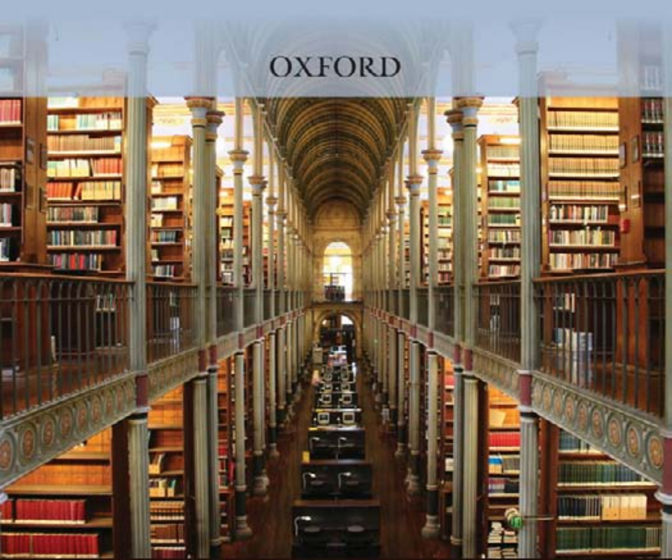


San Duanmu

Syllable Structure

The Limits of Variation

Syllable Structure

To my parents
端木道
DUANMU Dao
and
余承运
YU Chengyun

Syllable Structure
THE LIMITS OF VARIATION

San Duanmu

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2008

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–926759–0

1 3 5 7 9 10 8 6 4 2

CONTENTS

This page intentionally left blank

PREFACE

My interest in syllable structure started twenty years ago, by way of a different field of inquiry. When I entered graduate school at MIT in 1986, tone was a hot topic. The prevailing view was that contour tones (rise and fall) are made of level tones L (low) and H (high), so that a rise is L+H and a fall is H+L. Being a native speaker of a tone language where rises and falls seldom split into level tones, I found the prevailing view completely counterintuitive. However, when I started to work on Shanghai Chinese, I began to see the merit of the theory. In Shanghai, a rise splits into L-H when a second syllable is added, and a fall splits into H-L, exactly as the theory predicts.

But there remained a puzzle. Why do contour tones split in Shanghai and not in other Chinese languages? A common response was that there are different kinds of contour tones: some split and some do not. However, this was a restatement of the fact, not an explanation.

The answer emerged when I noticed a difference in syllable structure between Shanghai Chinese and other Chinese languages. Syllables in Shanghai are essentially CV (a consonant plus a vowel), whereas most syllables in other Chinese languages are CVV or CVC, where VV is a diphthong or a long vowel. The relation between syllable structure and tone split is somewhat elaborate (see Chapter 7), but here is an outline: a CVV or CVC syllable has two moras (rhyme units); therefore, it forms a moraic foot and has inherent stress. In addition, we know independently that stressed syllables can keep their lexical tones but unstressed ones cannot. As a result, in a string of CVV or CVC syllables, everyone has stress and their own lexical tones and there is little chance for tone split. In contrast, in a string of CV syllables, most of them lack inherent stress or lexical tones, and this creates the condition for tone split.

It became clear to me that syllable structure is a crucial link among several parts of phonology—features, sounds, stress, and tone. But while the concept of the syllable seems obvious in some languages, it is far from being so in others. For example, people do not agree on where syllable boundaries are in English, or how large a syllable can be, or whether syllables are real at all. Indeed, if the syllable is dubious in English, would it not be so in other languages as well? Clearly, the stakes on this issue are high. In this book I offer a systematic study of syllable structure in the hope that some questions can be clarified and a better understanding can be achieved, in regard both to syllable structure itself and to phonological theory in general.

ACKNOWLEDGMENTS

I am grateful to many people for their comments on various aspects of this book, in particular Steve Abney, Rusty Barrett, Bill Baxter, Pam Beddor, Jose Benki, Toni Borowsky, Rob Burling, Raung-fu Chung, Nick Clements, Andries Coetzee, Abby Cohen, John Davey, Stuart Davis, Sam Epstein, Heinz Giegerich, Morris Halle, John Harris, Jeff Heath, Yuchau Hsiao, Hsin-I Hsieh, Jim Huang, Rene Kager, Michael Kenstowicz, Hyoyoung Kim, Aman Kumar, Bob Ladd, Susie Levi, Yen-hwei Lin, Victor Manfret, Mike Marlo, John McCarthy, Tsulin Mei, Lesley Milroy, Andrew Nevins, David Odden, Marc Pierce, Donca Steriade, Sally Thomason, Moira Yip, some anonymous reviewers, and audiences at several schools and conferences.

I thank John Davey and other staff at the Oxford University Press for their experience and assistance throughout the writing of this book, Sarah Barrett for copy-editing, the College of Literature, Science, and the Arts, University of Michigan, for covering some subvention costs, and Xinting Zhang for making the Index.

Many thanks to Yan, Youyou, and Alan for their understanding when I often took too much time away from family, and for their insightful comments from an outsider's perspective.

NOTES ON TRANSCRIPTION

Phonetic transcriptions are given in square brackets at any level of detail. For example, (1) shows three ways to transcribe the word for 'melon' in Standard Chinese.

(1) [kua] indicating the phonemes
[kʷa] indicating the sounds but not their (predictable) lengths
[kʷaa] indicating the sounds and their (predictable) lengths

Similarly, the Chinese word for 'peace' can be transcribed as [an], [æ̃n], or [æ̃:]. Which degree of detail is transcribed will be noted when relevant.

The phonetic transcription of an example may be influenced by one's analysis or assumptions. For example, if one makes no distinction between [aa] and [a:], one might use either. Similarly, in *queen*, if one thinks the part before the vowel is a sequence of two sounds, one might transcribe the word as [kwi:n]. On the other hand, if one thinks the part before the vowel is a single sound (with double-articulation), one might transcribe the word as [kʷi:n]. One might also assume that *queen* is [kwi:n] at the underlying level but [kʷi:n] at the surface level. Before a theoretical position is introduced, any of the alternative transcriptions might be used.

Unless noted otherwise, IPA symbols are used for phonetic transcription. In the text, the spelling of cited examples is given in italics, the phonetic symbols are given in square brackets, and the translation is given in quotation marks (as in the Chinese example *mao* [mau] 'cat'). English examples are not provided with a gloss (as in the English example *cat* [kæt]). When relevant, both a word-for-word translation and a regular translation (in parentheses) are given for non-English examples (as in the Chinese example *da yi* 'big idea (careless)').

A hyphen is sometimes used to indicate morpheme or affix boundaries. For example, *see-ing* shows that the word has a suffix *-ing*, and *un-interest-ing* shows that the word has a prefix *un-* and a suffix *-ing*.

In a numbered non-English example, spelling is not italicized. The translations are given either on the same line, as in (2), where the regular translation is in parentheses, or on separate lines, as in (3), where only the regular translation is in quotation marks.

(2) gao-xing ‘high-mood (glad)’

(3) gao-xing
high-mood
‘glad’

English examples are usually italicized both in the text and in numbered examples.

Syllable boundaries in IPA transcription are often indicated either by a dot or by brackets. For example, [hæ.pi] and [hæ][pi] are the same. Similarly, ‘p]’ indicates a syllable-final [p] and ‘p][’ indicates a syllable-final [p] before another syllable. A hyphen is sometimes used to indicate syllable boundaries in spelling, especially for a polysyllabic word or compound. For example, *Chi-ca-go* L-H-L shows that the word has three syllables and that their tones are, respectively, L (low), H (high), and L.

An affricate is usually represented with two IPA symbols, such as [ts] and [tʃ]. To distinguish an affricate from a sequence of two sounds, the second symbol of an affricate is sometimes shown in superscript, such as [tˢ] and [tᶴ]. Similarly, [tʲ] and [kʷ] are single (complex) sounds, whereas [tj] and [kw] can be pairs of two sounds.

This page intentionally left blank

1

Introduction

To many people it may seem obvious what syllables are. For example, the English word *buy* is a syllable, so is the Chinese word *ni* 'you'. The English word *city* has two syllables, so does the Chinese word *mayi* 'ant'. The word *potato* has three syllables; the word *syllabification* has six syllables; and so on.

However, the clarity can be deceptive. While in some words it is easy to count syllables, in others the answer is not so obvious. For example, is the word *hour* or *shower* one or two syllables? In addition, while we can often count the number of syllables, it is not always clear where syllable boundaries are. For example, where is the syllable boundary in *city* and *happy*? Are all the sounds in *smile* in one syllable, or are [s] and [l] outside the syllable [mai]?

Linguists have wrestled with various answers. In (1) we see several views of syllable count in *hour*, *flour*, *flower*, and *shower*, which only differ in the initial consonant(s).

(1)

	hour	*flour*	*flower*	*shower*
Jones (1950)	1	1	1	1
Hanks (1979)	1	1	2	2
Baayen et al. (1993)	2	2	2	2
Kenyon and Knott (1944)	1	1	1 or 2	1 or 2
Kreidler (2004)	1	1	1	1
Merriam-Webster (2004)	1 or 2	1 or 2	1 or 2	1 or 2
Gussmann (2002)	2	2	2	2

Jones (1950), Hanks (1979), and Baayen et al. (1993) all describe British English, but their analyses are different. Similarly, Kenyon and Knott (1944), Kreidler (2004), and Merriam-Webster (2004) all describe American English, but their analyses are again different. Gussmann (2002) does not specify which English variety he is dealing with, but either way he would probably treat all the words as disyllabic.

Next consider syllable boundaries. Four analyses of *happy* are shown in (2), where the syllable boundary is indicated by a dot and the underlined [p̲] means that it belongs to both the first syllable and the second syllable (an "ambisyllabic" [p]).

(2) Analyses of *happy*

[hæ.pi]	Hayes (1995), Halle (1998), Gussmann (2002)
[hæp.i]	Selkirk (1982), Hammond (1999)
[hæp̲i]	Kahn (1976), Giegerich (1992), Kreidler (2004)
[hæp.pi]	Burzio (1994)

The medial [p] has been thought to belong to the first syllable, or the second syllable, or both syllables. In addition, the medial [p] has been treated as a single sound or as two sounds. Experimental studies do not always yield clear answers either. For example, Treiman and Danis (1988) found that, in VCV clusters where the stress is on the first V, subjects are not consistent with the syllabification, although Krakow (1989) and Turk (1994) found that, as far as phonetic cues are concerned, such clusters are syllabified as VC.V.

Next consider consonants at word edges, exemplified with *smile*. Three analyses are shown in (3).

(3) Analyses of *smile*

[smail]	Kahn (1976), Selkirk (1982), Blevins (1995), Coleman (1996)
[smai.l]	Burzio (1994)
[s.mai.l]	Harris (1994), Gussmann (2002)

Many people believe that all the sounds in *smile* are in one syllable. Some believe that [l] is in a separate syllable (e.g. a syllabic [ɫ̩]). Still others believe that both [s] and [l] are in separate syllables (supported by empty vowels).

The difficulties in defining the syllable are well known. Not surprisingly, some scholars have chosen a less definitive approach. For example, Jones (1950: 130–1) considers a syllable to be a prominence (or loudness) peak in a word. The word *cat* has one peak (on [æ]) and is one syllable; the word *happy* has two peaks (on [æ] and [i]) and are two syllables. This approach avoids the need to specify syllable boundaries, and essentially equates the number of syllables to the number of vowels (and certain sonorant consonants). But if syllables are derivable from vowels or loudness peaks, there is perhaps no need to assume them at all. Thus, some scholars believe that syllable structure only plays a secondary role in phonology, if at all (Chomsky and Halle 1968, Steriade 1999, Blevins 2003).

Nevertheless, it would be hasty to dismiss the notion of syllables just because there are difficulties. Indeed, the notion of syllable is arguably just as important as the notion of consonants and vowels. Some scholars even argue that the syllable is a more fundamental concept than consonants and vowels. For example, Ladefoged (2001: 170) argues that words are stored in memory "as wholes, or at least as whole syllables, in which the consonants and vowels are not separate items." Similarly, in regard to the invention of writing systems, Ladefoged (pp. 172–3) argues that "alphabetic writing has almost certainly been invented only once, whereas there are many independent inventions of systems for writing down syllables," and that this "shows that the division of the syllable into vowels and consonants is not a natural one."

If syllables *are* real, we want to know what their structures are and to what extent they can differ from language to language. For example, if [m] can be a syllable in English (as in *prism*), should we expect [mdok] 'color' in the Tibeto-Burman language Jiarong, *mgła* [mgwa] 'fog' in Polish, and [mba] 'be' in the African language Mende, to be two syllables each as well? More generally, is the analysis of a string of sounds, such as [mdok], the same for all languages, or can the analysis vary from language to language?

In this book I explore answers to such questions. Since syllables are made of sounds and sounds are made of features, I start in Chapter 2 with a discussion of sounds and features, with a focus on the notion of complex sounds. In Chapter 3 I review theories of syllable structure and argue for a new proposal, which I call the CVX theory, according to which the maximal syllable size is CVX (CVV or CVC) and extra consonants at word edges are explained by morphology.

A study of this kind can take one of two approaches: it can make a superficial survey of a large number of languages, or it can make an in-depth analysis of a small number of languages. I have chosen the latter because the syllable structure of a language is not always obvious at first sight, even in well-known languages. For example, most analyses of English assume very large syllables, yet an in-depth analysis will show that the maximal English syllable is CVX. Therefore, a superficial survey would have led to quite different and erroneous conclusions.

In Chapters 4–7 I discuss syllable structure in Standard Chinese and Shanghai Chinese. The focus is on the constraints that govern syllable-internal sound combinations. For example, the maximal Chinese syllable is $C^{G}VX$, where X can be C, G (glide), or V. If sounds can freely combine, there are over 2,000 C^GVX forms in Standard Chinese, yet only about 400 are used. Why are so many combinations not used? Are there general constraints that exclude the unused syllables? Are there physical reasons for the constraints? How does syllable structure interact with stress and tone? These questions will be addressed.

In Chapters 8–10 I examine English and German, whose large consonant clusters seem to require syllables larger than CVX. I show that word-medial syllables are no larger than CVX and extra consonants at word edges can be independently accounted for by morphology.

In Chapter 11 I discuss the Tibeto-Burman language Jiarong, which is rich in initial consonant clusters, such as [rp-], [zd-], [bst-], [nzd-], and [ʒngr-], which are not found in many languages. The analysis of Jiarong has implications for the analysis of other languages with large initial consonant clusters.

The big picture that emerges is that (i) word-edge consonants can be accounted for by morphology, and (ii) all the languages have just two basic syllable structures, shown in (4), where σ is the syllable node, (O) an optional onset, R the rhyme, and X a timing slot (a position for a single sound).

```
(4) Syllable type:  Light    Heavy
    Structure:        σ        σ       Syllable node
                     /\       /\
                   (O) R    (O) R      Onset/Rhyme
                    |  |     |  /\
                    X  X     X X X     Timing slots
    Shorthand:     (C) V    (C) VX
```

I use the shorthand (C)V for light (or unstressed) syllables and (C)VX for heavy (or stressed) syllables, where (C) is an optional onset and X is the coda. In (C)V, V can be a short vowel (such as [ə]) or a syllabic consonant (such as [m̩]). In (C)VX, VX can be VC (such as [in] or [ip]), VV (such as [i:] or [ai]), or CC (such as [m̩:]). Each of the three slots in CVX can be filled by a simple sound or a "complex sound." For example, [kw] and [bl] can each merge into a complex sound ([k^w] and [b^l]) to fill the onset slot C. In Chapter 2 I shall discuss specific conditions on the formation of a complex sound, which are more strict than other proposals.

A reviewer asks whether CVX is a phonetic or phonological unit. The answer is that it is both. It is phonological in the sense that it is a structural unit above the level of phonemes (sounds). It is phonetic in the sense that it places timing restrictions on how phonemes are realized in a syllable, such as whether two consonants must be pronounced simultaneously.

While this study aims to offer a general theory of syllable structure, equal emphasis is placed on data description. For each language, quantitative data will be provided from entire lexicons or entire syllable inventories, so that the reader can examine them in alternative theoretical frameworks.

For the most part, I shall focus on one theoretical question: What is the maximal syllable size in human languages? Equally interesting is another question: Which syllables within the maximal size are or are not used in different languages, and why? I shall leave the second question for Chapter 12, along with other theoretical issues.

1.1. COMMON TERMS DESCRIBING SYLLABLE STRUCTURE

Here is a review of terms on syllable structure that are commonly used in the literature.

Syllable. The term can refer to several different things. Casually, it can refer to a word that has one vowel (including a diphthong or a triphthong). For example, *cat* is a syllable, so is *smiles*. Such words are also called monosyllabic words. In another usage, the term refers to a specific unit in a given theory. For example, according to Lowenstamm (1996), every syllable is CV, where C is a consonant, V is a short vowel, and either C or V can be empty (represented by Ø). Thus, *cat* has two syllables [kæ.tØ] and *smiles* has five [sØ.ma.Øi.lØ.zØ].

In some languages, such as Chinese, the syllable can include tone. For example, [ma1] 'mother' and [ma2] 'hemp' are different words in Standard Chinese, where [ma1] has tone 1 (high or H) and [ma2] has tone 2 (rise or LH). Therefore, [ma1] and [ma2] are different syllables. However, we can also say that the syllable [ma] can carry different tones in Chinese; in this usage the term "syllable" does not include tone.

I shall use the term "syllable" loosely, without limiting it to a specific shade of meaning. Fortunately, it is often obvious in context which shade of meaning the term refers to.

Onset and **rhyme.** A syllable can be divided into an onset and a rhyme. The rhyme is the common part between two syllables that rhyme. The part before the rhyme is the onset. For example, [pɪn] *pin* and [twɪn] *twin* rhyme in English. The common part is [ɪn]. Therefore, [p-] and [tw-] are onsets and [-ɪn] is the rhyme. Clements and Keyser (1983: 13) use the term "nucleus" to refer to the part [ɪn] in *pin* or [æt] in *bat*, which we call the rhyme here.

The size of the onset or the rhyme of a syllable depends on one's view of what a syllable can contain. For example, if one believes that the maximal English syllable is CCCVXC (where X is V or C), then the onset of the syllable in *strike* [straik] is [str] and the rhyme is [aik]. On the other hand, if one believes that the maximal English syllable is CCVX, then the onset of the syllable in *strike* is [tr] and the rhyme is [ai], and the initial [s] and final [k] must be accounted for separately.

The rhyming part can be larger than the rhyme. For example, in English, *biker* and *hiker* rhyme: the rhyming part is [aikɚ], which consists of a rhyme [ai] and a syllable [kɚ]. This means that rhyming does not tell us where a syllable ends. Thus, while we know that *showers* and *flowers* rhyme, we cannot conclude that [auɚz] is a syllable rhyme. Nevertheless, the rhyming part always starts at an onset-rhyme juncture, which justifies dividing the syllable into two parts. Some linguists, such as Davis (1988), have argued against the rhyme as a unit. Others, such as Breen and Pensalfini (1999), have proposed that in some languages every syllable starts with a vowel and there is no onset. However, I do not pursue these proposals.

Nucleus (peak) and **coda.** If a rhyme has one sound, it is called the nucleus (or the peak). If the rhyme has two sounds, the second is usually a consonant and is called the coda. For example, in [pɪn] *pin*, [ɪ] is the nucleus and [n] is the coda, and in [bʌtn̩] *button*, the first nucleus is [ʌ] and the second nucleus is a syllabic nasal [n̩].

Diphthongs and long vowels pose some complications. For example, in the word *how*, if you consider the diphthong to be a single vowel [au], it is in the nucleus. If you consider the diphthong to be a vowel plus a glide [aw], and if you consider the glide [w] to be a consonant, then [a] is in the nucleus and [w] is in the coda. If you consider the diphthong to comprise two vowels [au], then you may consider [a] to be in the nucleus and [u] to be in the coda, or you may consider both vowels to be in the nucleus (a branching nucleus). Long vowels pose a similar problem. For example, in the word *bee*, if you consider the long vowel [i:] to be one sound, then it is in the nucleus. But if you consider [i:] to be two sounds [ii], and if you think the coda does not have to be a consonant, then you might consider the second part of [i:] or [ii] to be in the coda. Or if you consider [i:] to be [ij], where [j] is a consonant, then [j] is also in the coda.

If you think a rhyme can be as large as VVC (where VV is a diphthong or a long vowel), then C should be in the coda and so VV ought to be in the nucleus. For example, in [aik] (as in *like*) and [i:z] (as in *bees*), [k] and [z] should be in the coda, and so [ai] and [i:] ought to be in the nucleus. However, if the maximal rhyme size is VX (where X is C or V), as I shall argue, then we can simply call V the nucleus and X the coda, regardless of what fills X.

Weight: **light**, **heavy**, **superheavy.** The weight (or length) of a syllable refers to whether it is heavy (long) or light (short). A light syllable is one whose rhyme is V, where V is a short vowel or a syllabic consonant. A heavy syllable is one whose rhyme is VX, where X is the second half of a long vowel (as in [i:]), the second half of a diphthong (as in [ai]), or a consonant (as in [an]). Syllable weight is related to stress, where heavy syllables attract stress and light syllables do not. In most languages the onset of a syllable has no effect on syllable weight—a fact that remains a mystery (for reported cases where onsets do affect stress and possible analyses of them, see Everett and Everett 1984, Davis 1988, Goedemans 1998, and Gordon 2005).

When the rhyme of a syllable is VXC (VVC or VCC), it is said to be superheavy. Superheavy syllables are rare in non-final positions. I shall argue in Chapter 8 that all non-final superheavy syllables in English are simple heavy syllables.

Open and **closed.** An open syllable ends in a vowel, and a closed syllable ends in a consonant. Diphthongs again present a problem. For example, in the word *how*, if you think the diphthong ends in a vowel [au], then it is an open syllable, but if you think the diphthong ends in a consonant [aw], then it is a closed syllable. Similarly, if you think that the nucleus in the American English word *fur* is a vowel [ɝ], then syllables like [biɚ] *beer* and [foɚ] *four* are open, but if you think the nucleus in *fur* is a syllabic consonant [r̩], then syllables like [bir] *beer* and [for] *four* are closed. Syllabic consonants pose another problem. For example, is the syllable [n̩] (reduced form of *and*) open or closed? It seems to be open in the sense that the nucleus is not followed by another sound. On the other hand, the syllable ends in a consonant, similar to other closed syllables.

Mora. The mora is an alternative unit for measuring syllable weight (or length). A heavy syllable has two moras and a light syllable has one mora. If we focus on basic syllable structures, where the maximal rhyme has two slots, then each mora corresponds to a rhyme slot.

It is not clear whether the syllable onset is part of a mora. Most people agree that the presence of the onset does not affect how many moras a syllable has. However, some people believe that the onset shares a mora with the nucleus (e.g. Hyman 1985), whereas others believe that the nucleus has a mora to itself. The difference is shown in (5) for the word *pie*, where S is a syllable and M is a mora.

(5)


Onset sharing mora   Onset not sharing mora

At word edges extra consonants may occur, and some unusual structures have also been proposed for them, such as syllables without moras, or moras without syllables (Fery and van de Vijver 2003). I shall argue in Chapter 3 that word-edge consonants can be accounted for by morphology. Therefore, there is no need to propose unusual mora structures.

Appendix. An appendix is an extra consonant that does not belong to the basic syllable structure but is attached to it. The idea is used to account for extra consonants at word edges (e.g. Borowsky 1989). For example, *help* has the basic structure [hɛl] (because a basic syllable has only two rhyme slots), and the final [p] is an appendix to it, as shown in (6).

(6)


The use of appendixes allows one to maintain a simple basic syllable structure. However, to assume appendixes is to assume that (i) sounds cannot exist on their own but must belong to a syllable, and (ii) a syllable must have an overt vowel or a syllabic consonant. If not all sounds must depend on a syllable, we can maintain a simple syllable structure without appendixes. For example, *help* can be analyzed as [hɛl.(p)], where [p] is not in the preceding syllable. Similarly, if we assume that a syllable does not need to have an overt vowel but can have an empty vowel [Ø] (Burzio 1994, Harris 1994), we can also maintain a simple syllable structure. For example, *help* can be analyzed as [hɛl.pØ]. The same comments can be made of Vaux (2004), according to which extra consonants are appended not to the syllable but to the word.

Ambisyllabic. A sound is ambisyllabic if it belongs to two syllables at the same time. In the ambisyllabic analysis of American English word [sɪɾi] *city*, the first syllable is [sɪɾ] and the second syllable is [ɾi] (e.g. Kahn 1976, Clements and Keyser 1983, Giegerich 1992), where a single sound [ɾ] belongs to both the first syllable and the second syllable. An ambisyllabic sound is different from a true geminate consonant (a long consonant), such as [ss] in [ðɪssup] *this soup* and [tt] in [ðættaim] *that time*. A geminate consonant counts as two sounds, one in each syllable, and so neither part is ambisyllabic.

1.1.1. **Summary of syllable terms**

As will be discussed in Chapter 3, if we account for extra word-edge consonants by morphology, it suffices to assume just two syllable structures, shown in (7), where “M” indicates a timing slot that counts as a mora, and where the onset is optional.

```
(7)  Light    Heavy
       σ        σ
      / \      / \
    (O) R    (O) R     Onset/Rhyme
     |  |     |  / \
     X  M     X M  M   Timing slots/Moras
    (C)V     (C)VX     Shorthand
```

In (8) I show the interpretation of common syllable terms using the structures in (7).

(8) Common terms describing syllable structure:

appendix	(Not used in the present analysis)
closed	Having a consonant coda
coda	The second slot in the rhyme (the part after the nucleus)
heavy	Having two slots in the rhyme, or (C)VX
light	Having one slot in the rhyme, or (C)V
mora	A slot in the rhyme
nucleus	The first slot in the rhyme
onset	The optional first slot in the syllable (the part before the rhyme)
open	Having no coda, or where the coda is not a consonant
rhyme	The part for rhyming in a syllable
weight	The length of the rhyme

2

Features, sounds, complex sounds, and the No Contour Principle

If a language allows CVC syllables, would [bait] fit into such a syllable? The answer depends on what a sound is. If [ai] is a single sound, the answer may be yes, but if [ai] counts as two sounds, the answer may be no. Similarly, would [kwɪk] and [kɪts] fit into CVC? Since we know that [k] and [w] can merge into a single sound [k^w] and [t] and [s] can form and affricate [t^s], we expect the answer to be yes, i.e. [k^wɪk] and [kɪt^s]. Now what about [srɪk] and [frɪk]? The answer now is less obvious, unless we have a theory of which sound pairs can merge into a single sound (to be called a "complex sound") and which cannot. In this chapter I discuss what speech sounds are, what they are made of, and, most importantly, under what conditions two sounds can form a complex sound. I shall propose that two sounds can form a complex sound only if there is no violation of the No Contour Principle (Duanmu 1994).

2.1. WHAT IS A SOUND?

Speech is, at some level of abstraction, made of consonants and vowels, or speech sounds (sometimes called "segments"). For example, the English word [kæt] *cat* is made of three sounds. While the notion of a sound may seem simple, it is in fact far from obvious. I discuss two such problems. The first concerns the notion of what a minimal sound is. The second concerns the question of whether words or syllables can be "sliced" into sounds.

2.1.1. **What is a minimal sound?**

Speakers of alphabetic languages often have the impression that each alphabetic symbol represents a sound, but this is not always the case. In English, for example, the letter *x* in *ax* represents two sounds [ks]. The letters *sh* in *she* represents one sound [ʃ]. The letters *ch* can represent [k] as in *character* or [ʧ] as in *church*. The sound [k] is represented by *c* in *cat*, *q* in *queen*, *k* in *king*, and *ch* in *character*. And so on.

Phonetic symbols can represent sounds better, but the relation between symbols and sounds is not always one to one either. For example, the first sound in *church* is represented by two symbols [ʧ], so is the first sound in *pan*, which is [$\mathrm{p^h}$]. Sometimes a single symbol is used to represent a long vowel, such as [i] for the vowel in *bee* (Ladefoged 2006: 39), although a long vowel may count as two sounds for some purposes.

Phonologists often define a sound as a minimal unit that can contrast meaning. A pair of words that differ in one sound is called a minimal pair. An example is shown in (1).

(1) A minimal pair
 [bai] *by*
 [dai] *die*

The pair differ in the first part ([b] vs. [d]), which leads to a contrast in meaning. Therefore [b] and [d] are sounds. On the other hand, the words in (2) are not a minimal pair.

(2) Not a minimal pair
 [mɪst] *mist*
 [mɪlk] *milk*

Although the words differ in the last part ([st] vs. [lk]), which leads to a contrast in meaning, [st] and [lk] are not single sounds, because they can be sliced up further. For example, we can show that [st] is made of [s] and [t] by comparing [rɪst] *wrist* and [rɪsk] *risk* and then comparing [rɪst] *wrist* and [rɪpt] *ripped*.

But how do we know whether something is already small enough or whether it should be sliced up further? In the above discussion, there is a tacit assumption that we do know what a minimal sound is, but the assumption is not obvious. For example, consider the pair in (3).

(3) A minimal pair?
 [hau] *how*
 [hi:] *he*

The words are a minimal pair if you think that each vowel or diphthong is a single sound, which many people do (e.g. Kreidler 2004, Ladefoged 2006). But if we compare [au] and [ai], then it is possible to split a diphthong into two ([au] as [a] + [u] and [ai] as [a] + [i]). Similarly, the vowel [i:] is nearly twice as long as a short vowel (compare [ti:n] *teen* and [tɪn] *tin* in English), so [i:] is like two sounds in terms of duration. And if we think of [i:] as [ii] and compare it with [ai], then it is reasonable to think of [ii] as [i] + [i], too.

Some phonologists (e.g. Chao 1934, Odden 2005) believe that we cannot define sounds in a hard-and-fast way. Instead, we can use different definitions for different languages. For example, although [st] is analyzed as two sounds in English, Padgett (1995) analyzes it as one sound in Kabardian. Even for the same language there can be multiple solutions in defining sounds. For example, in analyzing Chinese, most people consider [p^h] to be one sound, but some consider it to be two sounds, [p] + [h], such as Hockett (1947) and Martin (1957). Similarly, some phonologists consider entire rhymes in Chinese to be single sounds (called "rhyme phonemes"), such as [ai], [au], [an], and [aŋ] (You et al. 1980).

If there is no universal definition of sounds, what are the bases for choosing among alternative definitions? A common consideration is phonemic economy, by which the size of the resulting sound inventory should be as small as possible. For example, suppose a language has three aspirated stops [p^h, t^h, k^h] and three unaspirated stops [p, t, k]. If we consider aspirated stops to be single sounds, the language has six sounds. But if we consider aspirated stops to be two sounds each [ph, th, kh], then the language has just four sounds [p, t, k, h].

However, phonemic economy is often in conflict with other priorities. For example, are [f] and [v] minimal units that cannot be further decomposed? The answer is no. Each sound is made of a number of articulatory features, shown in (4).

(4)

	[f]	[v]
Closure	fricative	fricative
Articulator	labial-dental	labial-dental
Vocal cords	voiceless	voiced

Both sounds have a narrow vocal opening (fricatives) and make the closure with the upper teeth and the lower lip (labial-dental). The sounds differ only in the state of the vocal cords, which are vibrating in

[v] (voiced) but not in [f] (voiceless). In fact, it is possible to represent all sounds with just a dozen or so features. Should we consider features to be sounds and treat [f] and [v] as three sounds each? By doing so we can surely achieve better phonemic economy by reducing the number of "sounds" in English from around forty to just a dozen. There are two reasons for not doing so. First, a sound should be pronounceable, but a single feature, such as "fricative" or "voiced," cannot occur by itself but must be accompanied by other feature(s). Second, we would like to consider sounds to be sequenced in time: we do not consider simultaneous elements that occur in the same time unit as different sounds.

The difficulty in defining what a minimal sound is may have caused some linguists to doubt the reality of consonants and vowels. For example, Ladefoged (2001) says that "words are not stored as sequences of sounds. They are stored as wholes, or at least as whole syllables, in which the consonants and vowels are not separate items," and that "the division of the syllable into vowels and consonants is not a natural one."

2.1.2. **Slicing or not?**

Goldsmith (1976) uses the term "Absolute Slicing Hypothesis" to refer to the assumption that words can be cut into consonants and vowels—an assumption that is fundamental to traditional phonology and the IPA. Goldsmith offers two arguments against the slicing analysis, and advocates an alternative model that he calls "autosegmental phonology."

The first argument is that speech is not made of a single tier of elements (consonants and vowels), but from multiple tiers of elements, each being independent of the others. In other words, speech is made of multiple tiers of "autonomous segments," hence the term "autosegmental phonology." For illustration, let us consider two tiers, a tonal tier and a consonant–vowel tier (although the latter may itself be made of multiple tiers). The independence of each tier is shown by the fact that the deletion of elements on one tier does not necessarily lead to the deletion of elements on another tier. In (5) we see that a tone can be deleted without deleting a syllable, and in (6) we see that a syllable can be deleted without deleting the tone.

(5) Deleting the tone and keeping the syllable (Shanghai)

Tone	LH	LH	→	LH		→	L	H
Syllable	wã	dɤ		wã	dɤ		wã	dɤ
	'yellow bean (soy bean)'							

(6) Deleting the syllable and keeping the tone (Standard Chinese)

Tone	L	LH	→	L	LH	→	LH
Syllable	wo	mən		wom			wom
	'we'						

In (5), the second syllable of a compound loses its tone, yet the syllable remains. In (6), which occurs in casual speech, the second syllable loses its rhyme and merges with the first syllable, yet its tone (at least the H part) is kept and reappears on the resulting syllable.

The second argument against the slicing analysis is that the relation between elements on different tiers is not always one-to-one, which makes it hard to make cuts across all tiers. For example, Goldsmith argues that it is possible for one tone to link to two vowels, as in (7a), or for two tones to link to one vowel, as in (7b).

(7) a. One tone to two vowels

```
 H
 /\
bai
```

b. Two tones to one vowel

```
HL
\/
ba
```

In such structures, the slicing analysis would have to decide how to cut across different tiers in order to obtain "segments" that make sense for all tiers.

Pulleyblank (1986: 12–14) points out a problem in the autosegmental model. If elements between any two tiers can link to each other in the way Goldsmith proposes, then we can encounter a contradiction in timing. For illustration, consider three tiers A, B, and C, each having two elements, A1, A2, B1, B2, C1, and C2. The timing relation between tiers A and C is shown in (8a) and those between tiers A and B and between tiers B and C are shown in (8b). On each tier, the left-to-right order defines the order in time. Between tiers, a linking line means that the linked elements are pronounced at the same time.

```
(8)  a. Timing for tiers A–C        b. Timing for tiers A, B and tiers B, C
        A1   A2                        A1    A2       Tier A
         \                                   |
          \                            B1    B2       Tier B
           \                                /
       C1   C2                         C1    C2       Tier C
```

Assuming that the timing relation allows identity substitution—i.e. if X and Y are identical in time and Y is before Z, then X is before Z—we obtain contradictory conclusions from (8), which are shown in (9).

(9) a. In (8a), because A1 and C2 are simultaneous (defined by the linking line) and A1 is before A2 (by linear order), therefore C2 is before A2 (by identity substitution).
b. In (8b), because A2, B2, and C1 are simultaneous (defined by the linking lines) and C2 is after C1 (by linear order), therefore C2 is after A2 (by identity substitution).

Since C2 cannot be both before A2 and after A2 at the same time, there is a contradiction. It shows a lack overall temporal coordination in the model of Goldsmith (1976).

The solution Pulleyblank proposes is to incorporate a special tier to coordinate timing, which can be the CV-tier (McCarthy 1979a) or the X-tier (Kaye and Lowenstamm 1984, Levin 1985). Let us call it the "timing tier," whose units are "timing slots." All other tiers are directly linked to the timing tier, instead of linking to each other. With the use of the timing tier, the notion of a sound or segment re-emerges: a sound is a set of features that are linked to a given timing slot. In addition, the timing tier offers a solution to the problem of defining what a minimal sound is (to be discussed next).

2.1.3. **Defining the speech sound**

Although the timing tier seems to offer a solution for the definition of what a sound is, there is still a problem: it is not clear what the minimal time unit is. For illustration, let us consider the relation between tones and vowels again.

Many linguists, including Goldsmith (1976) and Pulleyblank (1986), believe that a short vowel can be linked to two (or more) tones that are articulated in sequence. For example, the structure in (10) represents a short rising tone.

(10)

Tone tier	LH
	\/
Timing tier	X
	|
Vowel tier	a

The vowel is short because it is linked to one timing slot. The question for this structure is: Does X represent a minimal unit of time? If it does, how can there be additional division of time on the tone tier, since in order to get a rise, L must occur before H? If X does not represent a single minimal time unit but two (one for each tone), then is [a] a minimal vowel or two minimal vowels?

A solution to the problem is offered by Duanmu (1994), according to which there is no compelling evidence for structures like (10). Instead, when a vowel carries two (or more) tones, the vowel is lengthened, in which case it can count as two (or more) sounds. The conclusion echoes a proposal made by Woo (1969), according to which a short vowel can only carry one tone. If Duanmu (1994) is correct, phonological structures are simpler than proposed by Goldsmith (1976). For example, it suffices for timing to be encoded on the timing tier alone, and there is no need for other tiers to encode timing (or sub-timing).

Under these considerations, I define a speech sound by two factors, shown in (11).

(11) A speech sound is a set of features such that (a) the features are made in one unit of time and (b) no feature is made twice by the same articulator

The two factors are related, because an articulator cannot complete two gestures in one unit of time (around 70 ms). In other words, since there is just one unit of time, no feature can be made more than once by a given articulator. By this definition, [i:] is two sounds because it takes two time units, and [ai] is two sounds because it takes two time units and two gestures for the height of the tongue, first low and then high. [p^h] is one sound because there is no clear evidence that it needs two time units (e.g. [p^hai] and [pai] have similar durations, where the length of the vowels compensates the length difference in aspiration). In addition, the gesture for [^{h}] (spread glottis) is made at the same time as [p], even though the aspiration continues after the release of [p]. I return to more examples below.

2.2. FEATURES

Most phonologists assume that sounds are made of features. There are at least four reasons for the assumption. First, features indicate how sounds are made. For example, [p] is made when the vocal cords are not vibrating (voiceless) and the lips (labial) are closed (stop). Second, features can show similarities and differences between sounds. For example, [p] is "voiceless labial stop" and [b] is "voiced labial stop." Thus, the two sounds are similar in two features and differ in one. Third, features can reveal natural classes of sounds. For example, the English plural suffix is [s] when added to *map*, *cat*, *back*, *fourth*, etc., and [z] when added to *job*, *food*, *mug*, *pen*, *mom*, *pill*, *bee*, *cow*, etc. The former set of words end in [p, t, k, θ], which belong to the class "voiceless," and the latter set of words end in [b, d, g, n, m, l, i, u], which belong to the class "voiced." Finally, features offer better representations of possible and impossible sounds, to be discussed below.

Since speech sounds are made by articulators, the simplest way to view features is that they are articulatory gestures. Some linguists believe that some features should also be defined in acoustic or auditory terms (e.g. Jakobson et al. 1952, Ladefoged 1980, 1992). However, for the present discussion, it is sufficient to consider articulatory features only.

If we view features as articulatory gestures, then for every feature we should specify the articulator that performs the gesture. Sometimes I will informally call a feature a gesture, although a gesture may involve two or more features. For example, the gesture to make a velar closure may involve the features [−back] and [+stop] for the Dorsal articulator. In this study I assume six articulators, following Ladefoged and Halle (1988), Halle (1992, 1995, 2005), and references therein. They are shown in (12).

(12)	Abbreviation	Articulators
	VC	Vocal-cords
	SP	Soft palate (velum)
	TR	Tongue root
	Dor	Dorsal (body of the tongue)
	Cor	Coronal (tip of the tongue)
	Lab	Labial (lips)

The names of articulators reflect the fact that they are moveable parts in the vocal tract. For example, the name "Vocal-cords" is better than the common name “Glottal” or “Glottis,” because the latter is the result of vocal cords movement. Similarly, alveolar and palatal are not articulators, because they cannot move; closures made at these places are initiated by Coronal and Dorsal.

Ideally, the names of features should also reflect the action of the gesture. For example, in a nasal sound, Soft-palate is lowered, and so [lowered] is a better feature name than [nasal]. In (13) I list the features to be assumes in this study (except tonal features), along with their articulators, where features are placed in brackets and articulator names are capitalized.

(13)	Features	Articulators	Other names
	[stop]	various	
	[fricative]	various	(edge closure)
	[spread]	Vocal-cords	aspirated
	[slack]	Vocal-cords	voiced
	[lowered]	Soft-palate	nasal
	[advanced]	Tongue-root	tense
	[high]	Dorsal	
	[low]	Dorsal	
	[back]	Dorsal	
	[anterior]	Coronal	
	[lateral]	Coronal	
	[round]	Labial	

The feature [stop], also called [continuant], refers to a complete closure; it can be made by different articulators, such as Labial, Coronal, or Dorsal. The feature [fricative] can also be made by different articulators. The action of [fricative] is somewhat difficult to define, and some studies simply represent fricatives as [−sonorant, −stop] (Chomsky and Halle 1968, Halle and Clements 1983). I consider [fricative] to be a closure made by the edges of an articulator. For example, in [s] the edges of Coronal are pressed against the roof of the mouth while the center is left open. Such a closure allows a narrow and strong airstream that causes turbulence, giving the characteristic fricative noise. The feature [spread] refers to whether Vocal-cords are spread apart; it determines whether a sound is aspirated. The feature [slack] refers to whether Vocal-cords are slack or stiff, which determines whether they can vibrate in a consonant (Halle and Stevens 1971). The feature [lowered] refers to the position of Soft-palate; it determines whether a sound is nasal. It is possible that [−lowered] is the same as [+stop], both meaning that the nasal cavity is closed by Soft-palate.

