes is a difficult task, particularly in automation, one must look upon his results with caution. Is the accuracy reported a good reflection of how well this algorithm performs on a more standardized list, such as the Brown Corpus? The assumption made here is that it is not clear how accurate this system is, and therefore one can not compare these results to those obtained by others.

### **Current Methodology**

As was stated previously, two methods of evaluation were adopted for evaluation of the algorithm developed here. One was used to compare the stress assignment algorithm developed here with other systems. A second was used to determine how accurately the algorithm functioned on a body of text.

The method of comparative evaluation adopted here follows that set out by Bernstein and Nessly (1981). With the exception of Bernstein and Nessly (1981), no attempts were made to avoid the inclusion of monosyllabic words into the corpus used for testing purposes. In the case of monosyllabic words there can be no error in the placement of stress. This will indicate a higher level of accuracy than might be otherwise obtained as was the case for Hunnicutt (1967) and Church (1985). Therefore the current testing procedure, only polysyllabic words were chosen.

The corpus of words used for testing purposes was taken from the Brown Corpus of most frequent words. The list contains 485 of the most common polysyllabic words,<sup>10</sup> and is similar to that used by Bernstein and Nessly (1981).

The second method for evaluating the accuracy of the stress-assignment algorithm was to measure its accuracy on a sample of running text. The text used came from an information delivery system, and contained 1318 words.<sup>11</sup>

These texts were entered into the program and the results were checked against the Canadian Gage dictionary. After an initial run of the corpus of words, to determine a correct encoding of the algorithm, the corpus of words was tested again. In both testing procedures, errors made due to a faulty conversion of letters to sounds were ignored.

---

<sup>10</sup> This list can be found in Appendix C.

<sup>11</sup> This text can be found in Appendix C.

### Level of Accuracy

The results obtained for the testing of the sample of words taken from the Brown Corpus are as follows. Of the 485 words, it was found that 10 were noun-verb homographs that are differentiated by stress. These are words that can have stress located on either the first or the second syllable and still be considered correct. These were excluded from the sample in the determination of the accuracy because no matter where stress was assigned, it would be correct. This brought the number of the sample down to 475. Of the remaining 475 words, 413 were assigned stress correctly. This results in an accuracy rate of 86.9% at the current state of the algorithm.

In regards to the testing of the sample of text, it was found that of the 1318 words, there were 145 unique occurrences of polysyllabic words. Of these 145 words, 124 were assigned stress correctly. This yields an accuracy level of 85.5%. However, on the entire text, with monosyllabic words and multiple occurrences included, the system produces an overall accuracy of 94%.

When comparing the algorithm presented here with other algorithms, it functions at the same level of accuracy or better than all reported on in the literature to date. The system described here is more accurate than that reported on by Bernstein and Nessly (1981) (75%) even though it involved a similar algorithm for assigning stress. The increased accuracy is a result of the inclusion of a morphological analysis. It is felt that the results obtained are valuable in providing an accurate reflection of the performance of the stress-assignment algorithm on both a body of text and a corpus of words.

## **Error Analysis**

The discussion of the results obtained will focus on the errors that the stress-assignment algorithm made on items taken from the Brown Corpus. This will be done for two reasons. The first reason is that understanding the errors that the algorithm makes will assist in increasing its accuracy. Secondly, it will provide some insights into English primary stress assignment in general.

## **Results**

Of the 475 words examined, 62 were assigned primary stress incorrectly. In order to examine these errors in more detail, they were broken down into groups, based on the number of syllables in each word. These groups were disyllabic, trisyllabic and polysyllabic (greater than three syllables).

### **Disyllabic Words**

The majority of errors of stress placement occurred in disyllabic words. This is due to the fact that the majority of words tested were disyllabic. English exhibits much variety in stress placement within this group that can not be explained in simple terms. There were 45 errors made out of 320 disyllabic words. This results in a 14.1% error rate among disyllabic words, and accounts for 72.6% of the total error. It is interesting to note that, without including information about part of speech, the accuracy of 85% is maintained among disyllabic words.

When examining the errors made for disyllabic words, there was a further subclassification. Words that were incorrectly assigned stress on the first syllable totalled 41, and words that were incorrectly assigned stress on the second syllable totalled 4. The list of these words is provided below.

(1) Disyllabic words incorrectly assigned stress on the first syllable.

About BEcause Against  
 Around Upon UNTil  
 Away Enough TOWard  
 Among Along ALthough  
 PERhaps Above TOday  
 Across Ago CONtrol  
 PROvide OUTside Alone  
 REquired COMplete SUpport  
 BEyond INside INstead  
 Amount REceived ENTire  
 INVolved THROUGHout CONCerned  
 APpeared Direct HOtel<sup>12</sup>  
 Effects DEScribed ACcount  
 POLice DESigned

By examining the errors in the previous list, some observations were made regarding theories of stress assignment as well as models and algorithms for assigning stress in English. Because it has been noted that part-of-speech is significant in determining the location of the main stress, a closer analysis was made of disyllabic words.

Of the 41 errors made on disyllabic words, 14 were verbs.<sup>13</sup> These include:

---

<sup>12</sup> The program correctly assigns stress to 'appear' but not to 'appeared'. This is a result of incorrect morpheme decomposition. That is to say that the algorithm needs to know that if a stress-neutral suffix is added to a word, it is still necessary to check for the presence of prefixes. This inconsistency is a result of an error in the program and not the algorithm.

<sup>13</sup> Note that some of these are homographs that are not differentiated by stress.

control    provide    designed  
 required   complete   support  
 received   involved   concerned  
 appeared   direct effects  
 described   account

In the corpus of 475 words tested, there were 65 disyllabic verbs. There exist two possible solutions for the correct assignment of stress to these verbs. One is to include part-of-speech in the basic stress rule as a factor for stress placement. This is the widely accepted approach. A second possibility is to examine more closely the morphological composition of the words. Note that the preceding verbs are all composed of prefix + bound root, where the prefix is stress-repellent.

con+trol    pro+vide    de+signed  
 re+quired   com+plete    sup+port  
 re+ceived   in+volved    con+cerned  
 ap+peared   di+rect    ef+fects  
 de+scribed   ac+count

While it is not safe to reject the solution which is sensitive to part-of-speech, it seems plausible that the pattern of main stress on the final syllable can be handled by a very accurate morpheme decomposition.<sup>14</sup> This decomposition was not accomplished because there is no access to bound roots and the number of stress-affecting prefixes needs to be increased to include many more. Many words

---

<sup>14</sup> This observation is meant only as an observation and does not represent empirical evidence that part-of-speech is insignificant in determining stress in English.

also have the similar pattern of stress that are not verbs. The majority of the words in error are prepositions and adverbs. (about, again, before, and behind). While the list of words against which this algorithm was tested can not be considered a comprehensive test of English, it can be said to be a reflection of the usage of words in English.

(2) Disyllabic words incorrectly assigned stress on the second syllable.

seCOND

reCENT

marKET

seVEN

The errors made to disyllabic words in which the second syllable was incorrectly assigned main stress were the result of a faulty morpheme decomposition. The prefix se-, when attached to a bound root such as -cure, acts like a stress-repellent prefix. However, in the analysis of the words second and seven, the program incorrectly assumes that -cond and -ven are bound roots and assigns stress incorrectly to the final syllable. A similar analysis occurred with the word recent. These errors can be alleviated if a listing of bound morphemes were to be included. With the word market, the algorithm incorrectly locates the autostressed suffix -et and causes stress to be located on the final syllable.

## Trisyllabic Words

Many insights were gained by examining the errors made to trisyllabic words. It is on these words that the accuracy of the basic stress rule was tested. Therefore the discussion of the errors made to trisyllabic words will focus on two areas. The first will be a discussion of the errors. The second area will be a discussion of the errors made when vowel tensity was included and subsequently excluded as a condition for syllable weight.

Of the 102 trisyllabic words, 21 were assigned stress incorrectly. The trisyllabic words which were in error were examined further regarding which syllable the stress was incorrectly placed on. Eighteen were incorrectly assigned stress on the first syllable, two were incorrectly assigned stress on the second syllable, and one word was incorrectly assigned stress on the last syllable. These words are provided below.

Another    HOWever    United  
ALready    TOgether    DEpartment  
INcluding    DEveloped    Equipment  
CONdidered    DEcided    DIrectly  
ACcording    REMember    PROvided  
CONsider    DETermined    COMpletely  
  
proPERty    unDERstand  
  
committEE

Many of the errors found were based on a faulty morpheme decomposition. With the exception of property, however, together, and understand, all the words can be analyzed to provide the structure prefix + bound root. As was stated previously, analyzing the words as having this structure can predict the stress patterns of the words without reference to part-of-speech. This was not done because there was no access to bound morphemes during the course of development of the algorithm.

Note that there are some older English words, however and together, that were not assigned stress correctly. It would seem that these words are either exceptions to all rules of stress-assignment, or have stress assigned based on some historically significant factor.

Examining the errors made to trisyllabic words is the most significant method of determining if the QS stress rule is able to predict the stress pattern of larger words. When vowel tensity was included as a factor for determining syllable weight, the following words, containing a tense vowel in the penultimate syllable, were assigned stress incorrectly.

area	ratio	
religious	William	
Soviet	radio	
serious	obviously	
particle	vehicle	principle <sup>15</sup>

---

<sup>15</sup> These last three words were assigned stress incorrectly for two reasons. The first is that the letter-to-sound rules incorrectly converted the vowel in the penultimate syllable (i) to AY. This vowel is classified as a tense vowel. Therefore when the basic stress rule examined this syllable to determine its weight, it classified it as a heavy syllable. These errors are then the result of a faulty conversion of letters to sounds.