If so there is no need to use [lowered]. The feature [advanced] refers to the position of Tongue-root; in English it determines whether a vowel is tense (as in *beet*) or lax (as in *bit*) (Halle and Stevens 1969). The features [high], [low], and [back] refer to the position of Dorsal. The feature [anterior] refers to whether Coronal (the tongue tip) is positioned forward (towards the teeth) or backward (retroflex). For languages that have a three-way Coronal contrast such as dental-alveolar-retroflex, another feature may be needed. The feature [lateral] might be better called [narrow (tongue tip)]; the gesture would leave the sides of the tongue open. Finally, the feature [round] is made by Labial.

The features [stop] and [fricative] have been called manner features, stricture features, or articulator-free features. They can be made by more than one articulator. However, not every articulator can make [stop] or [fricative]. In particular, Tongue-root cannot make [stop], and Soft-palate cannot make [fricative] for the nasal cavity (the velar fricative [x] is made by Dorsal in the oral cavity, not by Soft-palate).

Some phonologists believe that stricture features belong to the sound as a whole, in the sense that each sound can only take one value for each stricture feature (Clements 1985, Sagey 1986, Halle 1992, 2005). For example, the sound [k^w] will be specified as [+stop], even though the labial articulator for [^{w}] does not have a stop closure. This view follows from the traditional classification of sounds into stops, fricatives, sonorants, etc. Other phonologists believe that stricture features are properties of articulators, so that each articulator has its own values of stricture features (Browman and Goldstein 1989, Keyser and Stevens 1994, Padgett 1995). For example, the sound [k^w] will have [+stop] for the Dorsal articulator and [−stop] for the Labial articulator. In this study I follow the latter view, because stricture features are ultimately carried out by individual articulators.

In the IPA tradition, there is no distinction between articulators and features. For example, [p] is thought to have three features, "voiceless," "labial," and "stop." In Sagey (1986), Ladefoged and Halle (1988), and Halle (1992), a distinction is made between articulators and features. For example, [p] has the feature [+stop], made by the articulator Labial, and the feature [−voice], made by the articulator Vocal-cords. Interestingly, Halle (2005) shows a return to the IPA tradition, in that articulators are again treated as features. For example, [p] is represented as [−voice], [Labial], and [+stop], where [Labial] is a mono-valued feature. In this study I follow Sagey (1986), Ladefoged and Halle (1988), and Halle (1992) and maintain a distinction between features and articulators throughout.

A number of common feature terms do not appear in the present feature list. In (14) I list some of them and their interpretation in feature structure.

(14)	Traditional terms	Feature structure
	retroflex	Coronal-[−anterior]
	palatal/distributed	Coronal and Dorsal-[−back]
	consonantal	(position in syllable)
	affricate	[+stop, +fricative]
	sonorant	[−stop, −fricative]

The feature [retroflex] is represented by [−anterior] of Coronal. The feature [distributed] is often used to describe palatals; in the present analysis a palatal is a combination of Coronal and Dorsal, often with [−back] for Dorsal. The feature [consonantal] is often used to distinguish glides from high vowels; in the present analysis the distinction is represented by different positions in the syllable. Finally, [sonorant] is replaced by [−stop, −fricative].

There are some other interesting questions about features that I do not pursue here. For example, can [lateral] be made by other articulators? Does [+low] imply Tongue-root? Does [−back] imply Coronal? And so on. Also, some phonologists argue that all features and articulators are organized into a tree structure called feature geometry (e.g. Clements 1985, Sagey 1986, Ladefoged and Halle 1988, McCarthy 1988, Halle 1995, Padgett 1995), although Halle (2005) argues that the tree can be considerably simplified. In this work I do not discuss any tree structure above articulators.

2.3. REPRESENTING AFFRICATES

The analysis of affricates has been a problem in feature theory. The difficulty lies in how to represent affricates with simultaneous gestures, instead of sequential gestures. Sequential gestures (or contour features) involve two or more values of a feature made by the same articulator, such as [+nasal, −nasal] by Soft-palate, [−round, +round] by Labial, or [−high, +high] by Dorsal. Such gestures usually take up twice as much time as a single sound does. Also, if sequential gestures are allowed, we predict too many possible sounds (Duanmu 1994). Halle and Clements (1983) and Sagey (1986), among others, represent affricates as [+stop, −stop], which are sequential gestures. Chomsky and Halle (1968) use the feature [delayed release], but it is hard to interpret the feature as a simple gesture. Jakobson, Fant, and Halle (1952), Steriade (1989), and Clements (1999) propose that affricates are strident stops, but it is unclear how to interpret [strident] as a gesture. Lombardi (1990) proposes that affricates are made with the gestures [+stop] and [−stop] simultaneously, but it is unclear how an articulator can make two opposite gestures at the same time.

I propose that affricates are made with the simultaneous gestures [+stop] and [+fricative]. The [+stop] gesture makes a complete closure. The [+fricative] gesture adds additional force at the edges of the articulator. This additional force at the edges causes a brief fricative effect at the release of the closure (in that the airstream breaks the center closure before the edge closure). In other words, the articulatory gestures for an affricate are made simultaneously, but the phonetic results are sequential, owing to the nature of articulatory aerodynamics.

2.4. ARE FEATURES INNATE?

A controversial question is whether features are innate—representing the most natural gestures, such as walking, laughing, coughing, swallowing, and blinking—or whether they are learned. Proponents of innate gestures often point to general patterns that are found across languages, such as a contrast between stops and fricatives, or a contrast between voiced and voiceless consonants. Critics of innate gestures, such as Ladefoged (1992) and Pierrehumbert (2001), often point to small differences between what seem to be corresponding sounds in different languages or dialects, such as [θ] in Californian English (where tongue protrusion is common) and that in Southern British English (where tongue protrusion is not common). Interestingly, although Sapir (1921) believes that walking is an innate ability, he also believes that linguistic abilities are not innate but learned, because speech seems to vary "without assignable limit."

Critics of innate features seem to assume that if a gesture is innate, it should be realized in exactly the same way in all languages. However, the assumption is not necessary. For example, supposing that walking is an innate ability, people who live in a mountain may still acquire a different style of walking than people who live on a plain, and people who often carry heavy weight may acquire a different style of walking than people who do not. This means that the realizations of innate abilities can be influenced by the environment. If so, the realizations of innate features can also be fine-tuned by the physical and linguistic environment. For example, the [t] in a language in which it is more frequently used may not be exactly the same as the [t] in a language in which it is less frequently used. Similarly, if the [t] in language A mostly occurs before [i] but the [t] in language B occurs equally before [i], [u], and [a], it is possible that [ti] in language A will be somewhat different from [ti] in language B. As Bybee (2001) and Pierrehumbert (2001) have argued, many factors, such as the phonological patterns of a language (the phonemic inventory and the phonotactics) and word frequencies, can affect the realization of a phonological entity. If so, the idea of innate features is compatible with the fact that the same feature can vary from language to language, or from person to person, just as the style of walking can vary from person to person.

A related question is whether features are discrete (categorical) or gradient. Proponents of innate features often believe that features are discrete. Critics of innate features often believe that features are gradient. Other things being equal, a theory that assumes discrete features is better, because it predicts fewer possible sounds and the prediction is easier to falsify. Despite the disagreement on whether features are innately discrete, there is general agreement that that features are discrete and categorical in the adult language.

2.5. SAMPLE SOUNDS IN FEATURES

Most sounds use just two or three features. For illustration, consider the consonants in (15). For ease of reading, I use some traditional feature names, such as [voice] instead of [slack] and [nasal] instead of [lowered]. I also use a dash (—) between an articulator and the features it executes. Following Keyser and Stevens (1994) and Padgett (1995), I assume that each articulator can have its own values for stricture features ([stop] and [fricative]).

(15) Sound Feature structure
[p] Labial—[+stop]
Soft-palate—[−nasal]
Vocal-cords—[−voice]
[m] Labial—[+stop]
Soft-palate—[+nasal]
[s] Coronal—[+fricative]
Vocal-cords—[−voice]
[ʃ] Coronal—[+fricative, −anterior]
Dorsal—[+fricative]
Vocal-cords—[−voice]
[ʧ] Coronal—[+stop, +fricative, −anterior]
Dorsal—[+stop, +fricative]
Vocal-cords—[−voice]
[l] Coronal—[+lateral]

Unspecified features are either implied or left to be determined by context. For example, we may assume that any sound unspecified for [fricative] is [−fricative] and any sound unspecified for [stop] is [−stop]. Similarly, we may assume that any fricative not specified for Soft-palate has a closed nasal cavity, namely Soft-palate—[−nasal] (or Soft-palate—[−lowered]). On the other hand, some features are determined by context. For example, in American English, [p] is [+aspirated] in *pan* but [−aspirated] in *happy*. The theory of underspecification predicts the extent to which features are specified, to be discussed below. For now it suffices to say that not all features are specified for every sound. The same is true for vowels. Some English vowels are shown in (16), where I interpret tense vowels as having the feature [+advanced] under Tongue-root.

(16) Sound Feature structure
[a] Dorsal—[+low, +back]
[i] Dorsal—[+high, −back]
Tongue-root—[+advanced]
[ɪ] Dorsal—[+high, −back]
Tongue-root—[−advanced]
[u] Dorsal—[+high, +back]
Labial—[+round]
Tongue-root—[+advanced]

In general, only contrastive features are are specified. In English, for example, there is no contrast between voiced and voiceless vowels, and so [voice] is unspecified for vowels. On the other hand, there is a contrast between tense and lax vowels, and so [advanced] is specified.

In all representations, each articulator only performs one value of a given feature. For example, Labial only performs [+round] or [−round] but not both. A feature can occur twice in a sound, but only if it is made by two different articulators simultaneously. For example, in [ʃ], [+fricative] is made by both Coronal and Dorsal; in [kʷ], [+stop] is made by Dorsal and [−stop] is made by Labial; and in [m], [+stop] is made by Labial and [−stop] is made by Soft-palate (if we consider [+nasal] to be the same as [−stop]).

2.6. COMPLEX SOUNDS AND THE NO CONTOUR PRINCIPLE

We can think of a complex sound as the merger of two (or more) sounds. The merger can be seen as gestural overlap (Browman and Goldstein 1989). There are three cases, shown in (17)–(19).

(17) Different articulators ([F] can be any feature, same or different)

Time-1	Time-2		Time-1
Art_a-[F]		→	Art_a-[F]
	Art_b-[F]		Art_b-[F]

Example: [k] (Dor-[+stop]) + [w] (Lab-[+round]) → [kʷ]

(18) Same articulator, different features

Time-1	Time-2		Time-1
Art_a-[F_i]	Art_a-[F_j]	→	Art_a-[F_i, F_j]

Example: [t] (Cor-[+stop]) + [s] (Cor-[+fricative]) → [tˢ] (Cor-[+stop, +fricative])

(19) Same articulator, different values of the same feature

Time-1	Time-2		
Art_a-[+F_i]	Art_a-[-F_i]	→	?

Example: [m] (SP-[+nasal]) + [b] (SP-[−nasal]) → [mb] (SP-[+nasal, −nasal])

In (17), two articulators that make separate gestures can do so simultaneously. In (18), an articulator that makes two gestures separately can also do so simultaneously. In (19) an articulator is making opposite values of the same feature. The case in (17) is not controversial. Most people also accept the case in (18), although there are different interpretations of what an affricate is. The case in (19) is controversial.

Sagey (1986) distinguishes two kinds of complex sounds. In the first, a complex sound involves simultaneous use of two or more oral articulators, where "oral articulators" refers to Labial, Coronal, and Dorsal; this is the case in (17). In the second kind (which Sagey calls "contour segments"), a complex sound is essentially a quicker pronunciation of two sounds in sequence. Sagey believes that both (18) and (19) belong to the second kind. For example, in her analysis the affricate [ts] (or [t^s]) involves the sequential gestures Coronal—[+stop, −stop]; the pre-nasalized stop [^{m}b] involves the sequential gestures Soft-palate—[+nasal, −nasal]; and [â] (a short falling tone on [a]) involves the sequential gestures Vocal-cords—[+H, −H].

However, there is no compelling evidence that pre-nasalized stops and short falling (or rising) tones exist as single sounds, as I have argued in Duanmu (1994). In addition, I have argued above that there is no need to assume sequential gestures for affricates. If so, we can maintain a simpler theory in which complex sounds observe the same constraints as regular sounds: no feature is made twice by the same articulator (see definition in (11) above). The constraint is called the No Contour Principle in Duanmu (1994), which I state in (20).

(20) No Contour Principle
An articulator cannot make the same feature (F) twice within one sound.

*Articulator	*Articulator	*Articulator	*Articulator
/\	/\	/\	/\
[+F][−F]	[−F][+F]	[+F][+F]	[−F][−F]

The constraint probably reflects a physical limitation that each gesture takes up a certain amount of time that cannot be further compressed.

The No Contour Principle assumes that all features in a complex sound are simultaneous, in the sense that no sequential timing difference can be used distinctively. For example, the merger of [p]+[l] and [l]+[p] would be the same, and that there would be no contrast between [p^l] and [^{l}p]. However, I shall discuss in Chapter 3 why the merger of [l]+[p] rarely occurs, and why the merger of [p]+[l] rarely occurs in the coda.

A complex sound is usually represented with two phonetic symbols, but sometimes it is represented with just one. For example, in English [s] and [j] can merge into [ʃ] (as happens in *this year*), which is one symbol (although the graph [ʃ] implies both [s] and [j]). It is also worth noting that a complex sound need not use more articulators than a simple sound. For example, [t^{s}] is a merger of [t] and [s], yet it uses just one oral articulator, which is Coronal; in contrast, [u] is usually thought to be a simple sound, yet it uses two oral articulators, Labial and Dorsal.

Feature structure offers a better prediction of possible and impossible complex sounds than IPA symbols. For example, IPA symbols allow us to merge any two symbols together, such as those in (21).

(21) Possible sounds as allowed by IPA symbols
[p^{k}], [k^{p}], [k͡p]; [k^{m}], [^{k}m], [k͡m]

However, not all combinations are possible sounds. For example, there is no evidence that [p^{k}], [k^{p}], and [k͡p] can contrast with each other. In feature analysis, there is only one way to represent a combination of [p] + [k], which is shown in (22), where Vocal-cords features are omitted.

(22)	[k]	Dorsal—[+stop]
	[p]	Labial—[+stop]
	[k] + [p]	Dorsal—[+stop], Labial—[+stop]

In [k] + [p] the Dorsal and Labial actions are made simultaneously, and therefore the proper IPA symbols should be [k͡p]. Similarly, consider [k^{m}], [^{k}m], and [k͡m]. There is again no evidence that they can contrast with each other. Indeed, none of them is a possible sound in feature structure. There are four possible results of [k] + [m], shown in (23), none of which is [k^{m}], [^{k}m], or [k͡m].

(23)	[k]	Dorsal—[+stop], Soft-palate—[−nasal]
	[m]	Labial—[+stop], Soft-palate—[+nasal]
	[ŋ͡m]	Dorsal—[+stop], Labial—[+stop], Soft-palate—[+nasal]
	[k͡p]	Dorsal—[+stop], Labial—[+stop], Soft-palate—[−nasal]
	*[k͡p͡ŋ͡m]	Dorsal—[+stop], Labial—[+stop], Soft-palate—[−nasal, +nasal]
	*[ŋ͡m͡k͡p]	Dorsal—[+stop], Labial—[+stop], Soft-palate—[+nasal, −nasal]

Since [k] is [−nasal] and [m] is [+nasal], there is a conflict. If we go with [+nasal], the result of [k] + [m] is [ŋ͡m] (same as [m͡ŋ]), which is a possible sound. If we go with [−nasal], the result is [k͡p] (same as [p͡k]), which is also a possible sound. According to Sagey (1986), we can also keep both values of [nasal] and order them in sequence. If the order is [−nasal, +nasal], the result is [k͡p͡ŋ͡m]. If the order is [+nasal, −nasal], the result is [ŋ͡m͡k͡p]. In the present analysis, [k͡p͡ŋ͡m] and [ŋ͡m͡k͡p] are not possible sounds, because in each case the feature [nasal] occurs twice under Soft-palate. It can also be shown that, in the present analysis, neither "falling diphthongs," such as [ou] and [ai], nor "rising diphthongs," such as [ia] and [ua], can be represented as single complex sounds, although in Sagey's theory they all can.

2.6.1. **Affricates**

Although the present feature theory allows fewer complex sounds than Sagey (1986) in some cases, there is a case where the present theory allows more complex sounds than Sagey's. The case involves affricates. Consider the representations in (24) and (25).

(24) Affricate with the same articulator

Sounds	Feature structure
[t]	Coronal—[+stop]
[s]	Coronal—[+fricative]
[t^s]	Coronal—[+stop, +fricative]

(25) Affricates with different articulators

[p^s]	Labial—[+stop], Coronal—[+fricative]
[p^x]	Labial—[+stop], Dorsal—[+fricative]
[k^s]	Dorsal—[+stop], Coronal—[+fricative]
[$k^ʃ$]	Dorsal—[+stop, +fricative], Coronal—[+fricative]

The affricate in (24) involves the same articulator (Coronal). Such affricates are allowed by both the present theory and Sagey (1986). In addition, in the present theory each articulator can have its own values of [stop] and [fricative], following Keyser and Stevens (1994) and Padgett (1995). Therefore, it is possible for one articulator to make a stop gesture and another articulator to make a fricative gesture, as those in (25). In other words, any stop can combine with any fricative to make an affricate, as proposed by Prinz and Wiese (1991) and Wiese (1996). In contrast, for Sagey (1986), if one articulator makes a stop, the other articulator must either make a stop, such as [p͡t], or an approximant, such as [p^j], but not a fricative, such as [p^s]. The present theory is more compatible with the analysis of syllable structure. For example, the sound [p^x] occurs in Standard Chinese (often transcribed as [p^h]). The recognition of [p^s] and [k^s] as complex sounds will be relevant to the analysis of the syllable in German. The recognition of [$k^ʃ$] is relevant to the analysis of the syllable in Hindi (Kumar 2005).

2.6.2. **Palatals and consonant–approximant clusters**

Feature structure can also show subtle differences among sounds that have similar articulations. For example, all the sounds in (26) involve a palatal quality, which is represented by the use of both Coronal and Dorsal. They differ in the values of [stop] and [fricative] under each articulator.

(26)
[j]	Coronal—[−stop, −fricative] Dorsal—[−stop, −fricative]
[ç]	Coronal—[−stop, −fricative] Dorsal—[−stop, +fricative]
[ɕ]	Coronal—[−stop, +fricative] Dorsal—[−stop, +fricative]
[s^j]	Coronal—[−stop, +fricative] Dorsal—[−stop, −fricative]
[t^j]	Coronal—[+stop, −fricative] Dorsal—[−stop, −fricative]
[ts^j]	Coronal—[+stop, +fricative] Dorsal—[−stop, −fricative]
[tɕ]	Coronal—[+stop, +fricative] Dorsal—[−stop, +fricative]
[k^j]	Coronal—[−stop, −fricative] Dorsal—[+stop, −fricative]

Next, we consider complex sounds that are made of a consonant plus an approximant, commonly found in the syllable onset. Some examples are shown in (27).

(27)
Sounds	Feature structure
[r^w]	Labial—[+round] Coronal—[−anterior]
[k^l]	Dorsal—[+stop] Coronal—[+lateral]
[k^{rw}]	Labial—[+round] Dorsal—[+stop] Coronal—[−anterior]
[t^{rw}]	Labial—[+round] Coronal—[−anterior, +stop, +fricative]

All these sounds occur in English. In particular, the English [r] is [r^w] in syllable-initial position, which some speakers (especially children) pronounce as just [w]. The other three complex sounds are usually transcribed as [kl], [kr], and [tr], thought to be clusters of two sounds each. However, as far as feature structure is concerned, they can all be represented as complex sounds. Of interest is the fact that [l] and [r] are voiceless in [kl], [kr], and [tr] and voiced in [gl], [gr], and [dr], in agreement with the complex sound analysis.

Since [tr] ([t^{rw}]) has the features [+stop, +fricative] under Coronal, it is an affricate, which is a fairly common view (Jones 1950, Abercrombie 1967, Gimson 1970, Wells 1990, Lawrence 2000). Rob Burling (p.c.) also told me that a friend of his once insisted that *try* should be spelled as *chry*, where the word starts with an affricate. The present analysis is in agreement with such intuition.

Less obvious are the clusters [tl] and [dl]. It seems possible to represent them as complex sounds in English. This is shown in (28).

(28) [d] Coronal—[+stop]
[l] Coronal—[+lateral]
[dl] Coronal—[+stop, +lateral]

Since [l] is the only lateral sound in English, its stricture features can be unspecified. This makes it possible to represent [dl] (and [tl]) as a complex sound. An obvious question is: Why does English not use [dl] and [tl] as onsets? A possible answer is that English does not need to use all possible complex sounds. For example, [ts] and [dz] are possible complex sounds, but English does not use them as onsets (except in occasional cases, such as *Tswana*), although Chinese does. Similarly, [pw] and [bw] are possible complex sounds, represented as Labial—[+stop, +round], but English does not use them either, although Spanish does.

2.6.3. **VNC clusters**

Another case of interest is VNC clusters, where the nasal N and the consonant C are homorganic (sharing the same place of articulation). In this case, VNC can be analyzed as ṼC. An example is given in (29), where features are shown only for sounds of interest.

(29) VNC → ṼC (e.g. [tɛnt] → [tɛ̃t] *tent*, [pʌmp] → [pʌ̃p] *pump*)
[ɛ] Dorsal—[−back, −high, −low]
[n] Coronal—[+stop]
Soft-palate—[+nasal]
t[ɛ̃]t Dorsal—[−back, −high, −low]
Soft-palate—[+nasal]

VNC can be pronounced as three sounds, if one speaks carefully, or it can also be pronounced as two sounds, where VN merges into one sound. The merger VN → Ṽ is found in Chinese and English. In some Chinese dialects, such as Shanghai, there is no place contrast in the nasal coda and so only the nasality of N needs to be preserved on the vowel. In English, as seen in *tent* and *pump*, the place feature of the nasal is the same as that of the following consonant, and so the nasality of N is again the only feature that needs to be preserved on the vowel.

2.6.4. **[Vʔ] clusters**

It is also possible for a vowel and a glottal stop to merge into a complex sound. The representation is shown in (30).

(30) [Vʔ] → [V^{ʔ}], e.g. [aʔ] → [a^{ʔ}]

[a]	Dorsal—[+back, +low]
[ʔ]	Vocal-cords—[+constricted]
[a^{ʔ}]	Dorsal—[+back, +low]
	Vocal-cords—[+constricted]

[Vʔ] → [V^{ʔ}] is found in Shanghai, where [Vʔ] can be pronounced as one sound [V^{ʔ}] if one speaks fast, or as two sounds if one speaks carefully.

2.7. CONSONANTS VS. VOWELS

We have not used the traditional feature [consonantal], which separates consonants from vowels. Still, we can distinguish several useful categories. Consider those in (31) for English.

(31)

Categories	Feature structure
vowels	Dorsal—[−stop, −fricative] (traditional vowels)
syllabic sounds	[−stop, −fricative] (vowels, nasals, [r] and [l])
stressable sounds	[−stop, −fricative, −lateral] (vowels, [r], and nasals)

Traditional vowels can be represented as Dorsal—[−stop, −fricative]. Syllabic sounds in English include traditional vowels, nasals, [r], and [l], which can be represented as [−stop, −fricative] for either the oral or the nasal air-passage. Among the syllabic sounds, all except [l] can be stressed in American English and Chinese, and we can specify this group as [−stop, −fricative, −lateral] (stressed nasals are not common but can occur in interjections, such as [hm]). In previous analyses it was difficult to answer whether the syllabic sound in the American English words *fur* and *bird* is a vowel [ɝ] or a consonant [r]. In the present analysis, the different categorizations of vowel-like sounds can be represented without the traditional feature [consonantal] or [vocalic].

2.8. LENGTH AND DIPHTHONGS

Sometimes a language has both short and long sounds, such as [t] vs. [tt] (or [t:]), and [a] vs. [aa] (or [a:]). The length of a sound can be represented by timing slots, shown in (32), where each feature structure is abbreviated to a phonetic symbol.

(32) Representing length

X	XX	X	XX	Timing slots
|	\/	|	\/	
t	t	a	a	Feature structure (abbreviated)
[t]	[tt]	[a]	[aa]	Transcription

The timing slot representation indicates length directly: a short sound has one timing slot and a long sound has two (and sometimes three). Diphthongs are also represented with two timing slots each, as shown in (33).

(33) Representing diphthongs

```
XX   XX
| |  | |
e i  a u
```

The analysis reflects the fact that diphthongs involve sequential gestures, such as [−high, +high], and so they cannot be represented as a single sound. In addition, diphthongs behave like long vowels. For example, in English long vowels and diphthongs can appear in a stressed monosyllable without a final consonant, but short vowels cannot.

2.9. UNDERSPECIFICATION

The basic claim of underspecification theory is that some features can be unspecified. Halle (1962) proposes that sounds are maximally underspecified in the lexicon, because predictable features can be filled in by redundancy rules; this way we achieve better economy by using the fewest symbols possible. There are two other, more direct, arguments for underspecification. The first is the transparency effect, whereby in a string of sounds ABC, A can affect C, or vice versa, as if B is transparent. A classic example is Turkish, shown in (34).

(34)	Root	Plural	Genitive	Plural-Genitive	
	ip	ip-ler	ip-in	ip-ler-in	'rope'
	son	son-lar	son-un	son-lar-ɯn	'end'

In the plural column, the suffix vowel gets its values for [back] from the vowel of the root: when the root vowel is [−back] the suffix vowel is [−back] (the [i]–[e] sequence in 'rope'), and when the root vowel is [+back] the suffix vowel is [+back] (the [o]–[a] sequence in 'end'). In the genitive column, the suffix vowel gets its values for both [back] and [round] from the vowel of the root: when the root vowel is [−back, −round] the suffix vowel is [−back, −round] (the [i]–[i] sequence in 'rope'), and when the root vowel is [+back, +round] the suffix vowel is [+back, +round] (the [o]–[u] sequence in 'end'). According to the underspecification analysis, features can pass through sounds that are not specified for them, but not through sounds that are. In (34) consonants are unspecified for [back] and [round], so these features can pass through them. It is interesting to note that, in the [o]–[a]–[ɯ] sequence of the plural-genitive form of 'end', the last vowel does not assimilate to the [+round] of [o]. This is because the intervening [a] is specified for [−round], which is itself spread onto the third vowel.

The second phenomenon is that specified features resist change. For example, in Turkish the vowel of the plural suffix remains [−high, −round], regardless of the preceding vowel, and the vowel of the genitive suffix remains [+high], regardless of the preceding vowel. In general, a sound does not change the features it is specified for; it only adopts features it is unspecified for from the neighboring sounds.

There remains some disagreement on how much a sound can be underspecified, both at the underlying level and at the surface level (Steriade 1987, Archangeli 1988, Keating 1988, Rice 1992, Inkelas 1994). However, the controversy need not concern us here.

2.10. SOUNDS VS. PHONEMES

The definition of sounds given above is similar to the traditional notion of phones. Not every sound or phone is felt to be different by the native speaker, though. For example, in American English the 'p' in *pen* is different from that in *open*: in the former it is [pʰ] and in the latter it is [p]. However, American English speakers usually consider [pʰ] and [p] to be the same. A set of related sounds that the native speaker considers to be the same are allophones, and each set of allophones represents a phoneme.

Sometimes two phonemes can be realized in the same way. For example, in English, [s] can become [ʃ] before [j], exemplified in (35).

(35) [sj] → [ʃj]
[ðɪs jeɚ] → [ðɪʃ jeɚ] *this year*
[ou ɛs ju] → [ou ɛʃ ju] *OSU* (Ohio State University)

For such speakers, the phonemes [s] and [ʃ] can both be realized as [ʃ]. As a result, when we see the sounds (phones) [ʃj], we are not sure whether the first comes from the phoneme [s] or the phoneme [ʃ], and we need more information about the language to know the answer. For example, English speakers know that there is a word *this* [ðɪs] but no word **thish* [ðɪʃ], and so the first word in (35) must be *this*.

Yet another complication is that sometimes a sound (or phone) is the result of the merger of two phonemes. For example, many speakers of Standard Chinese have the sound [ɲ], which comes from the merger of [n] and [j], exemplified in (36).

(36) [nj] → [ɲ]
[njan] → [ɲan] 'year'

The sound [ɲ] occurs only when we expect [n] and [j]. Similarly, English has the affricate [tˢ], which is a complex sound, but it occurs only when we expect [t] and [s].

The examples show that the interpretation of minimal pairs is more complicated than it appears. For example, given a pair of words [AXB] and [AYB], where X and Y are single sounds and [A__B] is the environment, what we can conclude is that X and Y do not belong to the same phoneme (ignoring free variation), but we cannot conclude that X (or Y) belongs to a single phoneme, because X (or Y) could belong to two possible phonemes (the [s]/[ʃ] case), or X (or Y) could be the merger of two phonemes (the [ɲ] = [n] + [j] case). Therefore, as argued by Halle (1962) and Chomsky (1964), the analysis of phonemes may depend on the consideration of the entire sound system of a language, including all the words in its lexicon.

2.11. SUMMARY

In this brief discussion of features and speech sounds, I have assumed only articulatory features, the fewest number of articulators that all analyses need, fewer features than proposed by other analyses (e.g. no [consonantal] or [sonorant]), and the No Contour Principle (i.e. no sound can contain two or more values of the same feature under the same articulator). I have shown that this minimalist feature theory is powerful enough to represent many possible complex sounds and restrictive enough to exclude many others, and that feature theory makes precise predictions of which complex sounds are possible and which impossible. The understanding of complex sounds will be crucial for the discussion of syllable structure in the following chapters.

3

Theories of syllable structure

3.1. DEFINING THE SYLLABLE

The phonetic definition of the syllable is notoriously difficult. A common view is that a syllable is a prominence peak, but the definition says little about where syllable boundaries are. Also, it is unclear why some phonetic peaks are not treated as syllables, such as the [s] in *stop*, *extra*, and *cats*. Another common definition is that a syllable is related to a chest pulse, or a pulse of air pressure. However, this definition again says little about where syllable boundaries are. In addition, as Gimson (1970) puts it, it is doubtful whether a double chest pulse will be evident in a word like *seeing* [siːɪŋ]. Nor can the pulse theory decide whether a word like *beer* [biɚ] in American English is one or two syllables.

Such problems have led to a certain skepticism about whether syllables are real linguistic units. For example, Chomsky and Halle (1968) do not consider the syllable to be a relevant phonological entity. Similar reservations are expressed by Gimson (1970), Steriade (1999), and Blevins (2003).

Nevertheless, in some case it is fairly clear what syllables are. For example, all people agree that *Canada* has three syllables and *America* has four, and many people agree that *after* can be divided as *af-ter*, *cactus* as *cac-tus*, and *whiskey* as *whis-key*. In addition, just as the lack of a definition of life (or death) does not prevent biologists from studying living things, the lack of a definition of the syllable should not prevent us from studying syllables. I shall argue that many questions about the syllable can be addressed, and reasonably answered, such as what the maximal syllable size is, what a possible onset is, and how to determine syllable boundaries. Such results constitute concrete progress towards the understanding of the syllable.

3.2. MAXIMAL SYLLABLE SIZE AND WORD-EDGE CONSONANTS

Theories of syllable structure often assume a maximal syllable size for a given language. Syllables within this size are in principle good. For example, if the maximal size is CCVCC, then CVCC, CCVC, CVC, and CV are generally good.

It is well known that word-medial syllables are generally quite simple, but extra consonants can occur at word edges. This is the case, for example, in Greek (Steriade 1982), English (Borowsky 1986), German (Giegerich 1985, 1989), Bella Coola (Bagemihl 1991), Spokane Salish (Bates and Carlson 1992), Polish (Bethin 1992), Georgian (Butskhrikidze 2002), and Jiarong (Lin 1993). Therefore, the maximal syllable size is mainly related to how we treat word-edge consonants. For example, consider word-final consonants in English. Several approaches are shown in (1).

(1) Rhyme size and the treatment of word-edge consonants

Analysis	*texts*	*helped*	Maximal rhyme
All-in	[tɛksts]	[hɛlpt]	VCCCC
Suffix-out	[tɛkst]s	[hɛlp]t	VCCC
Coronals-out	[tɛk]sts	[hɛlp]t	VCC
Medial-based	[tɛk]sts	[hɛl]pt	VC
Two-step	[[tɛk]sts]	[[hɛl]pt]	VCCCC

The “all-in” analysis is assumed by Jones (1950), Abercrombie (1967), Haugen (1956a, b), Fudge (1969), Hoard (1971), Kahn (1976), Hammond (1999), Hall (2002a), and Blevins (2003), and many others. However, it has two problems. First, it must explain why medial syllables remain small. Second, it must explain why all sounds must be in a syllable. One might think that it is not possible to pronounce a consonant without a syllable, but English does have consonant interjections, such as *shh* [ʃ], *pff* [pf], and *psst* [ps], and English speakers have no trouble saying [s] or [f] alone. If such utterances are not syllables, then consonants can be pronounced without being in a syllable, and there is no need for the all-in analysis. If such utterances are syllables, there is another problem: a word like *texts* need not be one syllable but can be three [tɛk][s][ts], because [s] can be a syllable by itself (or along with [t]).

The “suffix-out” analysis is assumed by Selkirk (1982), which yields smaller syllables than the all-in analysis. The “coronals-out” analysis is assumed by Kiparsky (1981), among others. It is based on the observation that most word-final consonants in English are coronals [s, z, t, d, θ]. If we exclude them, the maximal rhyme is VCC. The "medial-based" analysis is proposed by Giegerich (1985) and Borowsky (1986). It is based on the observation that medial rhymes are limited to VX (VV or VC). Therefore, more word-edge consonants are excluded.

While the suffix-out, coronals-out, and medial-based analyses yield a smaller syllable size, they face a common question, which is how to account for extra consonants at word edges. I shall argue below that they can be accounted for by morphology. If so, there is no need to assume a stretched syllable size at word edges. Instead, we can maintain a consistent syllable size for both edge and non-edge positions.

The "two-step" analysis has been proposed by Kiparsky (1981), Steriade (1982), Borowsky (1989), and Giegerich (1992). It assumes that syllables are built in two steps. In the first, only core syllables are built and some word-edge consonants are excluded. In the second, word-edge consonants are absorbed by an adjacent syllable. Thus, Kiparsky's analysis of *texts* is [[tɛk]sts], where the inner brackets indicate the core syllable and the outer brackets indicate the final syllable. However, it is not obvious what the advantage of the two-step analysis is. For example, it does not explain why certain consonants can be absorbed but others cannot. In addition, if we can account for word-edge consonants by morphology, there is no need to assume that they have to be absorbed by or appended to a syllable. Moreover, if a syllable can take in more consonants, why does it not do so word-medially? By the same argument, there is no advantage in assuming that word-edge consonants are appended to the word (Vaux 2004).

In summary, if we can account for word-edge consonants independently, we can maintain a consistent maximal syllable size, similar to that in word-medial positions.

3.3. EMPTY ELEMENTS AND THE CV-ONLY ANALYSIS

The maximal syllable size is also related to whether we assume empty elements. There are two kinds of empty elements in phonology. The first has phonetic realization. A well-known example is the pause at major phrase boundaries. Another example is the empty beat between stressed syllables, or at the end of a verse line. An example is shown in (2), where an empty beat is indicated by Ø.

(2) Empty elements that have phonetic realization (pauses)
(Ding Ø) (dong Ø) (bell Ø) (Ø Ø)
(Kitty's) (in the) (well Ø) (Ø Ø)

The rhythm in (2) is common in children's verse cross-linguistically (Burling 1966). If one taps the lines, each line has four feet and each foot has two beats. The second beat of the first two feet in line 1 is empty, in that it does not correspond to a written syllable. Similarly, the last three beats of each line are also empty. While empty beats are not indicated in orthography, they have phonetic realization, either as a pause or as the lengthening of the preceding syllable.

The second kind of empty elements have no phonetic content. For example, Lowenstamm (1996) and Scheer (2004) propose that both C and V can be empty, anywhere in a word. In addition, all sounds are syllabified into CV syllables. Two examples are shown in (3), where Ø is an empty C or V. In the CV-only analysis, most words that used to be called monosyllabic are now polysyllabic.

(3) The CV-only analysis
mix [mɪ][kØ][sØ]
spiked [sØ][pa][Øi][kØ][tØ]

The empty elements are purely abstract because there is in principle no contrast between [spa] and [sØ][pa], between [pai] and [pa][Øi], or between [kt] and [kØ][tØ]. The CV-only analysis necessarily requires [spa] to be represented as [sØ][pa], [pai] as [pa][Øi], and [kt] as [kØ][tØ]. In this regard, the use of Ø is theory internal (or circular).

The CV-only analysis has two problems. First, it must explain why only some word-edge Cs can occur with an empty V but others cannot. For example, why can [sØ] occur word-initially in English but [fØ] cannot (e.g. *stop* is found but **ftop* is not)? Similarly, it must explain why extra word-final Cs are mostly [s, z, t, d, θ] in English (e.g. *text* is good but **texp* is not). The CV-only analysis must say, as other theories do, that [s] is special initially and [s, z, t, d, θ] are special word-finally. But once we agree that these sounds are special, it becomes unnecessary to assume that they must also occur with an empty V, or be syllabified at all.

Another problem with the CV-only theory is that it opens up alternative solutions that are hard to choose from. For example, consider three analyses of *spiked*, shown in (4).

(4) Possible analyses of *spiked* with purely empty elements

CVX-only	[sØØ][pai][kØt]
CV-only	[sØ][pa][Øi][kØ][tØ]
VC-only	[Øs][Øp][aØ][ik][Øt]

The CVX-only solution uses a larger syllable size but fewer syllables and fewer empty elements. The CV-only and VC-only solutions use the same syllable size, the same number of empty elements, and the same number of syllables. All the solutions can represent any word without problem. It is not obvious how one should choose among the alternatives.

To conclude, there is no obvious advantage in using empty elements that do not have phonetic realization, nor is there any advantage of the CV-only theory.

3.4. THE CVX THEORY

The CVX theory proposes that a word has the schematic structure in (5).

(5) The CVX analysis of word structure: C_mCS . . . SCC_m

S . . . S	one or more syllables, whose maximal size is CVX
C	an optional C supported by a potential V (and anti-allomorphy)
C_m	one (or more) optional affix or affix-like consonant

Let us begin with non-edge syllables. The CVX theory claims that in word-medial positions the maximal syllable contains three sounds, which can be CVV, such as [hau] *how* and [bi:] *bee*, or CVC, such as [bɛt] *bet*. I further assume that VX forms a constituent, which can be called the rhyme (although this constituent has been questioned by Davis 1988). There are three reasons for the constituent. First, it is well known that, while there are restrictions within the onset or within the rhyme, there are fewer restrictions between onsets and rhymes (Kessler and Treiman 1997). Second, the onset usually does not affect stress, whereas the coda often does (see Goedemans 1998, who argues that the onset never affects stress). Third, VX is the unit in syllable rhyming, so that *buy* [bai] rhymes with *why* [wai], to the exclusion of the onsets [b] and [w]. In contrast, there is no evidence that CV forms a unit to the exclusion of the coda in any phonological process. The structure of CVX is therefore given in (6).

(6) The maximal syllable: CVX

```
     σ
    /\
   O  R     Onset/Rhyme
   |  /\
   X X X    Timing slots
```

CVX is the shorthand for the structure in (6). The full structure itself does not indicate that the second slots must be filled by a vowel. Therefore, it is possible that the second slot is filled by a syllabic consonant. Some examples are shown in (7).

(7) Sample syllables

Transcription	C V X	
[bæt]	[b æ t]	*bat*
[bai]	[b a i]	*buy*
[wai]	[u a i]	*why*
[jæm]	[i æ m]	*yam*
[bi:]	[b i i]	*bee*
[æn]	[æ n]	*Ann*
[n:]	[n n]	'fish' (Shanghai Chinese)
[sz:]	[s z z]	'four' (Standard Chinese)

There are several points of interest. First, the C slot can be filled by a consonant, as in *bat*, or a high vowel (transcribed as a glide), as in *why* and *yam*. Second, the V slot can be filled with a syllabic consonant, as in [n:] 'fish' in Shanghai Chinese and [sz:] 'four' in Standard Chinese. Third, the onset slot need not be filled, as in [æn] *Ann* and [n:] 'fish' in Shanghai Chinese; I shall return to this issue below. Finally, [ii], [nn], and [zz] are not repetitions of the same sound but a sound doubly linked to two timing slots, represented in (8). A doubly linked sound has one articulatory configuration that lasts for two units of time.

(8)

```
CVX    VX    CVX
| \/   \/    | \/
b  i    n    s  z
[bi:]  [n:]  [sz:]
```

In what follows I discuss the distribution of sounds inside CVX and sound strings that seem to exceed CVX.

3.4.1. **Sonority, the peak rule, and the C rule**

In this section I discuss the distribution of sounds within CVX. For example, given the sounds [ni], can they form a syllable [nni], where

[n] is doubly linked to CV and [i] is linked to X? Following a long tradition, I assume that sonority, or the loudness of a sound, governs the distribution of sounds inside a syllable. According to Jespersen (1904), speech sounds can be ranked along the sonority scale in (9), where > means "has greater sonority than."

(9) Sonority scale (Jespersen 1904: 192):
low vowels > mid vowels > high vowels > r-sounds > laterals > nasals > voiced fricatives > voiced stops > voiceless fricatives > voiceless stops

It is possible to derive the sonority scale from phonological features. For example, among vowels, [+low] is more sonorous than [−low] and [−high] is more sonorous than [+high]. However, I shall omit the derivation (see Hooper 1976a, Kiparsky 1981, Steriade 1982, and Rice 1992 for such proposals). Within a syllable, the most sonorous sound should occur in the nucleus. Let us call it the peak rule, stated in (10).

(10) The peak rule
The most sonorous sound should fill the V slot of CVX.

The most sonorous sound can also fill C or X, as long as it fills V. Some examples are shown in (11), where Ø indicates an unfilled onset.

(11) Illustration of the peak rule

Input	CVX analysis	Peak rule violation
[bi]	[bii], *[bbi]	[i] >[b]
[sz]	[szz], *[ssz]	[z] > [s]
[ia]	[iaa], *[Øia], *[iia]	[a] > [i]
[i]	[Øii], [iii]	
[n]	[Ønn], [nnn]	
[æn]	[Øæn], *[ææn]	??

In the first three cases, the bad forms violate the peak rule, because the sound in V is less sonorous than the sound X. For example, in *[ssz], V is filled by [s], which is less sonorous than [z]. Also, the most sonorous sound can extend to C, as in [iii], and to X, as in [iaa] and [iii]. Finally, some input forms allow two (or more) analyses. For example, the input [i] can be syllabified as [Øii] or [iii]. In fact, [i] can also be syllabified as [Øi], where X is left out; the same is true for [n]. However, *[ææn] cannot be ruled out by the peak rule. If *[ææn] is bad, we need a separate rule, which I call the C rule, stated in (12).

(12) The C rule
The C slot of CVX cannot be filled by a non-high vowel.

Vennemann (1988: 70) makes an interesting observation that [rn] can be syllabified in two ways: either [r] is the nucleus, as in *lantern* [læn.tr̩n], or [n] is the nucleus, as in *apron* [ei.prn̩]. If [r] has greater sonority than [n], the peak rule should rule out [prn̩]. In the CVX analysis, there is an explanation. Since there is only one onset slot, [prn̩] must be [p^{r}n̩], where [p^{r}] is a complex sound. If the sonority of a complex sound is the same as the articulator that has greater closure (i.e. the sonority of [p^{r}] is the same as that of [p]), then [p^{r}n̩] no longer violates the peak rule.

The peak rule and the C rule suffice to determine the distribution of sounds inside CVX. In contrast, in other analyses, sonority is also used to analyze onset clusters, which I discuss next.

3.4.2. **Onset clusters: sonority analysis vs. complex-sound analysis**

Many languages allow onsets that are often transcribed as an obstruent–approximant cluster, such as [pr, pl, br, bl, fr, fl, kj, kw, . . .]. Most analyses assume that onset clusters are governed by a sonority requirement (e.g. Selkirk 1982, Steriade 1982), which I rephrase in (13), illustrated with English examples in (14).

(13) The sonority analysis of onset clusters
The sonority in an onset cluster must show a sufficient rise.

(14) Examples of the sonority analysis of onset clusters

Cluster	Comment
pr	good: enough sonority rise
fr	good: enough sonority rise
θr	good: enough sonority rise
sr	good: enough sonority rise
*fn	bad: not enough sonority rise

The degree of sonority rise is measured by the sonority scale, such that the farther apart two sounds are on the scale, the greater sonority rise there is between them. More discussion of the sonority analysis of English onsets is given in Chapter 8.

In the CVX theory, there are no onset clusters. What appears to be an onset cluster is a complex sound (see Chapter 2). The proposal is given in (15) and exemplified in (16).

(15) The complex-sound analysis of onset "clusters"
Two sounds can fit into the C slot of CVX only if they can form a complex sound.

(16) a. Clusters that can form a complex sound: pr, fr, pl, kr, . . .
b. Clusters that cannot form a complex sound: θr, ʃr, ʃm, ʃn, sr, sl, sm, sn, st, sp, sk, sf, . . .

In the sonority analysis, [θr] and [sr] are both good, because they have the same sonority rise as [fr]. In contrast, *[θr] and *[sr] are bad in the complex-sound analysis, even though [fr] is good. The reason is that [f] and [r] use different articulators and so can form a complex sound, whereas [θ] and [r] use different feature values of the same articulator, which violates the No Contour Principle (Chapter 2) and so they cannot form a complex sound. The feature structures of [fr] and [θr] are shown in (17) and (18).

(17) [f] Labial—[+fricative]
[r] Coronal—[−fricative]
[f] + [r] Labial—[+fricative], Coronal—[−fricative]

(18) [θ] Coronal—[+fricative, +anterior]
[r] Coronal—[−fricative, −anterior]
[θ] + [r] *Coronal—[+fricative, −fricative, +anterior, −anterior]

For the same reason, [sr] (*Coronal-[+fricative, −fricative]) and [fn] (*Soft-palate-[−nasal, +nasal]) cannot form a complex sound or serve as an onset, nor can other clusters in (16b).

Since the sonority analysis and the complex-sound analysis make different predictions, we can test their claims. In particular, the sonority analysis predicts that clusters like [θr, ʃr] are possible word-medial onsets, because they have a proper sonority rise. In addition, clusters like [ʃm, ʃn, sr, sl, sm, sn, st, sp, sk, sf] are possible word-medial onsets, because [s] (and [ʃ]) is exempt from the sonority requirement. In contrast, the complex-sound analysis predicts that such clusters are not possible onsets and will not be found word-medially. We shall see in Chapter 8 that evidence supports the complex-sound analysis.

3.4.3. **VVN and VNC rhymes**

Although most non-final rhymes in English are limited to VX, sometimes VXC rhymes have been found (Borowsky 1989). Two well-known cases are VNC, where C is voiceless, such as [ɪmp] in *symptom*, and VVN, such as [aun] in *council*. In both cases the N has the same place of articulation as the following C. For example, in [mp] both sounds are Labial and in [ns] both sounds are Coronal.

Borowsky (1989) and Hall (2001) propose that an NC cluster with the same place of articulation can count as one sound. However, in feature theory NC clusters like [mp] and [ns] cannot be represented as a complex sound, because they contain the ill-formed structure Soft-palate-[+nasal, −nasal], which violates the No Contour Principle (Chapter 2). Therefore, Borowsky (1989) and Hall (2001) have not explained the VNC and VVNC problem. Goldsmith (1990: 151) proposes that English allows an extra sonorant between the nucleus and the coda, but the reason again remains unclear.

While NC cannot form a complex sound, VN can form a complex sound Ṽ, and so VNC becomes ṼC and VVNC becomes ṼṼC. This analysis has three merits. First, the representation of VNV as ṼC is compatible with feature theory (Chapter 2). Second, the analysis agrees with a well-known fact that English vowels are nasalized before a nasal coda. Third, the analysis agrees with independent phonetic judgments that VNC is often realized as ṼC, such as *simple* [sĩpl̩], *sinker* [sĩkə], *symptom* [sĩptəm], and *council* [kãũsl̩] (e.g. Malécot 1960, Bailey 1978, Fujimura 1979, Cohn 1993). I return to this issue in Chapter 8.

3.4.4. **V:C rhymes**

Another kind of VXC rhyme that appears in medial positions is V:C, such as [i:s] in *aesthetic* [i:sθɛtɪk] and [ɒ:l] in *also* [ɒ:lso] in American English. The vowel in V:C is tense, which can count as two sounds phonologically. For example, the rhyme of a stressed monosyllable must be either VC (as in *bit* [bɪt] and *bet* [bɛt]) or V: (as in *bee* [bi:] and *law* [lɒ:]), but not just V (e.g. *[bɪ] or *[bɛ]). This shows that a tense vowel is equivalent to a lax vowel plus a consonant.

If we treat a V:C rhyme as having three sounds and expand the rhyme size to VXC, we face two problems. First, there is a general lack of medial VCC rhymes. Second, there is a general lack of medial VVC rhymes, where VV is a diphthong (see Chapter 8). Therefore, V:C is a special case, not a general one.

Rather than stretching the rhyme size, a better solution is to assume that, while tense vowels can be long, they need not in all environments (Pike 1947, Jones 1950, Abercrombie 1967, Giegerich 1985, Alcantara 1998). On this view, V:C can be represented as VC, where V is tense and short but still distinct from a lax vowel. For example, *aesthetic* is [isθɛtɪk], where [i] is short (compared with [i:] in *bee*) yet distinct from [ɪ]. Similarly, *also* is [ɒlso], where [ɒ] is short compared with [ɒ:] in *author*. If this analysis is correct, the maximal rhyme is still VX.

3.4.5. **Morphology and word-edge consonants**

Let us now turn to extra consonants at word edges. First, consider the C beyond a final VX, such as [p] in [hɛlp] *help* and [haip] *hype*. There are three proposals, shown in (19), where square brackets indicate syllable boundaries.

(19) Analyses of final VXC

[. . . VXC]	Large rhyme
[. . . VX]<C>	Extrasyllabic
[. . . VX][C(V)]	Onset (of a potential vowel)

In the first analysis the entire VXC is in a large rhyme (e.g. Kahn 1976, Kiparsky 1981, Giegerich 1992, Harris 1994, Blevins 1995, Hall 2001). The problem is that non-final rhymes are limited to VX (Giegerich 1985, Borowsky 1989), instead of VXC. Hall (2001) proposes that the rhyme is VX non-finally but VXC finally, but the proposal is a restatement of the fact, not an explanation. In addition, should there be VXC rhymes, one would expect them to be more prominent than VX rhymes, but there is no such evidence.

In the second analysis a word-final C is extrasyllabic (e.g. McCarthy 1979b, Hayes 1982, Borowsky 1989, Giegerich 1989, Goldsmith 1990, and Gussmann 2002). As a result, we obtain a consistent maximal rhyme VX for both final and non-final syllables. The question, though, is why the final C should be excluded from syllabification and why it is allowed to stay.

In the third analysis the final C is the onset of a following V. There are two views of what the V is. In the first, the V can be completely abstract (Burzio 1994). In the second, the V can be real—it is provided by a V-initial suffix (Giegerich 1985, Borowsky 1986). Since a word does not always have a suffix, I shall refer to the V as a "potential V." For example, the [p] in *help* is an extra C when the word occurs alone but it is the onset of the following V in *helper* and *helping*. The question for this analysis is why the final C is kept when there is no V-initial suffix, such as the [p] in *help*, *helpful*, and *helpless*. Here I appeal to a requirement known as paradigm uniformity or "anti-allomorphy" (Burzio 1996), according to which one aims to keep a morpheme in the same shape regardless of the environment. In (20) I state the conditions that support the final C, which are illustrated in (21).

(20) Conditions for the final C

a. Potential V: A word-final C can become the onset of a potential V, which comes with a V-initial suffix.

b. Anti-allomorphy: Keep a morpheme in the same shape regardless of the environment.

(21)

Final C	Supported by potential V	Supported by anti-allomorphy
[hɛl]p	[hɛl][pɪŋ], [hɛl][pɚ]	[hɛl]p, [hɛl]p[fʊl], [hɛl]p[lɪs]
help	*helping, helper*	*help, helpful, helpless*
[rɪs]k	[rɪs][kɪŋ], [rɪs][ki]	[rɪs]k, [rɪs]k[fri]
risk	*risking, risky*	*risk, risk-free*

The analysis correctly predicts that a final C may behave as if it is extrasyllabic, in the sense that it is allowed even if the preceding rhyme is full, and that a final VC may behave like a final V (a light rhyme). In addition, since we can explain the existence of the extra C, there is no need to assume that it needs to join the preceding syllable. As a result, we can maintain a consistent maximal rhyme of VX for both final and non-final syllables.

A reviewer points out that in a word like *file*, the [l] is velarized (a "dark" [l]), which is an indication that it is in the rhyme. Therefore, the syllabification should be [fail]. In the CVX analysis, [ail] is not a possible rhyme; instead, there are two other solutions. First, while it is true that [l] is "dark" in the rhyme (for all speakers) and "clear" in the onset (for some speakers), it is unclear whether [l] is dark or clear when it is unsyllabified. Second, a more likely answer is that *file* has two syllables [fai][ɫ], where [ɫ] is syllabic, in the rhyme, and hence dark.

The potential-V analysis also predicts that, if a language has CV prefixes, an extra C may occur as a "potential coda" in word-initial position. This is schematically shown in (22).

(22) Initial C supported by potential V and anti-allomorphy

Root	Supported by potential V	Supported by anti-allomorphy
CCVC	[CV-C][CVC]	C[CVC]

I shall argue in Chapter 11 that the Tibeto-Burman language Jiarong is such a case. Consider the examples in (23), from Lin (1993: 36), where [tɕʰ] is an affricate.

(23)

Root	Supported by potential V	Supported by anti-allomorphy
ntɕʰok	[kɐ-n][tɕʰok]	
'dip'	'dip'	
ʒba	[tə-ʒ][ba]	ʒ[ba-n][tɕʰok]
'face'	'face'	'face dip (dimple)'

The root for 'dip' is CCVC. When it follows a CV prefix (or a vowel-final word), the root-initial [n] can serve as the coda of the preceding V. Similarly, the root for 'face' is CCV. When there is a CV prefix, the root-initial C can serve as the coda of the preceding V. When there is no prefix, as in 'face dip (dimple)', the root-initial C is supported by anti-allomorphy.

The potential-V analysis makes specific predictions for whether and where an extra C may occur in a given language. In a language that has V-initial suffixes, an extra C may occur in root-final position. In a language that has V-final prefixes, an extra C may occur in root-initial position. This is true for the languages we examine in the following chapters.

Given a potential V and anti-allomorphy, the maximal final string should be VXC. However, English words can take additional consonants beyond VXC. Some examples are shown in (24).

(24) Final consonants beyond VXC

Word	Sounds	Beyond VXC
saved	[seivd]	[d]
risked	[rɪskt]	[t]
text	[tɛkst]	[t]
texts	[tɛksts]	[ts]
sixths	[sɪksθs]	[θs]

It has been observed that final consonants beyond VXC are limited to [t, d, s, z, θ] in English. Phonetically, such sounds are all made with the Coronal articulator. One might suggest, therefore, that Coronal sounds are special. Let us call it the coronal hypothesis and state it in (25).

(25) The coronal hypothesis
Coronal obstruents do not need to be supported by a syllable, or they can be exceptionally attached to a syllable, even if the syllable is full.

On the other hand, it can be observed that [t, d, s, z, θ] are used as consonant suffixes in English; indeed they are the only consonant suffixes in English. This fact suggests a different explanation for the final consonants: they can survive not because they are in a syllable but because they are suffixes. A similar point is made by Goldsmith (1990: 127), who proposes that consonant morphemes are "licensed" to occur even if they are not in a syllable. I shall call it the affix rule and state it in (26).

(26) The affix rule (preliminary version)
Affix sounds can be pronounced, regardless of whether they fit into a syllable.

The affix rule is intuitively natural. According to the CELEX lexicon of English (Baayen et al. 1993), among the 54,447 basic words, 41,911 end in C. This means that in most English words the final syllable is full. If consonant suffixes are not pronounced when the preceding syllable is full, then most of the time we would not be able to tell whether there is a suffix or not, and the use of suffixes would become quite irrelevant.

Whether one assumes the coronal hypothesis or the affix rule, there is no need to expand the rhyme size VX, because the sounds are already accounted for. Therefore, it is redundant to assume that they must also be absorbed into or hosted by a syllable.

The coronal hypothesis and the affix rule make different predictions. According to the former, we should find extra coronal consonants not only word-finally but also word-medially. In contrast, according to the affix rule, we should find extra coronals only at word edges. Evidence from English, German, and Jiarong supports the affix rule, to be seen in Chapters 8, 10, and 11.

In English, not all final consonants beyond VXC are suffixes; for example, the final [t] in *text* [tɛkst] is not a suffix. Similarly, the final [st] in the German word *Herbst* [ɛrpst] 'autumn' are not suffixes. The coronal hypothesis could say that such extra consonants are allowed because they are coronals. If we assume the affix rule, what would the explanation be?

One proposal is that, since both German and English use coronal consonants as suffixes, all final coronals are perceived as suffixes (Fujimura 1979, Pierrehumbert 1994). Another way to look at the matter is that, because we are used to hearing coronal suffixes, other final coronals would not sound bad to the ear. The two proposals are stated in (27) and (28).

(27) Perceived affixes
Sounds that resemble affixes can be treated as affixes.

(28) Phonetic familiarity
If morphology requires one or more sounds C_m to occur in a given phonological environment (e.g. word-final position beyond VXC), then we accept C_m in the same phonological environment even if the morphological requirement is absent.

If either proposal is correct, we can revise the affix rule as in (29).

(29) The affix rule (final version)
Affix or affix-like sounds can be pronounced, whether they can fit into a syllable or not.

I shall show in Chapters 8 and 10 that many words in English and German can be accounted for by the affix rule.

3.4.6. **Can [lp] be an onset or [pl] be a coda?**

I suggested above that a word like *plum* is not CCVC but CVC, because [p] and [l] can form a complex sound [p^l] (Chapter 2). Several questions can be raised. First, why is there no word *lpum*? The word ought to fit into CVC, if [l] and [p] can form a complex sound. The answer is that all articulatory features in a complex sound are simultaneous, so that [p]+[l] and [l]+[p] are identical. Therefore, *lpum* would be the same as *plum*.

One might also ask why *help* is not realized as [hɛp^l], where [l] and [p] form a complex sound [p^l] in the coda. The answer is that [p] is already supported by "anti-allomorphy" (see above), and so there is no need to incorporate it into a syllable. Besides, the sequence [lp] is more similar to that in *hel.per*, and hence is a better way to satisfy anti-allomorphy.

Next, one might ask why there are no such words as *tikl* or *tepl* in English, where [k^l] and [p^l] are complex sounds in the coda. The answer is that [l] can be syllabic, and *tikl* and *tepl* would be disyllabic. Indeed, there are such words in English, such as *tickle* and *nipple*. This point was made by Lamontagne (1993).

Finally, one might ask why there is no medial coda [p^l] or [k^l], which could result from a hypothetical word *neplsa* or *tiklny*. A possible answer is that, because [l] can be syllabic, *neplsa* or *tiklny* can be trisyllabic (likely to be spelled as *neppelsa* or *tickelny*). Another possible answer is that languages generally use fewer codas than onsets. For example, Cantonese uses about fifteen consonants in the onset but six in the coda: [p, t, k, m, n, ŋ]. Similarly, Standard Chinese allows about twenty consonants in the onset but two in the coda: [n, ŋ]. Therefore, it is possible that English simply allows [p^l] and [k^l] in the onset but not in the coda.

3.4.7. **Summary**

There is little doubt that in many languages the maximal syllable size is at least CVX. What I have proposed is that CVX is also the upper limit on syllable size, where C, V, and X can each be a complex sound. Extra consonants can be found at word edges, but they can be explained by morphology. In particular, a word has the schematic structure C_mCSCC_m, where C_m is one or more affix or affix-like consonants, C is a consonant supported by a potential V from an affix, and S is one or more syllables whose maximal size is CVX.

English has many V-initial suffixes and a small number of V-final prefixes. In addition, English has consonant suffixes but not consonant prefixes. Therefore, the maximal English monosyllable is shown in (30).

(30) Maximal monosyllable in English
$CSCC_m$
S: maximally CVX, where C, V, and X can be a complex sound

A similar analysis applies to German. Comprehensive quantitative data from English and German will be provided in Chapters 8 and 10 in support of the analysis.

As in previous analyses, the CVX theory predicts cross-linguistic variations in the maximal size of a monosyllabic word. However, unlike previous analysis, which attributes the variations to syllable parameters, the CVX theory derives the variations from morphology. Let us call the CVX theory the "morphological" approach and the traditional analysis that assumes syllable parameters the "phonological" approach. In (31) I compare the two, where the "phonological" approach is similar to that of Blevins (1995).

(31) Accounting for variations in the size of monosyllables (S is maximally CVX)

Max size	Morphological (CXV)	Phonological (parameters)
S	No V-final prefixes, V-initial suffixes, or C affixes	No branching onset or branching coda
CS	Having V-final prefixes	Branching onset
C_mS	Having C prefixes	Branching onset; appendix
SC	Having V-initial suffixes	Branching coda
SC_m	Having C suffixes	Branching coda; appendix
CSC	Having V-final prefixes and V-initial suffixes	Branching onset and branching coda
SCC_m	Having V-initial suffixes and C suffixes	Branching coda and appendix

etc.

The main claim of the CVX theory is that there is a correlation between the morphology of a language and the maximal size of monosyllabic words. The correlation has not been proposed before, but I shall show that it is true for the languages discussed in the following chapters. If the correlation is true in general, then the "phonological" approach becomes redundant, because the morphological properties must be recognized anyway. In (32) I summarize the differences between the two approaches.

(32)

	CVX	Traditional (syllable parameters)
Predicting variation	yes	yes
Accounting for extra Cs	yes	yes
Morphological variation	yes	(yes)
Syllable parameters	no	yes
Consistent syllable size	yes	no
Over-prediction	no	yes

Both analyses account for the cross-linguistic variation in the size of the monosyllable and for extra consonants at word edges. However, the CVX analysis assumes no syllable parameters and achieves a more consistent syllable size. In contrast, since traditional analyses must assume morphology variation anyway, their assumption of syllable parameters becomes redundant. In addition, traditional analyses assume a wider range of syllable sizes and a much greater degree of over-prediction.

3.5. SYLLABLE BOUNDARIES

If a language allows CVC, CV, and VC syllables, as English does, how would CVCVC be syllabified? Both CV.CVC and CVC.VC seem to be possible. But if a given word is syllabified in just one way, we need a way to resolve the ambiguity. I discuss several proposals.

3.5.1. **Speaker intuition**

According to Chomsky (1957), a grammar is a set of rules that define possible and impossible linguistic structures. In addition, knowing the grammar, a native speaker has the intuition to judge whether a structure is or is not good in his or her language. Similarly, Haugen

(1956a, b) proposes that speakers can break up words into syllables if they are asked to speak slowly, and Halle (1962) argues that speakers have the intuition to judge whether a sound sequence is well formed, even if it is not a real word. For example, [bɪk], [θod], [gnait], and [vnig] are (or were) not real English words but the first two are possible words while the last two are not.

However, recent studies (e.g. Frisch et al. 2000, Myers and Tsay 2005, Zhang 2007) show that speaker judgment on possible words is not always clear-cut. In addition, many studies have noted that consistent judgment on syllable boundaries can be hard to obtain (Gimson 1970, Treiman and Danis 1988, Giegerich 1992, Hammond 1999, Steriade 1999, Blevins 2003).

Giegerich (1992) proposes that speaker judgment can be probed with a better test. In the test, speakers are asked to pronounce each syllable of a word twice. One can begin with words whose syllable boundaries are unambiguous, such as *after*, which is pronounced as *af*-*af*-*ter*-*ter*. More difficult words can then be presented to see how speakers deal with them. One such word is *apple*, whose output is reported to be *ap*-*ap*-*ple*-*ple*. Giegerich takes the result to mean that *apple* is syllabified as *ap*-*ple*, where the [p] is ambisyllabic.

There are several questions about this proposal. First, if the output for *apple* is *ap*-*ap*-*ple*-*ple*, instead of *ap*-*ap*-*le*-*le*, does it mean that the syllables are *ap*-*ple*, or does it mean that *le*-*le* [l̩-l̩] is an unusual sequence of syllables which the speaker would avoid? Second, consider the word *text*. If the result is *tek*-*tek*-*st*-*st*, one might conclude that the syllable is [tɛk] and [st] is outside the syllable. However, if the result is *text*-*text*, does it mean that the syllable is [tɛkst], or does it mean that the speaker is trying to avoid repeating a non-syllable cluster [st]? Finally, Giegerich does not discuss whether speaker judgment is always clear. My own test with some native speakers shows that the judgment can vary. For example, the output for *city* can be *cit*-*cit*-*ty*-*ty* or *ci*-*ci*-*ty*-*ty*. Therefore, the test does not seem to provide conclusive answers.

But suppose there is a total lack of speaker intuition, should we conclude that there are no syllable boundaries? I suggest the answer is no. The reason is that not everything is intuitively obvious. For example, we are not aware of how we see colors, how we digest food, or how we walk. Similarly, if we rely on intuition, we would wrongly conclude that the earth is flat.

3.5.2. The Law of Initials (LOI) and the Law of Finals (LOF)

Since a word often begins with a syllable, it is natural to expect word-medial onsets to resemble word-initial onsets. Similarly, since a word often ends with a syllable, it is natural to expect word-medial rhymes to resemble word-final rhymes. As Haugen (1956a: 196) put it, the idea can be used for "making syllabic cuts on the basis of initial and final clusters." Two formulations of the idea are shown in (33) and (34). For familiarity, I use "syllable onsets" for what Vennemann calls "syllable heads."

(33) Vennemann (1988: 32–3)
  a. Law of Initials: Word-medial syllable onsets are the more preferred, the less they differ from possible word-initial syllable onsets of the language system.
  b. Law of Finals: Word-medial syllable codas are the more preferred, the less they differ from possible word-final syllable codas of the language system.

(34) Blevins (2003: 203)
  a. If C_0 is possible word-initially, then C_0 is possible syllable-initially.
  b. If C_0 is not possible word-initially, then C_0 is not possible syllable-initially.
  c. If C_0 is possible word-finally, then C_0 is possible syllable-finally.
  d. If C_0 is not possible word-finally, then C_0 is not possible syllable-finally.
  e. If V_q is possible word-initially, then V_q is possible syllable-initially.
  f. If V_q is not possible word-initially, then V_q is not possible syllable-initially.
  g. If V_q is possible word-finally, then V_q is possible syllable-finally.
  h. If V_q is not possible word-finally, then V_q is not possible syllable-finally.

According to Vennemann, word-medial onsets and rhymes should be similar to those at word edges, but not vice versa. For example, the word-final rhyme VCCC in *text* does not have to resemble a word-medial rhyme, presumably because no VCCC rhyme is found word-medially. In contrast, Blevins (2003) expects all word-initial consonant clusters to occur as word-medial onsets and all word-final consonant clusters to occur as word-medial codas. For example, because [-kst] occurs word-finally in *text*, Blevins expects it to be a possible coda word-medially, but no such medial codas are found. In this regard, Blevins makes many over-predictions.

If we exclude extra consonants at word edges, we can maintain a two-way resemblance relation between word-edge syllables and those in word-medial positions. Therefore, I define the relation in (35).

(35) The relation between word-edge syllables and those in word-medial positions
   a. Law of Initials (LOI): Word-initial onsets and word-medial onsets should resemble each other.
   b. Law of Finals (LOF): Word-final rhymes and word-medial rhymes should resemble each other.

The LOI and LOF can interact with other factors. For example, if a language does not have final stress, it may disallow word-final C, yet still have CVC.CV words, where the first syllable violates the LOF. However, if we leave aside such factors, we would expect the LOI and LOF to be true overall.

The new definition can help syllabification both word-medially and at word edges. For example, because medial rhymes are limited to VX, the rhyme in *text* should not include [st]. Similarly, because there is no word-initial onset [ft] or word-final [ɛ], we would syllabify *hefty* as *hef.ty*, not *he.fty*. Some additional examples are shown in (36).

(36) Syllabification by LOI and LOF

[sɪŋ.ɚ] *singer*	OK: *sing*, *Urbana*.
*[sɪ.ŋɚ] *singer*	Violating LOI: [ŋ-] is not a word-initial onset Violating LOF: [ɪ] is not found word-finally
[sɪɾ.i] *city*	OK: *sit, east* (note: word-final [t] can flap, as in *sit in*).
*[sɪ.ɾi] *city*	Violating LOF: [ɪ] is not found word-finally
[æt.ləs] *atlas*	OK: *at, Las Vegas*.
*[ætl.əs] *atlas*	Violating LOF: [-tl] is not a word-final coda.
*[æ.tləs] *atlas*	Violating LOI: [tl-] is not a word-initial onset.
*[wɪ.ski] *whiskey*	Violating LOF: [-ɪ] is not a word-final rhyme.
[wɪs.ki] *whiskey*	OK: *miss, Kyoto*.
*[wɪsk.i] *whiskey*	OK: *risk, east*, but rhyme exceeds VX.
[tai.ni] *tiny*	OK: *tie, need*.
*[tain.i] *tiny*	OK: *fine, east*, but rhyme exceeds VX.

The LOI and the LOF can unambiguously determine syllabification in *singer*, *city*, and *atlas*. In *whiskey* and *tiny*, there is ambiguity if the rhyme length is not limited. To resolve the ambiguity, Vennemann (1988) proposes some additional laws. Of interest are the Onset Law (Head Law), which favors a single-sound onset, the Nucleus Law, which favors a single-sound nucleus, and the Coda Law, which favors a single-sound coda. These laws in effect favor a CVX syllable and can resolve the ambiguity in *whiskey* and *tiny*.

However, the LOI and the LOF cannot resolve another kind of ambiguity, even when CVX is assumed. This is shown in (37).

(37) Ambiguity under LOI and LOF

*[kæ.nəd.ə] *Canada*	Violating LOF: [-æ] is not a word-final rhyme.
[kæn.əd.ə] *Canada*	OK: *can, admitted, a*.
[kæn.ə.də] *Canada*	OK: *can, a, Dakota*.

To resolve the ambiguity in (37), additional assumptions are needed.

3.5.3. **Maximizing stressed syllables**

Hoard (1971) proposes that "stressed syllables attract a maximum number of segments." Bailey (1978) also proposes that there should be "more consonants in a cluster to be syllabified with the heavier-accented of two surrounding nuclei as the tempo increases." Similar proposals are made by Wells (1990) and Hammond (1999). Some examples are shown in (38).

(38) Maximizing stressed syllables

Bailey (1978): *tiger* [taig.ɚ], *capital* [kæp.ət.l̩], *ability* [ə.bɪl.ət.i], *multiply* [mʌlt.ə.plai]

Wells (1990): *city* [sɪt.i], *dolphin* [dɒlf.ɪn], *cauldron* [kɔːldr.ən]

Hammond (1999): *fealty* [fɪlt.i], *bulky* [bʌlk.i], *alcove* [ælk.kov]

There are two problems with this approach. The first is that it predicts many large rhymes that exceed VX, but such clusters are generally lacking (see Chapter 8). Secondly, it yields ambiguity when syllabifying unstressed syllables. For example, *Canada* can be [kæn.əd.ə] or [kæn.ə.də]. In both cases the stressed syllable is maximized, but the theory does not say whether an unstressed syllable can or cannot take a coda.

3.5.4. **Maximal Onset**

The Maximal Onset rule was proposed by Pulgram (1970) and Kahn (1976), according to which consonants between vowels should be syllabified as the onset of the following vowel as far as possible. Under the Maximal Onset rule, there is no ambiguity in syllabification. Some examples are shown in (39).

(39) Syllabification by Maximal Onset

[sɪŋ.ɚ] *singer*	*[sɪ.ŋɚ], because [ŋ-] is not a possible onset
[æt.ləs] *atlas*	*[æ.tləs], because [tl-] is not a possible onset
[tai.ni] *tiny*	*[tain.i], because [n-] is a possible onset
[lɛ.mən] *lemon*	*[lɛm.ən], because [m-] is a possible onset
[kæ.nə.də] *Canada*	*[kæn.əd.ə], because [n-] and [d-] are possible onsets

Steriade (1982: 77) and Clements and Keyser (1983: 37) further assume that VCV is syllabified as V.CV universally, a view that is shared by many others.

One problem with the Maximal Onset analysis is that it does not respect the Law of Finals. For example, in [lɛ.mən] *lemon* the first syllable ends in [ɛ], which is not found word-finally. To avoid this problem, Pulgram (1970) and Kahn (1976) added a condition that in VCV, if the first V is lax and stressed, the syllabification is V<u>C</u>V, where <u>C</u> belonging to both the first and the second syllables. Similarly, Selkirk (1982) proposes a rule to change V.CV to VC.V if the first V is lax and stressed. However, many other studies disregard the Law of Finals and use the Maximal Onset rule strictly (e.g. Halle and Vergnaud 1987, Baayen et al. 1993). A second problem with the Maximal Onset analysis is that there is no clear evidence for it.

For example, it is well known that vowel-initial words often start phonetically with a glottal stop. Two examples in English are shown in (40).

(40) *out* [ʔaut]
Ann [ʔæn]

The fact is often interpreted as evidence that every syllable needs an onset, and when there is no consonant onset, a glottal stop is added.

However, there are two ways to interpret the initial glottal stop. One is that it is an intended sound. The other is that it is an unintended gesture: the vocal tract cannot assume the vowel gesture all of a sudden, and the glottal stop reflects an unintended state before the vowel is pronounced. According to the first interpretation, the glottal stop should be present whether there is a preceding word or not. According to the second, the glottal stop should be absent when there is a preceding word. Evidence supports the second view, as the example in (41) shows.

(41) No glottal stop
ran out [rænaut]/*[rænʔaut]

When a vowel-initial word follows another word, the glottal stop cannot be added (unless one is speaking very slowly). One might suggest that perhaps the final consonant of the first word has been shifted to the onset of the second word. However, English generally does not resyllabify across words. For example, there is a distinction between each pair of expressions in (42), cited in Ladefoged (1972) and Wells (1990).

(42) (*I'm going to get my*) *lamb prepared* vs. (*I'm going to get my*) *lamp repaired*
nitrate vs. *night-rate*
plum pie vs. *plump eye*

Similarly, word pairs like those in (43) remain distinct (the first pair is from Kiparsky 1979). In the examples, vowel length is not shown and only relevant syllable boundaries are shown.

(43)	*at ease*	[æɾ.iz] / *[æ.tʰiz] / *[æt.ʔiz]
	a tease	[ə.tʰiz]
	beef eater	[bif.iɾɚ]/*[bif.ʔiɾɚ]
	bee feeder	[bi.fiɾɚ]
	beat owls	[biɾ.aulz] / *[bi.tʰaulz]
	be towels	[bi.tʰaulz]

In American English [tʰ] and [ɾ] are allophones of [t], where [tʰ] is used in the onset of a stressed syllable. In *at ease*, [tʰ] cannot be used, which means that [t] remains in the first syllable; [ʔ] cannot be added either, which means that the onset is not required. The difference between *beef eater* and *bee feeder* makes the same point: there is no cross-word syllabification of [f] in the former. If there were, the two compounds should sound the same (the phonetic difference seems to lie in the length of the first [i], which is longer in *bee* than in *beef*). In addition, there is no [ʔ] in *beef eater* either, which shows again that the onset is not required. The same analysis applies to the difference between *beat owls* and *be towels*.

In summary, there is no clear evidence for the Maximal Onset rule. Kahn (1976) argues that the Maximal Onset analysis is supported by other allophonic rules, such as aspiration and flapping. We will see below that the evidence is not compelling for this argument either.

3.5.5. **The Weight-Stress Principle (WSP)**

The WSP has been proposed in various forms by many people, such as Prokosch (1939), Fudge (1969), Hoard (1971), Bailey (1978), Selkirk (1982), Murray and Vennemann (1983), Kager (1989), Prince (1990), Wells (1990), Hammond (1999), and Duanmu (2000). There is also some phonetic evidence for the WSP (Krakow 1989, Turk 1994, Redford and Randall 2005). Counterexamples to the WSP have been reported (e.g. Davis 1988: 85–8), but the evidence does not seem compelling and is not discussed here. The WSP is stated in (44) and illustrated in (45), where V is a short stressed vowel, X can be C or V, and v is an unstressed vowel.

(44) The Weight-Stress Principle (WSP)
  a. Stressed syllables should be heavy (rhyme being VX).
  b. Unstressed syllables should be light (rhyme being v or a syllabic C).

(45)

Good	Bad	Reason	Example (relevant part underlined)
VC.v	*V.Cv	WSP-a	*city*
VC.VX	*V.CVX	WSP-a	*rabbi*
v.CVX	*vC.VX	WSP-b	*attack*
v.Cv	*vC.v	WSP-b	*Canada*
VV.Cv	*VVC.v	Max VX	*cola*
VC.Cv	*VCC.v	Max VX	*whisper*

The WSP offers unambiguous syllabification for the first four cases. In the last two cases there is no ambiguity either, because the VX limit on rhyme size can exclude VVC.v and VCC.v. It can be seen, too, that the WSP analysis satisfies both the Law of Initials and the Law of Finals. Finally, we shall see below that the WSP analysis can account for allophonic variations, too.

The WSP raises a question: How much information is there in the lexicon? Some studies assume that the English lexicon contains no syllables or stress; instead, we use rules to build syllables and assign stress (e.g. Halle 1962, Halle and Vergnaud 1987, Zec 1988). Other studies assume that the lexicon can contain a lot of memorized information (e.g. Vennemann 1974, Selkirk 1980, Clements and Keyser 1983, Giegerich 1992, Bromberger and Halle 1989, Burzio 1996, Hayes 1995, Halle 1997, Pater 2000, Ladefoged 2001, Pierrehumbert 2001, Vaux 2003, Port 2006, Colemen and Pierrehumbert 2007). I follow the latter and assume that syllable structure or stress can be part of memorized information in the lexicon.

3.5.6. Aspiration, flapping, and sounds at word-medial syllable boundaries

Kahn (1976) argues that syllabification can help account for allophonic rules, such as aspiration and flapping in American English.

His analysis is summarized in (46) and exemplified in (47), where C̲ is ambisyllabic, which means that it belongs to both the first and the second syllable.

(46) Kahn's rules for syllabification, aspiration, and flapping
VCV → V.CV   Maximal Onset rule
V.Cv → VC̲v   Ambisyllabic rule (v is unstressed vowel)
Aspiration: [p, t, k] are aspirated when they start a syllable.
Flapping: Intervocalic [t] or [d] is flapped when ambisyllabic.

(47) Analysis of *potato*
[pə.té:.to]   Maximal Onset rule
[pə.té:t̲o]   Ambisyllabic rule
[p^hə.t^hé:ɾ̲o]   Aspiration and flapping

In *potato*, the final vowel [o] is unstressed, and so the ambisyllabic rule applies to the second [t], giving [pə.tét̲o]. The analysis correctly predicts aspiration and flapping. However, Kahn's analysis faces a problem with *af.ter*, where [t] starts the second syllable (because [ft] is not an onset in English), yet [t] is not aspirated. Kahn proposes that *after* can be syllabified as *a.fter*, where [ft] is an onset, but there is little evidence for the claim.

Borowsky (1986) abandons the ambisyllabic rule and adopt a different resyllabification rule and a different interpretation for flapping. Her analysis is shown in (48) and (49).

(48) Borowsky's rules for syllabification, aspiration, and flapping
VCV → V.CV   Maximal Onset rule
V.Cv → VC.v   Resyllabification (v is unstressed vowel)
Flapping: Intervocalic [t] or [d] is flapped when it is in the coda.

(49) Analysis of *potato*
[pə.té:.to]   Maximal Onset rule
[pə.té:t.o]   Resyllabification
[p^hə.t^hé:ɾ.o]   Flapping

Because [o] is unstressed, resyllabification moves the second [t] to the coda of the preceding vowel, giving [pə.té:t.o]. Flapping then applies to the second [t] but not the first, which is in the onset. It is worth noting that Borowsky's resyllabification rule can violate the VX rhyme size, which she otherwise assumes. For example, the second rhyme in *potato* is [e:ɾ], which is VVC, because a tense vowel counts as VV. Similarly, the [t] in *mighty* is flapped, and so the syllabification must be [maiɾ.i], where the first rhyme again exceeds VX. The VVC rhyme has also been proposed by Hoard (1971), Selkirk (1982), Wells (1990), and Hammond (1999).

Jensen (2000) proposes yet another analysis of aspiration and flapping (see also Iverson and Salmons 1995, Davis 1999, Davis and Van Dam 2001), in which there is no need for ambisyllabic sounds or resyllabification. Instead, the Maximal Onset rule is sufficient. His analysis is shown in (50) and (51).

(50) Jensen's rules for syllabification, aspiration, and flapping
VCV → V.CV   Maximal Onset rule
Aspiration: [p, t, k] are aspirated word-initially or foot-initially.
Flapping: Intervocalic [t] or [d] is flapped when it is not word-initial or foot-initial.

(51)	*potato*	*at ease*	
	[pə.té:.to]	[æt.íz]	Syllabification by Maximal Onset
	[pʰə.tʰé:.ɾo]	[æɾ.íz]	Aspiration and flapping

In *potato*, [p] is aspirated because it is word-initial. The first [t] is aspirated because it is foot-initial, assuming that each stressed syllable starts a foot. The second [t] is flapped because it is neither word-initial nor foot-initial. Similarly, in *at ease*, the [t] is in the coda of the first vowel, assuming no resyllabification across words. The [t] is flapped because it is neither word-initial nor foot-initial. Thus, [t] can be flapped whether it is in the onset, as in *potato* [pʰə.tʰé.ɾo], or in the coda, as in *at ease* [æɾ.íz], and whether it precedes a stressed vowel, as in *at ease* [æɾ.ís], or an unstressed vowel, as in *hit it* [híɾ.ɪt]. Jensen has no problem with *after* [æf.tɚ] either, because the [t] is not foot-initial, so it is unaspirated.

As Kahn (1976), Borowsky (1986), and Jensen (2000) observe, aspiration and flapping are sensitive to stress. If we recognize the role of stress, we can derive syllabification without the Maximal Onset rule. The solution, shown in (52) and (53), is to use the Weight-Stress Principle (WSP) for syllabification, and the same rules for aspiration and flapping as Jensen (2000). A comparison between the WSP analysis and others is shown in (54).

(52) Syllabification by WSP, aspiration, and flapping
WSP: VCv → VC.v,   vCV → v.CV
Aspiration: [p, t, k] are aspirated word-initially or foot-initially.
Flapping: Intervocalic [t] or [d] is flapped when it is not word-initial or foot-initial.

(53)	*potato*	*at ease*	
	[pə.té:.to]	[æt.íz]	Syllabification by WSP
	[p^{h}ə.t^{h}é:.ɾo]	[æɾ.íz]	Aspiration and flapping

(54) Comparison of four analyses

at ease	*city*	*potato*	
[æɾ.íz]	[síɾ̱i]	[p^{h}ə.t^{h}é:ɾ̱o]	Kahn (1976)
[æɾ.íz]	[síɾ.i]	[p^{h}ə.t^{h}é:ɾ.o]	Borowsky (1986)
[æɾ.íz]	[sí.ɾi]	[p^{h}ə.t^{h}é:.ɾo]	Jensen (2000)
[æɾ.íz]	[síɾ.i]	[p^{h}ə.t^{h}é:.ɾo]	WSP

All the analyses agree that in *at ease* there is no resyllabification across words, but they differ in the analysis of *city* and *potato*. In *city*, [ɾ] is ambisyllabic for Kahn, in the coda for Borowsky and the WSP analysis, and in the onset for Jensen. In *potato*, [ɾ] is ambisyllabic for Kahn, in the coda for Borowsky, and in the onset for Jensen and the WSP analysis. In particular, in the WSP analysis, the first [t] in *potato* is not a coda, because the first syllable is unstressed and must be light. The second [t] is not a coda either, because [e:] is long and has filled up the VX rhyme. While the WSP analysis offers different syllabification, it accounts for aspiration and flapping just as well.