When vowel tensity was removed as a condition for determining syllable weight, these errors were not made. Also, there were no errors introduced because of this change. That is to say that no words were incorrectly assigned stress on the first syllable that had tense vowels in penultimate position.

All words with stress incorrectly placed on the first syllable should have received stress on the second syllable. Of these words, none that were monomorphemic had a tense vowel on the second syllable. In other words, there were no words that should have received main stress on the second syllable which contained tense vowels. Even though tensity was not a factor in determining if a syllable was heavy, or light, it appears that this is not a significant feature in determining the placement of stress. The errors made were not a result of excluding vowel tensity as a condition for syllable weight.

### **Polysyllabic Words**

The following polysyllabic words were incorrectly assigned stress.

DEvelopment	AVailable
APparently	unDERstanding

These errors account for 6.4% of the total error in the stress assignment algorithm. These errors can be said to be a result of an incorrect morpheme decomposition. The word development was analyzed to be composed of develop + ment. Because the penultimate syllable is light, stress is located on the first syllable. Develop is a verb and, according to SPE, should have stress located on the penultimate syllable as opposed to the first syllable. However, as stated in

the discussion of trisyllabic words, the stem develop can be further analyzed to form the constituents de+velop. Because de- is a stress-repellent suffix, stress will be correctly located on the second syllable. Similarly, the word understanding was analyzed as understand + ing. However, because understand can be broken down into under+stand, and the morphemes under and stand function together in the word-formation process, a rule should be developed that will account for the stress patterns of words of this nature.

Also, the words apparently and available have been assigned stress incorrectly due to an inadequate specification of the affixes. The stress-repellent prefix ap- was not recognized by the algorithm because the pre-stressed 2 suffix -ent was located. This caused the main stress to fall incorrectly on the first syllable. The word available was incorrectly assigned stress because it contains the affix -able, which is classified as a mixed suffix. The suffix is stress-neutral when attached to stems which are free forms and pre-stressed 2 when attached to bound forms. If available is analyzed as avail+able, where avail is a free form, then stress will be correctly located on the second syllable due to the presence of the stress-repellent prefix a-.

### **Generalizations**

The generalizations that can be made by the testing of this algorithm can be classified into two groups: further improvements to the algorithm, and general tendencies that can reflect the course of future work on theories of lexical stress assignment in English.

## Improvements to Algorithm

By far the most problematic aspect of the algorithm is in the decomposition of a word into its morphemes. This is problematic because it affects how stress is placed. When an error is made in the decomposition of a word, this usually results in an error in the placement of primary stress, as well as conversion from letters to sounds. These errors in morpheme decomposition can be said to be a result of two deficiencies of the algorithm. There is no access to a comprehensive morpheme dictionary. Further, rules about word structure in English need to be incorporated to assist in decomposing words more efficiently and accurately. In particular, it should be noted which morphemes exert an influence on the stress and phonology of a word at the word creation stage and at later stages in the addition of suffixes.

However some of the characteristics of bound roots were handled by utilizing information about the phonotactic constraints of syllable structure in English. Phonotactic rules are those which specify what are grammatical phonological sequences in a language.

Rules reflecting the phonotactic constraints in English can be used to ensure accurate affix stripping. In particular, a prefix must be followed by a phonotactically correct sequence for syllable onsets in English. Given the word absent and the two prefixes a- and ab-, the following structures can be analyzed: a+bsent and ab + sent. If absent is analyzed as a+bsent, this will be rejected because the phonological sequence bs is not a grammatical syllable onset in English. Therefore the correct analysis will be ab+sent and stress will be assigned according to the accentual properties of ab-. Yet, the matter is still more

complex because the accentual property of ab- is such that it is stress-repellent for verbs, but not for adjectives. More succinctly, it is a 'mixed' prefix when the form it is attached to is homographic.

In a similar way, the prefix be- has been included to handle the words: before, because, between, began, behind, became, believe, beyond and below. It is clear that all the words beginning with be- in the corpus tested receive stress on the following syllable, and can not be described by any alternate classification, such as its grammatical part of speech. Therefore be- has been classified as a stress-repellent prefix. Words such as better will not be decomposed as be+tter, because the syllable tter does not meet the structural description of onsets of English syllables. Therefore the prefix be- will not be stripped off the word and better will be assigned stress according to the basic stress rule. It should be pointed out that many of the rules for grammatical onsets are based on the orthography of the word, as this is the input in this text-to-speech conversion system for the decomposition of a word into its morphemes.

### **Implications for Future Theoretical Work**

In terms of the implications for future theoretical work in the area of word-stress assignment for English, it is felt that this algorithm, and the ability to test it in the manner it has been tested has provided some interesting insights. One of the most important observations made in this thesis is that vowel tensity was not used as a criterion of whether a syllable should be considered heavy or light.

The second observation is that by proposing a morphological analysis as a solution to the location of stress, one is able to accurately assign stress to many disyllabic words without requiring information about part-of-speech.

In regard to the necessity of information about part-of-speech, it seems clear from the above discussion that part-of-speech information is necessary only when encountering homographs. In other instances, there is no need to specify the part-of-speech for the determination of stress patterns in English words. This can be handled more effectively by acquiring a comprehensive morpheme dictionary that contains all information about affixes and bound morphemes.

### Summary

It is hoped that two issues have emerged from the discussion set forth in this chapter. The first is that it is both valuable and necessary to adopt a standard method of evaluating stress-assigning algorithms. The second is that the accuracy of such algorithms is highly dependent on a very good set of rules for decomposing English words into their constituent morphemes. While it is hoped that the first issue has been addressed adequately, it is clear that there is much work to be done on both the compilation of a morpheme dictionary and the development of word formation rules for English as part of text-to-speech conversion systems.

Examining the errors has led to the improvement of the algorithm. When first tested, it produced an accuracy of 82%. After vowel tensity was removed as a condition of syllable weight, and a few common prefixes were introduced, the accuracy improved to 86.9%.

While there are some words incorrectly assigned stress that can not be explained in terms of this approach, it is felt that the algorithm developed provides an adequate explanation and solution to the problem of automatic stress-assignment. The following chapter will represent an attempt to synthesize the major points made throughout this thesis.

## CONCLUSION

It is the purpose of this final chapter to summarize all that has been said previously, and put it into a perspective that can help us to understand the importance of algorithms of this type to theories of stress-assignment in linguistics. Because much of the information has been described in greater detail in the preceding chapters, the point of this chapter is brevity and clarity.

In terms of the major areas of discussion in the previous chapters, this conclusion will be divided into several areas. These are: the algorithm, further developments, and linguistic insights.

### The Algorithm

It is hoped that it has been shown that an algorithm can be briefly, economically, and explicitly stated that will accurately assign stress to a large majority of English words. The algorithm previously discussed has been based, not on theoretical concerns, but on pragmatic concerns, and it is felt that certain insights have been gained from this approach. While the algorithm does not have a theoretical basis to its design, it is hoped that a few points have been made through its testing and its development, that can be related to linguistic theory.

There are two components to the algorithm. The first stage involves the location of prefixes and suffixes. The second stage involves the application of a basic stress rule. If stress cannot be determined based on the accentual properties

of these morphemes, then a basic stress rule is applied to the stressable portion of the word. Both of these components reflect a simple analysis that can be arrived at quite efficiently.

Even though the algorithm has achieved an acceptable degree of accuracy of 86.9%, it has been designed so that any improvements that may be made to it can be done in a straight-forward and expedient manner.

### Further Developments

In terms of further developments to be made to the algorithm, there are a number of areas that require further research and development.

One area that necessitates further work is in developing an accurate resolution to the problem of morpheme decomposition. There are two aspects that need to be addressed in developing such a model. The first is to develop word-formation rules that will assist in decomposing words correctly. The second is in providing a more detailed specification of each morpheme. The information that needs to be specified will need to include: its accentual property, its possible part-of-speech as well as its etymology (Latin, Germanic). Many prefixes in English function in a variety of ways based on the etymology of the word to which they are attached.

Further, if an etymology recognizer was developed, to apply to words and names, then other, language specific stress rules may apply to create a more accurate algorithm. In particular it is hoped that rules can be developed to account for the large number of non-Indoeuropean loan words that have entered into the English language more recently.

It has been suggested that the inclusion of a VCV template will account for many of the loan words in English, as well as words that exhibit penultimate stress that cannot be accounted for in other terms. This template would function like an autostressed suffix, attracting stress onto the first vowel of the template. This would correctly locate stress on the following loan-words and place names: Arizona, Ticonderoga, linguini, legato, and tomato.

One point of expansion to the algorithm that will soon be undertaken is the inclusion of the mixed suffixes to the suffix dictionary. As was stated previously, this has been avoided because there was no access to bound roots. With the addition of bound roots to the algorithm, the use of mixed suffixes will improve the accuracy of the algorithm even further.

During the course of investigation it was found that there were some generalizations that can be made regarding mixed suffixes. For example, the suffix -able is stress-neutral if it is attached to a free root, and pre-stressed 2 if attached to a bound root. It was noted that suffixes that operate in this manner are always stress-neutral when attached to free roots. That is to say that the majority of mixed suffixes are stress-neutral when attached to words and may vary when attached to bound roots. This generalization can be taken advantage of in the following way. Mixed suffixes should be marked in the lexicon as mixed. The various accentual properties associated with the mixed suffix when it attaches to a bound root should also be indicated in the entry. If a mixed suffix is encountered then it should be necessary to check first if it is attached to a bound root. If it is attached to a bound root, then stress is located based on the accentual properties of the suffix. If it is not attached to a bound root, then it

must be attached to a word. The suffix then functions as a stress-neutral suffix and stress will be assigned based on this fact.<sup>16</sup> This capability will be incorporated into the algorithm in the future.