Hoard (1971) observes some other examples: *motto*, where the [t] is flapped, and *veto*, *Hittite*, and *satire*, where the medial [t] is not flapped. The WSP analysis is shown in (55).

(55)	*motto*	[mɑ:][ɾo] or [mɑɾ][o]
	veto	[vi:][t^{h}o:]
	Hittite	[hɪt][t^{h}ai](t)
	satire	[sæ:][t^{h}ai][ɚ] or [sæt][t^{h}ai][ɚ]

In *motto*, the second syllable is unstressed and so [t] is not foot-initial. In contrast, the second syllable in *veto*, *Hittite*, and *satire* is stressed and so its onset [t] is foot-initial and aspirated. The analysis echoes Hoard's proposal that sometimes a geminate consonant can occur in English, such as [tt] in *Hittite*.

Blevins (2003) argues that some words cannot satisfy the Law of Initials or the Law of Finals, no matter how we syllabify them. One example is *lemon*, shown in (56).

(56)	Syllabification	Problem
	[lɛ.mən]	Violating LOF: No word-final [ɛ]
	[lɛm.ən]	Violating LOI: No word-initial [ə]?

In [lɛ.mən] there is a syllable-final [ɛ], which is not found word-finally. In [lɛm.ən] there is a syllable-initial [ə], which Blevins claims is not found word-initially, because she assumes that word-initial vowels are preceded by a glottal stop. However, Blevins's claim is true only for words pronounced in isolation, not for words pronounced with another word. For example, while *out* can start with a glottal stop in isolation, there is no glottal stop in *ran out* [rænaut]. Thus, if we consider words in context, then *lem.on* satisfies both the LOI and the LOF and is the only correct syllabification. Similarly, consider the analysis of *Betty*, shown in (57).

	Syllabification	Word in isolation	Word in context
(57)	[bɛɾ.i]	Final [ɾ]; initial [i]	OK: *at ease*
	[bɛ.ɾi]	Final [ɛ]; initial [ɾ]	Violating LOF: no word-final [ɛ]

If we look at words in isolation, both syllabifications are problematic, because [ɾ] and [ɛ] are not found word-finally, and [i] and [ɾ] are not found word-initially. However, if we look at words in context, then [bɛɾ.i] is not problematic but [bɛ.ɾi] is. In [bɛɾ.i], the syllable-final [ɾ] and the syllable-initial [i] are found at word edges, as in *at ease*. In contrast, in [bɛ.ɾi], the syllable-final [ɛ] is not found word-finally.

There are, however, some cases where word-edge patterns do not match word-medial patterns. An example is *nostalgic* in British English, cited in Wells (1990) and shown in (58).

	Syllabification	Problem
(58)	a. [nɒ.stæl.ʤɪk]	Violating LOF: no word-final lax [ɒ]
	b. [nɒs.tæl.ʤɪk]	Violating LOI: no unaspirated word-initial [t]

In (58a), there is a syllable-final lax vowel, which is not found word-finally. In (58b), there is a syllable-initial unaspirated [t] before a stressed vowel, which is not found word-initially. Thus, neither analysis fits word-edge patterns. The word *gestation* poses the same dilemma ([ʤɛs.tei.ʃn̩] vs. [ʤɛ.stei.ʃn̩]). Such examples show that sounds at word-medial syllable boundaries may be subject to different conditions from those at word-edges. It also shows that, if *nostalgic* is syllabified as [nɒs.tæl.ʤɪk], the lack of aspiration in [t] is not triggered by an initial [s] in the same syllable. Therefore, in a word like *stay*, [s] need not be in the onset.

3.5.7. **Summary**

In (59) I compare four approaches to syllabification with regard to whether the rhyme exceeds VX (VX+), whether there is ambiguity in syllabification, and whether word-medial syllable boundaries agree with word-edge patterns.

(59)

	Ambiguity	VX+	Edge patterns
LOI and LOF	yes	?	yes
Max stress	yes	yes	yes
Max onset	no	no	no
WSP	no	no	yes

The Law of Initials and the Law of Finals are based on word-edge patterns, but there is ambiguity in syllabification (e.g. *whis.key* and *whisk.ey*) and rhymes can in principle exceed VX (e.g. *whisk.ey*), although additional rules can be proposed to limit them to VX (Vennemann 1988). The approach that maximizes stressed syllables also has ambiguity in syllabification (e.g. *Can.ad.a* and *Can.a.da*) and allows rhymes to exceed VX (e.g. *whisk.ey*). The Maximal Onset analysis is unambiguous and would usually not produce large rhymes, but it allows lax vowels to occur at the end of a syllable (e.g. *le.mon*), which does not match word-edge patterns. The approach using the Weight-Stress Principle is the most desirable in all respects.

3.6. THE *A/AN* ALTERNATION AND THE LINKING [r]

The English article is *a* before a C-initial word and *an* before a V-initial word. One might argue that the alternation supports the need for syllables to have an onset, as shown in (60).

(60) The [n]-insertion analysis of *a/an* alternation
[n] added as onset *an apple*, *an attack*
[n] not needed *a pear*, *a potato*

However, there are three problems with the analysis. First, if [n] is added to provide an onset for the following V, why does the article itself not need an onset? Second, why is [n] not added in other contexts, such as *be_on time*, *go_again*, *buy_apples*, *panda_act*, and *three_Iranians*? One might suggest that there is a glide [j] after *be*, *buy*, and *three*, a glide [w] after *go*, and an epenthetic [r] after *panda*, and these can serve as the onset of the following vowel. However, why is [r] not used after *a*, such as *a [r] apple*? And why is [j] not used before [i], such as *a [j] Iranian*? A third problem is that the original form of the article is not *a* but *an*, which comes from ‘one’ etymologically. Therefore, what we need to explain is not the insertion of [n] before V but the deletion of [n] before C, which has nothing to do with the need for an onset.

Given the etymological fact that the original form of the article is *an*, we can explain its alternation with the Weight-Stress Principle. Since *an* is unstressed, it should be light and cannot keep its coda [n]. The [n] can be moved to the next word, or else it can be deleted. The analysis is shown in (61).

(61) The [n]-deletion analysis of *a/an* alternation

	[n] moved	[n] unmoved	[n] deleted
an apple	[ə][næp][l̩]	*[ən][æp][l̩]	*[ə][æp][l̩]
		Unstressed VX	No need to
a bee	*[ə][nbi:]	*[ən][bi:]	[ə][bi:]
	Illegal [nb]	Unstressed VX	Best option

Both the insertion analysis and the deletion analysis assume that syllabification can occasionally cross a word boundary (especially when one word is completely unstressed). In addition, they both predict that *an aim* and *a name* should sound the same. The prediction is correct in casual speech but not in careful speech (Halle 1972, Kahn 1976, Kiparsky 1979). Regardless of how careful speech is analyzed, there is no obvious advantage in assuming the [n]-insertions analysis or the onset requirement.

A similar case of relevance is the linking [r] in English, such as *better [r] off* and *for [r] it*, and the intrusive [r], such as *law [r] and order* and *idea [r] of*. It may seem that [r] is added because syllables need an onset. On the other hand, the linking [r] is underlyingly present and so its variation is a matter of deletion rather than insertion. In particular, for speakers who do not use a word-final [r], it can be moved to the next word, unless the next word cannot accept it. This is shown in (62).

(62) The [r]-deletion analysis

	[r] moved	[r] unmoved	[r] deleted
for us	[fo][rʌs]	*[for][ʌs]	*[fo][ʌs]
		Final [r]	No need to
for her	*[fo][rhɝ]	*[for][hɝ]	[fo][hɝ]
	Illegal [rh]	Final [r]	Best option

The intrusive [r] can be analyzed as the result of analogy; for example, from [lɔ:] *lore* as to [lɔ:(r) ənd] *lore [r] and*, one gets [lɔ:] *law* as to [lɔ:(r) ənd] *law [r] and*. Therefore, what we can conclude is that syllables can have an onset, but they do not require one.

3.7. EPENTHETIC VOWELS BETWEEN CONSONANTS

In some languages, epenthetic vowels can appear in consonant clusters. For example, while Dell and Elmedlaoui (1985) argue that Berber words can be made entirely of consonants, Coleman (2001) argues that they contain epenthetic vowels that serve as the syllable nucleus.

Similarly, the Tibeto-Burman language Jiarong reportedly has unusual onset clusters, yet Lin (1993) points out that such clusters may contain epenthetic vowels. For example, [kə-n-mʧʰə] 'early' can be pronounced as [kə-nə-mʧʰə]. In the former case, the syllabification seems to be [kən][mʧʰə], which contains an unusual onset cluster [mʧʰ]. In the latter case, the syllabification could be [kə][nəm][ʧʰə], and there is no unusual onset cluster.

Polish is also known for having unusual initial clusters, such as [lgɲ] in *lgnąć* 'cling to' and [mkn] in *mknąć* 'scurry'. However, in the pronunciation of Sylvia Suttor, a native speaker of Polish, such clusters seem to contain epenthetic vowels, namely, [ləgəɲ] and [məkʰɲ] respectively. As a result, *lgnąć* sounds like three syllables to English speakers and *mknąć* sounds like two. Clearly, the presence of epenthetic vowels adds different possibilities for the analysis of syllable structure.

In the present study, I shall discuss one language that has epenthetic vowels (Jiarong in Chapter 11), and show how they affect the analysis of syllabification.

3.8. ARE THERE PARAMETERS FOR THE MAXIMAL SYLLABLE SIZE?

Most phonologists believe that the maximal syllable size can vary from language to language. In generative phonology, the standard approach to variation is to propose a set of parameters (Chomsky 1981). If we can determine the parameters and their possible values, we can predict how many possible human languages there are and what properties they have.

Blevins (1995) proposes six binary parameters for the maximal syllable size, rephrased in (63), along with the values (or "settings") for

English. Similar parameters have been proposed before by Clements and Keyser (1983: 28–30).

(63) Binary parameters for the maximal syllable size (Blevins 1995: 219)

Parameters	English settings
Can the onset contain two sounds?	Yes
Can the nucleus contain two sounds?	Yes
Is the coda allowed?	Yes
Can the coda contain two sounds?	Yes
Can extra C occur initially?	No (Yes)
Can extra C occur finally?	Yes

The two parameters for the coda are dependent and yield three choices: no coda sound, a single coda sound, and two coda sounds. The other four parameters are independent. Therefore, there are forty-eight possible kinds of maximal syllable structures.

In Blevins's proposal, to choose the simplest syllable, a language would make a "no" choice for all the parameters, and to choose the most complicated syllable a language would make a "yes" choice for all parameters. In this regard, English has one of the most complicated syllable structures. In particular, if we consider the initial [s], as in *split*, to be an extra C in English, then English has made "yes" choices for all parameters. Such a language would allow CCVVCC syllables in non-edge positions and CCCVVCCC syllables in monosyllables. In fact, Clements and Keyser (1983: 32) assume that C and V can each repeat at least three times, so that a maximal possible syllable is at least CCCVVVCCC.

The CVX theory predicts that such super-sized syllables are not only impossible but unnecessary. In particular, I shall show in Chapter 8 that word-medial syllables in English are limited to CVX, and apparent CC clusters are in fact complex sounds. In addition, extra Cs at word edges are not the result of parameter choices for the maximal syllable size, but the result of morphology: V-final prefixes support an initial C, V-initial suffixes support a final C, and consonant affixes constitute additional Cs. If other languages can be analyzed in the same way, then there is no need to assume parameters for the maximal syllable size. Rather, all languages have the same maximal syllable size, which is CVX.

One might note that if all the parameters are set to "no," the maximal syllable size is CV. Blevins (1995) cites the American Indian language Cayuvava as such a case, although it has some CC onsets, such as [tr, pr] (Key 1961), which in the present analysis can be seen as complex sounds. Let us suppose then that there are languages whose syllables are made of CV; in fact, I shall argue in Chapter 6 that Shanghai Chinese is such a case. Would we then need a parameter for the maximal syllable size, namely, CVX vs. CV? I suggest the answer is no. The reason is that when a CV syllable is stressed, it is likely to be lengthened, so that it is realized as [CV:] or CVV. This is true in Shanghai (Zhu 1995), and probably true in Cayuvava as well. Therefore, unless there is phonetic evidence to the contrary, there is no need to assume that the maximal syllable can be smaller than CVX. I shall discuss CV syllables again in Chapter 12.

3.9. THE "SPOTTY-DATA" PROBLEM

I use the term "spotty-data" to refer to the fact that there are often not enough data for making reliable generalizations, even if we examine the entire lexicon of a language. To see the problem, let us consider the ratios between possible words and actual words in American English. The data on CVC and CVCVC words are shown in (64).

(64) The "spotty-data" problem in English

Word form	Possible	Used	% used
CVC	2,415	615	25.5
CVCCVC	5,832,225	6,000	0.1

Excluding affixes and homophones, English has about 3,000 uninflected monosyllabic words (see Chapter 9), which include CVVC (842), CVC (615), CCVVC (453), and CCVC (326). The second most frequent type, CVC, includes 615 syllables. Now, given twenty-four consonants (excluding [ŋ] from the onset and [h, tr, dr] from the coda) and five short (lax) vowels, there are $23 \times 5 \times 21 = 2{,}415$ possible CVC syllables. This means that just 25% of all possible CVC words are used. In dialects that have more short vowels, the percentage of occurring syllables could be even lower. If we consider disyllabic words, the percentage of occurring syllables becomes diminishingly small. For example, if any two CVC syllables can form a disyllabic word, there are about 6,000,000 possible disyllabic words, yet English only uses around 6,000 uninflected disyllabic words. This means that just 0.1% of all possible disyllabic words are used.

Why are so many possible words not used in English? One might suspect that there are phonological constraints that rule out most of the disyllabic words, but there are no known phonological constraints that would rule out 99% of the disyllabic combinations. The real answer, in my view, is that a language simply does not need many morphemes, which make up words. In particular, I shall show in Chapter 9 that English and Chinese use about 10,000 morphemes each, and many of them are infrequent. If a language only needs 10,000 morphemes, they often constitute just a very small fraction of possible words.

So if a language only needs 1% (or a few percent) of all possible words, which ones would be chosen? There are two possibilities: either the words are chosen more or less arbitrarily, or they are chosen according to phonological principles. It is often possible to look at the lexicon of a language and make various phonological generalizations. But without knowing whether a lexicon is an arbitrary or systematic collection of words, we cannot be sure whether the generalizations are real or merely artifacts.

In Chinese, whose morphemes are mostly monosyllables, the ratio between actual words and possible words is higher (see Chapter 5). Still, actual words only constitute a small fraction of possible words. Therefore, whatever patterns one might see in Chinese, questions can be raised. For example, Yip (1988: 82) suggests that Cantonese Chinese has a restriction against syllables that have two labial consonants, one in the onset and one in the coda, such as [pim] and [map], but she notes a few exceptions, such as [pəm] 'pump'. Now, should we say that Cantonese disallows syllables with two labial consonants, or should we say that Cantonese happens not to use any (or many) such words, but is in principle open to their use? The judgment of native speakers on non-words does not help much either, because it might simply reflect whether a word is or is not in the language, or how similar a non-word is to an existing word, or how many existing words a non-word is similar to, rather than what phonological principles are.

The spotty-data problem shows that it is often difficult to obtain true phonological generalizations, and that sometimes we may not be sure what kind of rules a language has.

3.10. SUMMARY

When a language seems to have large syllables, it is often because there are extra consonants at word edges. If we assume that all consonants must be syllabified, we have to assume very large syllables, such as CCCCVCCC (Hooper 1976a: 229) or CCCVVVCCC (Clements and Keyser 1983: 32), and then face the problem of over-prediction, because we cannot explain why non-edge syllables are much smaller. On the other hand, if we exclude extra consonants at word edges, we can maintain a smaller and consistent syllable size for both edge and non-edge positions; but we have to explain why extra consonants can occur at word edges, a question that has not been satisfactory answered before.

The CVX theory makes two claims. First, the maximal size of non-edge syllables is CVX (either CVC or CVV). Second, extra consonants at word edges are predictable from morphology. In languages with V-initial suffixes, an extra C is allowed in word-final position, because it can serve as the onset of the suffix V. Similarly, in languages with V-final prefixes, an extra C is allowed in word-initial position, because it can serve as the coda of the prefix V. Moreover, if a language has consonant affixes, they can be added regardless of whether they fit into a syllable. The claims make empirical predictions, which will be tested in the following chapters.

The concept of complex sounds (Chapter 2) explains how many underlying sounds can fit into each of the three CVX slots. In the extreme case, six underlying sounds can fit into a CVX syllable. An example is shown in (65).

(65)

	Word	Sounds	CVX
	prints	[prɪnts]	[p^{r} ɪ̃ t^{s}]

The word *prints* has six underlying sounds, which can merge into three complex sounds, where [p^{r}] is formed from [p] and [r], the nasalized vowel [ɪ̃] is formed from [ɪ] and [n], and the affricate [t^{s}] is formed from [t] and [s].

A critic may ask why, if more than three sounds can fit into a syllable, we do not say that a syllable can contain more sounds, such as six. The answer is that not all sound combinations can fit into a syllable. The CVX theory predicts that only those combinations that can reduce to three or fewer complex sounds can fit into a syllable, and we shall see evidence in the following chapters. In contrast, other theories need to explain why most combinations cannot fit into a syllable except those that can form three or fewer complex sounds.

I have also discussed several approaches to syllabification. I have argued that there is little evidence that a syllable requires an onset, either word-initially or word-medially. Instead, two independent principles—the Weight-Stress Principle and the VX limit on rhyme size—suffice to yield unambiguous syllabification and a satisfactory account of allophonic variations and word-edge effects.

To test the CVX theory, we should ideally look at every language and examine whether every non-edge syllable is within CVX and whether every extra C at word edges is predictable from morphology. As a beginning step, we can start with a selection of languages for which there is a sufficient amount of data on both the phonology and the morphology. In the following chapters I shall analyze five languages in detail: Standard Chinese, Shanghai Chinese, English, German, and Jiarong. The first two are chosen because they have a small syllable inventory, which we can examine exhaustively in order to determine the constraints that govern syllable-internal sound combinations. In addition, a minimal difference in the syllable coda between Standard Chinese and Shanghai Chinese has led to a dramatic difference in tonal behavior, which I shall discuss. English and German are chosen for their large consonant clusters, which in previous analyses were thought to require extra-large syllables. Jiarong is chosen for its large initial consonant clusters, which are hard to account for in previous approaches. I shall argue that the analyses of these languages provide concrete evidence for the CVX theory.

4

Syllable structure in Chinese

4.1. SYLLABLE BOUNDARIES

Syllable boundaries in Chinese are mostly unambiguous, regardless of the dialect. The majority of Chinese words are monosyllabic. The maximal size of such a word is either CGVV or CGVC, where C is a consonant, G a glide, and VV either a long vowel or a diphthong. I shall refer to such syllables as CGVX, where CG is realized as a complex sound C^G. In polysyllabic words, which are often foreign names, syllable boundaries are also unambiguous. In particular, the durational patterns and allophonic changes provide cues for word-medial syllable boundaries. Some examples in Standard Chinese are shown in (1).

(1) Word-medial syllable boundaries in Standard Chinese

Sounds	Boundary	Gloss
[maajii]	[maa.jii]	ant
[maanau]	[maa.nau]	amber
[mæ̃nkuu]	[mæ̃n.kuu]	Bangkok

In 'ant' the length of [aa] indicates that the medial syllable boundary must be [maa.jii]. In 'amber' the length of [aa], its lack of nasalization, and its lack of fronting indicate that the medial syllable boundary must be [maa.nau]. In 'Bangkok' the first vowel is an allophone of [a]; its fronting to [æ] and its nasalization indicate that the medial syllable boundary must be [mæ̃n.kuu]. In addition, the lack of word initial [nk] supports the syllable boundary between them.

Chinese has no consonant prefixes. Consonant suffixes do occur in some dialects, but the syllable structure is unaffected. For example, in Standard Chinese there is a consonant suffix [ɚ] (diminutive), which usually replaces the coda of its host syllable, as in [njau] + [ɚ] → [njaɚ] '(little) bird'. The resulting syllable is still within CGVX.

Most Chinese words are stressed monosyllables; they are also called full or heavy syllables. Unstressed syllables are usually grammatical words; they are also called weak or light syllables. Stressed syllables carry lexical tones and are longer than unstressed syllables, which do not carry lexical tones. In addition, unstressed syllables undergo rhyme reduction (to be discussed later). I shall argue that stressed and unstressed syllables have different structures, in that stressed syllables are heavy and unstressed syllables are light. Syllable structure can also be affected by casualness of speech and lengthening at prepause boundaries. In addition, CGVVC syllables have been reported in some dialects. Such issues will be discussed, too. I begin with two controversial issues: (a) whether an onset is always needed, and (b) whether G is an independent sound.

4.2. THE ONSET: OBLIGATORY OR OPTIONAL?

When a syllable begins with a vowel, an onset of some sort is added, most commonly [ʔ]. Two example are shown in (2), where [Ø] is what Chao (1968) calls a "true vowel" onset.

(2) The zero onset effect in Standard Chinese
[ʔɤɤ] / [ɣɤɤ] / [ŋɤɤ] / [Øɤɤ] 'goose'
[ʔæ̃n] / [ɣæ̃n] / [ŋæ̃n] / [Øæ̃n] 'peace'

In initial position, [ʔ, ɣ, ŋ] only occur in syllables like 'goose', optionally, and they do not contrast with each other. (There may be speakers who use [ŋ] for such syllables throughout; for them [ŋ] is a real sound and there are no vowel-initial syllables.) The presence of such sounds has led some linguists to assume that every syllable has an onset, and that those syllables which do not have a regular initial C or G have a "zero onset" (Chao 1948, 1968, F. Li 1966). I also proposed in Duanmu (1990, 2000) that the zero onset effect is the result of an obligatory onset slot in the syllable structure, which needs to be filled with something.

In Chapter 3 I have argued that there is no compelling evidence that syllables need an onset. In particular, the initial glottal stop in vowel-initial words is an unintended gesture: the vocal tract cannot assume the vowel gesture all of a sudden, and the glottal stop reflects an unintended state before the vowel is pronounced. In addition, when there is a preceding word, a vowel-initial word is pronounced without a glottal stop (e.g. [rænaut] / *[rænʔaut] *ran out*).

In Chinese, when an unstressed vowel-initial syllable follows a consonant-final syllable, there is no glottal stop between the consonant and the vowel either (to be discussed shortly). However, Chao (1968: 20) observes that when a stressed vowel-initial Chinese syllable follows a consonant-final syllable, the consonant does not directly link with the vowel; instead, the zero onset seems to intervene in between. Consider the example in (3), from Standard Chinese.

(3) [mjæ̃n] + [au] → *[mjæ̃nau] (Cf. English: *ran out* [rænaut])
'cotton jacket'

Here [au] cannot directly link with [n] (although it can in English). Duanmu (1990, 2000) suggests the analysis in (4):

(4) [mjæ̃n] + [Øau] → a. [mjæ̃nʔau]
'cotton jacket' b. [mjæ̃nɣau]
c. [mjæ̃ŋɣau]
d. [mjæ̃ŋŋau]
e. *[mjæ̃nau]

In the analysis, various realizations of the zero onset intervene between [n] and [au], so that [n] and [au] do not link with each other. In the case of (4c) and (4d), the zero onset further causes [n] to change to [ŋ].

However, as Xu (1986) and Wang (1993) point out, in connected speech none of the output forms in (4) is natural. Instead, when the words are spoken together the most natural pronunciation is (5), in which there is no nasal closure or any version of the zero onset.

(5) [mjæ̃n] + [au] → [mjæ̃:au]
'cotton jacket'

Xu further points out that there is a three-way contrast in a VNV sequence—VN.V, V.NV, and VN.NV, exemplified in (6), where 'overturn trouble' is a made-up phrase (semantically odd but phonologically natural) for the sake of parallel comparison. (Xu's original example for VN.NV was [pæ̃n.njæ̃n] 'half year', which has different tones from the other two examples.)

(6) Three-way contrast in VNV

V.NV	[fa:] + [næ̃n] →	[fa:.næ̃n]	'raise trouble'
VN.V	[fæ̃n] + [æ̃n] →	[fæ̃:.æ̃n]	'overturn case'
VN.NV	[fæ̃n] + [næ̃n] →	[fæ̃n.næ̃n]	'overturn trouble'

An oral closure is required for a nasal only when it is in the onset position. When a nasal coda occurs before a vowel, oral closure is not allowed. When a nasal coda occurs before a pause, oral closure is optional. When a nasal coda occurs before another nasal, it may add an additional duration for oral closure. Thus, there is no oral closure in VN.V, and the oral closure in VN.NV is longer than that in V.NV; in addition, VN.V and VN.NV differ in the length of the preceding vowel. In summary, there is no evidence for the zero onset, and the contrast between V.NV and VN.V shows that vowel-initial syllables cannot move a nasal coda into its own onset position. The examples in (7) also show the lack of the zero onset.

(7) No zero onset

[daaɤɤ] / *[daaʔɤɤ] / *?[daaɣɤɤ] / *[daaŋɤɤ]	'big goose'
[maaæ̃n] / *[maaʔæ̃n] / *?[maaɣæ̃n] / *[maaŋæ̃n]	'horse saddle'

Here, [ʔ] cannot be added (unless one is speaking slowly), nor is there clear evidence for the presence of [ɣ] (although it may be hard to tell). [ŋ] cannot be used either (except for those who always use [ŋ] for the zero onset).

There is one case when VN.V does have oral closure. It happens when the second V is unstressed, such as the interjection [a]. Some examples are shown in (8).

(8) Oral closure for N when the following V is unstressed

[næ̃n] + [a] →	[næ̃na]	'Hard!'
[mɑ̃ŋ] + [a] →	[mɑ̃ŋa]	'Busy!'

Duanmu (1990, 2000) suggests that in such cases the medial consonant is lengthened to a geminate that serves as both the coda of the first syllable and the onset of the second. He further suggests that the same is true for glides, so that VG.V→VG.GV, as shown in (9).

(9) The geminate analysis

[næ̃n] + [a] →	[næ̃nna]	'Hard!'
[mɑ̃ŋ] + [a] →	[mɑ̃ŋŋa]	'Busy!'
[xau] + [a] →	[xauwa]	'Good!'
[lai] + [a] →	[laija]	'Come!'

However, there are two problems. First, does the interjection have an onset slot? If it does, why is it not already filled with a zero onset? If it does not, why should the preceding sound be geminated? Second, in order to establish the geminate analysis, one needs to show that (a) the intervocalic nasal or glide is twice as long as a regular nasal or glide and (b) there is a contrast between VN.v and VN.Nv (where v is an unstressed vowel). But no evidence for (a) has been provided, nor is there evidence for (b). Therefore, a non-geminate analysis is sufficient, which is shown in (10).

(10) The non-geminate analysis

[næ̃n] + [a] →	[næ̃n.a]	'Hard!'
[mãŋ] + [a] →	[mãŋ.a]	'Busy!'
[xau] + [a] →	[xau.a]	'Good!'
[lai] + [a] →	[lai.a]	'Come!'

The non-geminate analysis is the same as that for *ran out* or *ran a (bell)* in English, discussed earlier. Still, there remains a question: why does the nasal coda have oral closure before an unstressed vowel but not before a stressed vowel in Standard Chinese? The answer seems to lie in assimilation. Assuming that a nasal with no oral closure is more sonorous than one with oral closure, and that a stressed vowel is more sonorous than an unstressed one, it is natural for a nasal coda to become more sonorous (lose oral closure) between two sonorous vowels.

One might still wonder why a nasal coda in English always has oral closure, whether the following vowel is stressed or not. I suggest that the answer lies in the frequency of use. In English, nearly all the consonants can occur in the coda. In contrast, in Standard Chinese, only two can. Therefore, each coda consonant is used much more frequently in Standard Chinese than is English. According to Fidelholtz (1975) and Bybee (2001), frequent words are more likely to undergo reduction than infrequent words, which explains the difference between Chinese VN (frequent and reduced) and English VN (not frequent and not reduced). It is relevant to note that many Chinese dialects have lost nasal codas altogether, such as Shanghai, in which VN has become V or nasalized V. Other Chinese dialects seem to be moving towards this direction as well.

In summary, the zero onset effect in utterance-initial position is probably unintended, and in medial positions there is no evidence for the zero onset or for resyllabification across a word boundary. The conclusions are similar to those we reached for English in Chapter 3.

4.3. THE ANALYSIS OF CG

It is clear that in CGVX, the rhyme includes VX only. This can be seen in words that rhyme in Standard Chinese, exemplified in (11), where the common rhyming part is VX, or [æ̃n].

(11) Words that rhyme in Standard Chinese

[mæ̃n] 'slow'
[mjæ̃n] 'noodles'
[jæ̃n] 'bright colored'
[tæ̃n] 'egg'
[tjæ̃n] 'shop'
[wæ̃n] 'ten thousand'
[twæ̃n] 'broken'

More controversial is the analysis of the CG part. Some consider CG to share one onset slot (Cheung 1986, Duanmu 1990, Ao 1992, Wang 1993), some consider G to share the nuclear slot with V (Bao 1990: 342, Goh 2000), and some consider G to have its own slot. Within the third view, some consider G to be structurally closer to C (Firth and Rogers 1937, Bao 1990: 328, Fu 1990), some consider it to be structurally closer to V (R. Cheng 1966, C. Cheng 1973, Wang 1957, Ji 1988, Lin 1989, Baxter 1992), and some consider it to be variably related to C or V (Chao 1934, Li 1983, Lin 1989, Bao 1990, Fu 1990). Evidence has been drawn from various sources, including phonemic economy, co-occurrence restrictions, and language games, but the interpretations of the evidence remain divided (see Duanmu 1990, 2000 for a review). In what follows I argue that CG forms a single sound, C^G.

The first argument comes from an observation by Chao (1934) that [sw] sounds quite different in English (as in [swei] *sway*) and Chinese (as in [swei] 'age'). In English, [sw] sounds like two separate sounds, in that the rounding of [w] starts after [s]. In Chinese [sw] sounds like a single sound, in that the rounding of [w] starts at the same time as [s]. The simplest way to represent the difference, therefore, is to use [sw] (two sounds) for English and [s^w] (one sound) for Chinese, or C^G for Chinese CG in general. The second argument is that every CG (or C^G) in Chinese can be represented as a complex sound (see Chapter 2). Thus, there is no need to represent CG as two sounds. The third argument is that the presence of G does not increase the length of a syllable in any appreciable way. Rather, CVX and CGVX are more or less similar in duration. Therefore, it is better to use the same syllable structure for both, where there is only one onset slot, which can be filled by C or C^G. The fourth argument is that sometimes CG alternates with C. For example, in Standard Chinese [nj] can alternate with [ɲ] (as in [njau]/[ɲau] 'bird'), and [sj] can alternate with [ɕ] (as in [sjau]/[ɕau] 'small'). In terms of feature structure, the analysis is straightforward, which is shown in (12).

(12) Sounds Feature structure

	Sounds	Feature structure
	[ɕ]	Cor—[+fricative]
		Dorsal—[+fricative]
	[s^j]	Cor—[+fricative]
		Dorsal—[−stop, −fricative]
	[ɲ]	Cor—[+stop]
		Dorsal—[+stop]
		Soft-palate—[+nasal]
	[n^j]	Cor—[+stop]
		Dorsal—[−stop, −fricative]
		Soft-palate—[+nasal]

The difference between [ɕ] and [s^j] is that the feature for Dorsal is [+fricative] in the former but [−stop, −fricative] in the latter. Similarly, the difference between [ɲ] and [n^j] is that the feature for Dorsal is [+stop] in the former but [−stop, −fricative] in the latter. Finally, the fifth argument for the C^G analysis is that there is simply no compelling evidence for assuming otherwise.

A common criticism against the C^G analysis is that it assumes too many phonemes. In particular, in the C^G analysis we assume not only [p, t, k, s, f, . . . , j, w] but also [p^j, t^j, t^w, k^w, s^w, s^j, f^j, . . .]. In contrast, if CG is made of two sounds, we only need to assume the first set. However, C^G is phonetically a single sound, and all analyses must acknowledge the fact that both [p, t, k, s, f, . . . , j, w] and [p^j, t^j, t^w, k^w, s^w, s^j, f^j, . . .] exist in the language. In addition, accepting the existence of [p^j, t^j, t^w, k^w, s^w, s^j, f^j, . . .] need not imply accepting them as additional phonemes. Rather, it is possible for two phonemes to merge into a single sound, as discussed in Chapter 2. Therefore, acknowledging [p^j, t^j, t^w, k^w, s^w, s^j, f^j, . . .] as complex sounds at the phonetic level need not imply an increase in the underlying phonemic inventory.

4.4. STRUCTURE OF STRESSED SYLLABLES

In traditional transcription a stressed Chinese syllable can range from one to four sounds. Some examples in Standard Chinese are shown in (13).

(13) Traditional transcription

C	V	GV	VC	CV	VG
[m]	[ɤ]	[wa]	[an]	[ta]	[ai]
‘yes?’	‘goose’	‘frog’	‘peace’	‘big’	‘love’
CVG	GVG	CVC	CGV	CGVC	CGVG
[fei]	[wai]	[tʰaŋ]	[kwa]	[xwaŋ]	[njau]
‘fly’	‘outside’	‘sugar’	‘melon’	‘yellow’	‘bird’

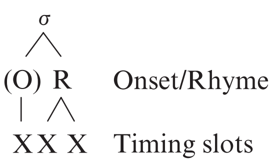
However, it is well known that all stressed syllables are long. In addition, I have argued that CG is a single sound C^G. Given such considerations, I offer the analysis in (14).

(14) Present analysis

[m:]	[ɤ:]	[wa:]	[ãn]	[ta:]	[ai]
‘yes?’	‘goose’	‘frog’	‘peace’	‘big’	‘love’
[fei]	[wai]	[tʰãŋ]	[kʷa:]	[xʷãŋ]	[nʲau]
‘fly’	‘outside’	‘sugar’	‘melon’	‘yellow’	‘bird’

All the syllables have the rhyme VX, although some have an onset and some do not. The full syllable structure is shown in (15).

(15) Structure for stressed syllables (onset is optional)



I use (C)VX as an abbreviation for the structure, although VX can be filled with a syllabic consonant. Some words are illustrated in (16).

(16) Sample representations

CVX	CVX	CVX	VX	VX	VX
| | |	| V	| | |	| |	V	V
nʲ a u	w a	f e i	a i	ɤ	m
‘bird’	‘frog’	‘fly’	‘love’	‘goose’	‘yes?’

When both CG are present, they share the onset slot. When only C or G is present, it occupies the onset slot alone. A single vowel is long because it occupies both rhyme slots. For the same reason, a syllabic consonant is also long. Syllabic consonants may seem to be marginal in Standard Chinese, because they only occur as interjections. However, in other dialects they can occur as lexical words. For example, in Shanghai, [n] can mean ‘fish’ or ‘five’ depending on its tone. In addition, according to Ramsey (1987) and Duanmu (2000), fricatives can be syllabic in Standard Chinese, too. For example, the analysis of [sz:] ‘four’ is shown in (17).

(17) Syllabic consonant

```
CVX
|  \/
s  z   'four'
```

Traditional analyses used to assume that every syllable must have a vowel, and the sound in the rhyme of 'four' has been thought to be a special vowel, called an "apical vowel." But if consonants can serve as the rhyme of stressed syllables, there is no need to assume apical vowels.

4.5. STRUCTURE OF UNSTRESSED SYLLABLES

In traditional transcription an unstressed Chinese syllable can also range from one to four sounds. Some examples in Standard Chinese are shown in (18), where ASP is an aspect marker.

(18) Traditional transcription

	Word		Example
V	[a]	interjection	[nan-a] 'hard-a (Hard!)'
CV	[lə]	ASP	[mai-lə] 'buy-ASP (bought already)'
CGV	[kwo]	ASP	[mai-kwo] 'buy-ASP (bought before)'
CVG	[tʰəu]	'head'	[mu-tʰəu] 'wood-head (wood)'
CGVC	[tʰjan]	'day'	[ʈʂʰwən-tʰjan] 'spring-day (spring)'

However, phonetic studies show that unstressed syllables are short and have a reduced rhyme (Woo 1969, Lin and Yan 1988). In particular, when VC and VG rhymes are unstressed, the coda is dropped and the rhyme duration is reduced by about 50 per cent. An example from Lin and Yan (1988) is shown in (19).

(19) Rhyme reduction in unstressed syllables
[ti:] + [fãŋ] → [ti:.fə̃] 'land-direction (place)'

When [fãŋ] is unstressed, the coda [ŋ] is dropped, and the vowel (still nasalized) is reduced to a schwa. Similar observations have been made before. For example, Gao and Shi (1963: 84–5) gave the examples in (20).

(20) [muu] + [tʰou] → [muu.tʰo] 'wood-head (wood)'
[nau] + [tai] → [nau.te] 'head-bag (head)'

The expressions in (20) are pseudo-compounds. A pseudo-compound is a compound in structure but a single word in meaning. The second syllable of a disyllabic compound is often unstressed, and when this happens, its rhyme is shortened. In the present analysis, an unstressed syllable loses the second rhyme slot. Its syllable structure is given in (21).

(21) Structure for unstressed syllables (onset is optional)

```
      σ
     / \
   (O) R    Onset/Rhyme
    |  |
    X  X    Timing slots
```

I use (C)V as an abbreviation for the structure, although V can be filled with a syllabic consonant. Some sample words are shown in (22).

(22) Sample representations

CV	CV	CV	V
| |	| |	| |	|
t^{h}o	t^{hj}ɔ̃	l ə	a
'head'	'day'	aspect	interjection

As with stressed syllables, when both CG are present, they share the onset slot. When only C or G is present, it occupies the onset slot alone. The rhyme for unstressed syllables has only one slot, though. The vowel is short and there is no coda.

4.6. CASUAL SPEECH AND VOWEL-LESS SYLLABLES

In casual speech some sounds can be deleted or changed, and this can create new syllables that do not occur in careful speech. For example, in careful speech no syllable in Standard Chinese ends in [m], but in casual speech such syllables are found. An example is shown in (23).

(23) wo mən → wom
I plural
'we'

In careful speech, 'we' is [woo.mən], but in casual speech the two syllables often merge into one [wom]. Similarly, devoicing of non-low vowels and syllabic consonants often happens for syllables that have an aspirated onset (including voiceless fricatives) and a low tone. As a result, many voiceless syllables can be created, such as [k^{h}ɤ] → [k^{h}x] 'can' and [tɕ^{h}i] → [tɕ^{h}ɕ] 'rise' in Standard Chinese. Voiceless or vowel-less syllables are not unique to Chinese, though: similar cases have been reported in Japanese and Berber.

4.7. FINAL VS. NON-FINAL POSITIONS

In the preceding discussion I proposed that Chinese has only two syllable types, the heavy (C)VX and the light (C)V. In a heavy syllable, the rhyme has just two slots. One might expect, then, that all heavy syllables have similar durations, and so do all light syllables. This seems to be the case, at least in controlled environments (Howie 1976, Lin and Yan 1988). However, it is also well known that syllables in pre-pause positions are longer. For example, Woo (1969) has shown that syllables with the third tone (T3) in Standard Chinese have an extra tone feature H in pre-pause positions, and their average rhyme duration is 50% longer than that of non-final syllables. Woo suggests that a final T3 syllable has three slots in its rhyme. The difference between final and non-final T3 syllables can be illustrated by the word [ma3] 'horse' in (24).

```
(24)  Non-final   Final
          L         LH
          /\        /\ |
         maa       maaa
```

Indeed, some linguists (e.g. Chao 1933: 132) point out that the full final T3 sounds like two syllables. Therefore, a lengthened pre-pause syllable can also be analyzed in (25).

```
(25)  CVX.V
      | | |  |
      ma a. a
```

In this analysis, there is a heavy syllable CVX and a light syllable V. Both agree with the proposed syllable types.

4.8. CVVC SYLLABLES

CVVC syllables have sometimes been reported in traditional descriptions. In this section I review two examples, Cantonese and Fuzhou. I show that CVVC syllables have limited distributions, and therefore that there is no need to assume this to be a separate syllable type.

4.8.1. **VVC Rhymes in Cantonese**

According to Huang (1970), Cantonese has seven vowels, all of which except [y] have a long and a short form. The rhyme inventory is shown in (26), grouped according to the coda. I have converted Huang's transcription into IPA symbols. Syllabic rhymes [m, ŋ] are not included.

(26) Cantonese rhymes including vowel length (Huang 1970)

Long	a:	a:i	a:u	a:m	a:n	a:ŋ	a:p	a:t	a:k
Short		ai	au	ʌm	ʌn	ʌŋ	ʌp	ʌt	ʌk
Long	ɛ:					ɛ:ŋ			ɛ:k
Short		ei							
Long	i:		i:u	i:m	i:n		i:p	i:t	
Short						eŋ			ek
Long	ɔ:	ɔ:i			ɔ:n	ɔ:ŋ		ɔ:t	ɔ:k
Short			ou						
Long	u:	u:i			u:n			u:t	
Short						oŋ			ok
Long	œ:					œ:ŋ			œ:k
Short		ɵy			ɵn			ɵt	
Long	y:				y:n			y:t	

Cantonese has preserved the historical codas [p, t, k] and [m, n, ŋ]. Of interest here are the large number of VVC rhymes, such as [a:m], [a:p], [i:m], which exceed the VX rhyme size. However, for all the vowels other than [a], the long and short forms are in complementary distribution. Even for [a], most long-short pairs differ in the quality of the vowel. The only pairs that appear to contrast in length are [a:i]-[ai] and [a:u]-[au]. However, in other studies, such pairs are transcribed with different vowels, such as [a:i]-[ɐi] and [a:u]-[ɐu] (Zee 2003). Therefore, vowel length plays no role in Cantonese phonologically and all VVC rhymes can be transcribed as VC rhymes. In (27) I show the transcription of Zee (2003).