As one final point for improvement, it is necessary to include a syntactic analysis of the words in order to determine the stress patterns that are noted in homograph pairs. While there are only 150 noun-verb homograph pairs that are differentiated by stress, these are common words in English usage. As was noted, in the list of 485 words tested, 10 homographs were included. It will be necessary to enter these 150 words in an exceptions dictionary to be referred to after a syntactic analysis is conducted on the text. When the correct part of speech is determined by this syntactic analysis, stress should be located on the final syllable for verbs and on the penultimate syllable for nouns. Further analysis of part of speech will also assist in locating stress correctly on adjectives as well.

### Linguistic Insights

In terms of the linguistic insights gained from the development and testing of this algorithm, it was felt that two points were of significance. The first of these is that vowel quality was not a significant factor in determining syllable weight. It was found when this condition was included in an initial testing of the algorithm, there were more errors in stress prediction than when the condition was removed. In regard to linguistic theory, this may imply that reference can be made to only the syllabic tier in determining syllable weight.

---

<sup>16</sup> It should be noted that a second possibility exists. That is that it can be possible to order the decomposition to look for word-forming suffixes first, and then suffixes that attach to free words.

The second observation is that stress in English can be said to be morphologically determined. There are a number of facts which lend support to this idea. Recall that many of the errors made can be decreased by the inclusion of an accurate morphological decomposition. Particularly, the verbs incorrectly assigned stress were found to be of the structure prefix + bound root. With the addition of bound roots, this will eliminate the necessity to specify grammatical part-of-speech. Recall also that many stress-affecting suffixes are sensitive to the presence of bound roots. Bound roots are those which require the addition of an element to create a word. The combining of elements in this manner is said to occur at the word-formation stage. Therefore, many of the factors that contribute to the assignment of lexical stress in English occur at the word-formation stage and are morphologically determined.

While it was known that word stress in English is sensitive to the presence of affixes, the extent to which the morphology influences stress assignment was greatly underestimated in the design of the algorithm. It is felt that in the future, theories of lexical-stress assignment for English, as well as applications such as text-to-speech conversion, will require a complete morphological analysis.

Therefore, it can be said that in this algorithm stress is morphologically determined in cases where the word is not monomorphemic. In cases where the basic stress rule applies, stress is determined with reference to only the syllabic tier.

## Summary

In summary, it is felt that this algorithm provides an efficient, accurate and flexible solution to the problem of assigning primary stress to words in English. The work described in this thesis also confirms that providing a complete morphological analysis for English is important, and it is hoped that future research in the field of speech technology will reflect this.

## BIBLIOGRAPHY

- Allen, J. 1973. "Speech Synthesis from Unrestricted Text." in Flanagan, J.L. and Rabiner, L.R. (eds). Speech Synthesis. Dowden, Hutchinson and Ross Inc. Strousberg, Pennsylvania: 416-428.
- Allen, J., Hunnicutt, S. and Klatt, D.H. 1987. From Text to Speech: The MITalk System. Cambridge University Press. Cambridge. UK.
- Aronoff, M. 1976. Word Formation in Generative Grammar. MIT Press, Cambridge, Mass.
- Bates, Dawn. 1988. Prominence Relations and Structure in English Compound Morphology. unpublished Ph.D. dissertation, University of Washington.
- Bernstein, J. and Nessly, L. 1981. "Performance comparison of component algorithms for the phonemicization of orthography." Proceedings of the 19th Meeting of the Association of Computational Linguistics: 19-22.
- Chomsky, N. and M. Halle. 1968. The Sound Pattern of English. Harper and Row. New York.
- Church, K.W. 1985. "Stress Assignment in Letter-to-Sound Rules for Speech Synthesis." Proceedings of the 23rd Meeting of the Association for Computational Linguistics: 246-253.
- Clements, G.N. and S.J. Keyser. 1983. CV Phonology The MIT Press. Cambridge Mass.
- Dresher, B.E. and J.D. Kaye.. 1989. "A computational learning model for metrical phonology." to appear in Cognition.
- Elovitz, H.S., Johnson, R.J., McHugh, A., and Shore, J.E. 1976. Automatic Translation of English Text to Phonetics by means of Letter-to-Sound Rules. Naval Research Laboratory. NRL Report 7948.
- Fudge, E.C. (ed). 1973. Phonology. Modern Penguin Linguistics Readings. England.
- Fudge, E.C. 1984. English Word-Stress. George Allen and Unwin Publishers. U.K.
- Fry, D.B. 1955. "Duration and intensity as physical correlates of linguistic stress." Journal of the Acoustical Society of America 27:765-768.

- , 1958. "Experiments in the perception of stress." Language and Speech 1:126-152.
- , 1965. "The dependence of stress judgements on vowel formant structure." Proceedings of the 5th International Congress of Phonetic Sciences, Munster, 1964 (Basel-New York: S. Karger). pp. 36-311.
- Halle, M. and J.R. Vergnaud. 1987. An Essay on Stress. The MIT Press. Cambridge Massachusetts.
- Hayes, Bruce. 1982. "Extrametricity and English Stress." Linguistic Inquiry 13: 227-276.
- Hertz, S.R. 1982. "From text to speech with SRS." Journal of the Acoustical Society of America 74:1155-1170.
- Hill, K. and Nessly, L. 1973. "Review of the Sound Pattern of English." Linguistics 106: 57-121.
- Hunnicut, Sharon. 1976. "Phonological Rules for a Text-to-Speech System." American Journal of Computational Linguistics. Microfiche 57. 1-72.
- Kaisse, E.M. and P.A. Shaw. 1985. "On the theory of lexical phonology." Phonology Yearbook 2: 1-30.
- Kilbury, James. 1976. The Development of Morphophonemic Theory. John Benjamins B.V. Amsterdam.
- Kingdon, Roger. 1958. The Groundwork of English Stress. Longmans, Green and Co. Ltd. London.
- Klatt, D.H. 1987. "Review of text-to-speech conversion for English." Journal of the Acoustical Society of America. 82(3). 737-793.
- Kreidler, C.W. 1987. "Stress differentiation in sets of English words." Word 38(2): 99-125.
- Kucera, H., and Francis, W.N. 1967. Computational Analysis of Presentday American English. Brown U.P. Providence, R.I.
- Lehiste, Ilse. 1970. Suprasegmentals. The M.I.T. Press. Cambridge, Massachusetts.
- Liberman, M. and A. Prince. 1977. "On stress and linguistic rhythm." Linguistic Inquiry 8:249-336.
- Mohanan, K.P. 1986. The Theory of Lexical Phonology. Reidel Publishing Company. Boston.

- Pei, Mario. 1966. Glossary of Linguistic Terminology. Anchor Books. New York.
- Poldauf, Ivan. 1984. English Word Stress: A Theory of Word-Stress Patterns in English. Pergamon Press, Oxford.
- McAllister, J.M. 1986. "The implementation of lexical stress rules in the CSTR text-to-speech system." Proceedings of the I.O.A. V8. pt. 7. pp. 403-407.
- O'Shaugnessy, D. 1987. "Specifying intonation in a text-to-speech system using only a small dictionary". Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing pp 1430-1433.
- Selkirk, E.O. 1982. The Syntax of Words. Linguistic Inquiry Monograph 7. Cambridge Mass. MIT Press.
- . 1984. Phonology and Syntax: The Relation between Sound and Structure. MIT Press. Cambridge Massachusetts.
- Sherman, D. 1975. "Noun-verb stress alternation: An example of the lexical diffusion of sound change in English." Linguistics 159: 43-71.
- Urbanczyk, S.C. 1988. "A preliminary investigation of the acoustic context of homograph pairs." Working Papers of the Linguistics Circle of the University of Victoria. Vol.7: 135-146.
- Urbanczyk, S.C. and S.J. Eady. 1989. "Assignment of syllable stress in a demisyllable-based text-to-speech synthesis system." Proceedings of the IEEE Pacific Conference on Communications, Computers and Signal Processing: 467-470.
- Urbanczyk, S.C., S.J. Eady and B.C. Dickson. 1987. "Durational adjustments in a demisyllable-based speech synthesis system." Proceedings of the Canadian Acoustical Association. October 1987: 79-84.
- Woo, L. J.A.W. Clayards, and S.J. Eady. 1988. "The Text-to-demisyllable system: A report on development of software for converting English orthographic text into demisyllable units." CSTR Technical Report No. SS8801.
- Woods, Howard, B. 1979. Syllable Stress and Unstress. Canadian Government Publishing Centre. Hull, Quebec.

APPENDIX A  
PRACTICAL ORTHOGRAPHY

REPRESENTATION	PHONEME
II	<u>beat</u>
I	<u>bit</u>
EI	<u>bait</u>
E	<u>bet</u>
AE	<u>bat</u>
UU	<u>boot</u>
U	<u>put</u>
OI	<u>boy</u>
OU	<u>boat</u>
A	<u>bought</u>
AI	<u>buy</u>
AU	<u>proud</u>
UE	<u>but</u>
OE	<u>about</u>
EOR	<u>her</u>
OR	<u>bore</u>
ION	<u>kitten</u>

P	pea
T	tea
K	key
B	bee
D	deed
G	<u>geese</u>
F	fee
TH	<u>theif</u>
V	vow
DH	the
S	sea
SH	she
CH	<u>cheap</u>
Z	soo
ZH	<u>beige</u>
JH	judge
H	he
M	me
N	knee
NG	<u>sing</u>
L	lee
R	raw
Y	ye
W	we

## APPENDIX B

### AFFIXES

#### Suffixes

The following list of suffixes contains the orthographic representation of the suffix (OR) followed by the abstract linguistic representation (AR) which contains both the phonemic representation and the stress predicting diacritic.

OR	AR	OR	AR
ABLE	OEBOEL	ABLY	OEBLII
AC	AK	ACEOUS	/EISHOES
ACRE	EIKEOR	ACY	OESII
AD	^AED	ADE	/EID
AE	II	AEAN	/IIOEN
AGE	OEJH	AIRE	/ER
AIS	/EI	AISE	/EIZ
AL	OEL	ALIA	/EILIIOE
ALITY	/AELITII	ALLY	LII
AM	OEM	AN	^OEN
ANCE	OENS	ANCY	OENSII
ANNE	/AEN	ANT	OENT

ANTHROPIC	>OENTHRAPIK	ANTHROPISE	/AENTHROEPAIZ
ANTHROPISM	/AENTHROEPIZOEM		
ANTHROPIST	/AENTHROPIST		
ANTHROPY	/AENTHROPII	AR	^OER
ARD	φEORD	ARIAN	/ERIIOEN
ARIUM	/ERIIOEM	ARIUS	/ERIIOES
ARY	ERII	ASSE	/AES
AST	<OEST	ATE	<EIT
ATIM	/EITIM	ATION	/EISHOEN
ATIVE	^OETIVE	ATOR	EITOR
ATORY	<OETORII	ATRIC	/AETRIK
ATURE	^OECHEOR	AX	φAEKS
BANK	BAENGK	BARROW	BEROU
BATCH	BAECH	BECK	BEK
BERG	BEORG	BORG	BORG
BOROUGH	BEOROE	BOURG	BURG
BOURNE	BEORN	BRIDGE	BRIJH
BROKE	BROUK	BROOK	BROEK
BURG	BEORG	BURGH	BUEROE
BURN	BEORN	BURNE	BEORN
BURY	BEORII	BY	BII
CASTER	KAESTEOR	CE	S
CESTER	CHOESTEOR	CESTER	STEOR
CHESTER	CHOESTEOR	CHROMATIC	>KROUMATIK
CHRONE	φKROEN	CHRONIC	/KRANIK

CHRONISM	ϕKROENIZOEM	CHRONOUS	ϕKROENOEES
CHRONY	ϕKROENII	CHURCH	CHEORCH
CIDAL	/SAIDOEL	CIDE	<SAID
COLOUS	ϕKOELOES	COMB	KOEM
COMBE	KOEM	COT	KAT
COTT	KAT	CRACY	ϕKROESII
CULE	<KYUUL	CULTURAL	/KUELCHEROEL
CULTURE	<KUELCHOR	CY	SII
D	OED	D	D
D	T	DALE	DEIL
DAY	DEI	DELL	DEL
DEN	DOEN	DOM	DOEM
DON	DOEN	DON	DAN
DONT	/DANT	EAN	OEN
EAN	/IIOEN	EDY	OEDII
EE	/II	EE	/EI
EEN	/IIN	EER	/IIR
ELLE	/EL	ENCE	^OENS
ENCY	^OENSII	ENDA	/ENDOE
ENDUM	/ENDOEM	ENE	<IIN
ENNE	/EN	ENSIS	/ENSIS
ENT	^OENT	ENTIAL	/ENSHOEL
ENTIOUS	/ENSHOES	EOUS	ϕOES
ERIE	ϕEORII	ES	IIZ
ESCE	/ES	ESCENCE	/ESOENS

ESCENT	/ESOENT	ESE	/IIZ
ESQUE	/ESK	ESS	ES
ET	/ET	ETH	OETH
ETTE	/ET	ETY	¢OETH
EUM	/IIOEM	EUR	/EOR
EUS	¢IIOES	EUSE	/OEZ
EX	¢EKS	EY	II
FACTION	/FAEKSHOEN	FICATION	>FIKEISHOEN
FIELD	FIILD	FOLD	FOLD
FOLK	FOEK	FORD	FEORD
FORM	<FORM	FUL	FOEL <sub>1</sub>
FY	<FAI	GARTH	GARTH
GATE	GEIT	GILL	GIL
GNON	/NYOUN	GOGY	GAJHII
GON	<GAN	GONAL	¢GOENOEL
GRAD	GRAED	GRAPH	<GRAEF
GRAPHER	¢GROEFEOR	GRAPHIC	/GRAEFIK
GRAPHIST	¢GROEFIST	GRAPHY	GROEFII
GUARD	GEORD	GUE	/G
HALL	HAL	HAM	OEM
HAMPTON	/HAEMPTOEN	HART	HART
HAY	HEI	HEAD	HED
HILL	HIL	HOLME	HOUM
HOME	HOUM	HOOD	HUD
HOPE	HOUP	HOUSE	HAUS

HURST	HEORST	I	ϕII
IA	ϕOE	IAC	ϕIIAK
IACAL	/AIOEKOEL	IAL	ϕOEL
IAL	ϕPIOEL	IAN	ϕOEN
IAN	ϕPIOEN	IANA	ϕIIAENOE
IAN T	ϕOENT	IAN T	IIOENT
IASIS	/AIOESOES	IBLE	^OEBOEL
IC	ϕISH	IC	ϕIS
IC	ϕIK	ICA	ϕIKA
ICE	<IS	ICIAN	/ISHOEN
ISIOUS	/ISHOES	ICITY	/ISOETH
ICUL	/IKYUUL	ID	ϕID
IDA	/OEDOE	IDAE	/OEDII
IDE	^AID	IDITY	/IDOETH
IDO	/IDOU	IENCE	ϕOENS
IENCE	ϕPIOENS	IENCY	ϕOENSII
IENCY	ϕPIOENSII	IENT	ϕOENT
IENT	ϕPIOENT	IER	/IIR
IERE	/YER	IFEROUS	/IFEOROES
IFIC	/IFISH	IFIC	/IFIS
IFIC	/IFIK	IFICATION	~OEFIKEISHOEN
IFICENCE	/IFISOENS	IFICENT	/IFOESOENT
IFY	ϕFIAI	ILE	^AIOEL
IN	IN	INE	^AIN
INE	/IIN	ING	ING

INOUS	<INOES	ION	ϕOEN
ION	ϕIOEN	IONAL	ϕOENOEL
IONAL	ϕIOENOEL	IONATE	ϕOENOET
IONATE	ϕIOENOET	IOUS	ϕOES
IOUS	ϕIOES	IQUE	/IIK
IS	^IS	ISE	/AIZ
ISH	ISH	ISM	IZOEM
ISSIM	/ISIM	ISSIMA	/ISHIMA
ISSIMO	/ISHIMO	IST	IST
ISTE	/ISTE	ISTER	OESTEOR
ITE	AIT	ITION	/ISHOEN
ITIOUS	/ISHOES	ITIS	/AITIS
ITIVE	ϕOETIV	ITOR	ϕITEOR
ITORY	ϕITORII	ITUDE	ϕITUUD
ITURE	ϕICHOER	ITY	ϕITII
IUM	ϕIOEM	IUM	ϕOEM
IUS	ϕOES	IUS	ϕIOES
IVE	^IV	IVITY	ϕIVITII
IZABLE	AIZOEBOEL	IZATION	>AIZEISHOEN
IZE	AIZ	KIN	KIN
KIRK	KEORK	LAND	LOEND
LATER	ϕLOETEUR	LATRY	ϕLOETRII
LAW	LA	LENT	LOENT
LESS	LOES	LET	LOET
LEY	LI	LIKE	LAIK

LING	LING	LITHIC	/LITHIK
LOGGER	¢LOEJHEOR	LOGIAN	/LOUJHOEN
LOGICAL	/LAJHIKOEL	LOGISM	¢LOEJHIZOEM
LOGIST	¢LOEJHIST	LOGUE	LAG
LOGY	¢LOEJHII	LOW	LOU
LY	LII	LYN	LIN
MAN	MOEN	MANN	MAN
MENT	/MAN	MENT	MOENT
MENTAL	/MENTOEL	MENTARY	/MENTRII
MERE	MIIR	METER	¢MOETEUR
METRIC	/METRIK	METRY	¢MOETRII
MONGER	MANGGEOR	MONGERY	MANGGEORII
MONIAL	/MOUNIOEL	MONIOUS	/MOUNIOES
MONY	<MOUNII	MOOR	MUR
MORE	MOR	MORPHIC	/MORFIK
MORPHISM	¢MORFIZOEM	MORPHOUS	/MORFOES
MORPHY	MORFII	MOST	MOUST
MOUTH	MOETH	N	OEN
N	N	N'T	IONT
NAUT	NAT	NESS	NOES
NOMIC	/NAMIK	NOMOUS	¢NOEMOES
NYMIC	/NIMIK	NYMITY	/NIMOETII
NYMOUS	¢NOEMOES	OID	ˆOID
OIR	<WAR	OIRE	<WAR
OIS	<WA	OISE	<WAZ