(27) Cantonese rhymes where vowel length is not contrastive (Zee 2003)

Long	a	ai	au	am	an	aŋ	ap	at	ak
Short		ɐi	ɐu	ɐm	ɐn	ɐŋ	ɐp	ɐt	ɐk
Long	ɛ					ɛŋ			ɛk
Short		ei							
Long	i		iu	im	in		ip	it	
Short						ɪŋ			ɪk
Long	ɔ	ɔi			ɔn	ɔŋ		ɔt	ɔk
Short			ou						
Long	u	ui			un			ut	
Short						ʊŋ			ʊk
Long	œ					œŋ			œk
Short		ɵy			ɵn			ɵt	
Long	y				yn			yt	

In this transcription, there are no VVC rhymes. I conclude that Cantonese constitutes no problem for the CVX theory.

4.8.2. **VVC rhymes in Fuzhou**

In Fuzhou, most rhymes have two alternating forms, depending on the tone of the syllable and whether the syllable occurs in a final position. According to Feng (1998), Fuzhou has the GVX inventory shown in (28). When a GVX has two alternating forms, they are separated by a slash.

(28) GVX inventory in Fuzhou

	a/ɑ	ɛ/ɑ	o/ɔ	œ/ɔ	ai/ɑi	au/ɑu	ɛu/ɑu
i/ei	ia/iɑ	ie/iɛ				iau	ieu/iɛu
u/ou	ua/uɑ		uo/uɔ		uai/uɑi		uoi/uɔi
y/øy			yo/yɔ	øy/ɔy			
	aŋ/ɑŋ						ŋ (n, m)
iŋ/ɛiŋ	iaŋ/iɑŋ	ieŋ/iɛŋ			ɛiŋ/aiŋ		
uŋ/ouŋ	uaŋ/uɑŋ		uoŋ/uɔŋ			ouŋ/ɔuŋ	
yŋ/øyŋ			yoŋ/yɔŋ	øyŋ/ɔyŋ			
	aʔ/ɑʔ	ɛʔ	oʔ/ɔʔ	œʔ			
iʔ/ɛiʔ	iaʔ/iɑʔ	ieʔ/iɛʔ			ɛiʔ/aiʔ		
uʔ/ouʔ	uaʔ/uɑʔ		uoʔ/uɔʔ			ouʔ/ɔuʔ	
yʔ/øyʔ			yoʔ/yɔʔ	øyʔ/ɔyʔ			

Feng calls the form before the slash the "tense" form and the one after the slash the "lax" form. The "lax" form is used when the syllable is in final position and carries the tone L, MHM, or MH. The "tense" form is used when the syllable is in non-final position, regardless of its tone, or carries the tone H, HM, or M, regardless of its position.

Many of the rhymes seem to be VVC, such as [aiŋ] and [aiʔ], which exceeds VX. Even if we ignore the "lax" forms, which occur in final position and could have been lengthened, there are still some "tense" rhymes that seem to be VVC, such as [ɛiŋ] and [ɛiʔ]. However, it is relevant to note that Fuzhou has just one nasal coda, which means that [VVŋ] can be analyzed as a nasalized rhyme [ṼṼ]. Similarly, [VVʔ] can be analyzed as a rhyme with a glottalized vowel [$\text{VV}^{ʔ}$]. Such a reanalysis is entirely compatible with feature theory (see Chapter 2). A true VVC rhyme can be established if we have contrasting pairs such as [ain]/[aim] or [aik]/[aip], but Fuzhou does not have such pairs. Therefore, Fuzhou does not pose a serious problem for the CVX theory.

The present analysis makes a couple of phonetic predictions. First, it predicts that in non-final positions, a [VVŋ] rhyme either has no closure duration for the nasal coda (i.e. [ṼṼ]), or the vowel will be simplified to a monophthong (i.e. [Vŋ]). Similarly, it predicts that in non-final positions, either a [VVʔ] rhyme has no closure duration for the glottal coda (i.e. [VVˀ]), or the vowel will be simplified to a monophthong (i.e. [Vʔ]). I am not aware of such phonetic studies in Fuzhou, and will leave the predictions open.

4.9. SUMMARY

In traditional analyses both full (tone-bearing) and weak (toneless) syllables have variable structures, ranging from a minimum of C or V to a maximum of CGVX. I have argued instead that there are two syllable types in Chinese: all full syllables are heavy (C)VX and all weak syllables are light (C)V. The onset C is optional; it can be absent, or fill by C, or by G, or by C^G (a complex sound formed by of C and G). The rhyme VX or V can sometimes be filled by consonants. As far as I am aware, the only argument for the traditional analysis is phonemic economy. However, the present analysis accounts for more phonetic and phonological facts, at no extra cost to phonemic economy.

5

Standard Chinese

According to a recent survey (Chinese National Committee on Language Affairs 2004), 53% of the people in Mainland China can speak Standard Chinese, and among them 20% can do so fluently. I venture to estimate that, among the fluent speakers, about one tenth can do so without any obvious accent. This puts the number of fluent speakers of Standard Chinese at one tenth of the Chinese population, or about 130 million, and the estimated number of accent-free speakers of Standard Chinese at one hundredth of the Chinese population, or about 13 million.

5.1. SOUND INVENTORY AND TONES

There are nineteen consonants in Standard Chinese (excluding glides), shown in (1). Sounds with limited use are shown in parentheses.

(1) Consonants in Standard Chinese

Labial	Dental	Palatal	Retroflex	Velar
p, p^h	t, t^h			k, k^h
	ts, ts^h	(tɕ, tɕ^h)	ʈʂ, ʈʂ^h	
f	s	(ɕ)	ʂ, ʐ	x
m	n			ŋ
	l			

Initial p, p^h, t, t^h, k, k^h, ts, ts^h, ʈʂ, ʈʂ^h, f, s, ʂ, ʐ, x, m, n, l, (tɕ, tɕ^h, ɕ)
Final n, ŋ

Most can occur in syllable-initial position, but only two can occur in syllable-final position The three palatals only occur when we expect a consonant-glide combination, or when the nuclear vowel is [i] or [y]. In addition, some speakers use [ts^j, ts^{hj}, s^j] instead of [tɕ, tɕ^h, ɕ]. Therefore, [tɕ, tɕ^h, ɕ] can be seen as variants of [ts, ts^h, s] when they are palatalized

(Duanmu 2000). The aspirated sounds [pʰ, tʰ, kʰ, tsʰ, ʈʂʰ] are often [pˣ, tˣ, kˣ, tsˣ, ʈʂˣ] before a back vowel. For most speakers [ŋ] does not occur initially. For some speakers, [ŋ] is used initially in some syllables that otherwise start with a vowel.

The sounds [ʈʂ, ʈʂʰ, ʂ, ʐ] are made with the tongue tip curled back past the alveolar ridge and are commonly called "retroflex" consonants. Lee and Zee (2003) assume a narrow definition of retroflex sounds, according to which the underside of the tongue should touch the roof of the mouth. Since they did not find such contact in the Chinese [ʈʂ, ʈʂʰ, ʂ, ʐ], they call them "post-alveolar" instead. Lee and Zee also replace the symbols [ʈʂ, ʈʂʰ, ʂ, ʐ] with [tʃ, tʃʰ, ʃ, ɹ̺]. Their transcription has two shortcomings. First, [ʃ] (by its resemblance to [j]) suggests a palatal sound, which [ʈʂ, ʈʂʰ, ʂ] are not. Second, as far as I am aware, there is no language in which there is a contrast between a Chinese-type retroflex sound (with a back-curled tongue tip) and a "true" retroflex sound (with not just a back-curled tongue tip but also a contact between the underside of the tongue and the roof of the mouth). If so, there is no need to use a new set of symbols.

The sound [ʐ] needs additional comment. Phonetically, it is different from the retroflex vowel [ɚ] (or its stressed version [ɝ]). In some studies, such as Duanmu (2000), both are transcribed as [r], as in [ʂr] 'thing/event' ([ʂʐ] in the present analysis) and its diminutive form [ʂər] ([ʂɝ] in the present analysis). Since the rhyme in the diminutive form is a single (long) sound (like the vowel in the American English word *fur*), it is better to use a single symbol for it, i.e. [ɝ]. Thus, there are two ways to distinguish the unsuffixed and suffixed forms: either [ʂr]-[ʂɝ] or [ʂʐ]-[ʂɝ]. I choose the latter for the reason that the relation between [ʂ] and [ʐ] is similar to that between [s] and [z]. For example, in the syllable [sz] 'die', the rhyme is the voiced version of the onset, and in the syllable [ʂʐ] 'history', the rhyme is also the voiced version of the onset. In addition, in casual speech [sz] 'die' can undergo devoicing and become [ss], so can [ʂʐ] 'history' become [ʂʂ]. A criticism of [ʐ] is that it is the only voiced obstruent, because all other stops and fricatives are voiceless. A possible answer is that the distinction between [ʂ] and [ʐ] is not in voicing but in aspiration, i.e. [ʂʰ] and [ʂ] respectively. Another possible answer is that a phonological system need not always be perfectly balanced, and it is not impossible for one fricative to be voiced and the others not.

The syllables [sz] 'die' and [ʂʐ] 'history' contain syllabic consonants, so are the words [m] 'yes' and [hm] (showing contempt). Syllabic consonants are not included in (1) but will be discussed separately later.

The vowels in Standard Chinese are shown in (2). Sounds with limited use are shown in parentheses.

(2) Vowels in Standard Chinese

High vowels/glides	i	u	y		
Mid vowels	ə	(ɤ)	(e)	(o)	(ɚ)
Low vowels	a				
Diphthongs	ai	au	əi	əu	

When a high vowel is not in the nucleus, it can be treated as a glide; otherwise there is no difference between high vowels and glides. The mid vowel [ɤ] is long and [ə] is short. Therefore, [ɤ] can be seen as a long [ə]. The mid vowels [e, o] can occur in interjections, such as [e] 'OK' and [o] 'I see'. They can also occur as allophones of [ə], where [o] occurs after [w] in an open syllable (i.e. [uə] → [wo] 'I'), and [e] occurs after [j] or [y] in an open syllable (i.e. [iə] → [je] 'leaf' and [yə] → [ɥe] 'moon'). The diphthongs can be analyzed as combinations of two vowels each, and [əi, əu] can be realized as [ei, ou] respectively. The low vowel [a] is close to [æ] when the coda is [n] and close to [ɑ] when it is in an open syllable or when the coda is [ŋ]. Of the high vowels, [i] and [u] can occur in the coda but [y] cannot.

The retroflex vowel [ɚ] or [ɝ] is often written as [r]. It is the same as the vowel in the American English word *fur*. In Standard Chinese it occurs in two cases. The first is in the lone syllable [aɚ] (sometimes transcribed as [ɚ]), which represents a few words (with different tones), such as 'son', 'ear', or 'two'. In the second case, [ɚ] is a diminutive suffix, which replaces the coda of the original syllable, such as [p^{h}ai] + [ɚ] → [p^{h}aɚ] 'plaque', to be discussed later.

Each full syllable in Standard Chinese can carry one of four lexical tones. In addition, an unstressed syllable carries no lexical tone. In (3) I show several ways in which tones are represented in Standard Chinese, using the syllable *ma*.

(3) Tones in Standard Chinese

T1	T2	T3	T4	T5	Tone labels
H	LH	L	HL	(L)	Tone features
maa	maa	maa	maa	mə	IPA (without tones)
ma1	ma2	ma3	ma4	ma5	Pinyin (tones as numbers)
mā	má	mǎ	mà	ma	Pinyin (tones as diacritics)
'mother'	'hemp'	'horse'	'scold'	(interjection)	

The lexical tones on full syllables are often referred to as T1–T4 and the lack of lexical tone on an unstressed syllable is sometimes referred to as T5. In addition, [a] is often reduced to [ə] in an unstressed syllable. T1 is H (high), T2 is LH (rise), T3 is L (low) in non-final positions, and T4 is HL (fall). T5 is toneless, which is often realized as a low pitch. In the spelling system Pinyin, tones are represented either by the numbers 1–5 or by diacritics over the vowel.

In summary, Standard Chinese has nineteen consonants, of which all except [ŋ] can occur in the onset, and only [n] and [ŋ] can occur in the coda. In addition, if we leave aside interjections and the retroflex vowel, Standard Chinese has five vowels, which include three high vowels, one mid vowel, and one low vowel. Finally, a full syllable carries one of the four lexical tones and an unstressed syllable carries no lexical tone.

5.2. SYLLABLE INVENTORY

Each Chinese syllable is written as a separate orthographic unit called a "character." Therefore, the syllable inventory is closely related to the character inventory. In most cases, each character is also a separate word or morpheme.

The total number of characters ever used in Chinese is in the order of tens of thousands. The specific number is hard to determine because characters can be created, fall out of use, or change their shapes. For example, the Empress Wu Zetian (AD 625–705) created the character 曌 for herself, which shows a sun (日) and a moon (月) over the sky (空). The word is pronounced the same way as the word 'shine' ([ʈʂ̻au] in Standard Chinese). The intended idea is that she shines on the country (or the earth) as brightly as the sun and the moon combined. Similarly, the character '濆' ([fen] in Standard Chinese) used to mean 'riverside'; it is still used in the name of a river ('濆江') in Yaan, Sichuan, but it is rarely used elsewhere. The well-known *Kangxi Zidian* 'Kangxi Dictionary', sanctioned by the emperor Kangxi (Xuan Ye) and published in 1716, contains 47,035 different characters.

5.2.1. **Syllable frequencies and homophone density**

How many characters are used in Chinese today? Da (2004) examined a corpus of 258,852,642 character tokens, including both classic and modern texts, and found 12,041 different characters, or a quarter of those in the *Kangxi* dictionary. In addition, over 99% of the character tokens in the corpus are covered by the first 4,000 most frequent characters. The remaining 8,000 characters account for less than 1% of the occurrences. The data are shown in (4).

(4)

Most frequent	Cumulative coverage (%)
1,000	86.1740171074
2,000	95.5528960759
3,000	98.3248125395
4,000	99.304578085
5,000	99.7320610697
6,000	99.9267977338
7,000	99.9802177024
8,000	99.992322659
9,000	99.9966166078
10,000	99.9986162011
11,000	99.9995630719
12,000	99.9999841609

In modern texts, even fewer characters are used. This can be seen in (5), which shows the number of most frequent character types that cover 99% of all character tokens in each kind of corpora (Da 2004).

(5)

Corpus type	Total tokens	Characters for 99% coverage
Classic texts	65,348,624	4,433
Modern literature	87,249,603	3,050
Modern non-literature	106,254,415	2,445
All	258,852,642	3,591

The data show that to read modern texts, including modern literature, one only needs to know about 3,000 characters.

Next consider how many syllables the characters represent. According to *Xiandai Hanyu Cidian* ‘Modern Chinese Dictionary’ (Chinese Academy of Social Sciences Institute of Linguistics 1978), Standard Chinese has 1,334 syllables including tonal contrasts, or 413 syllables without tonal contrasts. However, some of the syllables represent infrequent characters, which occur only in classic texts or literary texts. According to the Chinese National Committee on Language Affairs (1988), there is a basic vocabulary of 2,500 common characters, which cover 99% of all occurrences in modern non-literature texts. The common characters represent just 1,001 syllables, as shown in (6).

(6) Syllables in Standard Chinese

Vocabulary type	All	Common
Characters	12,041	2,500
Syllables (with tones)	1,334	1,001
Syllables (without tones)	413	386

The numbers for the common characters were calculated by me, using the Chinese word processor NJ Star to convert characters to their pronunciations (in Pinyin spelling). Some characters have alternative pronunciations, with or without different meanings. For example, 長 can be *chang2* 'long' or *zhang3* 'to grow'; 熟 can be *shou2* or *shu2*, both meaning 'familiar' or 'ripe'; and 一 'one' can be *yi1* in final position, *yi2* before T4, and *yi4* before other tones. When a character has alternative pronunciations, only the first chosen by NJ Star is used, which is usually the more common pronunciation.

The data show that, although Standard Chinese has just 1,334 syllables, the average speaker still might not know all of them, or be sure of all of them. This is confirmed by a recent study by Myers and Tsay (2005), which found that acceptability judgments on possible syllables in Standard Chinese are gradient and influenced by lexical frequency.

Next consider homophone density. In Chinese each character has a different meaning, although some characters can have two or more meanings. For simplicity let us assume that each character represents one word or morpheme. Since there are many more characters than syllables, an average Chinese syllable represents quite a few words, or homophones, as shown in (7).

(7) Homophone density in Standard Chinese

Vocabulary type:	All	Common
Av. no. of homophones (with tones)	9.0	2.5
Av. no. of homophones (without tones)	29.2	6.5

Because the homophone load is not distributed evenly, some syllables represent more homophones than others. Let us use "homophone density value" to refer to the number of characters a given syllable represents. Some results are shown in (8) and (9), where tones are indicated by the digits 1–4 in the transcription.

(8) Top five and bottom two homophone density values in Standard Chinese, including tonal contrasts, and the number of syllables with these values, based on the 2,500 common characters (1,001 syllables)

Density value	No. of such syllables
20	1 [ʂʐ̩4]
15	1 [i4]
13	2 [fu4, tɕan4]
12	3 [tɕi2, tɕi4, y4]
11	2 [li4, tʂʐ̩1]
2	229 (23% of all)
1	432 (43% of all)

(9) Top ten homophone density values in Standard Chinese, excluding tonal contrasts, and the number of syllables with these values, based on the 2,500 common characters

Density value	No. of such syllables
38	1 [ʂʐ̩]
36	1 [tɕi]
33	1 [tʂʐ̩]
32	1 [i]
31	1 [tɕan]
29	1 [fu]
27	1 [y]
23	1 [ɕi]
22	4 [jan, ɕan, tɕʰi, li]
20	4 [tʂu, wəi, tɕe, tɕau]

Several interesting remarks can be made. First, as seen in (8), nearly half of the syllables do not have homophones. Second, what are often thought to be the most "natural" or "unmarked" syllables—those that children learn early or those that are thought to be most common in the world's languages, such as [ma], [pa], and [ta]—are not among those with high homophone density values. In particular, Hooper (1976a: 225) proposes that (a) the optimal syllable is CV, where C is a stop, and (b) the optimal syllable should also occur most frequently. However, the most frequent syllable in Standard Chinese is [ʂʐ̩], which is CC and whose onset is not a stop. In addition, [fu], [i], and [y] do not have a stop onset and so would be ranked as less optimal than [pa] and [ti], yet the former are far more frequent than the latter. Third, the most frequent syllable [ʂʐ̩] is made of [ʂ] and [ʐ̩], which are not very common cross-linguistically. Fourth, most English monosyllables represent just one word each. Since Chinese has so many homophones, one might wonder how Chinese avoids ambiguity in speech. The answer seems to be that most ambiguities are resolved by context. For example, although *sun* and *son* are homophones in English, there is hardly any context in which they would cause ambiguity.

5.2.2. **Phoneme frequencies**

Next consider phoneme frequencies in the basic lexicon of 2,500 common characters. If we count every underlying phoneme in the 2,500 characters, there are 7,014 phoneme tokens, based on the syllable analysis of Chapter 3 (see also Duanmu 2000). The frequency of a phoneme is the number of its occurrences in the 7,014 tokens. The results are shown in (10).

(10) Phoneme frequencies in the basic lexicon of 2,500 characters in Standard Chinese

Phoneme	Frequency	% of occurrences
i	1,001	14.27
a	963	13.73
u	940	13.40
ə	787	11.22
n	574	8.18
ŋ	468	6.67
ts	275	3.92
s	209	2.98
ts^h	156	2.22
l	155	2.21
tʂ	153	2.18
ʂ	141	2.01
t	126	1.80
y	126	1.80
p	121	1.73
x	114	1.63
k	112	1.60
$tʂ^h$	112	1.60
t^h	108	1.54
m	101	1.44
f	90	1.28
p^h	74	1.06
k^h	66	0.94
ʐ	42	0.60
All	7,014	100.00

Some observations can be made. First, most vowels are ranked high. The reason is that most syllables contain one consonant (excluding the nasal codas [n] and [ŋ]) and at least one vowel. Because Standard Chinese has just five vowels but nineteen consonants, vowels are expected to occur more frequently. Second, the vowel [y] is ranked much lower than other vowels, which seems to reflect the fact that [y] is an uncommon vowel cross-linguistically. Third, it may seem unexpected that [n] and [ŋ] are the most frequent consonants. The reason, however, is that they are the only consonants allowed in the coda. In the onset [n] occurs just forty-five times and [ŋ] does not occur at all. Fourth, coronal consonants are more frequent than labial and dorsal consonants, in agreement with the view that the tongue tip is the most active articulator. Fifth, the retroflex consonants [ʈʂ] and [ʂ] are ranked rather high, despite the fact that they are absent in many other languages. Finally, [m] is ranked rather low, although it is one of the most frequent sounds in the world's languages (Maddieson 1984).

5.2.3. **Onset and rhyme frequencies**

If we split each of the 2,500 common characters (morphemes) into an onset and a rhyme, we get 2,500 rhyme tokens and a slightly smaller number of onset tokens (some syllables do not have an onset). As discussed in Chapter 4, a pre-nucleus glide is in the onset. For example, in [nʲan] the onset is [nʲ] (or [ɲ] for some speakers) and the rhyme is [an]. In addition, according to Duanmu (2000), a high nuclear vowel will spread to the onset. For example, underlying [si] is realized as [sʲi], where the rhyme is [i] and the onset is [sʲ] for some speakers and [ɕ] for others. Given this analysis, there are fifty-five onsets, whose frequencies are shown in parentheses in (11). The symbol Ø indicates lack of an onset, which is found in thirty characters.

(11) Fifty-five onsets ranked by frequencies, in the basic lexicon of 2,500 characters in Standard Chinese.

tsʲ (177)	j (143)	sʲ (120)	ʈʂ (107)	ʂ (107)	ʈʂʰ (78)	w (77)
tsʰʲ (75)	p (75)	lʲ (75)	xʷ (72)	f (60)	kʷ (58)	ɥ (54)
t (54)	k (54)	m (53)	tʰ (49)	ʈʂʷ (46)	x (42)	ts (41)
l (41)	tsʰ (40)	pʲ (38)	tʷ (36)	tʲ (36)	ʈʂʰʷ (34)	ʂʷ (34)
tsᶣ (33)	pʰ (33)	kʰʷ (33)	kʰ (33)	sᶣ (32)	tʰʷ (30)	fʷ (30)
Ø (30)	tʰʲ (29)	s (29)	lʷ (29)	sʷ (28)	pʰʲ (27)	mʲ (26)
tsʷ (24)	tsʰᶣ (24)	ʐ (23)	mʷ (22)	nʲ (18)	n (18)	tsʰʷ (17)
ʐʷ (15)	pʰʷ (14)	lᶣ (10)	pʷ (8)	nʷ (8)	nᶣ (1)	

It is interesting, and perhaps surprising, to see that the plain coronal sounds [t], [tʰ], and [s] are not among the most frequent onsets, yet the retroflex sounds [ʈʂ], [ʈʂʰ], and [ʂ] are.

The 2,500 rhymes fall into twenty-one types, which can be grouped into three general types: VC, VG, and V or C. Their frequencies are shown in (12). Two of them are syllabic consonants [ʐ̩] and [z̩]. The rhymes [o, e, ɤ] are allophones of the phoneme [ə]. [o] occurs after [u] or a labial consonant. [e] occurs after [i] or [y]. [ɤ] occurs in other open heavy syllables.

(12) Twenty-one rhymes ranked by frequencies, in the basic lexicon of 2,500 characters

an (349)	u (206)	aŋ (194)	i (192)	au (189)	əŋ (178)	a (134)
əu (120)	əi (116)	ən (109)	uŋ (96)	o (96)	ai (93)	e (89)
ʐ̩ (84)	ɤ (78)	y (74)	in (71)	z (27)	aɚ (4)	ə (1)

Type	Count
VC	1,001 (40%)
VG	518 (21%)
V or C	981 (39%)
All	2,500

Some observations can be made. First, whereas [i] is the most frequent phoneme, [u] is a more frequent rhyme that [i]. Second, the vowels [i, a, u] are often thought to be the most basic, but none of them occurs as the most frequent rhyme by itself. Third, 61% of the rhymes have a coda or a diphthong, while just 41% of the rhymes are made of V or C.

5.2.4. **Tonal frequencies**

Next consider tonal frequencies in the basic lexicon of 2,500 common characters. The data are shown in (13).

(13) Tonal frequencies in the basic lexicon of 2,500 characters in Standard Chinese

Tone type	T1	T2	T3	T4	T5	All
Count	587	627	444	837	5	2,500
	23.5%	25.1%	17.8%	33.5%	0.2%	100.0%

Most syllables have T4 and only about half as many have T3. The T5 syllables include three interjections (*la*, *me*, and *ne*) and two grammatical particles (the aspect marker *le* and the nominal modification marker *de*), which usually occur in unstressed forms. One might get the impressions that most words in Chinese are stressed. But my own examination of a natural speech corpus (Duanmu et al. 1998) shows that about one third of all syllables are unstressed. This is because many full syllables can become unstressed in context (especially in compounds), in which case they lose their lexical tones and become T5. The data in (13) do not reflect de-stressing in natural speech.

If we ignore homophones, there are 1,001 syllables, which divide into the four full tones fairly evenly. The data are shown in (14). T2 syllables are slightly below the average and T4 syllables are slightly above the average, but not by much in either case.

(14) Tonal frequencies in the basic lexicon of 2,500 characters, excluding homophones

Tone type	T1	T2	T3	T4	T5	All
Count	249	220	245	282	5	1001
	24.9%	22.0%	24.5%	28.2%	0.5%	100.0%

In principle, every full syllable can take one of the four lexical tones. In fact, however, only a fifth of the syllables have four tones each. This is shown in (15).

(15) Tonal density in the basic lexicon of 2,500 characters, excluding homophones.

Tones per syllable	1	2	3	4	5	All
No. of such syllables	74	91	135	85	0	385
	19%	24%	35%	22%	0%	100.0%

Most syllables have three tones each and over 40% of the syllables have just one or two tones each. One might wonder why no syllable has all five tones. The reason is that only unstressed syllables have T5, and unstressed syllables may appear to have a different vowel. For example, the T5 syllable [mə] (an interjection) is spelled as *me*, instead of *ma*. If it is spelled as *ma*, then this syllable would have all five tones (ignoring vowel length), as seen earlier in (3).

5.3. ACCOUNTING FOR MISSING SYLLABLES

A full syllable in Standard Chinese can have up to four underlying sounds, CGVX, where C is a consonant, G a glide, V a vowel, and X an offglide of a diphthong or a consonant. In addition, each full syllable can carry one of four tones. In principle, there are 1,900 possible full syllables without tonal contrasts, or 7,600 including tonal contrasts. The calculation is shown in (16). Since [y] does not occur in the coda, and [ɚ] is mostly limited to suffixed words, they are not included in the choices for X. In addition, I have excluded unstressed syllables and consider there to be four possible tones per syllable.

(16) Possible combinations of syllables in Standard Chinese

	Choices	Notes
C	19	One of 18 Cs, or no C
G	4	One of [j, u, ɥ], or no G
V	5	One of five vowels
X	5	One of [i, u, n, ŋ] or no X
Total	1,900	without tonal contrasts
	7,600	with tonal contrasts (four tones per syllable)

The actual number of full syllables in Standard Chinese, as given in *Xiandai Hanyu Cidian* 'Modern Chinese Dictionary' (Chinese Academy of Social Sciences Institute of Linguistics 1978), is much smaller, as shown in (17).

(17) Actual and predicted numbers of full syllables in Standard Chinese

	Actual	Possible	% missing
Without tonal contrast	404	1,900	79
With tonal contrast	1,297	7,200	82

With or without tonal contrasts, just one fifth of the possible syllables are used.

The data raise an interesting question: Why are so many possible syllables missing? The question is especially puzzling because it is often thought that Chinese has such a shortage of syllabic contrasts that it has created many disyllabic words (Guo 1938, Wang 1944, Karlgren 1949, Lü 1963, Li and Thompson 1981, Chen 2000a). One might suggest that perhaps because Chinese can now use disyllablic words, there is no need to worry about a shrinking syllable inventory. But such reasoning is purely hypothetical. We need instead an analysis that can tell us the phonological reason for not using at least some of the missing syllables.

To simplify the discussion, let us focus on the GVX part. In Standard Chinese there are 100 possible GVX forms, calculated in (18). I have ignored tonal contrasts. In addition, I have omitted the vowel [ɚ] and syllabic consonants, which will be discussed separately later.

(18)

	Choices	Notes
G	4	One of [j, u, ɥ], or no G
V	5	One of five vowels
X	5	One of [i, u, n, ŋ] or no X
Total	100	

The 100 possible GVX forms are shown in (19). The first column indicates choices for X, the top row indicates choices for G, and 0 indicates lack of G or X. For clarity, high vowels are written as glides before the nuclear vowel.

(19)

		0-	j-	w-	ɥ-	
[-0]	i	+	(+)	–	–	ji = i
	u	+	–	(+)	–	wu = u
	y	+	–	–	(+)	ɥy = y
	ə	+	+	+	+	
	a	+	+	+	–	
[-n]	in	+	(+)	–	–	jin = in
	un	–	–	–	–	
	yn	+	–	–	(+)	ɥyn = yn
	ən	+	–	+	–	
	an	+	+	+	+	
[-ŋ]	iŋ	–	–	–	–	
	uŋ	+	+	(+)	–	wuŋ = uŋ
	yŋ	–	–	–	–	
	əŋ	+	+	+	–	
	aŋ	+	+	+	–	
[-i]	ii	(+)	(+)	–	–	ii = i, jii = ji
	ui	–	–	–	–	
	yi	–	–	–	–	
	əi	+	–	+	–	
	ai	+	+	+	–	
[-u]	iu	–	–	–	–	
	uu	(+)	–	(+)	–	uu = u, wuu = wu
	yu	–	–	–	–	
	əu	+	+	–	–	
	au	+	+	–	–	

Actual +	35
Missing – or (+)	65
Total	100

The symbol (+) indicates a form that does not contrast with another, as explained on the right-hand side. For example, there is no contrast between [ji] and [i]. In such cases, the longer form is marked with (+). As can be seen, many GVX forms are missing. There are two possible views on this, shown in (20).

(20) Two views on missing forms
- a. The missing forms are due to arbitrary choices or accidents.
- b. Some or most missing forms are due to systematic constraints.

On one view, the choice of syllables used in a language is arbitrary, and no further explanation is needed. On the second view, the missing forms indicate systematic constraints on syllable structures. The second view needs two kinds of evidence. First, one must show that there are reasonably natural constraints for most of the missing forms. Second, one must show that the constraints apply to other Chinese dialects (and to other languages). I shall argue for the second view. In particular, I propose four constraints, given in (21).

(21) Constraints on syllable structure

Rhyme-Harmony	VX cannot have opposite values in [round] or [back]. That is: *[+back][−back], *[−back][+back] *[+round][−round], *[−round][+round]
Merge	Two tokens of the same feature merge into one long feature. That is: XX → XX | | ∨ F_iF_i F_i
G-Spreading	A high nuclear vowel spreads to the onset C. That is: [Ci] → [C^ji], [Cu] → [C^wu], [Cy] → [$C^ɥ$y]
Y-Spreading	The nuclear [y] spreads to the pre-nuclear G. That is: [jy] → [ɥy], [wy] → [ɥy]

Rhyme-Harmony rules out opposite values of [back], but if a sound is unspecified for [back], it can combine with [+back] or [−back]. Thus, [+back][Ø back], [−back][Ø back], [Ø back][−back], [Ø back][+back] are allowed. The same is true for [round].

The phonetic reason for Rhyme-Harmony and Merge seems to be articulatory ease: it is hard for an articulator to move fast and execute two opposite gestures in a rhyme (such as [+back][−back]), or the same gesture twice in a rhyme (such as [−back][−back]). If so, one might find similar constraints in other languages, a topic that is beyond the scope of the present study.

G-Spreading is an anticipatory process. It does not apply when the onset is G or C^G. For example, G-Spreading does not require [wi] to become [ɥi] or [w^ji], or [t^wi] to become [$t^ɥ$i] or [t^{wj}i], or [ju] to become [ɥu] or [j^wu]. However, there seem to be some cases where G-Spreading occurs when the onset already has a glide, to be discussed below. Also, since the onset slot is optional (Chapter 3), G-Spreading does not apply when there is no onset. This explains why the English word [ist] *east* does not become [jist], the latter being a different word, *yeast*. The reason is that *yeast* has an onset but *east* does not.

Y-Spreading seems to be related to G-Spreading. The difference is that Y-Spreading occurs even if there is already a G, whereas G-Spreading does not.

Of the 25 VX rows in the GVX table, Rhyme-Harmony and Merge rule out 9, shown in (22), where I consider [n] to be [−back] in Standard Chinese and [ŋ] to be [+back]. Since Standard Chinese has only one mid vowel and one low vowel, they are unspecified for [back] or [round], and so they can combine with any glide or high vowel.

(22) Rows ruled out by Rhyme-Harmony and Merge

*row-[un]	differ in frontness
*row-[iŋ]	differ in frontness
*row-[yŋ]	differ in frontness
*row-[ui]	differ in rounding and frontness
*row-[yu]	differ in frontness
*row-[yi]	differ in rounding
*row-[iu]	differ in rounding and frontness
*row-[ii]	Merge, [ii] = [i:]
*row-[uu]	Merge, [uu] = [u:]

Because full syllables are heavy, a rhyme with just [i] or [u] are in fact [i:] and [u:] (Chapter 4). Therefore, there is no contrast between [i:] and [ii], because the latter will change to [i:] under Merge. Similarly, there is no contrast between [u:] and [uu].

There remain sixteen rows of rhymes, all of which satisfy Rhyme-Harmony. They are shown in (23). The mid vowel [ə] and the low vowel [a] are unspecified for rounding and frontness.

(23)

Row	Comment
[i]	
[u]	
[y]	
[ə]	
[a]	
[in]	both [−back]
[yn]	both [−back]; [n] unspecified for [round]
[ən]	[ə] unspecified for [back]
[an]	[a] unspecified for [back]
[uŋ]	both [+back]; [ŋ] unspecified for [round]
[əŋ]	[ə] unspecified for [back]

[aŋ]	[a] unspecified for [back]
[əi]	[ə] unspecified for [back]
[ai]	[a] unspecified for [back]
[əu]	[ə] unspecified for [back]
[au]	[a] unspecified for [back]

The rhyme [in] is [−back][−back], which will undergo Merge, as shown in (24).

(24)
```
 i  n → i  n
 |  |    \/
−B −B   −B    (B = [back])
```

The result is still [in], which does not overlap with another rhyme. Similarly, [yn] and [uŋ] will undergo Merge but still remain separate rhymes.

The sixteen rows contain sixty-four forms, listed in (25), where G-Spreading (G) covers six cases and Y-Spreading (Y) covers four. On the right are equivalent pairs under G-Spreading or Y-Spreading. For example, [u] and [wu] are equivalent, which means that [Cu] becomes [C^{w}u].

(25) Cells ruled out by G-Spreading (G)

	Ø-	j-	w-	ɥ-	
i	+	G	–	–	[ji] = [i]
u	+	–	G	–	[wu] = [u]
y	+	Y	Y	G	[jy] = [ɥy], [wy] = [ɥy], [ɥy] = [y]
ə	+	+	+	+	
a	+	+	+	–	
in	+	G	–	+	[jin] = [in]
yn	–	Y	Y	G	[jyn] = [ɥyn], [wyn] = [ɥyn], [ɥyn] = [yn]
ən	+	–	+	–	
an	+	+	+	+	
uŋ	+	+	G	–	[wuŋ] = [uŋ]
əŋ	+	+	+	–	
aŋ	+	+	+	–	
əi	+	–	+	–	
ai	+	+	+	–	
əu	+	+	–	–	
au	+	+	–	–	

The thirty-five occurring forms are indicated by +. There are nineteen non-occurring forms, indicated by −, which I list in (26). About half of them seem to be accountable in some ways, and I have made some comments on why they do not occur independently. For each pair of variants, such as [wi]-[wəi] or [ɥuŋ]-[juŋ], there seems to be no principled way to decide which variant is underlying and which is missing or derived.

(26)

Forms				Comments
wi				Variant of [wəi]
ju				Variant of [jəu]
win				Same as [ɥin] if G-Spreading applies to [win]
ɥuŋ				Same as [juŋ] if G-Spreading applies to [juŋ]
ɥəu	ɥau	wəu	wau	[+round]-[+round] between onset and coda
ɥi	ɥu	yn		
ɥəi	jəi			
jən	ɥən	ɥəŋ		
ɥaŋ	ɥai	ɥa		

In Standard Chinese, [wi] is a variant of [wəi] and [ju] is a variant of [jəu]. In [win] and [juŋ], there is a prenuclear G and a high nuclear vowel. If G-Spreading applies to the high nuclear vowel even if G is already occupied, these forms will become [ɥin] and [ɥuŋ], which explains why [win] and [ɥuŋ] do not occur independently. In [ɥəu], [ɥau], [wəu], and [wau], both the onset and the coda are [+round], and there seems to be a dissimilation effect against that. It is interesting to note that [wau] is marginal in English, too, and only occurs in three words, *powwow*, *bowwow*, and *wow*, but [wəu] occurs more frequently, such as *quote*, *quota*, *swollen*, *woeful*, and *wont*. So there does not seem to be a good reason to rule out [wau] and [wəu] completely. The syllable [jəi] is similar to [jei] in English, which only occurs in a marginal word *yea* and a few French borrowings such as *soigné*. Finally, it is interesting to note that twelve out of the nineteen non-occurring forms contain [y] or [ɥ].

If [ɥuŋ] and [iuŋ] are identical under G-Spreading, the underlying form could be either [ɥuŋ] or [iuŋ], although I have assumed that it is the latter. The ambiguity may explain some confusion in the literature. In traditional Chinese phonology, syllables are sometimes grouped according to the prenuclear glide. The word 'use' in Standard Chinese is sometimes thought to be [juŋ] and grouped with the [j]-group (Chao 1968), and sometimes thought to be [ɥuŋ] and grouped with the [ɥ]-group (Hsueh 1986). The present analysis suggests that there is no simple way to resolve the ambiguity, because both forms are possible.

Duanmu (2003) proposes a rule called Triphthong Raising, according to which a form with three vowels [high][mid][high] is raised to

[high][high][high]. The effect has been observed in Standard Chinese (Zee 2003) and can be understood in terms of articulatory effort reduction. The rule accounts for the alternation in the pairs [wəi]-[wi] and [jəu]-[ju]. It can also account for the lack of [ɥəu], [wəu], [ɥəi], and [jəi], all of which contain [high][mid][high]. However, Triphthong Raising does not hold for English, because English uses [wi] *we* and [wei] *way* contrastively. Also, English uses [wəu] quite often, as just seen. In (27) I summarize the analysis of GVX forms.

(27) Summary of the analysis of GVX forms

Reason	No. of forms
Rhyme-Harmony	28
Merge	8
G-Spreading	6
Y-Spreading	4
Other missing ones	19
Occurring	35
Total	100

70% of all missing forms are accounted for.
81% of all forms are accounted for.

One might attempt to offer various further accounts of the missing forms. However, it is unrealistic to expect that all or most missing syllables are due to phonological reasons. For example, I am not aware of any reason why some syllables in Standard Chinese have four tones but some do not, or why [m^{j}an], [p^{j}an], and [p^{hj}an] are used in Standard Chinese but [f^{j}an] is not. Therefore, I assume that some missing forms are just accidental gaps.

5.4. THE [ɚ]-SUFFIX

In unsuffixed words of Standard Chinese there is one syllable whose coda is [ɚ], which is [aɚ] or *er* in Pinyin. The syllable represents four words in the basic vocabulary of 2,500 common characters, which are *er2* 'son', *er2* 'but', *er3* 'ear', and *er4* 'two'.

Standard Chinese also has a suffix [ɚ], which often has a diminutive meaning. When [ɚ] is added to a syllable, it replaces the original coda if it is a coronal sound, including [i]; otherwise [ɚ] superimposes a retroflex feature on the rhyme. Ignoring the unstressed rhyme [ə], there are twenty rhymes in Standard Chinese. I list them in (28) together with their suffixed forms, based on the analysis of Duanmu (2000).

(28)	Unsuffixed (20 rhymes)	Suffixed (11 rhymes)
	a, ai, an	aɚ (a^r)
	ɑŋ	$ɑŋ^r$ ($ɑ̃^r$)
	au	au^r
	e	e^r
	əŋ	$əŋ^r$ ($ə̃^r$)
	əu	$əu^r$
	ɤ	$ɤ^r$
	o	o^r
	u	u^r
	uŋ	$uŋ^r$ ($ũ^r$)
	z, ʐ, əi, ən	ɚ
	i, in	(jɚ)
	y	(ɥɚ)
	ar	(aɚ)

In most cases (except [aɚ]), [ɚ] is not realized as a separate sound but a retroflex color on the entire rhyme, indicated by [^{r}]. For [i], [in], and [y], the rhyme of the suffixed form is [ɚ], which is the same as the suffixed form for [z], [ʐ], [əi], and [ən]. Also, the rhyme [ɚ] is a single vowel, similar to that in the word *fur* in American English. Therefore, it is better to use a single symbol [ɚ] rather than two [ər], which are used in Duanmu (2000). The rhyme [aɚ] does not have a suffixed form (or it is identical to the unsuffixed one). Therefore, there are a total of eleven suffixed rhyme types. In rhymes that end in [ŋ], the oral closure is optional. The coda [ŋ] will also change the vowel [a] to [ɑ].