OLAR	/OULEOR	OLE	OUL
OLEUM	/OULIOEM	LOGY	/ALOEJHII
OMA	/OUMOE	ON	AN
ON	^OEN	ONER	OENOER
ONNE	/AN	OO	/UU
OON	/UUN	OR	EOR
ORIAL	/ORIOEL	ORIAN	/ORIOEN
ORIOUS	/ORIOES	ORIUM	/ORIOEM
OROUS	ϕEOROES	ORY	EORII
OSE	<OUS	OSIS	/OUSOES
OSITY	/ASITII	OSSE	/AS
OTTE	/AT	OUIN	/WEN
OUR	^EOR	OUS	OES
OVER	OUVEOR	PARK	PARK
PAROUS	ϕPEOROES	PATHIC	/PAETHIK
PATHY	ϕPOETHII	PHOBIA	/FOUBIOE
PHOBIC	/FOUBIK	PHONIC	/FANIK
PHONIST	ϕFOENIST	PHONY	ϕFOENII
PHORIC	/FORIK	POLIS	POELIS
PORT	PORT	QUET	/KEI
R	EOR	RAMA	/RAEMOE
REL	ROEL	RICH	RICH
RY	RII	S	OEZ
S	Z	S	S
SALL	SAL	SCOPIC	/SKAPIK

SCOPY	¢SKOEPII	SEA	SII
SELF	/SELF	SELVES	/SELVZ
SET	SOET	SHAW	SHA
SHIP	SHIP	SHIRE	SHEOR
SIDE	SAID	SOME	SOEM
SOM	SOEN	SPHERIC	/SFERIK
ST	OEST	STABLE	STOEBOEL
STAN	/STAEN	STEAD	STED
STER	STEOR	STOKE	STOUK
STON	STOEN	STONE	STOUN
STOW	STOU	STREET	STRIIT
SY	SII	SY	ZII
TARN	TARN	TEEN	/TIIN
TH	TH	THORP	THORP
THORPE	THORP	THROP	THROEP
THWAITE	THWEIT	TOFT	TOFT
TON	TOEN	TWISTLE	TWISOEL
TY	TI	UAL	¢YUUOEL
UAL	¢UUOEL	UAL	¢WOEL
UANT	¢YUUOENT	UANT	¢UUANT
UATE	¢UUEIT	UATE	¢UUEIT
UBLE	¢YUUBOEL	ULA	¢YUULA
ULA	¢UULA	ULET	¢YUULOET
UM	OEM	UOUS	¢YUOOES
UOUS	¢UOOES	UR	/EOR

URE	YEOR	URE	EOR
US	OES	UTE	<YUUT
UTE	<UUT	VILLE	VIL
VOROUS	¢VEOROES	WALL	WAL
WARD	WEORD	WAY	WEI
WELL	WEL	WICH	IJH
WICK	WIK	WILL	WIL
WISE	WAIZ	WOLD	WOULD
WOOD	WUD	WORT	WEORT
WORTH	WEORTH	WORTHY	WEORDHII
WRIGHT	RAIT	Y	II
YARD	YARD	'D	D
'LL	OEL	'LL	L
'VE	V		

### Prefixes

PREFIX	ABSTRACT REPRESENTATION
A	/EI
A	>OE
AP	>AEP
BE	>BOE
DE	>DOE
EX	>EGZ

EX	>EKS
OB	>AB
OC	>AK
OF	>AF
OP	>AP
RE	>ROE
SE	>SOE
WITH	>WITH

**APPENDIX C**  
**CORPUS**

The two corpii used in the evaluation of the algorithm are provided in the following lists.

**Brown Corpus**

The following is a list of the corpus of words used in the testing of the algorithm. It contains the first 485 most frequent words from the Brown Corpus. Beside the word is the stress-pattern assigned by the algorithm if stress was assigned incorrectly.

WORD	WORD
ABOUT (ABout)	INTO
ONLY	OTHER
ANY	OVER
EVEN	AFTER
ALSO	MANY
BEFORE	BECAUSE (BECause)
PEOPLE	LITTLE
VERY	BETWEEN
BEING	UNDER
NEVER	ANOTHER (ANother)

AGAINST (AGainst)	HIMSELF
DURING	WITHOUT
AGAIN (AGain)	AMERICAN
AROUND (ARound)	HOWEVER (HOWever)
GENERAL	UPON
EVERY	UNITED (UNited)
NUMBER	UNTIL (UNtil)
ALWAYS	AWAY (AWay)
SOMETHING	WATER
PUBLIC	ALMOST
ENOUGH (ENough)	GOVERNMENT
SYSTEM	BETTER
NOTHING	GOING
LATER	PROGRAM
CITY	BUSINESS
TOWARD (TOWard)	SOCIAL
GIVEN	PRESENT
SEVERAL	ORDER
NATIONAL	POSSIBLE
RATHER	SECOND (seCOND)
AMONG (AMong)	IMPORTANT
OFTEN	EARLY
BECAME	WITHIN
ALONG (ALong)	CHILDREN
POWER	DEVELOPMENT (DEVelopment)

FAMILY	INTEREST
MEMBERS	COUNTRY
AREA	OTHERS
ALTHOUGH (ALthough)	OPEN
SERVICE	CERTAIN
PROBLEM	BEGAN
DIFFERENT	MATTER
TRAGEDY	CONSEQUENCE
ABSENT	YONDER
CHINESE	PICTURESQUE
HOSTESS	MODEST
BEFOREHAND	MANHOOD
VISIBLE	JUSTICE
PARTICLE	VERIFY
ALIMONY	WORKING
RATIO	JETTISON (jetTISon)
MORALITY	VEHICLE
KNOWLEDGE	HOPELESS
JUDGEMENT	HAPPINESS
KINDRED	FRIENDSHIP
HANDSOME	SOLITUDE
ATTITUDE	PERHAPS (PERhaps)
ITSELF	HUMAN
ABOVE (Above)	EXAMPLE
ACTION	COMPANY

LOCAL	HISTORY
WHETHER	EITHER
TODAY (TOday)	ACROSS (ACross)
TAKEN	ANYTHING
HAVING	EXPERIENCE
REALLY	ALREADY (ALready)
THEMSELVES	INFORMATION
COLLEGE	TOGETHER (TOgether)
MONEY	PERIOD
PROBABLY	BEHIND
CANNOT	POLITICAL
QUESTION	MAKING
OFFICE	SPECIAL
MAJOR	PROBLEMS
AGO (AGo)	BECAME
FEDERAL	MOMENT
STUDY	AVAILABLE (AVailable)
RESULT	ECONOMIC
POSITION	REASON
INDIVIDUAL	SOCIETY
AREAS	COMMUNITY
FUTURE	WANTED
DEPARTMENT (DEPartment)	CENTER
WOMAN	COMMON
CONTROL (CONtrol)	NECESSARY

POLICY	FOLLOWING
SOMETIMES	FURTHER
ABLE	MOTHER
MUSIC	PARTY
PROVIDE (PROvide)	EDUCATION
UNIVERSITY	EFFECT
LEVEL	STUDENTS
MILITARY	MORNING
TOTAL	OUTSIDE (OUTside)
FIGURE	CENTURY
USUALLY	WASHINGTON
THEREFORE	EVIDENCE
MILLION	VARIOUS
BELIEVE	SURFACE
VALUE	MODERN
TABLE	MINUTES
PERSONAL	PROCESS
SITUATION	ALONE (ALone)
ENGLISH	IDEA
INCREASE	WOMEN
AMERICA	LIVING
STARTED	LONGER
FINALLY	NATURE
PRIVATE	SECRETARY
SECTION	GREATER

EXPECTED (EXpected)

VALUES

PRESSURE

FATHER

SPIRIT

COMPLETE (COMplete)

CONDITIONS

SUPPORT (SUPport)

PARTICULAR

BEYOND (BEyond)

NATIONS

TAKING

INSIDE (INside)

REPORT

INSTEAD (INstead)

ADDED

FEELING

SINGLE

HUNDRED

INDUSTRY

DEVELOPED (DEVelped)

COMMITTEE (commitTEE)

EQUIPMENT (EQUipment)

ACTUALLY

RIVER

NEEDED

EVERYTHING

BASIS

REQUIRED (REquired)

UNION

EXCEPT

RETURN

ATTENTION

RECENT (REcent)

FORCES

PERSON

COMING

MATERIAL

DATA

LOOKING

AMOUNT (AMount)

FOLLOWED

BASIC

INCLUDING (INCluding)

RESEARCH

SIMPLY

DEFENSE

ISLAND

RELIGIOUS

RIVER

BEGINNING	CENTRAL
GETTING	DOING
RECEIVED (REceived)	TRYING
INDEED	MEDICAL
PICTURE	ADMINISTRATION
DIFFICULT	SIMPLE
SUBJECT	BUILDING
ESPECIALLY	HIGHER
MEETING	FOREIGN
PAPER	SIMILAR
FINAL	NATURAL
PROPERTY	TRAINING
COUNTRY	INTERNATIONAL
MARKET (marKET)	POLICE (POLice)
ENGLAND	WRITTEN
STORY	SUDDENLY
ANSWER	CONGRESS
ISSUE	CONSIDERED (CONsidered)
COUNTRIES	LIKELY
WORKING	ENTIRE (ENTire)
HAPPENED	LABOUR
PURPOSE	RESULTS
CASES	DIFFERENCE
PRODUCTION	WILLIAM
INVOLVED (INvolved)	EARLIER

INCREASED	PARTICULARLY
BELOW	EFFORT
KNOWLEDGE	LETTER
THINKING	USING
CHRISTIAN	CERTAINLY
IDEAS (IDEas)	INDUSTRIAL
READY	ADDITION
METHOD	MORAL
COLOUR	DECIDED (DECided)
DIRECTLY (DIRectly)	NEARLY
NEITHER	STATEMENT
THROUGHOUT (THROUGHout)	ANYONE
KENNEDY	QUESTIONS
READING	ACCORDING (ACcording)
NATION	PROGRAMS
SERVICES	PHYSICAL
REMEMBER (REMember)	MEMBER
RECORD	SOUTHERN
UNDERSTAND (unDERstand)	WESTERN
NORMAL	POPULATION
APPEARED (APpeared)	CONCERNED (CONcerned)
DISCTRICT	MERELY
TEMPERATURE	VOLUME
DIRECTION	MAYBE
SUMMER	TRIAL