The main effect of the [ɚ]-suffix is that it pushes away a coronal sound ([i], [n], [z], [ʐ], or [y]) from the coda position. In addition, it adds a retroflex feature to the sounds in the rhyme. Since the surviving sounds in the rhyme do not originally have a coronal component, the addition of the [ɚ] color does not incur violations in feature structure. In addition, the new rhyme incurs no violation of syllable constraints (i.e. Rhyme-Harmony and Merge).

5.5. SYLLABIC CONSONANTS

In transcribing words like [sz] 'four' and [ʂʐ] 'thing/event', I have used two syllabic consonants [z] and [ʐ]. The interjection [m] 'yes?' is another syllabic consonant.

Some phonologists believe that every syllable must have a vowel, and that there are no syllabic consonants (e.g. Luo and Wang 1957, Cheung 1986, Hsueh 1986, Coleman 1996, 2001). For them, interjections like [m] should be excluded as being marginal, and syllables like [sz] and [ʂʐ] must be reanalyzed so that they each contain a vowel. The special vowels are called "apical vowels," which are [ɿ] for the syllabic [z] and [ʅ] for the syllabic [ʐ].

In some dialects, however, syllabic consonants do occur as lexical words. For example, in Shanghai, [n] can mean 'fish' or 'five' depending on its tone, and in the disyllabic [m.ma] 'mom' the first syllable has a syllabic [m]. Since the oral tract remains completely closed in such syllables, there is no reason to think that they contain a vowel. But if one really believes that every syllable must have a vowel, one would argue that even in these cases there is a hidden vowel. Thus, Cheung (1986: 150) suggests that the Cantonese word [m] 'without' is [mi:m] underlyingly, where the hidden vowel [i:] is not pronounced. However, there is no evidence for such hidden vowels. If we recognize [n] and [m] as syllabic consonants, we can accept syllabic [z] and [ʐ], too, and there is no need to assume any special vowel or hidden vowels.

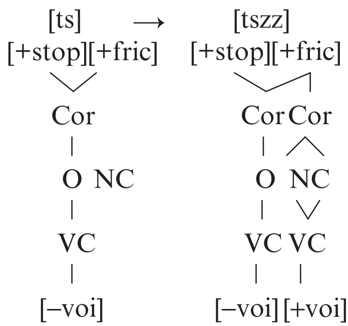
The distribution of syllabic [z] and [ʐ] is rather limited, as shown in (29). There are three syllables in which the syllabic [z] occurs and four syllables in which [ʐ] does.

(29) Occurrences of syllabic [z] and [ʐ]

[z]	tsz	tzʰz	sz	
[ʐ]	ʈʂʐ	ʈʂʰʐ	ʂʐ	ʐʐ

The onset of these syllables is either a fricative or an affricate, i.e. it contains the feature [+fricative] (see Chapter 2). In addition, the rhyme is identical or similar to the fricative component of the onset except that the rhyme is voiced. The fact that [z] and [ʐ] do not occur with other onsets suggests that they are not independent sounds. A simple analysis, therefore, is that syllabic [z] and [ʐ] are spread from the onset to the rhyme (Pulleyblank 1984, Lin 1989, Wiese 1997, Duanmu 2000). In (30) I show the analysis of [tszz] 'self', where the rhyme [zz] is long in a full syllable and O, N, and C are onset, nucleus, and coda respectively.

(30)



The word starts with the affricate [ts], which has two articulators, Vocal-cords (VC), which dominates the features [−voice], and Coronal (Cor), which dominates [+stop] and [+fricative]. Assuming that the nucleus does not allow [+stop] or [−voice], [ts] must be linked to the onset. Next, the empty slots in the rhyme trigger the spreading of [+fricative] (but not [+stop]). The rhyme also gets default [+voice]. The result is [ts] in the onset and [zz] in the rhyme.

The above analysis predicts that every fricative or affricate should have a syllabic fricative as its counterpart. The predictions for Standard Chinese are shown in (31).

(31)

Fricatives/affricates	Syllabic counterparts ([−stop, +voice])
[ts, tsʰ, s]	[z]
[ʈʂ, ʈʂʰ, ʂ, ʐ]	[ʐ]
[f]	[v] ([ʋ])
[x]	[ɣ] ([ɤ])

The predictions for the first two sets of sounds are found. The prediction for [f] overlaps with [fu], which is often realized as [fʋ], where the vowel (or syllabic [v]) is labiodental. The prediction for [x] overlaps with [xɤ], where [ɤ] is often realized as [ɯ] or [ɣ].

I have shown how the present analysis accounts for the distribution of the syllabic consonants [z] and [ʐ]. Now if they are apical vowels [ɿ] and [ʅ], as traditionally assumed, we should expect them to occur as freely as other vowels, which is not the case. A common story is that [ɿ] and [ʅ] are allophones of the vowel [i] (Dong 1958: 81). However, consider the words [sz] 'silk' and [si] (realized as [sʲi] or [ɕi]) 'west'. In the present analysis, they form a minimal pair of contrast. If so, [z] and [i] are not in complementary distribution and not allophones of the same phoneme. In the traditional analysis, the solution is not so obvious.

Lee and Zee (2003) use the symbol [ɹ̩] for both [z] and [ʐ], but it is not clear whether their transcription is meant to be phonemic or allophonic. If [ɹ̩] is meant to be a phoneme, then there is no need for it, because its occurrence is entirely predictable. If [ɹ̩] is meant to be an allophone, then it is not specific enough, because [z] and [ʐ] derive from different sets of consonants and their phonetic qualities are quite different.

In casual speech, high vowels are often devoiced when they occur in syllables that have an aspirated onset (including voiceless fricatives) and a low tone. As a result, additional syllabic consonants are created. Some examples are shown in (32), where vowel length is not shown. When a sound is devoiced, the tone cannot be heard, which is indicated by Ø.

(32)		HL-L	HL-Ø	
	[ʋ]→[f]	təu-fʋ →	tou-ff	‘tofu’
		HL-L	HL-Ø	
	[z]→[s]	ʂaŋ-tsʰz →	ʂaŋ-tsʰs	‘last time’
		HL-L	HL-Ø	
	[ʐ]→[ʂ]	li-ʂʐ →	li-ʂʂ	‘history’
		L-LH	Ø-HL	
	[ɤ]→[x]	kʰɤ-nəŋ →	kʰx-nəŋ	‘possible’
		HL-L	HL-Ø	
	[i]→[ɕ]	ji-tɕʰi →	ji-tɕʰɕ	‘together’
		H-L	H-Ø	
	[y]→[ɕʷ]	tʂəŋ-tɕʰʷy →	tʂəŋ-tɕʰʷɕʷ	‘strive for’
		L-H	Ø-H	
	[y]→[ɕʷ]	ɕʷy-tʷo →	ɕʷɕʷ-tʷo	‘many’
		L-HL	Ø-HL	
	[u]→[xʷ]	ʂʷu-tɕa →	ʂʷxʷ-tɕa	‘summer vacation’
		H-L	H-Ø	
	[u]→[xʷ]	ɕin-kʰʷu →	ɕin-kʰʷxʷ	‘working hard’
		L-HL	Ø-HL	
	[u]→[xʷ]	tʰʷu-tʲi →	tʰʷxʷ-tʲi	‘land’

Devoicing can happen to syllables in any position (initial, medial, or final). Devoiced [ʐ, i, ɤ, u, y] sound like [ʂ, ɕ, x, xʷ, ɕʷ] respectively. Devoiced syllables have similar durations to the originals (although rhyme length is not indicated in the above transcription), and therefore they still sound like separate syllables. One might suggest that the devoiced vowels are still vowels [i̥, ɤ̥, u̥, y̥]. But if we accept syllabic consonants, there is no need to assume voiceless vowels.

5.6. HOMOPHONE DENSITY, FREQUENCY, AND SYLLABLE LOSS

It is interesting to compare homophone densities in Chinese and English. For simplicity, let us assume that most syllables that occur in polysyllabic words also occur in monosyllabic words. Therefore, we can limit our examination of English to monosyllables.

Let us compare two electronic lexicons of English. The CMUDICT lexicon (Weide 1998, version 0.6) contains 127,008 entries. The CELEX lemma lexicon (Baayen et al. 1993) contains 52,447 entries. The difference in size results from different selection policies. CMUDICT includes as separate entries all forms that are spelled differently, such as *bank–banked–banks–bank's–banks'*, or pronounced differently, such as [housts]–[houss]–[hous] for *hosts*. CMUDICT also includes proper names. On the other hand, CMUDICT ignores letter case and meaning. For example, there is one entry *BILL* for *Bill* and *bill*, and there is one entry *BARK* for *bark* (of a tree), *bark* (a boat), *bark* (the sound of a dog), *bark* (to make a dog sound), and *bark* (to strip tree bark). In contrast, the CELEX lemma lexicon lists different meanings of words, but excludes proper names and regular inflections, such as tenses for verbs, plurals for nouns, or possessives for nouns. For example, there is *bank* but no *banks*, *banked*, *banks'*, or *bank's*, and there is *catch* but no *caught*. On the other hand, *bark* (of a tree), *bark* (a boat), *bark* (the sound of a dog), *bark* (to make a dog sound), and *bark* (to strip tree bark) are all listed as separate entries.

Let us define homophone density as the number of homophones per pronunciation. There are two ways to define homophones, though. In the first, they are words that are spelled differently but pronounced the same. In the second, they are words that have different meanings but are pronounced the same. In English the difference is not great. This can be seen in (33), which shows monosyllables in CELEX.

(33) Homophone density of monosyllables in the lemma lexicon of CELEX

All meanings	6,760
Non-repeated spelling	4,321
Non-repeated pronunciation	3,801
Homophone density by meaning	6,760 ÷ 3,801 = 1.78
Homophone density by spelling	4,321 ÷ 3,801 = 1.14

The data show that of 6,760 entries, 4,321 (64%) are spelled differently. Therefore, the homophone density based on meaning is not much larger than the homophone density based on spelling (1.78 vs. 1.14). In (34) I show homophone densities in CMUDICT, in comparison with those in the CELEX lemma lexicon, where the value of "homophone density by meaning" for CMUDICT is estimated, based on the fact that in CELEX the number of spelled forms is about 2/3 of the number of meanings.

(34)

	CMUDICT	CELEX
All monosyllables	16,479	6,760
Non-repeated spelling	15,901	4,321
Non-repeated pronunciation	10,253	3,801
Homophone density by spelling	1.55	1.14
Homophone density by meaning	(2.43)	1.78

The data show that, for monosyllables in English, the homophone density lies between 1 and 2.5. It is considerably lower than that in Standard Chinese, which is shown in (35).

(35) Homophone density of monosyllables in Standard Chinese

Vocabulary type	All	Common
Non-repeated characters (morphemes)	12,041	2,500
Non-repeated pronunciation with tone	1,334	1,001
Homophone density	9.03	2.50

If we consider all characters in use (Da 2004), the homophone density is 9.0. If we consider the 2,500 most commonly used characters, the homophone density is 2.5. Since some characters have two or more meanings, just as some spelled forms do in English, the homophone density in (35) corresponds to the "homophone density by spelling" in English. Therefore, the homophone density in Chinese is much higher than that in English. The comparison is shown in (36).

(36) Homophone density based on orthography

	English	Standard Chinese
Range	1.14–2.43	2.50–9.03
Mean	1.8	5.8

If we use the mean for the homophone density range, then the homophone density in Chinese is three times that in English (5.8 vs. 1.8). It can be seen that both English and Chinese have up to 10,000 monosyllabic words. Therefore, the higher homophone density in Chinese does not come from a difference in vocabulary size, but from a difference in the size of syllable inventory, i.e. the syllable inventory in Chinese is a small fraction of that in English.

Since Chinese has so many homophones, one might expect Chinese speakers to pronounce their syllables carefully and to maintain the already small syllable inventory. Surprisingly, neither seems to be the case. In a natural speech corpus of Chinese (Duanmu et al. 1998), about a third of all syllables are unstressed and reduced. In addition, while Middle Chinese (about AD 600) had over 3,000 syllables (including tonal distinctions), modern Standard Chinese has 1,334. Thus, over a period of 1,500 years Chinese lost more than half of its syllables. Moreover, the syllable inventory of modern Chinese continues to shrink. For example, Standard Chinese no longer uses [p, t, k, m] in syllable final position. In addition, Standard Chinese does not make use of such contrasts as [wi] vs. [wei], or [ji] vs. [i], which English does (consider *we* vs. *way*, and *yeast* vs. *east*). Moreover, about 200 of the 1,334 syllables in Standard Chinese are rarely used and probably on the way out. In contrast, I am not aware of a similar scale of syllable loss in English, either recently or currently.

It is a puzzle why a language with a small syllable inventory is undergoing massive syllable loss, whereas a language with a large syllable inventory is not. Many linguists believe that Chinese has an urgent need to preserve syllable contrasts in order to avoid ambiguities (e.g. Guo 1938, Wang 1944, Karlgren 1949, Lü 1963, Li and Thompson 1981, Chen 2000a), and the continued syllable loss comes as a surprise. I suggest that there are two explanations. First, ambiguities rarely arise in speech, because most ambiguities are clarified by context. For example, although *sun* and *son* are homophones in English, they hardly cause any ambiguity in context. Second, paradoxically, high homophone density may in fact speed up syllable loss. Studies on frequency effects suggest that frequent words are more likely to undergo reduction than infrequent words (Fidelholtz 1975, Hooper 1976b, Bybee 2001). If so, Chinese syllables are more likely to undergo reduction and loss of contrasts, because Chinese has fewer syllables than English and so Chinese syllables are used much more frequently.

5.7. SUMMARY

I have presented some quantitative data on the sounds, tones, and syllable structure in Standard Chinese. Of interest are the following points. First, the most frequent sounds or syllables are not always what are often thought to be the most common cross-linguistically. Second, the list of actual syllables in Chinese is only a small fraction of all conceivable combinations. Third, most of the missing syllables can be accounted for by four constraints—Rhyme-Harmony, Merge, G-Spreading, and Y-Spreading—which seem to be based on articulatory ease. Finally, although Chinese has only a tenth as many syllables as English, the Chinese syllable inventory continues to shrink at a fast rate. The fact is unexpected if there is a pressing need to preserve lexical contrasts or to avoid ambiguities. Instead, the fact supports the view that ambiguity is rarely a problem when words are used in context, and the proposal that frequently used forms are more likely to undergo reduction and loss of contrast than infrequent forms.

6

Shanghai Chinese

Shanghai Chinese is spoken in metropolitan Shanghai, the largest city in China. It is a member of the Wu dialect family of Chinese. The Wu family covers 8% of Chinese speakers and is second only to the Mandarin (or Northern) family, which covers over 70% of Chinese speakers. Shanghai is also one of the most documented Chinese dialects. The first detailed description appeared over 150 years ago (Edkins 1853).

Xu and Tao (1997) distinguish three contemporary varieties of Shanghai: Old Shanghai, spoken by people born before 1920, Mainstream Shanghai, spoken by people born between 1940 and 1965, and New Shanghai, spoken by people born after 1965. There may be few speakers of Old Shanghai left, and speakers of New Shanghai have become the majority. In recent years, however, there has been an increase of Shanghai residents who do not speak or use Shanghai, including some children of Shanghai-speaking parents.

As discussed in Chapter 4, the maximal Chinese syllable is CGVX. Traditional descriptions of Chinese usually do not list the consonants and vowels. Instead, they list the inventory of the initial C and the inventory of the remaining GVX. Indeed, You et al. (1980) argue that the Chinese syllable should not be analyzed in terms of consonants and vowels, but in terms of "onset phonemes" (the initial C inventory) and "rhyme phonemes" (the GVX inventory). While this position may seem unusual, it is echoed by Ladefoged (2001). Similarly, in their transcription of Standard Chinese, Lee and Zee (2003) list "vowels," "diphthongs" (such as [ei, ie, ou, uo]), and "triphthongs" (such as [uei, iou]). Evidently, they consider diphthongs (GV and VG) and triphthongs (GVG) as phonemes. In the present analysis, a sound has a specific definition (see Chapter 2), and diphthongs and triphthongs are each made of two or three sounds. Therefore, we shall interpret all GVX forms in terms of consonants and vowels.

A striking property of Shanghai is that its GVX inventory has been shrinking fast, although its initial C inventory has remained stable. This is evident if we compare the four varieties in (1), where Edkins is the variety described by Edkins (1853) (interpreted by Qian 1997), and Old, Mainstream, and New are the three varieties described by Xu et al. (1988) and Qian (1997). I have replaced [ər] with [ɚ] and the so-called 'apical vowel' with [z] and placed them in brackets with syllabic consonants.

(1)

Variety	Initial C	GVX inventory
Edkins	28	56 plus [z, zʷ, z̃ʷ, ɚ, m, ŋ]
Old	27	44 plus [z, zʷ, ɚ, m, n, ŋ]
Mainstream	28	38 plus [z, ɚ, m, n]
New	27	28 plus [z, ɚ, m, n]

A few sample words are given in (2), from Chen (2003: 187), which show the kind of mergers that have occurred as a result.

(2) Syllable mergers as a result of GVX reduction

<table>
<tr><th></th><th>客
'guest'</th><th>掐
'pinch'</th><th>磕
'knock'</th><th>刻
'carve'</th><th>渴
'thirsty'</th><th>哭
'weep'</th><th>壳
'shell'</th></tr>
<tr><td>Edkins</td><td>kʰɑʔ</td><td>kʰæʔ</td><td>kʰeʔ</td><td>kʰʌʔ</td><td>kʰøʔ</td><td>kʰoʔ</td><td>kʰɔʔ</td></tr>
<tr><td>Old</td><td colspan="2">kʰAʔ</td><td colspan="2">kʰəʔ</td><td>kʰøʔ</td><td>kʰoʔ</td><td>kʰɔʔ</td></tr>
<tr><td>Mainstream</td><td colspan="2">kʰAʔ</td><td colspan="3">kʰəʔ</td><td colspan="2">kʰoʔ</td></tr>
<tr><td>New</td><td colspan="5">kʰɐʔ</td><td colspan="2">kʰoʔ</td></tr>
</table>

As far as CVX theory is concerned, there is nothing unusual about Shanghai, in that its syllables do not exceed the maximal size of CVX. In particular, the G can form a complex sound with the initial C and the VX falls within the limit of rhyme size.

However, Shanghai Chinese has two interesting properties. First, it has no true codas. In other words, Shanghai lacks true VC and VG rhymes, where C is a consonant coda and VG is a diphthong. This has some interesting consequences, especially with regard to tone split. Second, the GVX inventory seems to be rather unbalanced, in the sense that there are many missing GVX forms that do not seem to violate any phonological constraints. In this chapter I illustrate these two points through an examination of New Shanghai, which is the dominant variety today.

6.1. CONSONANT INVENTORY

According to Qian (1997), New Shanghai has twenty-seven initial consonants, shown in (3).

(3) Consonants in New Shanghai (27)

	Labial	Dental	Palatal	Velar	Glottal
Stop	p, pʰ, b	t, tʰ, d		k, kʰ, g	ʔ
Affricate		ts, tsʰ	tɕ, tɕʰ, dʑ		
Fricative	f, v	s, z	ɕ		h, ɦ
Nasal	m	n	ȵ	ŋ	
Lateral		l			

There is some overlap between the dental series and the palatal series. For example, [ts, tsʰ, s, z, n] do not occur before [i] or [y]; instead, we find the palatals [tɕ, tɕʰ, ɕ, dʑ, ȵ].

As in other Wu dialects, Shanghai has a three-way contrast in Vocal-cord features for stops and affricates: [−voice, −aspirated], [−voice, +aspirated], and [+voice, −aspirated], such as [p, pʰ, b]. Syllables that start with a voiceless obstruent (stop, fricative, or affricate) have a clear and generally higher tone. Syllables that start with a voiced obstruent have a murmured and lowered tone, while the voiced obstruent itself becomes voiceless in foot-initial position. The correlation is a common interaction between onset voicing and tone and has been referred to as "voiceless-high" and "voiced-low." The same correlation can be seen in Korean, where the "aspirated," "tense," and "lax" stops correspond to voiceless aspirated, voiceless unaspirated, and voiced stops respectively in Shanghai (Kim and Duanmu 2004).

6.2. GVX INVENTORY

According to Qian (1997), New Shanghai has thirty-two GVX forms, considerably fewer than the forty-two in Mainstream Shanghai. The inventory is given in (4), where the forms present in Mainstream Shanghai but lost in New Shanghai are given in parentheses. I have replaced Qian's [A] with [a].

(4) GVX inventory: 32 in New Shanghai (42 in Mainstream Shanghai)

m	n	ɚ	z			
o	i	u	y			
a	ia	ua	ã	iã	uã	(iaʔ)

ɐʔ	uɐʔ		
ɔ	iɔ		
ɤ	iɤ		
ɛ	(iɛ)	uɛ	
ø	(uø)	(yø)	
(ɒ̃)	(iɒ̃)	(uɒ̃)	
ən	in	uən	yn
oŋ	ioŋ	oʔ	(ioʔ)
(əʔ)	(uəʔ)		
iɪʔ	yɪʔ		

The inventory invites many questions. For example, how should we interpret the transcription? How many vowels are there? How many final Cs are there? Can we account for the GVX forms that are used and those that are not? And so on. Unfortunately, Xu et al. do not decompose the GVX forms any further. In addition, there has been no systematic study of the syllable in Shanghai.

It is clear that some vowels do not contrast, such as [ɐ] and [ə]. Similarly, there is no contrast between [n] and [ŋ] in the coda. Therefore, I propose the interpretations in (5)–(8), and the resulting GVC inventory in (9).

(5) Questions about [ɛ]
There is no [e].
Proposal: [ɛ] = [e]

(6) Questions about [ɤ] and [ə]
[ɤ] and [ə] are in complementary distribution.
Proposal: [ə] = [ɤ]

(7) Questions about [ɪ]
[ɪ] only occurs in [iɪʔ] and [yɪʔ].
There is no contrast between [i] and [ɪ], or between [ɪ] and [y].
Proposal: [iɪʔ] = [iʔ], [yɪʔ] = [yiʔ]

(8) Questions about [n], [ŋ], and [˜]
[n], [ŋ], and [˜] are in complementary distribution.
Proposal: There is only one nasal N, unspecified for place.

(9) Revised GVX inventory

m	n	ɚ	z			
u						
a	ia	ua	aN	iaN	uaN	(iaʔ)
ɐʔ	uɐʔ					

(ɒN)	(iɒN)	(uɒN)			
o	oN	ioN	oʔ	(ioʔ)	
ɔ	iɔ				
ɤ	iɤ	ɤN	uɤN	(ɤʔ)	(uɤʔ)
e	ue	(ie)			
ø	(uø)	(yø)			
y	yN				
i	iN	iʔ	yiʔ		

Some vowels still have limited occurrences in New Shanghai. For example, [u, ø] do not combine with any G or X, [ɐ] does not occur alone, and [ɔ, e] occur just twice each. We can also see that with the loss of ten GVX forms, some reorganization is likely to occur. For example, since there is no [io], [iɔ] is likely to be realized as [io] and [ɔ] itself may be on the way out.

What this means is that the GVX inventory is not fully stable, in the sense that we do not expect to find a perfectly balanced distribution of GVX forms for all vowels. Therefore, the GVX inventory is unlikely to be describable, or generated, by a set of precise rules, as classic generative phonology would hope (Halle 1962). Rather, we should see the inventory as a system in transition, perhaps constantly so, at least for Shanghai in the last 150 years.

On the other hand, we would not expect the GVX inventory to be completely chaotic either. Instead, we would expect to find some phonological or phonetic regularity that guides the change. For example, we do not expect major constraints on syllable structure to be violated, such as Rhyme-Harmony, Merge, G-Spreading, and Y-Spreading discussed in Chapter 5. In addition, we might be able to account for the absence of some forms and perhaps also the likely or forthcoming loss of other forms.

6.3. ANALYSIS OF THE GVX INVENTORY

I propose that the GVX inventory in New Shanghai be interpreted along the points in (10). The revised inventory is given in (11), where I use a glide for a high vowel before the nucleus.

(10) [ɐʔ, uɐʔ] are [aʔ, uaʔ].
[iɔ] is [io].
[VN] is nasalized vowel [Ṽ].
[Vʔ] is glottalized vowel [Vˀ].

(11) Revised GVX inventory of New Shanghai (32 in all)

m	n	ɚ	z		
u	ø	ɔ			
y	ỹ				
i	ĩ				
o	jo	õ	jõ		
ɤ	jɤ	ɤ̃	uɤ̃		
e	we				
a	ja	wa	ã	jã	wã
iˀ	ɥiˀ	aˀ	waˀ	oˀ	

The nasal feature [˜] in [Ṽ] is the remnant of historical [m, n, ŋ], and the feature [ˀ] in [Vˀ] is the remnant of historical [p, t, k]. In some Chinese languages, such as Cantonese, [m, n, ŋ, p, t, k] still contrast in the coda position.

The idea that [VN] and [Vʔ] are [Ṽ] and [Vˀ] respectively is in agreement with phonetic descriptions, as noted by Xu et al. (1988). Under this analysis, Shanghai does not have true codas. Every rhyme is V or C, whose length or weight is determined by stress: the rhyme is long when stressed ([V:] or [C:]) and short when unstressed ([V] or [C]), as argued in Duanmu (1999).

I further propose that the vowel [ɔ] is on the way out, because it does not combine with any other sound in the GVX inventory and because some [ɔ] syllables have already merged with [o] syllables. For example, both 'below' and 'summer' have changed from [ɦɔ] to [ɦo] (Xu et al. 1988: 54). On the other hand, I propose that [ø] be kept, even though it does not combine with other sounds, because (as I shall argue below) most missing GVX forms that involve [ø] can be ruled out independently.

Under these considerations, the sounds that can fill G, V, and X in New Shanghai are shown in (12). The syllabic consonants [m, n, z], the retroflex vowel [ɚ], and the marginal vowel [ɔ] are not included.

(12)

G	4	[i, u, y] or no G
V	8	[i, y, e, ø, u, ɤ, o, a]
X	3	[ˀ, ˜] or no X

Total possible GVX forms: $4 \times 8 \times 3 = 96$

Since there are four choices for G, eight for V, and three for X, there are 96 possible GVX forms, of which only twenty-seven (or 28%) are found.

In Chapter 5 I used four constraints to account for the missing GVX forms: Rhyme-Harmony, Merge, G-Spreading, and Y-Spreading.

Since there is no true coda in New Shanghai, Rhyme-Harmony and Merge have no effect. The relevant constraints, therefore, are G-Spreading and Y-Spreading, under which fifteen GVX forms are ruled out because they are identical to others, shown in (13) and (14).

(13) Equivalent pairs under G-Spreading (9 in all)
ji = i, wu = u, ɥy = y
jĩ = ĩ, wũ = ũ, ɥỹ = ỹ
ji^ʔ = i^ʔ, wu^ʔ = u^ʔ, ɥy^ʔ = y^ʔ

(14) Equivalent pairs under Y-Spreading (6 in all)
jy = ɥy, wy = ɥy, jỹ = ɥỹ, wỹ = ɥỹ, jy^ʔ = ɥy^ʔ, wy^ʔ = ɥy^ʔ

There are still fifty-four other missing GVX forms to account for. The most obvious constraint, it seems, is a lack of contrast in vowel height in GV and VX. This can be seen in the data in (15). Each vertical pair of GVX forms differ only in the height of the nuclear vowel. GVX forms not used in New Shanghai are given in parentheses.

(15) Lack of contrast in nuclear vowel height in GV and VX

High	(ju)	(wi)	(ũ)	ĩ	(u^ʔ)	i^ʔ	...
Non-high	jo	we	õ	(ẽ)	o^ʔ	(e^ʔ)	...

In fact, if we interpret [ɤ] as [ɯ], we can divide the eight vowels in New Shanghai into two sets, high and non-high, shown in (16).

(16) Vowels in New Shanghai

	Unrounded		Rounded	
High	i	ɯ (ɤ)	y	u
Non-high	e	a	ø	o
	Front	Back	Front	Back

If there is no contrast in nuclear vowel height in GV and VX, we expect nearly half of all GVX forms to be missing. This is largely true, with just three apparent exceptions, where each pair of GVX forms contrast in vowel height. I show them in (17), where I interpret [ɤ] or [ə] as [ɯ].

(17)

High	jɯ (jɤ)	ɯ̃ (ən)	wɯ̃ (wən)
Non-high	ja	ã	wã

It is possible that even these are not all true exceptions. In particular, [ɯ̃] (or [ən] in the original transcription) could be analyzed as a syllabic nasal [n̩] and [wɯ̃] (or [wən]) could be analyzed as [wn̩]. One might recall that in the original transcription of Xu et al. (1988), there is already a syllabic nasal [n̩] in New Shanghai, but it can remain distinct. This is explained in (18).

(18)

Xu et al.	Analysis
[n̩]	[n̩] for syllables without an onset
[ən]	[n̩] for syllables with an onset
[uən]	[wn̩] for syllables with a pre-nuclear glide [w]

The original syllabic nasal [n̩] is used for syllables that do not have an onset, such as [n̩] 'fish'. In contrast, the new [n̩] (from [ən]) is used for syllables that have an onset, such as [fn̩] 'powder', and the new [wn̩] (from [uən]) is used for syllables that have a pre-nuclear glide [w], such as [wn̩] 'lukewarm' and [k^{hw}n̩] 'sleepy'. Therefore, there is no loss of contrast in the current analysis.

Given the fact that few GV or VX forms, if any, contrast in nuclear vowel height, I propose the language-particular generalization in (19) for New Shanghai.

(19) Generalization on vowel height in New Shanghai
  a. In GV, there is a lack of contrast in vowel height.
    e.g. ju = jo, wi = we
  b. In [Ṽ] and [V^{ʔ}], there is a lack of contrast in vowel height.
    e.g. ũ = õ, ĩ = ẽ, u^{ʔ} = o^{ʔ}, i^{ʔ} = e^{ʔ}

The lack of vowel height contrast in [Ṽ] may have a phonetic explanation. In [Ṽ], the velum is lowered, which reduces the space in the oral cavity for tongue height movement.

The result of the generalization, along with G-Spreading and Y-Spreading, is shown in (20), where + indicates an occurring form, (+) indicates an occurring form that may be reducible (see (18)), − indicates a form not found, G indicates a form that is the same as another under G-Spreading, and Y indicates a form that is the same as another under Y-Spreading. Also, H indicates a form that is the same as another owing to lack of contrast in height, and not covered by G or Y. For example, [ji] is the same as [i] under G and the same as [je] under H, and so [ji] is marked with G, not H. Similarly, [jy] is the same as [ɥy] under Y and the same as [jø] under H, and so [jy] is marked with Y, not H. When two GVX forms do not contrast under H, either the higher vowel or the lower vowel can be marked with H. For example, [ũ] and [õ] do not contrast, and I mark [ũ] with H and [õ] with +, because Xu et al. (1988) gives [õ] and not [ũ]. Similarly, [ĩ] and [ẽ] do not contrast, and I mark [ĩ] with + and [ẽ] with H, because Xu et al. (1988) gives [ĩ] and not [ẽ]. When neither form occurs, I mark the one with a lower vowel with H. For example, [jə^{ʔ}] and [ja^{ʔ}] do not contrast under (13) and neither one occurs, and I mark [ja^{ʔ}] with H.

(20)

		0-	j-	w-	ɥ-
[V]	i	+	G	H	–
	e	+	H	+	H
	u	+	H	G	–
	o	+	+	H	H
	y	+	Y	Y	G
	ø	+	H	H	H
	ɯ (ɤ)	+	+	H	–
	a	+	+	+	H
[Ṽ]	ĩ	+	G	–	–
	ẽ	H	H	H	H
	ũ	H	H	G	–
	õ	+	+	H	H
	ỹ	+	Y	Y	G
	ø̃	H	H	H	H
	ɯ̃ (ən)	(+)	H	(+)	–
	ã	+	+	+	H
[Vˀ]	iˀ	+	G	–	+
	eˀ	H	H	H	H
	uˀ	H	–	G	–
	oˀ	+	H	H	H
	yˀ	–	Y	Y	G
	øˀ	H	H	H	H
	ɯˀ	H	–	H	–
	aˀ	+	H	+	H

Occurring + = 25 (26%)
Occurring but possibly reducible (+) = 2 (2%)
G-Spreading G = 9 (9%)
Y-Spreading Y = 6 (6%)
No height contrast H = 41 (43%)
Missing – = 13 (14%)
Total = 96

Most of the missing forms are ruled out by H. The thirteen missing GVX forms are shown in (21).

(21) Unexplained missing forms (13 in all)

ɥi	ɥu	ɥɯ	wĩ	ɥĩ	ɥũ	ɥɯ̃
wiˀ	juˀ	ɥuˀ	yˀ	jɯˀ	ɥɯˀ	

It is interesting to note that nine of the thirteen missing forms involve [y] or its glide form [ɥ], which shows again the lack of productivity of [y] in combining with other sounds.

6.4. FREQUENCY DATA ON SYLLABLES IN MAINSTREAM SHANGHAI

Xu et al. (1988) provide a full syllable inventory for Mainstream Shanghai (although not for New Shanghai). In addition, they provide a complete lexicon of characters (monosyllabic morphemes) divided according to syllable composition and tone. In this section I offer some statistic results on syllable types, sound frequencies, and tone. All statistic data were calculated by me.

6.4.1. **Syllable frequencies and homophone density**

Let us first compare syllable and character counts in Mainstream Shanghai with those of Standard Chinese (SC), which is shown in (22). "T-Syllables" refers to syllable types including tonal contrasts, and "Syllables" refer to syllable types excluding tonal contrasts.

(22) Syllables and characters in Standard Chinese and Mainstream Shanghai

	SC-all	SC-basic	Shanghai
Characters (morphemes)	12,041	2,500	5,943
T-Syllables (incl. tones)	1,334	1,001	705
Syllables (excl. tones)	413	386	487

The 5,943 characters in Shanghai include alternative pronunciations of the same word. For example, there are two tokens for 'big': [du], which is the native pronunciation, and [dᴀ], which is the literary pronunciation influenced by Standard Chinese. Similarly, there are two tokens for 'tooth': the native [ŋᴀ] and the literary [ɦiᴀ]. For such words Xu et al. (1988) have indicated the literary pronunciation with double underline and the native pronunciation with single underline. However, some dual pronunciations are not indicated. For example, 'you' has the literary pronunciation [ni] and the native pronunciation [nõ], but neither was underlined. Some words have more than two pronunciations. For example, there are four tokens for 'moon', which are [ɦioʔ, ȵioʔ, ɦyɪʔ, ȵyɪʔ], and none of them was underlined.

According to my manual count, of the 5,943 characters there are at least 237 that have both a literary and a native pronunciation. If we assume that only common characters have both literary and native pronunciations and that only half of the 5,943 characters are commonly used, then about 8% of the common vocabulary in Shanghai is devoted to literary pronunciation. It would be interesting to know whether similar percentages exist in other dialects.

The influence of Standard Chinese is not limited to literary pronunciation but extends to the way native morphemes are combined to make words. For example, in native Shanghai 'today' is [tɕĩ-tso] 'present-morning' and 'yesterday' is [zoʔ-ɲeʔ] 'yester-sun', but young speakers use [tɕĩ-tʰi] 'present-sky' and [zoʔ-tʰi] 'yester-sky' instead, after the morphology of Standard Chinese. According to the estimate of Xu and Tao (1997), half of the words used by young speakers of Shanghai are modeled after the morphology of Standard Chinese, although most of the component morphemes are still native.

The 5,943 characters seem to be an exhaustive list of all characters used in Shanghai. Many seem to be rare words. Although Xu et al. (1988) do not indicate the frequency of use, it is likely that only about half of them, or 3,000 characters, are commonly used in Shanghai, a number comparable to those in Standard Chinese. In (23) I show the average number of morphemes each syllable represents in Standard Chinese (SC) and Mainstream Shanghai (SH). Values for the basic vocabulary in Mainstream Shanghai are estimated (indicated by *).

(23) Syllables in Standard Chinese and Mainstream Shanghai

	SC-all	SC-basic	SH-all	SH-basic
Characters	12,041	2,500	5,943	3,000*
T-Syllables (incl. tones)	1,334	1,001	705	600*
Syllables (excl. tones)	413	386	487	400*
Tones per syllable	3.2	2.6	1.4	1.5*
Words per syll. (incl. tones)	9.0	2.5	8.4	5.0*
Words per syll. (excl. tones)	29.2	6.5	12.2	7.5*

Excluding tones, Standard Chinese and Shanghai have similar numbers of syllables (386 vs. 400* for the basic vocabulary) and similar numbers of homophones per syllable (6.5 vs. 7.5* for the basic vocabulary). However, when tones are included, Standard Chinese has a lower numbers of homophones per syllable (2.5 vs. 5.0* for the basic vocabulary) than Shanghai, because it has more tones per average syllable than Shanghai (2.6 vs. 1.5* for the basic vocabulary).

As in Standard Chinese, the homophone density in Shanghai is not the same for all syllables. In (24) I list the top twenty syllables in terms of homophone density, and in (25) I show the syllables that have the two lowest numbers of homophone densities.

(24) Top twenty syllables that have the highest homophone density values (in parentheses), excluding tonal contrasts, based on 5,943 characters in Mainstream Shanghai

tɕi (82), tsz (79), zz (75), zən (64), li (61), ɕi (61), ɦi (56), tsən (53), tɕʰi (52), tɕin (50), sz (48), ɦu (47), zø (46), ȵi (45), lin (44), ɦy (43), lu (43), Øi (43), ɦuɛ (42), dʑi (41)

(25) Syllables that have the two lowest homophone density values, excluding tonal contrasts, based on 5,943 characters in Mainstream Shanghai

Density value	Syllables
2 (26 in all)	dʑiᴀ, dʑiã, gɛ, gəʔ, goʔ, gu, ɦm, ɦuã, Øã, Øən, Øn, Øuəʔ, vã, nən, nəʔ, ɕiᴀʔ, fɤ, kʰã, kuəʔ, hᴀ, kʰø, pʰø, pʰoŋ, pø, tʰã, tiᴀ
1 (34 in all)	gã, gã, gᴀʔ, gən, goŋ, guᴀʔ, ɦiã, ØiᴀʔØ, Øuã, Øuᴀʔ, vᴀ, viɔ, ʑiᴀʔ, ʑioŋ, lã, miɤ, ni, noʔ, ŋã, ŋã, ŋo, ŋø, ȵyɪʔ, fəʔ, hã, huəʔ, pʰəʔ, pʰɤ, pʰo, tã, tɕʰiᴀ, tɕʰioŋ, tɕʰyn, tiɤ

As in Standard Chinese, what are often thought to be the most "natural" or "unmarked" syllables, such as [pa] and [ta], are not among those with high homophone density values. Indeed, we do not see [a] at all in the top twenty list. In contrast, we see front rounded vowels [y] and [ø] and the syllabic consonant [z]. We also see that, although some syllables have very high homophone densities, over 10% of the syllables represent just one or two words each, including some syllables that seem to be quite common, such as [gu] and [ni].

6.4.2. **Sound frequencies**

The frequencies of the initial C in Mainstream Shanghai are shown in (26), based on the 487 syllables, excluding tones and homophone density. The sounds are ordered according to frequencies of occurrence.

(26) Frequencies (in parentheses) of the initial C in the inventory of 487 syllables in Mainstream Shanghai, excluding tones
ɦ (41), Ø (38), k (22), h (21), l (21), kʰ (20), m (20), t (19), pʰ (18), b (17), tʰ (17), p (16), d (16), g (15), ts (15), tsʰ (15), s (15), z (15), tɕ (15), n (14), tɕʰ (14), ɕ (14), v (13), dʑ (13), ȵ (13), ŋ (12), ʑ (11), f (10)

The highest frequencies are [ɦ] at 41 and [Ø] at 38. However, [ɦ] and [Ø] are not quite the same as other consonants: they are special symbols added for syllables that do not otherwise have an initial C. In particular, [ɦ] is added for C-less syllables that have a low-register tone and [Ø] is added for C-less syllables that have a high-register tone. For most of the consonants, the frequencies range from 10 to 20. Among the stops, velars seem to be slightly more frequent than labials and dentals.

The frequencies of sounds in GVX forms are shown in (27), ordered according to frequencies. Not every sound is an independent phoneme, though. For example, [ᴀ] and [a] seem to belong to the same phoneme (see analysis above).

(27) Frequencies (in parentheses) of sounds in GVX, based on the inventory of 487 syllables in Mainstream Shanghai, excluding tones
Vowels: i (123), o (72), ᴀ (69), u (62), ə (53), ɔ (37), ø (31), ɛ (31), y (30), ɤ (30), ã (30), ɑ̃ (29), ɪ (23)
Consonants: ʔ (114), n (49), ŋ (29), z (4)

The syllabic consonant [z] has a low frequency because it only occurs with the onsets [ts, ts^h, s, z]. The nasals [n, ŋ] do not contrast with each other, nor do they contrast with [˜]; instead, [n, ŋ, ˜] can be treated as allophones of the same phoneme. If so, their total frequency is 127, which is similar to that of [ʔ].

Most vowels have higher frequencies than onset consonants. The reason obviously is that there are more consonants than vowels. As in Standard Chinese, [i] seems to be the most frequent vowel. However, if we add the frequencies of [ᴀ, a, ɑ], which have merged in New Shanghai (see analysis above), the resulting frequency is 128, which is similar to that of [i].

6.4.3. **Tonal frequencies**

In Shanghai, syllables with a voiceless onset can have two tones, either a rise (LH) or a fall (HL), and syllables with a voiced onset can have a rise only (LH). The rise with a voiceless onset is higher and has a normal voice quality. The rise with a voiced onset is lower and has a murmured voice quality. The relation between onset voicing and tone is summarized in (28).

(28) Onset voicing and tone in Shanghai

	HL	LH	
Voiceless onset	yes	yes	(clear and higher tones)
Voiced onset	no	yes	(murmured and lower tone)

This means that most syllables have either one or two tones. The data in (29) show the details, where "Syllables" refers to syllable count excluding tones, "T-density" refers to the number of tones per syllable, "T-Syllables" refers to syllable count including tones, and "Characters" refers to the number of characters in this set of syllables.