TROUBLE	CONTINUED
EVENING	LITERATURE
ARMY	ASSOCIATION
GENERALLY	INFLUENCE
PROVIDED (PROvided)	CHANGES
FORMER	HUSBAND
OPENED	SCIENCE
STUDENT	AVERAGE
SERIES	DIRECT (DIRect)
EFFECTIVE	MYSELF
PLANNING	SOVIET
SYSTEMS	THEORY
CLEARLY	FREEDOM
MOVEMENT	BEAUTIFUL
CONSIDER (CONsider)	EFFORTS
MEANING	ORGANIZATION
SOMEWHAT	TREATMENT
HOTEL (HOTel)	APPARENTLY (APparently)
CARRIED	DEGREE
EASY	HERSELF
NUMBERS	RESPECT
MANNER	APPROACH
REACTION	IMMEDIATELY
LARGER	LOWER
RECENTLY	RUNNING

COUPLE	DAILY
PERFORMANCE	OPPORTUNITY
PERSONS	
UNDERSTANDING (unDERstanding)	
ADDITIONAL	DESCRIBED (DEScribed)
PROGRESS	RADIO
TECHNICAL	DECISION
DETERMINED	IMAGE
RELIGION	REPORTED
WINDOW	APPEAR
BRITISH	CHARACTER
EUROPE	MIDDLE
RESPONSIBILITY	ACCOUNT (ACcount)
WRITING	ACTIVITY
FISCAL	SERIOUS
ACTIVITIES	AUDIENCE
CORNER	FORWARD
LETTERS	NUCLEAR
OBTAINED (OBtained)	RETURNED (REturned)
SLOWLY	SPECIFIC
DESIGN	JUSTICE
LATTER	MOVING
OBVIOUSLY	QUALITY
FIGURES	FUNCTION
INCLUDE (INclude)	OPERATION

PATTERN	SAYING
SEVEN (seVEN)	WHATEVER
COMPLETELY (COMpletely)	EXTENT
HEAVY	HOSPITAL
STANDARD	WAITING
AHEAD	DEMOCRATIC
EFFECTS (EFFECTs)	INCOME
LANGUAGE	PRINCIPLE
VISIT	ANALYSIS
DESIGNED (DEsigned)	

### Running Text

The following is the running text used to evaluate the accuracy of the stress-assignment algorithm developed in this thesis.

TO HEAR THE INFORMATION, TOUCH THE BOTTOM LEFT OF THE SCREEN.

FOR ENGLISH, TOUCH THE BOTTOM RIGHT OF THE SCREEN. PLEASE TOUCH THE SQUARE NEXT TO THE ITEM YOU WANT.

TO OBTAIN A COPY OF THIS MAP FROM THE PRINTER, TOUCH THE BOTTOM CENTRE OF THE SCREEN.

TO RETURN TO THE MENU, TOUCH THE BOTTOM RIGHT OF THE SCREEN.

THIS IS THE MAIN MENU. TOUCH THE SQUARE NEXT TO THE ITEM YOU WANT. THE TOP ITEM IS AIRPORT INFORMATION. ITEM TWO IS BUSES, TAXIS, RENTALS. LAST ITEM, HOTEL INFORMATION.

AIRPORT INFORMATION. TOP ITEM, FLIGHT ARRIVALS. ITEM TWO, FLIGHT DEPARTURES. ITEM THREE, TICKET COUNTER INFORMATION. LAST ITEM, AIRPORT FACILITIES.

AIRLINE TICKET COUNTERS. TOP ITEM, U.S. DESTINATIONS. ITEM TWO, DOMESTIC DESTINATIONS. LAST ITEM, OVERSEAS DESTINATIONS.

U.S. DESTINATIONS. LEVEL TWO. TOP ITEM, AIR CANADA. ITEM TWO, AMERICAN AIRLINES. ITEM THREE, CANADIAN AIRLINES. TO CONTINUE THE LIST, TOUCH THE LAST ITEM.

THE AIR CANADA COUNTER FOR U.S. DESTINATIONS IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FIFTY PACES. AIR CANADA IS ON YOUR LEFT.

THE AMERICAN AIRLINES COUNTER IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WALK ABOUT FIFTY PACES. AMERICAN AIRLINES IS ON YOUR RIGHT.

THE CANADIAN AIRLINES COUNTER FOR U.S. DESTINATIONS IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FIFTY PACES. CANADIAN AIRLINES IS ON YOUR LEFT.

U.S. DESTINATIONS. LEVEL TWO. CONTINUED TOP ITEM, CONTINENTAL AIRLINES. ITEM TWO, DELTA AIRLINES. ITEM THREE, SAN JUAN AIRLINES. LAST ITEM, UNITED AIRLINES.

THE CONTINENTAL AIRLINES COUNTER IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FORTY PACES. CONTINENTAL AIRLINES IS ON YOUR LEFT.

THE DELTA AIRLINES COUNTER IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FIFTY PACES. DELTA AIRLINES IS ON YOUR RIGHT.

THE SAN JUAN AIRLINES COUNTER IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FORTY PACES. SAN JUAN AIRLINES IS ON YOUR RIGHT.

THE UNITED AIRLINES COUNTER IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FORTY PACES. UNITED AIRLINES IS ON YOUR RIGHT.

DOMESTIC DESTINATIONS. LEVEL THREE. TOP ITEM, AIR B.C. ITEM TWO, AIR CANADA. ITEM THREE, AQUILA AIR. FOURTH ITEM, BURRARD AIR. TO CONTINUE THE LIST, TOUCH THE LAST ITEM.

THE AIR B.C. COUNTER IS DIRECTLY BEHIND YOU AT SIX O'CLOCK. AIR B.C. MAKES USE OF THE AIR CANADA COUNTER.

THE AIR CANADA COUNTER IS DIRECTLY BEHIND YOU AT SIX O'CLOCK.

AQUILA AIR IS LOCATED AT THE SOUTH TERMINAL. DIRECT TELEPHONE AND BUS SERVICES TO THE SOUTH TERMINAL ARE AVAILABLE ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO THREE O'CLOCK. WALK ABOUT EIGHTY PACES. THE TELEPHONE IS TO THE RIGHT OF THE MAIN ENTRANCE.

DOMESTIC DESTINATIONS. LEVEL THREE. CONTINUED. TOP ITEM, CANADIAN AIRLINES. ITEM TWO, CANADIAN FORCES AIRLINES. ITEM THREE, SKYLINK AIRLINES. LAST ITEM, TIME AIR.

THE CANADIAN AIRLINES COUNTER IS LOCATED ON THIS LEVEL, AT TWELVE O'CLOCK. WALK ABOUT FIFTY PACES. CANADIAN AIRLINES IS ON YOUR LEFT.

THE CANADIAN FORCES AIRLINES COUNTER IS DIRECTLY BEHIND YOU AT SIX O'CLOCK. CANADIAN FORCES AIRLINES MAKES USE OF THE AIR CANADA COUNTER.

SKYLINK AIRLINES IS LOCATED AT THE SOUTH TERMINAL. DIRECT TELEPHONE AND BUS SERVICES TO THE SOUTH TERMINAL ARE AVAILABLE ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO THREE O'CLOCK. WALK ABOUT EIGHTY PACES. THE TELEPHONE IS TO THE RIGHT OF THE MAIN ENTRANCE.

THE TIME AIR COUNTER IS LOCATED ON THIS LEVEL AT TWELVE O'CLOCK. WALK ABOUT FIFTY PACES AND TURN LEFT. TIME AIR MAKES USE OF THE CANADIAN AIRLINES COUNTER.

OVERSEAS DESTINATIONS. LEVEL THREE. TOP ITEM, AIR CANADA. ITEM TWO, AIR CHINA. ITEM THREE, AIR NEW ZEALAND. FOURTH ITEM, BRITISH AIRWAYS. TO CONTINUE THE LIST, TOUCH THE LAST ITEM.

THE AIR CANADA COUNTER IS DIRECTLY BEHIND YOU AT SIX O'CLOCK.

THE AIR CHINA COUNTER IS DIRECTLY AHEAD OF YOU AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES AND TURN LEFT. AIR CHINA MAKES USE OF THE CANADIAN AIRLINES COUNTER.

THE AIR NEW ZEALAND COUNTER IS LOCATED ON THIS LEVEL AT TWELVE O'CLOCK. WALK ABOUT FIFTY PACES AND TURN LEFT. AIR NEW ZEALAND MAKES USE OF THE CANADIAN AIRLINES COUNTER.

THE BRITISH AIRWAYS COUNTER IS DIRECTLY AHEAD OF YOU AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES AND TURN LEFT. BRITISH AIRWAYS MAKES USE OF THE CANADIAN AIRLINES COUNTER.

OVERSEAS DESTINATIONS. LEVEL THREE. CONTINUED. TOP ITEM, CANADIAN AIRLINES. ITEM TWO, CANADIAN FORCES AIRLINES. ITEM THREE, CATHAY PACIFIC. FOURTH ITEM, JAPAN AIRLINES. TO CONTINUE THE LIST, TOUCH THE LAST ITEM.