(29) Onset and tonal density in Mainstream Shanghai

Onset	Syllables	T-density	T-Syllables	Characters
Voiceless	70	1	70	580
	181	2	362	2575
Voiced	159	1	159	1854
Sonorant	44	1	44	406
	29	2	58	442
	4	3	12	86
All	487		705	5943

As expected, syllables with voiced obstruent onsets have one tone each. Syllables with voiceless obstruent onsets can have one or two tones each.

Of interest are syllables with a sonorant onset, which can pattern with either a voiced onset or a voiceless onset. As a result, they can take up to three tones each, although such cases are rare. There are just four such syllables, shown in (30).

(30) Syllables with three tones each

	Fall	High-rise	Low-rise
[mi]	'sip (wine)'	'reluctant'	'rice'
[mɛ]	'every'	'pretty'	'coal'
[ȵiã]	'aunt'	'facing up'	'mother'
[mən]	'stuffy'	'fierce'	'door'

Even for these four syllables, there are not many words with fall or high-rise tones. In particular, [mi] has four words for the fall tone and two words for the high-rise tone, [mɛ] has two words for the fall tone and three words for the high-rise tone, and both [ȵiã] and [mən] have one word each for the fall and high-rise tones. In general, sonorant onsets are more likely to go with low-rise than with other tones, as the data in (31) show.

(31) Tonal frequencies of syllables with sonorant onsets

Onset	m	n	l	ȵ	ŋ	All
Fall/high-rise	30	9	43	15	4	101
Low-rise	224	55	353	132	69	833

Of the 934 syllables (including tonal contrast) that have a sonorant onset, 101 have fall or high-rise, and 833 have low-rise. This shows that sonorants mostly pattern with voiced consonants.

Fall and high-rise have been called "upper register" tones and low-rise has been called a "lower register" tone (Yip 1980). The overall frequencies of upper and lower register tones are shown in (32), where "Syllables" do not include tonal contrast and "T-Syllables" do.

(32)

	Syllables	T-Syllables	Characters
Upper tones	251 (51.5%)	432 (61.3%)	3,384 (57%)
Lower tone	236 (48.5%)	273 (38.7%)	2,559 (43%)
All	487	705	5,943

There are about the same number of syllables with voiceless onsets as there are syllables with voiced onsets (including sonorant onsets). However, there are more upper-register tones, because voiceless onsets can go with two possible tones, whereas voiced onsets can go with one. On the other hand, lower-register tones have a higher homophone density than upper-register tones (9.4 vs. 7.8), and so there are only 15% more characters with upper-register tones.

In (33) I show the number of characters for each tone. The tone LHˀ occurs on a syllable with a glottalized vowel. The tone is short and is traditionally listed as a separate tone, but phonologically it has the same behavior as LH.

(33) Character counts for tone types

Tones	Upper tone	Lower tone	All
HL	1,350 (22.7%)	0 (0.0%)	
LH	1,414 (23.8%)	2,154 (36.2%)	
LHˀ	620 (10.4%)	405 (6.8%)	
All	3,384 (56.9%)	2,559 (43.1%)	5,943

The data show that most characters in Mainstream Shanghai have LH and just 22.7% of the characters have HL.

6.5. COMBINATIONS BETWEEN C AND GVX

Mainstream Shanghai has 28 onset Cs and 42 GVX forms. If C and GVX are freely combinable, we expect 1,176 syllables, but just 487, or 40%, are used. In this section I examine what kind of syllables are missing.

Let us begin with the syllabic consonants and [ɚ]. Their complete occurrences are shown in (34), where the top row indicates the onset C and + indicates a form is found.

(34)

	ts	tsʰ	s	z	h	ɦ	Ø
z	+	+	+	+			
ɚ						+	
m						+	+
n							+
ŋ						+	+

The syllabic [ŋ] has merged with [n] in New Shanghai. The syllabic [z] is found only after [ts, ts^h, s, z], which supports the analysis that it is the prolongation of the onset C when there is no vowel. Other syllabic sounds combine with [ɦ, Ø] only, but not [h]. However, it is possible that [h] is not a glottal sound but a velar [x] instead. Also, as discussed earlier, [ɦ, Ø] are used for syllables that do not have an initial C, whereby [ɦ] is added for C-less syllables with a lower tone and [Ø] is added for C-less syllables with a higher tone.

Next we consider the palatals [tɕ, tɕ^h, dʑ, ȵ, ɕ, ʑ]. In (35) I list all occurrences of palatals, in comparison with velars and some dentals. I use + to indicate occurring forms and (+) to indicate limited occurrence. The top row shows the onset C. The first column shows the GVX, where X is the (optional) coda, V is any vowel other than [i, u, y], and [i, u, y] are in the nucleus. In the velar set I have included [x], which Xu et al. (1988) transcribe as [h].

(35) Distribution of palatals in Mainstream Shanghai

	n	ts, ts^h, s, z	tɕ, tɕ^h, dʑ, ȵ, ɕ, ʑ	k, k^h, g, ŋ, x
iX	(+)		+	
jVX			+	
yX			+	
ɥVX			+	
uX	+	+		+
wVX				+
VX	+	+		+

The palatals only occur before [i, y] or their glide forms, whereas [n, ts, ts^h, s, z] and [k, k^h, g, ŋ, x] do not. The only exception is the word [ni] 'you', which is the literary pronunciation of the native [noŋ]. If we ignore [ni], the palatals are in complementary distribution with [n, ts, ts^h, s, z] and [k, k^h, g, ŋ, x]. However, it is not obvious how to reduce the palatals. If we group them with [k, k^h, g, ŋ, x], we are short of a corresponding [ɣ] for [ʑ]. It is relevant to note, though, that [ʑ] has dropped out in New Shanghai. If we group the palatals with [n, ts ts^h, s, z], we are short of a corresponding [dz] for [dʑ]. However, after [ʑ] drops out, there is an extra [z], which could be the counterpart of [dʑ]. The two options are summarized in (36).

(36) Two ways to derive the palatals

a. Deriving palatals from velars

Before [i, y]	tɕ	tɕ^h	dʑ	ȵ	ɕ	(ʑ)
Elsewhere	k	k^h	g	ŋ	x	

b. Deriving palatals from dentals

Before [i, y]	tɕ	tɕʰ	dʑ	ȵ	ɕ	(ʑ)
Elsewhere	t	tʰ	z	n	s	

The choice between the two options is not very clear, and will be left open.

The final set of major gaps is shown in (37), where the top row shows the onset C and the first column shows the GVX.

(37)

	Labial	Dental	Palatal	Velar
uX	+	+		+
wVX				+
yX		(+)	+	
ɥVX			+	

Although labials, dentals, and velars can all combine with [u], only velars can combine with [w]. Also, only palatals can combine with [y, ɥ], except one syllable [ly], which represent thirteen characters (morphemes).

6.6. SUMMARY

As in Standard Chinese, the list of syllables used in Shanghai is a small fraction of all conceivable combinations; and most of the non-occurring syllables can be accounted for in terms of phonological generalizations. However, whereas Rhyme-Harmony, Merge, G-Spreading, and Y-Spreading rule out most missing forms in Standard Chinese, the main constraint in Shanghai is the lack of contrast in the height of the nuclear vowel in GV, [Ṽ] and [Vˀ] (see (19)).

Of special interest is the fact that Shanghai does not have true codas. In particular, [VN] and [Vʔ] rhymes in traditional transcriptions can be analyzed as [Ṽ] (a nasalized vowel) and [Vˀ] (a glottalized vowel) respectively. Thus, all Shanghai syllables are (C)V, where C is an optional onset and V can be a syllabic consonant. The (C)V-only syllable inventory has an interesting consequence: most Shanghai syllables cannot hold onto their lexical tones but often lose them, a property not found in many Chinese dialects. I discuss this effect in Chapter 7.

7

Syllable and tone

When a Chinese syllable is pronounced alone, it has a tone, often called the "citation tone." In many dialects the citation tone stays the same whether a syllable is pronounced alone or with other syllables. Examples from Standard Chinese are shown in (1).

(1)	Surface	H	H	H	LH	HL	H	HL	LH
	Citation	H	H	H	LH	HL	H	HL	LH
		san	pei	san	p^han	sz	pei	sz	p^han
		'three cups'		'three plates'		'four cups'		'four plates'	

However, not all Chinese languages behave this way. In particular, Shanghai Chinese differs from most others in two striking respects. First, if we ignore variations conditioned by onset voicing and glottalized vowels, Shanghai (the variety that Xu et al. (1988) call New Shanghai) has just two citation tones, LH and HL, the fewest among Chinese dialects. Second, non-initial syllables in a tonal domain lose their citation tones in Shanghai, where a tonal domain is a domain with initial stress (Duanmu 1999). The tonal patterns in Shanghai are schematically represented in (2), where T refers to any citation tone, 0 refers to lack of tone, and a hyphen indicates a syllable boundary.

(2)	Citation		Surface
	HL-T-T-. . .	→	H-L-0-. . .
	LH-T-T-. . .	→	L-H-0-. . .
	LH-. . .-T	→	L-. . .-H

The last pattern has limited use. It occurs when the first syllable has a voiced onset and a glottalized vowel. Let us focus on the first two patterns, which are more productive. They are exemplified in (3) with the same expressions as those in (1). The surface tones are shown above citation tones.

(3)

H	L	H	L	L	H	L	H	Surface
HL	HL	HL	LH	LH	HL	LH	LH	Citation
se	pe	se	pø	sz	pe	sz	pø	
‘three cups’		‘three plates’		‘four cups’		‘four plates’		

It can be seen that while citation tones are stable in Standard Chinese, they easily split in Shanghai. A longer example is shown in (4), in fairly broad transcription, where SC refers to Standard Chinese and | indicates a boundary between tonal domains in Shanghai.

(4) Tonal patterns in Shanghai and Standard Chinese (SC)

Surface	L-H	0 |	L-H |	H-L	0 |	L-H
Citation	LH-LH	LH	LH-LH	HL-HL	LH	LH-HL
Shanghai	ku-po^{ʔ}	lu	la^{ʔ}-la^{ʔ}	t^{h}i-se	lu	pã-pi
Gloss	Gubei	road	be-at	Tianshan	road	vicinity
SC	ku-pei	lu	zai	t^{hj}an-ʂ̩an	lu	p^{h}aŋ-p^{j}an
Citation	L-L	HL	HL	H-H	HL	LH-H
Surface	LH-L	HL	HL	H-H	HL	LH-H

‘Gubei Road is in the vicinity of Tianshan Road.’

A striking difference is again seen in the stability of citation tones. In Standard Chinese, citation tones remain unchanged, except for one rule, which changes L to LH before L, as seen on the first syllable. In contrast, citation tones are lost in Shanghai unless they occur in the initial position of a domain; in addition, each surviving citation tone is split between the first two syllables of a domain.

To account for the difference, some linguists propose that there is a typological difference between Shanghai and Standard Chinese. For example, Yue-Hashimoto (1987) suggests that Shanghai has left-dominant tonal domains but Standard Chinese does not. Similarly, Chen (2000b) suggests that tonal domains in Shanghai are determined by left-headed stress whereas those in Standard Chinese are not. However, the typological approach in effect restates the difference and offers no explanation why Shanghai behaves one way and Standard Chinese behaves another way.

What the typological approach has overlooked is an independent correlation between tone and syllable structure: Dialects that are like Shanghai in tonal behavior (with unstable citation tones) have no diphthongs or true codas, while dialects that are like Standard Chinese in tonal behavior (with stable citation tones) have diphthongs and/or true codas (Duanmu 1990, 1993, 1999). In other words, Shanghai only has “simple rhymes” while Standard Chinese has many “complex rhymes." Given the difference in rhyme structure between Shanghai and Standard Chinese, it is possible to explain their difference in tonal behavior.

7.1. SIMPLE AND COMPLEX RHYMES

I use "complex rhymes" to refer to VC and VG, where VG is a diphthong, and "simple rhymes" to refer to other rhymes. Examples are given in (5) and (6).

(5) Complex rhymes

	English	Standard Chinese
[VC]	[bæt] *bat*	[pan] 'half'
[VG]	[bai] *buy*	[pai] 'white'

(6) Simple rhymes

	English	Standard Chinese
[V]	[ðə] *the*	[lə] an aspect marker
[V:]	[bi:] *bee*	[mi:] 'rice'

A simple rhyme can also be [C], such as [n] '*n* (reduced *and*) in English or [C:], such as [n:] 'fish' in Shanghai Chinese.

Of interest for the present discussion is the fact that, while Standard Chinese (SC) has many complex rhymes, Shanghai has lost all of them. Some examples are shown in (7).

(7)

SC	Shanghai	
[mai]	[ma:]	'buy'
[lau]	[lo:]	'old'
[lai]	[le:]	'come'
[lan]	[le:]	'blue'
[fei]	[fi:]	'fly'
[kəu]	[kɤ:]	'dog'
[faŋ]	[fã]	'square'

Some Shanghai rhymes have been transcribed as VC, such as [paʔ] 'eight' and [ɕin] 'new', which can be analyzed as [pa^{ʔ}] and [ɕĩ] (or [ɕn̩]) respectively (see Chapter 6). Therefore, Shanghai has no real complex rhymes.

If a language has neither complex rhymes nor vowel length contrast, then its syllables have no inherent weight: they can be heavy or light, depending on the metrical environment. This is the case in Shanghai, where syllables are long when spoken in isolation or in a stressed position—such as the first syllable of a disyllabic word or compound—but otherwise the syllables are short. This will be a crucial piece of the puzzle in the analysis of tone.

7.2. RHYME STRUCTURE AND TONE LOSS

The relation between syllable structure and tone is mediated by stress, whereby heavy syllables have stress and stressed syllables can carry tone. The principle that governs syllable weight and stress is stated in (8), and the principle that governs stress and tone is stated in (9).

(8) Weight-Stress Principle
Stressed syllables are heavy (long) and unstressed syllables are light (short).

(9) Tone-Stress Principle
Stressed syllables can be accompanied by a lexical tone (pitch accent). Unstressed syllables are not accompanied by a lexical tone (pitch accent).

The Weight-Stress Principle has been proposed in various forms in the literature (e.g. Prokosch 1939, Kager 1989, Prince 1990, Hammond 1999). It follows if a heavy syllable has two moras and forms a moraic foot, and each foot has stress. The effect of the Weight-Stress Principle is illustrated in (10), with a compound meaning 'wood-head (wood)' in Standard Chinese. An apostrophe indicates stress and an asterisk indicates a bad form.

(10) Effect of the Weight-Stress Principle (WSP)

a.	[múu.tóu]	No violation
b.	*[múu.tou]	Violation by [tou]
c.	[múu.to]	No Violation
d.	*[múu.tó]	Violation by [tó]

In (10a) both syllables are stressed and heavy (long), and the Weight-Stress Principle is not violated. In (10b) the second syllable is unstressed but heavy, which violates the Weight-Stress Principle. In (10c) the stressed syllable is heavy and the unstressed syllable is light (short), and the Weight-Stress Principle is not violated. In (10d) the second syllable is stressed but light, so the Weight-Stress Principle is violated.

The Tone-Stress Principle has also been proposed in various forms in the literature (e.g. Liberman 1975, Clements and Ford 1979, Pierrehumbert 1980, Goldsmith 1981). It governs the use of lexical tones or "pitch accent," but not the use of "boundary tones." In Chinese, the Tone-Stress Principle is evidenced by the fact that unstressed syllables lose their lexical tones. In English, it is evidenced by the fact that only stressed syllables are assigned a pitch accent.

In Standard Chinese, many syllables are inherently heavy (with two slots in the rhyme), because they contain complex rhymes: either VG (a diphthong) or VC. The frequencies of rhyme types of Standard Chinese are shown in (11) based on the 2,500 most common characters.

(11) Rhyme types in Standard Chinese

V	120	(31%)	e.g. [ma] 'scold'
VC	169	(44%)	e.g. [man] 'slow'
VG	96	(25%)	e.g. [mai] 'sell'
All	385	(100%)	

Of the 385 syllables (excluding tonal contrasts), 69% have VG or VC rhymes and are inherently heavy. By the Weight-Stress Principle, these syllables always have stress. In addition, by the Tone-Stress Principle, they can retain their own tones. Therefore, most syllables in Standard Chinese show stable tones, whether they are spoken in isolation or together with other syllables. Interestingly, the African language Bench of Ethiopia shows the same property: Bench syllables are mostly CVC and Bench tones are mostly stable (Wedekind 1985: 883–4).

Next we consider syllables with V rhymes. For convenience let us call them CV syllables, although the onset C is optional and the rhyme can be a syllabic consonant. When a CV syllable occurs alone in a monosyllabic content word (e.g. noun, verb, or adjective), it is realized as [CV:], such as [ma:] 'scold', because content words usually have stress and a stressed syllable must be heavy. Therefore, CV syllables can keep their own tones.

But when a CV syllable occurs with other syllables, it may or may not have stress. In some positions, such as the first word of a compound or the object of a verb, the CV syllable has compound or phrasal stress, is lengthened to [CV:], and can retain its tone. In other positions, the CV syllable has no compound or phrasal stress. Will it remain long and stressed and retain its tone, or will it become short and unstressed and lose its tone? The fact in Standard Chinese is that CV syllables generally keep their stress and tones, as CVG and CVC syllables do. The question is why.

For CV syllables to surface as short and unstressed, there needs to be a de-stressing rule that requires content words without phrasal stress to be light. If this rule is applied generally, it causes a problem with CVG and CVC syllables: If these syllables are de-stressed, they would undergo rhyme reduction and coda deletion, as expected by the Weight-Stress Principle and as commonly happens to unstressed syllables (Lin and Yan 1988). This would cause a massive loss of segmental information or phonemic contrasts. Obviously, Standard Chinese favors the preservation of phonemic contrast, and the presence of many CVG and CVC syllables seems to have prevented a general use of the de-stressing rule.

Let us now consider Shanghai, in which all syllables are CV and no syllable is inherently heavy (see Chapter 6). When a monosyllabic content word occurs alone, it has stress, as content words generally do; the rhyme is lengthened to [V:], and the stress enables the syllable to retain its tone. The same is true for syllables that receive phrasal stress, such as the first word of a disyllabic compound, where the rhyme is indeed long (Zhu 1995).

When a Shanghai syllable occurs in an unstressed position, such as the second syllable of a disyllabic compound, it is short and unstressed and loses its tone. The question is why this occurs in Shanghai but not in Standard Chinese. A common view is that different languages have different rules: Shanghai has a de-stressing rule and Standard Chinese does not. If so, their difference in stress and tone loss needs no further explanation. In addition, their difference in stress and tone loss would have nothing to do with their difference in syllable structure.

However, there is a better analysis, in which Shanghai and Standard Chinese are governed by the same principles and their difference in stress and tone loss is derivable from their difference in syllable structure. In particular, both languages are subject to the need to preserve segmental contrasts and the influence of de-stressing. In Standard Chinese, however, a general de-stressing rule would lead to a massive loss of segmental contrasts. Therefore, de-stressing cannot be used as a general rule (although it sometimes applies to expressions that are frequently used, such as [muu.tou] → [muu.to] 'wood-head (wood)', [mʲan.hʷaa] → [mʲan.hʷə] 'cotton-flower (cotton)', and [ʈʂʰʷən.tʰʲæn] → [ʈʂʰʷən.tʰʲə̃] 'spring-day (spring)'). In contrast, in Shanghai there is no contrast between [CV] and [CV:] and the use of the de-stressing rule causes no loss of segmental contrast. Therefore, the de-stressing rule indeed applies generally.

The above analysis can be presented in terms of constraint interaction (Prince and Smolensky 1993). To simplify the discussion, let us assume that the Weight-Stress Principle and the Tone-Stress Principle are always satisfied. To account for Standard Chinese and Shanghai, we need three other constraints, defined in (12), where A >> B means A is ranked above B.

(12) Constraints on syllable and stress
  a. Segmental-Contrast: Preserve segmental contrasts.
  b. De-Stress: De-stress syllables that do not receive phrasal stress.
  c. Anti-Allomorphy: Avoid changing the phonological form of a morpheme.
  Ranking: Segmenal-Contrast ≫ De-Stress ≫ Anti-Allomorphy

We have discussed the first two constraints. The third one is proposed by Burzio (1996). It says, for the present purpose, that we should use the same syllable structure for a word whether it occurs alone or in a compound. The lack of De-Stress in Standard Chinese means that it is ranked below Segmental-Contrast. The presence of De-Stress in Shanghai means that it is ranked above Anti-Allomorphy. Thus, the three constraints are ranked as shown.

The analysis of Standard Chinese and Shanghai is shown in (13), where * indicates a violation, √ indicates a good option, and alternating syllabic forms are linked with a hyphen.

(13) Analysis of syllable and stress in Standard Chinese (SC) and Shanghai (SH)

			S-Contrast	De-Stress	A-A
√	a. SC without De-Stress	[CV:] [CVC] [CVG]		*	
	b. SC with De-Stress	[CV:]-[CV] [CVC]-[CV] [CVG]-[CV]	*		*
	c. SH without De-Stress	[CV:]		*	
√	d. SH with De-Stress	[CV:]-[CV]			*

In (13a), there is no syllable alternation in Standard Chinese; the result satisfies Segmental-Contrast (S-Contrast) and Anti-Allomorphy (A-A) but violates De-Stress. In (13b), there is syllable alternation in Standard Chinese; the result satisfies De-Stress but violates Segmental-Contrast and Anti-Allomorphy. Because Segmental-Contrast is more important than De-Stress, Standard Chinese uses the option in (13a). In (13c), there is no syllable alternation in Shanghai; the result satisfies Segmental-Contrast and Anti-Allomorphy but violates De-Stress. In (13d), there is syllable alternation in Shanghai; the result satisfies

Segmental-Contrast (because there is no contrast between [CV] and [CV:]) and De-Stress, but violates Anti-Allomorphy. Because Anti-Allomorphy is less important than De-Stress, Shanghai uses the option in (13d). In fact, if Anti-Allomorphy only prohibits changes in contrastive features but allows changes in non-contrastive features, (13d) does not violate Anti-Allomorphy either.

A couple of interesting observations can be made. First, the grammar treats De-Stress as a general rule: it either applies to all content words or not at all. In other words, it does not target specific syllable types. If the latter were the case, De-Stress would change [CV:] to [CV] (as it happens in Shanghai) but not CVG or CVC. To ensure that all content words are treated the same way, one might assume a constraint that requires the homogeneity of stress patterns in content words. Second, syllable simplification is a continuous trend in Chinese. It is reasonable to assume that before Shanghai lost CVG and CVC rhymes it also lacked de-stressing and tone loss, and that the process Shanghai has undergone is the one Standard Chinese is undergoing or will undergo. It would be interesting to know how the evolution takes place, i.e. how VG and VC get lost and the De-Stress rule comes into play. I hypothesize that rhyme simplification takes place first, and De-Stress follows soon after. Rhyme simplification is likely to be a long process, and to affect different rhymes in different ways. Some possible cases are shown in (14)–(16).

(14) [Vp, Vt, Vk] → [Vʔ] → [V^{ʔ}] → [V]

(15) [an, aŋ] → [æn, aŋ] → [æ̃, ã]

(16) [ən, əŋ] → [ən] → [ə̃]

Cantonese still has [Vp, Vt, Vk], which have become [Vʔ] or [V^{ʔ}] in Shanghai, which may further reduce to [V], which is the case in Standard Chinese. In fact, [Vʔ] and [V] alternate in Xiamen, whereby [Vʔ] occurs in final position and [V] occurs in nonfinal positions. The change of [an, aŋ] → [æn, aŋ] has occurred in Standard Chinese, where [æn, aŋ] are often realized as [æ̃, ã]. They can further become [æ, ã] or [e, ã]. For example, [mæn] 'slow' in Standard Chinese is [me] in Shanghai, and [maŋ] 'busy' in Standard Chinese is [mã] in Shanghai. The contrast between [ən] and [əŋ] is still kept in Standard Chinese, but the pair has merged to [ən] in many dialects, such as Chengdu. In Shanghai, [ən] is often realized as [ə̃].

Having discussed the relation between syllable structure and tone loss, let us consider syllable structure, in particular rhyme structure, and tone split.

7.3. RHYME STRUCTURE AND TONE SPLIT

I use "tone split" to refer to the split of a citation contour tone into level tones, where one of the level tones is shifted to another syllable. The recipient is typically an unstressed syllable that does not have its own tone. In Standard Chinese there is a general lack of tone split, whereas in Shanghai tone split is quite active. We can see the difference in heavy-light disyllables, shown in (17) and (18), where [lə] or [ləˀ] is a grammatical aspect marker and 0 indicates the lack of an underlying tone. Both L-H (or LLH) and LH are rising tones, but L-H is an extra long tone which starts as L and ends in H, whereas LH is a simple rise. In Standard Chinese LH is a high rise and L-H is a low rise. In Shanghai L-H (often transcribed as LH) can be a high rise (with a voiceless onset, as in [see]) or a low rise (with a voiced onset, as in [zee]).

(17) Lack of tone split in Standard Chinese

Surface	H	(L)	LH	(L)	L	H	HL	(L)
Citation	H	0	LH	0	L-H	0	HL	0
	fei	lə	lai	lə	mai	lə	mai	lə
	'flew'		'came'		'bought'		'sold'	

(18) Tone split in Shanghai

Surface	H	L	L	H	L	H
Citation	HL	0	L-H	0	L-H	0
	fii	ləˀ	see	ləˀ	zee	ləˀ
	'flew'		'broke'		'earned'	

In Standard Chinese, citation tones stay on the original syllables, except when the citation tone is L-H, to be discussed later. A toneless syllable is generally realized with a low pitch, which can be interpreted either as L or as 0, which I represent as (L). In Shanghai, the second half of a citation tone is always shifted to the following light syllable.

Let us consider why citation tones in Standard Chinese are stable. For example, why can HL split in Shanghai but not in Standard Chinese? We begin by asking what the underlying tones are. In Standard Chinese, they are H, LH, L, and HL (Duanmu 2000). H, LH, and HL are not controversial, since they are realized as such in both monosyllables and disyllables. The H after the underlying L can be attributed to a polarity requirement (to be discussed below).

In Shanghai, if we ignore variations conditioned by onset voicing and vowel glottalization, Shanghai has two citation tones, HL and L-H. There are two possible analyses of the underlying tones. I shall discuss both but not make a choice between them. In the first analysis the underlying tones are H and L, as shown in (19).

(19) Shanghai analysis 1: underlying tones are H and L

Surface	H	L	L	H
Citation	HL	0	L-H	0
Underlying	H	0	L	0
	fii	ləˀ	lee	ləˀ
	'flew'		'came'	

A polarity requirement would provide L for an initial H and H for an initial L (i.e. realizing H as HL and L as L-H). The merit of this analysis is that it minimizes the underlying tones. In addition, the underlying L is realized in the same way as that in Standard Chinese. A question for this analysis is why the citation form of the underlying H is HL, instead of H, as it is in Standard Chinese. The reason might be the need to preserve tonal contrast. In Standard Chinese there is already an underlying HL; for H to be realized as HL would lose a contrast between underlying HL and H. In Shanghai, on the other hand, there are just two underlying tones, H and L. Therefore, there is no loss of contrast whether H is realized as H or HL.

In the second analysis of Shanghai, the underlying tones are the same as the surface tones, namely, HL and LH, as shown in (20).

(20) Shanghai analysis 1: underlying tones are HL and LH

Surface	H	L	L	H
Citation	HL	0	L-H	0
Underlying	HL	0	LH	0
	fii	ləˀ	lee	ləˀ
	'flew'		'came'	

The merit of this analysis is that the underlying tones are "transparent," in the sense that they are the same as surface tones. As a result, there is no need to assume the polarity requirement. A question for this analysis is why underlying HL and LH are realized differently from those in Standard Chinese. The answer may again lie in contrast. Since Shanghai has just two underlying tones, whether HL is realized as HL or H, it is still distinct from L or LH. In Standard Chinese, HL cannot be realized as H, because there is a separate H already. Similarly, LH cannot be realized as L in Standard Chinese, because there is a separate L already.

The above analysis can be formulated in terms of constraint interaction. For illustration, let us assume that the underlying tones in

Shanghai are HL and LH. In addition, let us assume that Standard Chinese and Shanghai follow the same set of requirements. The constraints and their ranking are given in (21).

(21) Tonal constraints
  a. Tonal-Distinction: A stressed syllable must maintain its tonal distinction.
  b. Polarity: A foot-initial tone is followed by an opposite tone.
  c. Simple-Tone: Avoid contour tones.
  Tonal-Distinction ≫ Polarity ≫ Simple-Tone

Tonal-Distinction is sensitive to the number of tonal categories in a language. For example, if a language has two categories, H and L, H can be realized as HL without violating Tonal-Distinction, since HL is still distinct from L. However, if a language has three categories, H, HL, and L, then H cannot be realized as HL, otherwise the distinction between H and HL will be lost. Polarity requires foot-initial H to be followed by L and foot-initial L to be followed by H. It has been reported in African tone languages (Newman 1997), and can probably be related to the obligatory contour effect (Leben 1971). Finally, Simple-Tone can be related to articulatory effort, because a simple tone (H or L) may be easier to pronounce than a complex tone (HL or LH).

The ranking of the constraints is based on two facts. First, there is no output that violates Tonal-Distinction, while there are outputs that violate Polarity or Simple-Tone. Therefore, Tonal-Distinction is top-ranked. Second, there are cases where Polarity is ranked over Simple-Tone (see below).

First, let us examine monosyllables in Standard Chinese. There are four tones, which are shown in (22)–(25).

(22) Analysis of H in Standard Chinese

	H /fei/ ‘fly’	Tonal-Distinction	Polarity	Simple-Tone
√	H fei		*	
	HL fei	*		*

(23) Analysis of LH in Standard Chinese

	LH /lai/ 'come'	Tonal-Distinction	Polarity	Simple-Tone
	L Lai	*	*	
√	LH lai			*

(24) Analysis of L in Standard Chinese

	L /mai/ 'buy'	Tonal-Distinction	Polarity	Simple-Tone
√	L-H maai			
√	L mai		*	
	LH mai	*		*

(25) Analysis of HL in Standard Chinese

	HL /mai/ 'sell'	Tonal-Distinction	Polarity	Simple-Tone
√	HL mai			*
	H mai	*	*	

In (22), H violates Polarity while HL violates Tonal-Distinction (because HL would confuse with a real HL) and Simple-Time. The fact that H is the actual form follows from the given ranking. In (23), L violates Tonal-Distinction (because it would confuse with a real L) and Polarity, whereas LH violates Simple-Tone. The fact that LH is the actual form again follows from the given ranking. The case in (25) is similar to that in (23).

The case in (24) deserves comment. The best form is L-H (or LLH), where the syllable is lengthened to an extra long one, shown as [maai]. If [maai] counts as two syllables [maa.i], each syllable has a level tone, and there is no violation of Simple-Tone. Now besides L-H, L can also be used, even though L does not rank as the best in (24). The reason is that an extra-long syllable is available only in pre-pause position. This means that there is another constraint that governs syllable lengthening. The effect of the constraint is that L-H is preferred in pre-pause position but L is used otherwise. For simplicity I do not discuss how to incorporate the additional constraint.

Next we examine monosyllables in Shanghai. There are two cases, which are shown in (26) and (27).

(26) Analysis of HL in Shanghai

	HL /fii/ 'fly'	Tonal-Distinction	Polarity	Simple-Tone
	H fii		*	
√	HL fii			*

(27) Analysis of LH in Shanghai

	LH /lee/ 'come'	Tonal-Distinction	Polarity	Simple-Tone
	L lee		*	
	LH lee			*
√	L-H Lee-e			

In (26) H violates Polarity and HL violates Simple-Tone. The fact that HL is used shows that Polarity ranks above Simple-Tone, as we assume. The form in (27) needs some comments. The output is not a simple LH, but an extra long L-H. As far as the constraints are concerned, L-H is better than LH. However, the question is why the HL in (26) is not realized as an extra-long H-L, which can better satisfy the constraints. The answer may lie in articulatory effort. Simple-Tone assumes that a contour tone is harder to make than a simple tone. However, it is likely that HL is easier to make than LH, which is probably why HL is realized as it is but LH is realized as L-H. If so, the constraints on articulatory effort could be {Avoid-LH, Avoid-Extra-Long-Syllable} ≫ Avoid-HL, where {} indicates equally ranked constraints. For simplicity I omit such details.

Let us now consider heavy-light disyllables. The analysis of HL and LH in Shanghai is shown in (28) and (29).

(28) Analysis of HL in Shanghai

	HL 0 /fii ləʔ/ 'flew'	Tonal-Distinction	Polarity	Simple-Tone
√	H L fii ləʔ			
	H H fii ləʔ		*	
	HL 0 fii ləʔ			*

(29) Analysis of LH in Shanghai

	LH 0 /lee ləʔ/ 'came'	Tonal-Distinction	Polarity	Simple-Tone
√	L H lee ləʔ			
	LH H lee ləʔ		*	*
	LH 0 lee ləʔ		*	*

In (28), H-L satisfies all the constraints, H-H violates Polarity, and HL-0 violates Simple-Tone. Therefore, H-L is the best form. In (29), L-H satisfies all the constraints, but LH-H and LH-0 violate Polarity and Simple-Tone. Therefore, L-H is the best form.

Next we consider the analysis of heavy-light disyllables in Standard Chinese, which are shown in (30)–(33).

(30) Analysis of H in Standard Chinese

	H 0 /fei lə/ 'flew'	Tonal-Distinction	Polarity	Simple-Tone
√	H L fei lə			
	H H fei lə	*		
	HL 0 fei lə	*		*

(31) Analysis of LH in Standard Chinese

	LH 0 /lai lə/ 'came'	Tonal-Distinction	Polarity	Simple-Tone
	L H lai lə	*		
√	LH 0 lai lə			*

(32) Analysis of L in Standard Chinese

	L 0 /mai lə/ 'bought'	Tonal-Distinction	Polarity	Simple-Tone
√	L H mai lə			
	L 0 mai lə		*	
	LH 0 mai lə	*		*

(33) Analysis of HL in Standard Chinese

	HL 0 /mai lə/ 'sold'	Tonal-Distinction	Polarity	Simple-Tone
√	HL 0 mai lə			*
	H L mai lə	*		

In (30), H-L satisfies all the constraints. H-H violates Polarity. HL-0 violates Tonal-Distinction, because it is the pattern used by underlying /HL-0/. In addition, HL-0 violates Simple-Tone. Thus, the best choice is H-L.

In (31), L-H violates Tonal-Distinction, because it is the pattern used by underlying /L-0/. On the other hand, LH-0 violates Simple-Tone. Since Tonal-Distinction ranks above Simple-Tone, LH-0 is a better choice. For simplicity, we did not consider other candidates, such as LH-H or LH-L. These can be ruled out with an additional constraint that prohibits adding surface tones that do not relate to underlying tones.

In (32), L-H satisfies all the constraints. In contrast, L-0 violates Polarity, and LH-0 violates Simple-Tone and Tonal-Distinction (confusing with underlying /LH-0/).

In (33), H-L violates Tonal-Distinction, because it is the pattern used by underlying /H-0/. On the other hand, HL-0 violates Simple-Tone. Since Tonal-Distinction ranks above Simple-Tone, HL-0 is a better choice. Again, we did not consider other candidates, such as HL-H or HL-L, which can be ruled out with an additional constraint that prohibits surface tones that do not relate to underlying tones.

7.4. RHYME STRUCTURE AND TONAL INVENTORY

With just HL and LH, Shanghai has the smallest tonal inventory among all Chinese dialects. Let us consider why this is the case. The question might seem unnecessary. For example, it seems natural that some languages have more tones than others: Shanghai just happens to be one that has the fewest tones, and there is nothing to explain.

However, there is a problem. Until quite recently, Shanghai had many more tonal patterns. For example, excluding the influence of onset voicing and glottal vowels, there are six disyllabic patterns in what is called Old Shanghai, a variety still spoken by people who are in their 80s or older (Xu and Tao 1997). The patterns are shown in (34), based on Xu et al. (1988: 60), along with those in the currently dominant variety called New Shanghai.

(34) Disyllabic patterns in Old Shanghai: H-HL, H-H, H-L, L-H, LH-H, LH-HL
Disyllabic patterns in New Shanghai: H-L, L-H

It is quite striking that in a matter of a few decades Shanghai has lost all but two of its disyllabic tonal patterns, whereas Standard Chinese has not lost any.

Recall that New Shanghai also has an extensive tone deletion rule. This means that there are fewer tonal patterns in New Shanghai than in Standard Chinese. To see the effect, consider disyllabic compounds in a language that has four lexical tones, H, LH, L, and HL. As shown in (35), without tone deletion from the second syllable, there are sixteen tonal patterns, but with tone deletion there are just four.

(35) Disyllabic patterns with four lexical tones H, LH, L, and HL
No tone deletion: H-H, H-LH, H-L, H-HL; LH-H, LH-LH, LH-L, LH-HL; L-H, L-LH, L-L, L-HL; HL-H, HL-LH, HL-L, HL-HL
With tone deletion from the second syllable: H-0, LH-0, L-0, HL-0

Since Shanghai has tone deletion, it has far fewer disyllabic patterns than would otherwise be the case, and so there ought to be a greater need for it to preserve its tonal inventory, if it is important to maintain lexical contrast. However, the actual result is the opposite: dialects that do not have tone deletion, such as Standard Chinese and Cantonese, have maintained their tonal inventory, whereas Shanghai, which has tone deletion, has lost all but two of its tonal patterns.

One might suggest that perhaps the loss of tonal patterns in Shanghai is related to the fact that it has no complex rhymes. For example, Woo (1969) proposes that a simple tone H or L can be carried by a CV syllable, but a contour tone HL or LH must be carried by a long syllable, such as CVV or CVC. Since Shanghai only has CV, perhaps it cannot display too many tonal contrasts. This suggestion has some problems. First, a CV syllable is lengthened to CVV when it has stress, which happens in Shanghai, and CVV can carry either

HL or LH. Second, monosyllables in Shanghai do carry HL and LH, instead of H and L. Third, CV syllables in Standard Chinese can contrast four tones, such as [maa] H 'mother', [maa] LH 'hemp', [maa] L 'horse', and [maa] HL 'scold'. Therefore, there should be no problem for Shanghai to contrast at least four tones.

So what is the culprit responsible for the massive loss of tonal patterns in Shanghai? I suggest that, paradoxically, it is the lack of tonal patterns itself. In particular, there is a relation between frequency of use and reduction (Fidelholtz 1975, Hooper 1976b, Bybee 2001). For example, the first vowel is reduced to [ə] in *astronomy* but not in *gastronomy*, because the former is a frequent word but the latter is not. Similarly, the second vowel is reduced to [ɚ] in *information* but not in *importation*, again because the former is a frequent word but the latter is not. I suggest that the same is the case for tonal reduction, whereby high frequency patterns are more likely to undergo reduction or loss of contrast. For illustration, consider two languages, one having sixteen disyllabic tonal patterns and one having four. In the first, each pattern is used at one fourth the average frequency of that in the second language. Therefore, it is more likely for tonal reduction to occur in the second language, the one that already has fewer tonal patterns. This is the same analysis I used to account for syllable loss in Chinese and the lack of it in English (Chapter 5).

If this analysis is correct, the loss of tonal patterns in Shanghai is ultimately triggered by the loss of complex rhymes: the loss of VG and VC rhymes allows the de-stressing rule to apply, which leads to extensive loss of tones, which leads to a decrease of (disyllabic) word-tone patterns, which leads to an increase in the frequency of the word-tone patterns, which triggers reduction and loss of contrasts in the tonal patterns.

7.5. SUMMARY

I have shown how syllable structure—whether a language has CV syllables only or whether it has many CVX syllables (CVG and CVC)—can influence a number of other phonological properties, in particular the ability to carry stress and tone, the presence or absence of the de-stressing rule, whether contour tones split, and reduction of the tonal inventory. If a tone language has CV syllables only, it is likely to show tone split, where contour tones break into level tones; in addition, its tonal inventory is likely to reduce. In contrast, if a tone language has many CVX syllables, its tones are usually stable. I have shown that Shanghai is an example of the former and Standard Chinese is an example of the latter.

I have also shown that, to account for differences between languages, we do not need to assume different rules, constraints, or constraint rankings. Instead, we can use the same constraints that are ranked the same way, once we understand how a difference in syllable structure can lead to differences in other aspects of phonology.

8

English I: the maximal syllable size

In this chapter I examine the maximal syllable size in English, using the CELEX lexicon of English (Baayen et al. 1993). I shall argue for an analysis in which the maximal syllable size is CVX. Let me outline the proposal with a monosyllabic word shown in (1), where S is an initial C that is usually (but not always) [s] and X is V or C.

(1)	Maximal string	CCCVXCCC	C = consonant or glide; V = vowel
	Restricted string	SCCVXCTT	T = [s, z, t, d, θ]
	Syllable analysis	S[CRVX]CTT	CR = a possible complex sound

English allows up to three Cs before a vowel, such as [spl] in *splash*, and up to four Cs after a short vowel (V), such as [ksts] in *texts* and [ksθs] in *sixths*, or three Cs after a long vowel (VV), such as [lds] in *fields* and [nts] in *counts*. In principle, therefore, English allows a maximal string of CCCVXCCC for a monosyllabic word. However, initial CCC and final CCCC are highly restricted. In particular, in the initial CCC the first C is [s] and the last C is an approximant (indicated as R in the restricted string). In addition, the final CC after VXC must be one of the coronals [s, z, t, d, θ], although the C right after VX can be any consonant. The restricted string is therefore SCRVXCTT. I argue that both S and CTT can be accounted for by morphology and need not be included in a syllable. In addition, the CR can form a complex sound (see Chapter 2). Therefore, the maximal syllable size is CVX.