THE CANADIAN AIRLINES COUNTER IS LOCATED ON THIS LEVEL, AT TWELVE O'CLOCK. WALK ABOUT FIFTY PACES. CANADIAN AIRLINES IS ON YOUR LEFT.

THE CANADIAN FORCES AIRLINES COUNTER IS DIRECTLY BEHIND YOU AT SIX O'CLOCK. CANADIAN FORCES AIRLINES MAKES USE OF THE AIR CANADA COUNTER.

THE CATHAY PACIFIC COUNTER IS LOCATED ON THIS LEVEL AT SIX O'CLOCK. WALK ABOUT FIFTY PACES. CATHAY PACIFIC IS ON YOUR RIGHT.

THE JAPAN AIRLINES COUNTER IS DIRECTLY AHEAD OF YOU AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES AND TURN LEFT. JAPAN AIRLINES MAKES USE OF THE CANADIAN AIRLINES COUNTER.

OVERSEAS DESTINATIONS. LEVEL THREE. CONTINUED. TOP ITEM, KLM AIRLINES. ITEM TWO, KOREAN AIRLINES. ITEM THREE, LUFTHANSA. FOURTH ITEM, QANTAS AIRLINES. TO CONTINUE THE LIST, TOUCH THE LAST ITEM.

THE KLM AIRLINES COUNTER IS LOCATED ON THIS LEVEL AT TWELVE O'CLOCK. WALK ABOUT FIFTY PACES AND TRUN LEFT. KLM AIRLINES MAKES USE OF THE CANADIAN AIRLINES COUNTER.

THE KOREAN AIRLINES COUNTER IS LOCATED ON THIS LEVEL AT SIC O'CLOCK. WALK ABOUT FIFTY PACES. KOREAN AIRLINES IS ON YOUR IRGHT.

THE LUFTHANSA COUNTER IS LOCATED ON THIS LEVEL AT SIX O'CLOCK. WALK AB OUT FIFTY PACES. LUFTHANSA IS ON YOUR RIGHT.

THE QANTAS AIRWAYS COUNTER I LOCATED ON THIS LEVEL AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES AND TURN LEFT. QANTAS AIRWAYS MAKES USE OF THE CANADIAN AIRLINES COUNTER.

OVERSEAS DESTINATIONS. LEVEL THREE. CONTINUED. TOP ITEM, SINGAPORE AIRLINES. LAST ITEM, WARDAIR.

THE SINGAPORE AIRLINES COUNTER IS LOCATED ON THIS LEVEL AT SIX O'CLOCK. WALK ABOUT FIFTY PACES. SINGAPORE AIRLINES IS ON YOUR RIGHT.

THE WARDAIR COUNTER IS LOCATED ON THIS LEVEL AT NINE O'CLOCK. WALK ABOUT FIFTY PACES. YOU WILL PASS THE STAIRWELL ON YOUR RIGHT.

AIRPORT FACILITIES. TOP ITEM, FOOD AND DRINK. ITEM TWO, ELEVATORS, ESCALATORS AND STAIRS. ITEM THREE, WASHROOMS. FOURTH ITEM, TELEPHONES. LAST ITEM, CURRENCY EXCHANGE.

FOOD AND DRINK. TOP ITEM, RESTAURANT. ITEM TWO, CAFETERIA AND SNACKS. LAST ITEM, COCKTAILS.

THE SEA ISLAND DINING ROOM IS LOCATED ON THIS LEVEL AT TEN O'CLOCK. WALK ABOUT EIGHTY PACES. YOU WILL PASS THE STAIRWELL AND THE NEWS STAND ON YOUR RIGHT.

THE JOLLY CHEF CAFETERIA IS LOCATED ON THIS LEVEL. WALK ABOUT FIFTY PACES AT NINE O'CLOCK. YOU WILL PASS THE STAIRWELL ON YOUR RIGHT. AT THE WARDAIR COUNTER, TURN LEFT AND WALK ABOUT FIFTY PACES. THE JOLLY CHEF CAFETERIA IS ON YOUR RIGHT.

THE SEA ISLAND LOUNGE IS LOCATED ON THIS LEVEL, AT TEN O'CLOCK. WALK ABOUT EIGHT PACES. YOU WILL PASS THE STAIRWELL AND THE NEWS STAND ON YOUR RIGHT.

THE CHIEFTAIN LOUNGE IS LOCATED ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO SIX O'CLOCK. WALK ABOUT FIFTY PACES. YOU WILL PASS THE STAIRWELL ON YOUR RIGHT.

THE SNACK BAR. LEVEL ONE. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE SNACK BAR IS AT ONE O'CLOCK. WALK ABOUT FORTY PACES.

THE HAPPY LANDING LOUNGE. LEVEL ONE. TAKE THE CLOSEST ELEVATOR WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR THE HAPPY LANDING LOUNGE IS AT TWO O'CLOCK. WALK ABOUT FORTY PACES.

ELEVATORS, ESCALATORS AND STAIRS. TOP ITEM, ELEVATORS. ITEM TWO, ESCALATORS. LAST ITEM, STAIRS.

ELEVATORS. THE CLOSEST ELEVATOR IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. THIS IS LEVEL THREE. WHEN LEAVING THE ELEVATOR, ON LEVEL ONE, AND LEVEL TWO. THE DOORS OPEN ON THE OPPOSITE SIDE.

THE DESCENDING ESCALATOR IS BESIDE YOU AT NINE O'CLOCK. WALK ABOUT TWENTY PACES, FOLLOWING THE HANDRAIL ON YOUR RIGHT. AT THE END ON THE HANDRAIL, TURN AROUND. THE DESCENDING ESCALATOR IS TO THE RIGHT OF THE STAIRWELL.

THE STAIRWELL IS BESIDE YOU AT NINE O'CLOCK. WALK ABOUT TWENTY PACES, FOLLOWING THE HANDRAIL ON YOUR RIGHT. AT THE END OF THE HANDRAIL, TURN AROUND. ESCALATORS ARE ON BOTH SIDES OF THE STAIRWELL. THE STAIRS ARE IN THE MIDDLE OF THE STAIRWELL.

WASHROOMS. TOP ITEM, GENERAL PUBLIC. LAST ITEM, WHEELCHAIR ACCESS.

WASHROOMS. LEVEL ONE. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE CLOSEST WASHROOMS ARE AT ONE O'CLOCK. WALK ABOUT FORTY PACES.

WASHROOMS LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WALK ABOUT SIXTY PACES. YOU WILL PASS THE GOURMET GRILL ON YOUR RIGHT.

WASHROOMS. THIS LEVEL. FOR THE CLOSEST WASHROOMS, WALK ABOUT FIFTY PACES AT NINE O'CLOCK. AT THE WARDAIR COUNTER, TURN RIGHT, AND WALK ABOUT EIGHTY PACES. THE WASHROOMS ARE ON YOUR RIGHT.

WASHROOMS WITH WHEELCHAIR ACCESS. LEVEL ONE. TAKE THE CLOSEST ELEVATOR WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE CLOSEST WASHROOMS ARE AT ONE O'CLOCK. WALK ABOUT FORTY PACES.

WASHROOMS WITH WHEELCHAIR ACCESS. LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO SIX O'CLOCK. WALK ABOUT SIXTY PACES. YOU WILL PASS THE GOURMET GRILL ON YOUR LEFT.

WASHROOMS WITH WHEELCHAIR ACCESS. THIS LEVEL. FOR THE CLOSEST WASHROOMS, WALK ABOUT FIFTY PACES, AT NINE O'CLOCK. AT THE WARDAIR COUNTER, TURN RIGHT, AND WALK ABOUT EIGHTY PACES. THE WASHROOMS ARE ON YOUR RIGHT.

TELEPHONES. TOP ITEM, GENERAL PUBLIC. LAST ITEM, THE AIRPORT TDD..

TELEPHONES. LEVEL THREE. THE CLOSEST TELEPHONES ARE BEHIND THE AIR CANADA COUNTER, AT SEVEN O'CLOCK. WALK ABOUT FORTY PACES. YOU WILL PASS THE AIR CANADA COUNTER ON YOUR LEFT.

THE AIRPORT TDD IS LOCATED ON THIS LEVEL AT THE TRANSPORT CANADA INFORMATION BOOTH, WHICH IS DIRECTLY AHEAD OF YOUR, AT ELEVEN O'CLOCK.

TELEPHONES. LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE CLOSEST TELEPHONES ARE AT EIGHT O'CLOCK. WALK ABOUT TWENTY PACES.

TELEPHONES. LEVEL ONE. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE CLOSEST TELEPHONES ARE AT ELEVEN O'CLOCK. WALK ABOUT THIRTY PACES.

CURRENCY EXCHANGE. THIS LEVEL. FOR THE CLOSEST CURRENCY EXCHANGE, WALK ABOUT FIFTY PACES, AT NINE O'CLOCK. AT THE WARDAIR COUNTER, TURN RIGHT, AND WALK ABOUT SEVENTY PACES. THE CURRENCY EXCHANGE IS DIRECTLY AHEAD OF YOU.

CURRENCY EXCHANGE. LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU, AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO NINE O'CLOCK. WALK ABOUT FORTY PACES. THE CURRENCY EXCHANGE IS ON YOUR LEFT.

BUSES, TAXIS, RENTALS. TOP ITEM, TASIC. ITEM TWO, HOTEL BUSES. ITEM THREE, BC TRANSIT BUSES. FOURTH ITEM, CAR RENTALS. LAST ITEM, PARKING LOT BUSES.

TAXIS. TOP ITEM, REGULAR TAXIS. LAST ITEM, WHEELCHAIR TAXIS.

REGULAR TAXIS. LEVEL TWO. TAXIS ARE AVAILABLE TO THE LEFT OF THE MAIN ENTRANCE ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU, AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE MAIN ENTRANCE IS AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES. FARE TO DOWNTOWN IS FIFTEEN TO EIGHTEEN DOLLARS. DRIVING TIME IS THIRTY MINUTES.

WHEELCHAIR TAXIS. LEVEL TWO. CALL VANCOUVER TAXI AT TWO FIVE FIVE, SEVEN THREE TWO TWO. WHEELCHAIR TAXIS ARE AVAILABLE TO THE LEFT OF THE MAIN ENTRANCE ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE MAIN ENTRANCE IS AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES. FARE TO DOWNTOWN IS FIFTEEN TO EIGHTEEN DOLLARS. DRIVING TIME IS THIRTY MINUTES.

BC TRANSIT BUSES SERVE GREATER VANCOUVER FROM LEVEL THREE. TOP ITEM, DOWNTOWN VANCOUVER. ITEM TWO, RICHMOND AND LADNER. ITEM THREE, NEW WESTMINSTER. LAST ITEM, PORT COQUITLAM.

DOWNTOWN VANCOUVER. TAKE BUS ONE HUNDRED, WHICH LEAVES FROM THE RIGHT OF THE MAIN ENTRANCE, THIS LEVEL. THE MAIN ENTRANCE IS AT TWO O'CLOCK, WALK ABOUT TWENTY PACES. TAKE BUS ONE HUNDRED TO GRANVILLE AND SEVENTIETH AVENUE. TRANSFER TO

BUS TWENTY ONE. BUS FARE IS ONE DOLLAR SEVENTY FIVE DURING PEAK HOURS. YOU MUST HAVE THE EXACT CHANGE.

RICHMOND AND LADNER. TAKE BUS FOUR HUNDRED AND FOUR, OR FOUR HUNDRED AND FIVE, WHICH LEAVES FROM THE RIGHT OF THE MAIN ENTRANCE, THIS LEVEL. THE MAIN ENTRANCE IS AT TWO O'CLOCK, WALK ABOUT TWENTY PACES. BUS FARE IS ONE DOLLAR SEVENTY FIVE DURING PEAK HOURS. YOU MUST HAVE THE EXACT CHANGE.

NEW WESTMINSTER. TAKE BUS ONE HUNDRED, WHICH LEAVES FROM THE RIGHT OF THE MAIN ENTRANCE, THIS LEVEL. THE MAIN ENTRANCE IS AT TWO O'CLOCK. WALK ABOUT TWENTY PACES. BUS FARE IS ONE DOLLAR SEVENTY FIVE DURING PEAK HOURS. YOU MUST HAVE THE EXACT CHANGE.

PORT COQUITLAM. TAKE BUS ONE HUNDRED, WHICH LEVES FROM THE RIGHT OF HTE MAIN ENTRANCE, THIS LEVEL. THE MAIN ENTRANCE IS AT TWO O'CLOCK, WALK ABOUT TWENTY PACES. BUS FARE IS ONE DOLLAR SEVENTY FIVE DURING PEAK HOURS. YOU MUST HAVE THE EXACT CHANGE.

CAR RENTALS. LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, CAR RENTALS ARE ON YOUR RIGHT.

PARKING LOT BUSES. LEVEL ONE. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. PARKING LOT BUSES HAVE WHEELCHAIR ACCESS AND LEAVE FROM THE MAIN ENTRANCE EVERY TEN TO FIFTEEN MINUTES. THERE IS NO CHARGE FOR THIS SERVICE.

ASSOCIATIONS FOR THE DISABLED. TOP ITEM, THE WESTERN INSTITUTE FOR THE DEAF. ITEM TWO, THE CNIB. LAST ITEM, THE KINSMEN REHABILITATION FOUNDATION.

THE WESTERN INSTITUTE FOR THE DEAF OFFERS SERVICES FOR THE HEARING IMPAIRED. FOR INFORMATION BY VOICE, CALL SEVEN THREE SIX, SEVEN THREE NINE ONE. FOR TDD SERVICE, CALL SEVEN THREE SIX TWO FIVE TWO SEVEN. TOUCH THE LAST ITEM TO LOCATE THE AIRPORT TDD.

THE CNIB OFFERS SERVICES FOR THE VISUALLY IMPAIRED. FOR INFORMATION CALL THREE TWO ONE TWO THREE ONE ONE.

THE KINSMAN REHABILITATION FOUNDATION HELPS PEOPLE WITH ANY PHYSICAL DISABILITY. CALL SEVEN THREE SIX, EIGHT EIGHT FOUR ONE. FOR TDD SERVICE, CALL SEVEN THREE EIGHT ZERO SIX ZERO THREE. TOUCH THE LAST ITEM TO LOCATE THE AIRPORT TDD.

HOTEL INFORMATION. TOP ITEM, HOTEL INFORMATION BILLBOARD. ITEM TWO, LOCAL HOTELS. LAST ITEM, THE AIRPORT EXPRESS.

THE HOTEL INFORMATION BOARD. LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, TURN TO THREE O'CLOCK. WALK ABOUT FIFTY PACES. YOU WILL PASS CAR RENTALS ON YOUR LEFT. DIRECT TELEPHONES ARE AVAILABLE ON THE HOTEL INFORMATION BOARD.

LOCAL HOTELS. TOP ITEM, ABERCORN MOTOR INN. ITEM TWO, AIRPORT INN RESORT. ITEM THREE, DELTA RIVER INN. TO CONTINUE THE LIST, TOUCH THE LAST ITEM.

LOCAL HOTELS. CONTINUED. TOP ITEM, GRANADA MOTOR INN. ITEM TWO, RICHMOND INN. LAST ITEM, SKYLINE INN.

ABERCORN INN. FOR BUS OR RESERVATIONS, CALL TWO SEVEN ZERO, SEVEN FIVE SEVEN SIX. A TELEPHONE IS BESIDE YOU AT TWO O'CLOCK.

AIRPORT INN RESORT. FOR BUS OR RESERVATIONS, CALL TWO SEVEN EIGHT, NINE SIX ONE ONE. A TELEPHONE IS BESIDE YOU AT TWO O'CLOCK.

DELTA RIVER INN. FOR BUS OR RESERVATIONS, CALL TWO SEVEN EIGHT ONE TWO FOUR ONE. A TELEPHONE IS BESIDE YOU AT TWO O'CLOCK.

GRANADA MOTOR INN. FOR BUS OR RESERVATIONS, CALL TWO SEVEN ZERO, SIX ZERO THREE ZERO. A TELEPHONE IS BESIDE YOU AT TWO O'CLOCK.

RICHMOND INN. FOR BUS OR RESERVATIONS, CALL TWO SEVEN THREE, SEVEN EIGHT SEVEN EIGHT. A TELEPHONE IS BESIDE YOU AT TWO O'CLOCK.

SKYLINE INN. FOR BUS OR RESERVATIONS, CALL TWO SEVEN EIGHT, FIVE ONE SIX ONE. A TELEPHONE IS BESIDE YOU AT TWO O'CLOCK.

THE AIRPORT EXPRESS LEAVES FROM THE MAIN ENTRANCE ON LEVEL TWO. TAKE THE CLOSEST ELEVATOR, WHICH IS DIRECTLY BEHIND YOU AT FIVE O'CLOCK. WHEN LEAVING THE ELEVATOR, THE MAIN ENTRANCE IS AT TWELVE O'CLOCK. WALK ABOUT THIRTY PACES. THE AIRPORT EXPRESS SERVES DOWNTOWN HOTELS AND LEAVES EVERY FIFTEEN MINUTES BETWEEN SIX FIFTEEN AM AND TEN THIRTY PM. ONE WAY FARE IS SIX DOLLARS AND FIFTY CENTS.

ONE TWO THREE FOUR FIVE SIX SEVEN EIGHT NINE TEN ELEVEN  
TWELVE TWENTY THIRTY FORTY FIFTY SIXTY SEVENTY EIGHTY NINETY  
HUNDRED

## VITA

Surname: Urbanczyk Given Names: Suzanne Claire

Place of Birth: Vancouver, B.C. Date of Birth: August 3, 1962

### Educational Institutions Attended:

University of Victoria, British Columbia	1987-89 1980-85
Caledonia Sr. Secondary School, Terrace, British Columbia	1978-80

### Degrees Awarded:

B.Sc. in Linguistics (with distinction)	1985
High School Diploma	1980

### Honours and Awards:

Graduate Research Engineering and Technology Award The Science Council of British Columbia	1988-89 1987-88
Award for Outstanding Paper Presented by a Student (CAA)	1987

### Publications:

Urbanczyk, S.C. and S.J. Eady. (1989). "Assignment of syllable stress in a demisyllable-based text-to-speech synthesis system." Proceedings of the IEEE Pacific Conference on Communications, Computers and Signal Processing: 467-470.

Urbanczyk, S.C. (1988). "A preliminary investigation of the acoustic context of homograph pairs." Working Papers of the Linguistics Circle of the University of Victoria. Vol. 7: 135-146.

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis (or dissertation) 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 of this thesis for scholarly purposes 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/Dissertation: An Algorithm for the Assignment of Lexical  
Stress in Converting Text to Speech

---

---

Author



(Signature)

Suzanne C. Urbanczyk  
(Name in Block Letters)

August 31, 1989  
(Date)



National Library  
of Canada

Bibliothèque nationale  
du Canada

Canadian Theses Service    Service des thèses canadiennes

Ottawa, Canada  
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-53721-3