There are two arguments for the present analysis. First, in morpheme-medial (i.e. non-edge) positions the maximal syllable is CVX. Second, extra consonants at word edge can indeed be accounted for by morphology. Both arguments will be presented. Apparent exceptions to the CVX limit, such as those noted by Borowsky (1989), will be shown to be accountable by the notion of "complex sounds."

The phonetic symbols I use are conventional, with a couple of minor exceptions. I use [ɝ] for the stressed retroflex vowel and [ɚ] for the unstressed one in American English, instead of [r̩] or [r]. For example, *fur* is [fɝ] (not [fr̩]), *worker* is [wɝkɚ] (not [wr̩kr̩]), and *four* is [foɚ] (not [for]). Also, I use [ɒ] for the low back round vowel, instead of [ɔ]. For example, *law* in American English is [lɒ], not [lɔ].

8.1. RHYMES IN NON-FINAL POSITIONS

A rhyme is final if it is before a word boundary, whether there is a following suffix or not. For example, in #*abstract*#, where # is a word boundary, the second rhyme is final. In #*abstract*#*ness*#, both the second and the third rhymes are final; the second is final because *abstract* is a word, even though it has a suffix. Similarly, the rhyme in *boast* is final whether the word occurs alone or with a suffix, such as *boasting* and *boastful*. According to this definition, a prefix-final rhyme can also be treated as final. For example, the rhyme in *dis*- of *dislike* is final, because it is before a word boundary. This definition differs from that of Hall (2001), for whom a word-final syllable is not final if there is a vowel-initial suffix, such as *boast* in *boasting*.

A rhyme can be as small as one sound, either a short vowel, as in the first syllable of *about*, or a syllabic consonant, as in the second syllable of *little* (although some linguists believe that so-called syllabic consonants are preceded by the vowel [ə]). It is less obvious how large a rhyme can be in English. Borowsky (1989) reports that in non-final positions the maximal rhyme size is mostly VX, which is either VV (a long vowel or a diphthong) or VC (a short vowel and a C). Therefore, she argues that VX is the basic rhyme size for English, although extra consonants can be appended to it. In contrast, other people adopt a larger rhyme size based on the final syllable, such as VXCCC (Haugen 1956b), VXCC (Selkirk 1982) or VXC (Kiparsky 1981, Giegerich 1992, Blevins 1995, Hall 2001). I show that non-final rhymes are indeed limited to VX. I also argue that apparent exceptions can be explained without expanding VX.

The CELEX lexicon of English contains 160,595 entries, of which 52,447 are uninflected (called "lemmas"). I shall focus on the uninflected list, which includes *cat* (noun) but not *cats* (noun), *catch* (verb) but not *catches* (verb) or *caught* (verb), and so on. The list includes some compounds, such as *fireworks*, *absent-minded*, and *capital gains*. Some plural words are included if they are so used commonly, such as *thanks*, *fireworks*, *assets*, *ranks*, and *capital gains*. Word categories and derivational affixes are included. For example, *catch* (verb), *catch* (noun), *catching* (adjective), *catcher* (noun), *catchy* (adjective), and *catchily* (adverb) are listed as separate words. The lexicon is quite comprehensive, but not exhaustive. In particular, it lacks some infrequent words, such as *gules* and *deictic*, and it generally does not include proper names, such as *Albert*, *Alzheimer*, and *Patrick*. However, it is reasonable to assume that by and large the syllable structures of uncollected words are fairly represented by the words in CELEX.

CELEX follows the Maximal Onset rule (Kahn 1976) and the "all-in" approach to syllabification, and syllable boundaries are indicated by brackets. For example, *larynx* is syllabified as [læ][rɪŋks] [CV][CVCCC], although Borowsky (1989) would initially syllabify it as [CV][CVC]CC. For convenience I continue to use the terms "syllable," "onset," and "rhyme" to refer to the units under the CELEX syllabification.

By searching the syllabification information, we can extract all words that have one or more rhymes that exceed VX. The results are shown in (2). When a word has more than one pronunciation, only the first (a primary one) is included. For example, *actress* has three pronunciations [ˈæktrɪs] (primary), [ˈæktrəs] (secondary), and [ˈæktrɛs] (secondary), and only [ˈæktrɪs] is included. Alternative pronunciations will be examined only when relevant.

(2)

Step	Result (no. of words)	
Start	52,447	
Remove compounds	42,089	(based on space or hyphen)
Words exceeding VX	4,193	(8%)
Remove affixed words	166	
Remove repetitions	146	("exceptions")

After removing compounds based on an internal space or hyphen, there are 42,089 words left, of which 4,193 contain one or more non-final rhymes that exceed VX. Many such words contain an internal word boundary. For example, in *abruptness* the second syllable is [brʌpt], whose rhyme [ʌpt] exceeds VX. But since *abrupt* is a word itself, the rhyme [ʌpt] is word-final, and so it is not what we are looking for here. To collect true non-final rhymes that exceed VX, I removed words with common affixes, which leaves 166 words. The list still contains some repetitions. For example, *empty* is listed three times, because it can be an adjective, a verb, or a noun. After removing repetitions, there are 146 words remaining.

On the basis of manual examination, I grouped the 146 words into several cases, shown in (3). When a word belongs to two (or more) cases, it is counted under just one case. For example, the first rhyme in *laundry* falls under [VVN] and [V:C], and it is counted under [VVN] only.

(3)

Case	Count	Example
Compound/affix	27	*coltish, dachshund, WTO*
Initial *ex*-	8	*exchange*
Syllabification	5	*scherzo*
VVN	53	*council*
[V:C]	27	*almost*
VNC	19	*empty*
Others	7	*arctic*
Total	146	

Twenty-seven items are compounds, affixed words, or abbreviations, which failed to be removed in earlier steps. Of the rest, there are eight words that start with *ex*-, which some linguists consider to be a prefix (Pierrehumbert 1994); if so such words should be removed too (because the offending rhyme [ɛks] or [ɪks] is before a word boundary). It is also possible to treat [ks] as an affricate [k^s] (Chapter 2); if so [ɛk^s] or [ɪk^s] does not exceed VX. Of the rest, there are five words that can be syllabified differently so that there is no offending rhyme. For example, *scherzo* is syllabified as [skɛət][səʊ] (American [skeɚt][so]), where [ɛət] exceeds VX. However, if we syllabify it as [skɛə][t^səʊ] (American [skeɚ][t^so]), where [t^s] is an affricate, there is no offending rhyme. Similarly, *Sexagesima* is syllabified as [sɛks][ə][ʤɛ][sɪ][mə], where [ɛks] exceeds VX. However, if [ks] is an affricate [k^s], or if the syllabification is [sɛk][sə][ʤɛ][sɪ][mə], there is no offending rhyme. Of the rest, there are fifty-three words whose offending rhyme is [VVN], where VV is a diphthong. Of the rest, there are twenty-seven words whose offending rhyme is [V:C], where [V:] is a tense vowel. Of the rest, there are nineteen words whose offending rhyme is [VNC], where NC is a homorganic cluster. In summary, if we exclude the first three cases, there are only 106 words (the last four cases) that seem to be exceptions to the VX size.

According to Bromberg and Halle (1989), phonology and lexicon are mostly stored in memory and can tolerate exceptions. In a lexicon of 52,447 words, a total of 106 exceptions, or 0.00002%, seems remarkably trivial and should not obscure a robust generalization. Similarly, Borowsky (1989: 158) assumes that some words "are entered in the lexicon with their aberrant syllabification."

Borowsky (1989) notes, though, that [VVN] and [VNC] make up most of the exceptions, where the N is homorganic with the following C. She suggests that homorganic NC can count as one C. A similar proposal is made by Hall (2001). However, in feature theory NC cannot form a complex sound (Chapter 2). Hall (2001) also suggests that homorganic LC clusters such as [ls, ld, lt] can count as one sound, and medial VLC rhymes should be possible; but such rhymes are not found, nor do nonsense words with them seem very natural, such as *filtky* or *sheldmy*.

I argue that there is a simpler analysis of [VVN], [VNC], and [V:C], which is shown in (4) and illustrated with American pronunciation.

(4)

CELEX	Analysis	Example
[VVN]C	[ṼṼ]C	*council* [kaun][səl] → [kãũ][səl]
[VNC]	[ṼC]	*empty* [ɛmp][ti] → [ɛ̃p][ti]
[V:C]	[VC]	*almost* [ɒ:l][mo:st] → [ɒl][most]

As discussed in Chapter 3, the representation of VNC as ṼC has been proposed by Bailey (1978), who transcribes *simple* as [ˈsĩpɫ] and *sinker* as [ˈsĩkə]. Phonologically, there is no loss of contrast because the place feature of N is the same as that of the following C. Phonetically, it is well known that English vowels are nasalized when they follow a nasal coda and that the nasal itself often has little duration of its own (Malécot 1960, Fujimura 1979). The representation of [V:C] as [VC] is also phonologically possible because a tense vowel differs from a lax one in quality, i.e. in the feature [tense], and therefore even if a tense vowel is shortened it remains distinct from a lax one. The idea that tense vowels can be long or short has also been proposed by others (Pike 1947, Jones 1950, Abercrombie 1967, Giegerich 1985, Alcantara 1998).

If the present analysis is correct, we are left with just seven problematic words, along with two additional words, *deixis* and *deictic*, which are not included in CELEX. I argue that such words can also be accounted for. The analysis makes use of what Fujimura (1979) calls "phonetic" affixes or what Pierrehumbert (1994) calls "perceived" affixes. A phonetic or perceived affix is one that has the same form as a real affix. An example is shown in (5).

(5) Phonetic or perceived affixes (Fujimura 1979, Pierrehumbert 1994)

Real suffix	Perceived suffix
drama-tic	*arc-tic*

Since -*tic* is a real suffix in English, words like *arctic* may be perceived as having the same suffix *arc*-*tic*. This is not to claim that speakers necessarily think that *arctic* has a suffix. Instead, the claim is that if real suffixes do not make a word sound bad, then a word with an ending looking like a suffix will not sound bad, provided the rest of the word does not sound bad (i.e. the *arc* part of *arctic*). With this in mind, I offer the analysis of the nine problematic words in (6), based on American pronunciation.

(6) American pronunciation

Word	Pronunciation	Analysis
arctic	[ɑɚk][tɪk]	perceived suffix (-*tic*, as in *dramatic*)
deictic	[daɪk][tɪk]	perceived suffix (-*tic*, as in *dramatic*)
deixis	[daɪk][sɪs]	perceived suffix (-*is*, as in *analysis*)
dextrose	[dɛks][trəʊz]	[k^s] as affix?
maestro	[maɪs][trəʊ]	??
ordnance	[oɚd][nəns]	??
parsnip	[pɑɚs][nɪp]	??
poultice	[pol][tɪs]	OK
seismic	[saɪz][mɪk]	perceived suffix (-*ic*, as in *rhythmic*)

Four of the words can be analyzed in terms of perceived suffixes. It is worth noting that many American speakers pronounce *arctic* as [ɑɚ][tɪk], in which there is no offending rhyme. The word *poultice* has no offending rhyme in American English because the tense vowel [o] need not be long. This leaves us with four words, *dextrose*, *maestro*, *ordnance*, and *parsnip*. One might offer some speculations about them, too. For example, could [ks] in *dextrose* be an affricate (see Chapter 2, Prinz and Wiese 1991, Wiese 1996)? Could *ordnance* be [o˞d][nəns], with a retroflexed [o˞], or [oɚ][dn][əns], with a syllabic [n]? Could *parsnip* be a perceived compound, or could it be [pɑ˞s][nɪp], with a retroflex [ɑ˞]? Could *maestro* be a perceived compound? However, I shall leave the answers open.

It is interesting to consider the analysis of the same words in British pronunciation, which is shown in (7).

(7) British pronunciation

Word	Pronunciation	Analysis
arctic	[ɑːk][tɪk]	OK: [ɑk][tɪk]
deictic	[daɪk][tɪk]	perceived suffix (-*tic*, as in *dramatic*)
deixis	[daɪk][sɪs]	perceived suffix (-*is*, as in *analysis*)
dextrose	[dɛks][trəʊz]	[k^s] as affix?
maestro	[maɪs][trəʊ]	??
ordnance	[ɔːd][nəns]	OK: [ɔd][nəns]

parsnip	[pɑːs][nɪp]	OK: [pɑs][nɪp]
poultice	[pəʊl][tɪs]	perceived suffix (-*is*, as in *analysis*)
seismic	[saɪz][mɪk]	perceived suffix (-*ic*, as in *rhythmic*)

Three of the words now have no offending rhyme, if tense V can be short. In addition, four of the words can be analyzed in terms of perceived suffixes. This leaves us with two remaining words: *dextrose*, whose first syllable could be [dɛkˢ], and *maestro*, which may be a perceived compound.

To summarize, the VX limit on non-final rhymes is stricter than previously thought. An exhaustive search of the English lexicon shows no compelling counterexample.

8.2. WORD-FINAL RHYMES

Next consider word-final rhymes. I summarize the results in (8), where S is the CELEX notation for a syllabic consonant and [C] and [CC] are words not thought to be proper syllables.

(8) Word-final rhymes in the CELEX English lexicon

Type	Count	Example
V	10,536	*delta* [dɛl][tə]
VC	19,949	*campus* [kæm][pəs]
S	4,193	*able* [eɪ][bl̩]
[C]	6	*shh* [ʃ], *'ve* [v]
[CC]	1	*psst* [ps]
SC	326	*fabled* [feɪ][bl̩d]
VX+	17,436	*self* [sɛlf]
All	52,447	

There are 17,436 words whose final syllable exceeds VX. They are detailed in (9). The column under [t, d, s, z, θ]# indicates how many words end in one of those coronal consonants.

(9) Counts of word final rhymes that exceed VX

Type	Count	[t, d, s, z, θ]#	Example
VVC	11,167	5,074 (45%)	*circulate*
VCC	4,390	3,794 (86%)	*repent*
VVCC	1,663	1,571 (94%)	*astound*
VCCC	189	189	*larynx*
VVCCC	23	23	*whilst*
SCC	3	3	*presents*
VCCCC	1	1	*precincts*
All	17,436		

Two well-known generalizations can be observed, stated in (10).

(10) a. Consonants after VXC are limited to the coronals [t, d, s, z, θ].
b. The first C after VX can be any consonant (except [h] and [ŋ]).

The first generalization can be seen in VVCC, VCCC, and VCCCC. The second generalization can be seen in VVC and VCC, in which the final C is often not one of the coronals; apparent exceptions will be discussed below.

There are different views on how to account for extra sounds at word edges. Ito (1986) proposes that all sounds must belong to a syllable, unless they are at word edges, but she does not explain why word-final consonants can be tolerated even though they are not in a syllable. Borowsky (1989) proposes that word-edge consonants are ignored at first in syllabification, but are added to the preceding syllable later. The question for her is why word-edge consonants can be ignored in the first step, why they need to be adjoined to a syllable in the second step, and why a rhyme can be expanded beyond VX. In yet another proposal, Hall (2001) suggests that word-medial rhymes are maximally VX, but word-final rhymes are maximally VXC. In addition, word-final coronals can be "appended" to the preceding syllable. It is not clear why a word-final rhyme can or should be longer, nor is it clear why word-final coronals are or should be appended to the preceding syllable. Indeed, Hall's analysis predicts that there can be languages whose non-final syllables are maximally VXC; but such languages remain to be shown.

Clearly, if we can explain word-edge consonants, we do not need to stretch the rhyme size, but can maintain a consistent limit VX for both final and non-final rhymes. As discussed in Chapter 3, the solution is morphological, which uses three concepts, given in (11).

(11)

	Concept:	Accounting for:
	The affix rule	Final coronals [t, d, s, z, θ]
	Potential vowel	Final C (immediately after VX)
	Anti-allomorphy	Final C (immediately after VX)

First, the coronals [t, d, s, z, θ] can all be suffixes in English. According to the affix rule (section 3.3.6), if a language has consonant suffixes, they must be used regardless of whether the preceding syllable is full. If so, we no longer need to include final coronals into a syllable. Recall also that the affix rule covers both real affixes and "perceived" affixes (Fujimura 1979, Pierrehumbert 1994). Therefore, it does not matter whether a final coronal is a real suffix (as the [t] in *helped*) or not (as the [t] in *text*).

Next, we consider the C immediately after VX, which can be any consonant except [h] and [ŋ]. In (12) I show consonants found in the C position of word-final VVC. The CELEX symbol [r*] occurs in [@r*], which is pronounced as [ɚ] in American English; for example, *ear* [iɚ] is [I@r*] in CELEX. The CELEX symbol [~:N] indicates a nasal vowel, which occurs in French-accented words only.

(12) Word final C in VVC (total = 11,167)

Type	Count	Example
[t, d, s, z, θ]	5,074	*out, cloud, rouse, house, mouth*
n	1556	*down* [daun]
l	1119	*eel* [i:l]
r*	1068	*ear* [iɚ]
k	585	*stroke* [strouk]
m	548	*time* [taim]
p	282	*hype* [haip]
f	236	*loaf* [louf]
v	211	*five* [faiv]
ʤ	139	*age* [eiʤ]
ʧ	131	*couch* [kauʧ]
b	90	*robe* [roub]
ʒ	32	*massage* [məsɑ:ʒ]
ð	29	*clothe* [klouð]
g	27	*plague* [pleig]
ʃ	21	*leash* [li:ʃ]
~:N	19	*salon* [səlɒ̃:]

If we consider this C to be in the preceding rhyme (e.g. Kiparsky 1981, Selkirk 1982, Giegerich 1992, Blevins 1995, Hall 2001), we cannot explain why non-final rhymes are limited to VX. If we consider this C to be outside of a syllable (McCarthy 1979b, Hayes 1982, Ito 1986), we must explain why it is allowed to occur.

As discussed in Chapter 3, the extra C can again be accounted for by morphology. Since English has V-initial suffixes, such as *-ing*, *-er*, *-ic*, *-y*, *-able*, they provide a "potential vowel," of which the root-final C can serve as the onset (Giegerich 1985, Borowsky 1986). When this happens, there is no extra C. To explain why the word-final C is kept when there is no following V, we can use the notion of "paradigm uniformity" or "anti-allomorphy" (Burzio 1996), which requires a morpheme to keep the same shape regardless of the environment. For example, because the word *help* occurs in *helping* and *helper*, we would like to keep it in the same shape, even when there is no following V, as when it occurs by itself or in *helps* and *helpless*. It is interesting to note that sometimes extra consonants do get deleted, despite anti-allomorphy. For example, Gimson (1970: 238) offers such examples: *exactly*, *restless*, *facts*, *mostly*, and *fifths*, where the underlined consonants are deleted in casual speech. Similarly, Wells (1990) offers the example *first rate*, where deletion again happens to an unsyllabified consonant. In the present analysis, where the maximal rhyme is VX, such deletion is expected. It should be noted that the deletions in *tl*, *ts*, *ths*, and *tr* are real, because they involve the same articulator, Coronal (from two gestures to one gesture); therefore, they cannot be explained in terms of gestural overlap (Browman and Goldstein 1989).

Having introduced the analysis, I show how it accounts for the data. First, VVC and VCC present no problem, because the final C is supported by a potential V and anti-allomorphy.

Next, consider VCCC. The present analysis predicts that the final C is one of [t, d, s, z, θ], which is true, shown in (13). The present analysis also predicts that the penultimate C is not limited to [t, d, s, z, θ], which is true, too. The remaining part is VX.

(13) Analysis of VCCC (total = 189)

Final C	t (92), s (68), z (22), θ (5), d (2)
Penult C	s (59), k (46), t (30), d (19), p (22), f (4), v (4), ʧ (3), θ (1), ʤ (1)

Next, we consider VVCC. Since the first C after VV can be any consonant, we focus on the final C, which is predicted to be one of [t, d, s, z, θ]—which is mostly true, as shown in (14).

(14) Final C in VVCC (total = 1,663)

[t, d, s, z, θ]:	1,571 (t = 600, d = 587, s = 209, z = 154, θ = 21)
Other cases:	92

There are 92 exceptional cases, which fall into two types, [VVNC] and [V:CC]. We have seen similar cases in non-final positions, where [VVN] is analyzed as [ṼṼ] and [V:C] is analyzed as [VC]. We can use the same analysis here, shown in (15).

(15)

Rhyme	Analysis	Count	Example
V:CC	VCC	30	*grasp*
VVNC	ṼṼC	62	*change, launch*

In this analysis, the final C need not be a coronal, because it is the first C after VX. It is worth noting that in the case of [V:CC], while

British pronunciation uses a long vowel (such as [ɑ:] in *grasp*, the vowel in American pronunciation is [æ], which is often thought to be short.

Next, consider VVCCC. It is predicted that the final C is one of [t, d, s, z, θ], which is true. The penult C should also be one of [t, d, s, z, θ], which is mostly true, as shown in (16).

(16) Analysis of VVCCC (total = 23)

Final C	[t, d, s, z, θ] = 23	
Penult C	[t, d, s, z, θ] = 20	
	Others = 3 (*masked, unasked, arranged*)	

There are three words whose penult C is not a coronal, all of which end in [V:CCC]. Once again, they can be analyzed in terms of [V:CC] → [VCC] (for *masked* and *unasked*) and [VVNC] → [ṼṼC] (for *arranged*). The word *arranged* can also be analyzed as [V:NC] → [ṼNC], if the second vowel is not pronounced as a diphthong [ei]. Finally, the first C after VV can be any consonant. Thus, VVCCC rhymes again present no problem.

The remaining two cases are SCC (where S is a syllabic consonant) and VCCCC. There are altogether four such words, shown in (17).

(17) SCC *presents, thousandth* (noun), *thousandth* (adjective)
VCCCC *precincts*

In SCC, the last two consonants are both coronals. In VCCCC, we predict the last two consonants to be coronals, which is true. Thus, none of the words presents a problem.

In summary, word-final consonants in English can be explained by morphology. The C immediately after VX can be accounted for by a "potential V," which is from a vowel-initial suffix (Giegerich 1985, Borowsky 1986). When this V is absent, the final C is kept because of paradigm uniformity or anti-allomorphy (Burzio 1996). Consonants after VXC can be accounted for by the affix rule (section 3.3.6), according to which extra final Cs are either suffixes or suffix-like (Fujimura 1979, Pierrehumbert 1994). Apparent exceptions can be accounted for by the idea that [V:C] can be analyzed as [VC] (Pike 1947, Jones 1950, Giegerich 1985) and [VVN] can be analyzed as [ṼṼ] (Malécot 1960, Bailey 1978, Fujimura 1979). As a result, the maximal size of word-final rhymes is still VX, as it is for non-final rhymes.

8.3. ONSET CLUSTERS IN WORD-INITIAL POSITIONS

A summary of word-initial onset clusters is given in (18), based on the CELEX "lemma" lexicon, where # is a word boundary, [C] and [CC] are non-syllabic words, and [S] is a syllabic C.

(18)

	All entries	52,447	
	Compounds	10,358	
	Non-compounds with #CCCV	468	
	Non-compounds with #CCV	7,217	
	Non-compounds with #CV	26,606	
	Non-compounds with #V	7,789	
	Non-compounds with #[C]	6	(e.g. *'ve* [v], *'s* [s])
	Non-compounds with #[CC]	1	(*psst* [ps])
	Non-compounds with #[S]	1	(*hem* [m])
	Non-compounds with #[CS]	1	(*ahem* [hm])

There are thirteen types of initial CCC onset clusters, shown in (19), along with sample words for those that do not start with [s], in the pronunciation given in CELEX. The affricates [ʤ] and [ʧ] are treated as single sounds in CELEX, but [ts], [Cl], [Cr], [Cj], and [Cw] are treated as clusters. Some words occur more than once, because they have more than one word category or meaning. For example, *cloisonné* can be a noun or an adjective, and *Tswana* can be the language or the people in Botswana.

(19) Word-initial CCC clusters (13 types, 468 tokens)
str (171), skr (97), skw (68), spr (50), spl (39), stj (21), skj (7), spj (6), skl (2), klw (2),
psj (2), tsw (2), krw (1)
[klw] *cloisonné* [klwɑːzɔneɪ] (twice)
[psj] *pseudo* [psjuːdəʊ] (twice)
[tsw] *Tswana* [tswɑːnə] (twice)
[krw] *croissant* [krwʌsɑ̃ː]

The clusters [klw], [psj], [tsw], and [krw] each involves one word form. In addition, [klw], [psj] and [krw] are usually pronounced as [kl], [sj], and [kr] (unless the speaker wants to show a French or Greek accent). Moreover, if we treat [ts] as an affricate, [tsw] is not CCC but CC. Thus, we have a generalization that in all word-initial CCC clusters the first sound is [s]. This means that the initial [s] is a special case, as noted in previous studies. If we exclude the initial [s], the remaining CC clusters also occur independently, as we shall see.

There are fifty-eight types of initial CC clusters in non-compound words, which are shown in (20), ordered by their frequency of occurrence.

(20) Word-initial CC clusters (total = 7,217)
pr (929), st (559), tr (510), kr (428), br (356), gr (347), sp (328), sk (324), fl (313), fr (301), kl (300), pl (279), bl (269), sl (239), kw (222), dr (195), sw (168), sn (146), gl (133), nj (107), sm (92), pj (73), tw (71), θr (68), kj (66), dj (65), mj (61), fj (58), tj (46), hj (41), ʃr (34), bj (22), sf (9), vj (9), dw (7), gw (6), lj (4), θw (4), sj (3), θj (2), ʃm (2), ʃn (2), ʃp (2), ts (2), pw (2), pf (1), ph (1), zl (1), km (1), kn (1), sr (1), ps (1), pʃ (1), kv (1), sv (1), mw (1), ʃw (1), vw (1)

In most cases, the second sound is one of the approximants [l, r, j, w]. Those whose second sound is not an approximant are listed in (21).

(21) Word-initial CC where the second C is not an approximant (total = 1,474)
st (559), sp (328), sk (324), sn (146), sm (92), sf (9), ʃm (2), ʃn (2), ʃp (2), ts (2), pf (1), ph (1), km (1), kn (1), ps (1), pʃ (1), kv (1), sv (1)

In all except fifteen tokens, the first sound is [s]. The special status of [s]-C clusters has been noted before (e.g. Selkirk 1982, Lamontagne 1993, Treiman et al. 1992), to which I return later. The fifteen special tokens are detailed in (22).

(22) Word-initial CC where the first C is not [s] and the second C is not an approximant
ʃm (*schmaltzy, schmaltz*), ʃn (*schnapps, schnitzel*), ʃp (*spiel* (twice)), ts (*tsetse, zeitgeist*), pf (*pfennig*), ph (*pooh*), km (*Khmer*), kn (*Knesset*), ps (*psoriasis*), pʃ (*pshaw*), kv (*kvass*)

The clusters [pf], [ps], and [pʃ] are often pronounced without [p]. The cluster [ph] seems to be an error and should be [p]. The cluster [km] is often pronounced as [kəm]. The cluster [ts] is an affricate. However, [ʃm], [ʃn], and [ʃp] usually do occur as a cluster.

The question why some onset clusters occur and some do not will be discussed later. Before then, let us examine onset clusters in non-initial positions.

8.4. ONSET CLUSTERS IN NON-INITIAL POSITIONS

A syllable is initial if it is the first after a word boundary, whether there is a prefix or not. For example, in *sprinkle*, [sprɪŋ] is initial. In *re-sprinkle*, both [ri] and [sprɪŋ] are initial; the latter is initial because *sprinkle* is a word, even though it has a prefix. By this definition, a suffix-initial syllable is also initial. For example, in *mob-ster*, [stɚ] is initial, because it is after the word boundary of *mob*. To obtain non-initial onset clusters, I exclude all compounds and monosyllables. The results are shown in (23). When a word has more than one pronunciation, only the first is included.

(23)

Type:	Count:
All words	52,447
Non-compounds	42,089
Polysyllables	35,329
Polysyllables with non-initial onset clusters	7,933

After removing compounds (based on internal space or hyphen), monosyllables, and words with no non-initial onset cluster, there are 7,933 words left (affix boundaries will be considered later). The information on their onset clusters is given in (24).

(24) Non-initial onset clusters in English

All such words	7,933
Non-initial syllables in them	18,452
Non-initial clusters	8,376
Cluster types	72

Excluding the first syllable from each word, there are 18,452 syllables and 8,376 onset clusters, which fall into seventy-two types. The seventy-two types of clusters can be divided into three groups, shown in (25), where R is an approximant and where frequencies are shown in parentheses. Again, the affricates [ʤ] and [ʧ] are treated as single sounds.

(25) Non-initial onset clusters in English

CCC (12 types, 510 tokens)

[str] (273), [skr] (92), [spl] (35), [spr] (28), [skj] (27), [stj] (24), [skl] (11), [spj] (10), [skw] (7), [gjw] (1), [frw] (1), [trw] (1)

CR (49 types, 5,905 tokens)

[tr] (762), [gr] (401), [bl] (399), [pr] (378), [nj] (330), [kj] (291), [kr] (280), [pl] (280), [kw] (275), [tj] (249), [br] (233), [dj] (212), [kl] (212), [dr] (186), [fl] (150), [lj] (134), [mj] (127), [fr] (124), [pj] (113), [bj] (105), [gl] (90), [fj] (68), [gj] (68), [vj] (57), [sj] (52), [sl] (50), [zj] (49), [θr] (46), [gw] (46), [sw] (29), [ʧr] (18), [vr] (16), [tw] (14), [hj] (11), [θj] (10), [ʃj] (8), [ʃr] (5), [ʤj] (4), [ʒj] (4), [dw] (3), [zw] (3), [ʒw] (3), [pw] (2), [nw] (2), [θw] (2), [ðr] (1), [rj] (1), [sr] (1), [vw] (1)

Other CC (11 types, 1,961 tokens)

[st] (1,366), [sp] (362), [sk] (172), [ʃn] (22), [sn] (14), [sm] (13), [sf] (5), [ts] (4), [ʧn] (1), [ʃp] (1), [ʃt] (1)

CR clusters are generally accepted in most analyses; they are discussed in more detail below. For now let us focus on CCC and other CC clusters, all of which start with [s], except eight cluster types. The cluster [ts] (four tokens) can be treated as an affricate. The remaining seven types are shown in (26), according to the CELEX transcription.

(26) Non-initial, non-CR onset clusters that do not start with [s]

[ʃn] (22)	*optionally* [ɔp][ʃnə][lɪ] ([ʃn̩][l̩][ɪ])
	vacationist [və][keɪ][ʃnɪst] ([ʃn̩][ɪst])
[ʧn] (1)	*unfortunate* [ʌn][fɔ:][ʧnət] ([ʧə][nət])
[gjw] (1)	*multilingual* [mʌl][tɪ][lɪŋ][gjwəl] ([gjʊ][əl], [gwəl])
[frw] (1)	*sangfroid* [sɑ̃:][frwɑ:]
[trw] (1)	*octroi* [ɔk][trwɑ:]
[ʃp] (1)	*glockenspiel* [glɔ][kən][ʃpi:l]
[ʃt] (1)	*gestalt* [gə][ʃtælt]

All tokens of the onset [ʃn] are optional pronunciations of the syllabic [ʃn̩] (and all of them occur in suffixed words). For example, *optionally* [ɔp][ʃnə][lɪ] can be pronounced as [ɔp][ʃn̩][l̩][ɪ] (or [ɔp][ʃə][nə][lɪ]) and *vacationist* [və][keɪ][ʃnɪst] can be pronounced as [və][keɪ][ʃn̩][ɪst]. Similarly, [ʧn] occurs in *unfortunate* [ʌn][fɔ:][ʧnət], which can be pronounced as [ʌn][fɔ:][ʧn̩][ət]. The cluster [gjw] occurs in the last syllable of *multilingual*, which can also be pronounced as [gwəl] or [gjʊ][əl]. The clusters [frw] and [trw] occur in two French words. Finally, [ʃp] and [ʃt] occur in two German words, where *glockenspiel* 'bell-play' is a compound in which [ʃp] is word initial, and *gestalt* may have a perceived prefix *ge*- or can be syllabified as [gəʃ][tælt].

We have seen earlier that [s] seems to be an extra sound in word-initial clusters. Let us take a close look at onset clusters that start with [s] in non-initial positions. The CELEX cluster [sj] (fifty-two tokens), which is often pronounced as [s] (as in *insular*) or [ʃ] (as in *social*) in American English, presents no problem in any analysis and so I shall ignore it. Based on manual examination, I divide the other [s]-clusters into several cases, shown (27), where [V] is a short vowel and [V:] is a long vowel (but not a diphthong).

(27) Non-initial [s]+CC and [s]+C onset clusters, excluding [sj]

Types (18) [str, skr, spl, spr, skj, stj, skl, spj, skw, st, sp, sk, sl, sr, sw, sn, sm, sf]

Case	Count	Example
After [V]	1,422	*aspirin, mascot*
After [V:]	157	*thirsty, basket, Easter, auspice*
Prefixed	490	*inscribe, transcribe, construct, destruct, obstruct*

Compound/Suffixed	387	*doorstep, roadster, Yugoslav, coastal*
Others	63	
Total	2,519	

In the analysis, [V:] does not include the diphthongs [ei] (as in *tasty*) and [ou] (as in *coastal*), or what is [Vɚ] in American English, such as *corselet* [kɔ:][slɪt] ([koɚ][slɪt] in American English). Also, if a word falls under two or more cases, it is counted under just one, usually an earlier one in the order given. For example, *aspiration* is counted under "After [V]" and not "Suffixed," *thirsty* is counted under "After [V:]" and not "Suffixed," *awestruck* is counted under "After [V:]" and not "Compound," and *obstruction* is counted under "Prefixed" and not "Suffixed."

Of the 2,519 clusters, 1,422 occur after [V] and 157 occur after [V:]. In such cases, the [s] can be analyzed as the coda of the preceding syllable. For example, *mascot* [mæ][skət] can be analyzed as [mæs][kət], *Easter* [i:][stɚ] can be analyzed as [is][tɚ], *basket* [bɑ:][skɪt] ([bæ][skɪt] in American English) can be analyzed as [bɑs][kɪt] ([bæs][kɪt] in American English), and *auspice* [ɔ:][spɪs] can be analyzed as [ɔs][pɪs] ([ɒs][pɪs] in American English). Of the rest there are 490 tokens where the [s]-cluster occurs at a prefix-word boundary and are not non-initial. For example, in *inscribe* and *transcribe*, [skr] is initial for the word *scribe*. In fact, many such prefixes do not present a problem. For example, *inscribe* can be [ĩs][kraib], where [s] is not in the second onset. Finally, there are 387 tokens where the [s]-cluster again occurs after a word boundary in a compound or suffixed word. For example, in *doorstep* [st] occurs after the word *door* and in *roadster* [st] occurs after the word *road*. Such [s]-clusters are not non-initial either. This leaves us with sixty-three tokens of non-initial onsets that involve [s]-clusters.

The sixty-three clusters occur in sixty-three words. On the basis of manual examination, I grouped them into several cases, shown in (28).

(28) Remaining non-initial onset [s]-clusters

Case	Count	Example
Repeats	10	*bolster (noun), bolster (verb)*
With -*ster*	13	*bolster, cloister, hamster, holster, huckster, lobster, maltster, minster, monster, oyster, spinster, ulster, upholster*
Suffixed	9	*boisterous, corselet, dexterity, dexterous, dextrous, ecstatic, Trotsky, Trotskyist, Trotskyite*

Compounds	11	*coxswain, feldspar, feldspar, larkspur, lodestar, maelstrom, rheostat, solstice, Telstar, tungsten, wainscot*
Prefixed	4	*proscribe, sextant, sextet, sexton*
Others	16	
Total	63	

In the sixty-three words, ten are repeats of another word. Of the rest there are thirteen that end in -*ster*, which can be treated as a perceived suffix (Fujimura 1979, Pierrehumbert 1994), because it is identical to the suffix in *mobster*, *roadster*, *gangster*, etc. The word *dexter* 'on the right' did not occur in CELEX; if it did, it could fall into this group, too. Of the rest there are nine words that have a suffix. For example, *corselet* has the suffix -*let* 'small' and *Trotsky* has the suffix -*sky* 'estate of', although *Trotsky* can also be analyzed as [trat^s][ki], where [t^s] is an affricate. Of the rest there are eleven compounds, some of which being foreign words. For example, *tung*-*sten* means 'heavy-stone' and *mael*-*strom* means 'grind-stream'. Of the rest there are four words that have a prefix or a bound root. For example, *sex*- in *sextant* and *sextet* means 'six' and *sexton* may have a perceived prefix or root *sex*-. Also, *sex*- can be [sɛk^s], where [k^s] is an affricate (Chapter 2); if so there is no problem anyway. This leaves us with sixteen words still to be accounted for, which I list in (29).

(29) Words with non-initial onset [s]-clusters still to be accounted for (16 in all) *borstal, capstan, ecstasy, eisteddfod, hamstring, Holstein, juxtapose, mangosteen, menstrual, menstruate, minstrel, monstrance, Pleistocene, Sanskrit, Shakespeare, textile*

This is a very short list out of an entire dictionary. In addition, further analyses seem possible. For example, *capstan*, *hamstring*, *mangosteen*, and *Shakespeare* may be treated as compounds or perceived compounds. Some words may have a perceived affix, such as *borstal* (*al*), *ecstasy* (*ex*-), *Holstein* (-*stein*), and *textile* (-*ile*). The words *menstrual, menstruate, minstrel, monstrance,* and *Sanskrit* involve the sequence [Vn.sCrV], which can be analyzed as [Ṽs.CrV]. For example, *Sanskrit* [sæn.skrɪt] can be analyzed as [sæ̃s.krɪt] (Malécot 1960, Bailey 1978, Fujimura 1979). The words *hamstring* and *mangosteen* can perhaps be analyzed in terms of [VN.st . . .] → [Ṽs.t . . .] or as perceived compounds. If [ps, ks] are a non-homorganic affricate (Wiese 1996), then we can exclude *capstan*, *ecstasy*, *juxtapose*, and *textile*. However, I leave the ultimate analysis of these words open.

In summary, non-initial onsets are more restricted than initial onsets. An exhaustive examination of the English lexicon shows that only CR clusters can occur in non-initial onsets, where R is one of [l, r, j, w]. In addition, the only [s]-cluster that occurs in non-initial onsets is [sj]. There is no compelling counterexample to the generalizations.

8.5. ANALYZING ONSET CLUSTERS

Chomsky (1957) proposes that the grammar of a language is a set of rules that can generate all and only good forms that can occur in a given language. Halle (1962: 60–1) makes a similar proposal. He argues that English speakers know that [bɪk], [θod], and [nɪs] are possible words but [tsaim], [gnait], and [vnɪg] are not.

Most linguists would agree with Chomsky and Halle that not all well-formed syllables are used in a language, such as [bɪk] and [nɪs]. However, it is not obvious whether speaker intuition is clear enough to decide whether a phonological form is or is not good. In fact, there is evidence that speakers are not always sure. For example, Frisch et al. (2000) have shown that native judgment on made-up English words is not clear cut. The same has been shown for Chinese (Myers and Tsay 2005, Zhang 2007). For existing words, judgment is not always clear either. For example, some speakers pronounce *Tswana* as [tswɑnə], while others pronounce it as [swɑnə]. For the latter speakers, [ts] does not occur in word initial position. For the former speakers, [ts] is a possible onset form, and they might be willing to accept other words that start with [ts]. Similarly, in American English [gj] can occur in non-initial onsets, as in *argue* [ɑɚ.gju]. However, it rarely occurs in word-initial onsets. Thus, if a speaker does not have the word *gules* [gjulz] (or pronounces it as [gulz]) and pronounces the word *gewgaw* as [gu:gɒ], instead of [gju:gɒ], then [gj] is not a good initial onset, although it is a good non-initial onset.

One might wonder whether a distinction should be between frequent words and infrequent ones, but the line is hard to draw, because there is a wide range of frequencies. For example, word-initially [pr] (as in *price*) occurs 929 times, while [tw] (as in *twice*) occurs just 71 times, but there is no evidence that *price* is a better form than *twice*. Indeed, forms with even lower frequencies are not necessarily judged bad. For example, the onset [dw] only occurs in three unaffixed words, *dwarf*, *dwell*, and *dwindle*, yet there is no evidence that such words are felt to be less than fully well-formed.

One might also wonder whether a distinction should be made between English words and borrowed foreign words; but it is not always easy to decide what is foreign and what is not, because English has been continually borrowing large numbers of words. For example, when did (or will) words like *sphinx*, *schnapps*, *schnitzel*, *schnorkel*, *Alzheimer*, *pizza*, *Tzwana*, *Buenos Aires*, and *zloty* become fully English?

Given these considerations, it is relevant to ask whether all clusters seen in CELEX need to be accounted for, and if not, which ones should be.

8.5.1. **Which onset clusters should be accounted for?**

For ease of exposition, I shall use a few simple criteria to decide whether a form should be accounted for or not. They are given in (30).

(30) Which forms should be included in the analysis?
  a. If a form only occurs in an interjection, it is not included.
  b. If a form contains a foreign sound, it is not included.
  c. If a form only occurs as an alternative pronunciation in CELEX, it is not included.
  d. If a form is used as the only pronunciation of a word, it is included, even if the form is infrequent.

Examples of (30a) are the syllables [ps] and [ʃ], which only occur in the interjections *psst* and *sh*, and so they are excluded. Similarly, the onset [pʃ] only occurs in the interjection *pshaw* and it is excluded, too. An example of (30b) is *sangfroid* [sɑ̃:][fʁwɑ̃:], where [ʁ] and [ɑ̃] are French sounds. Therefore, [fʁw] is not included as an onset cluster. An example of (30c) is the coda [x], which only occurs in words like *loch* [lɔx] and *Reich* [raɪx], where it can be pronounced as [k], and so [x] is not included as a coda. Similarly, the onset [ps] only occurs in words like *pseudo* [psjudo], where it can be pronounced as [s], and so [ps] is not included as an onset. Still another example is [ðr], which only occurs in *brethren* [brɛ][ðrən], which can be syllabified as [brɛð][rən]; therefore [ðr] is not included as an onset. An example of (30d) is the word-initial onset [ʃw], which is included, even though it only occurs in the word *schwa* [ʃwɑ]. Similarly, the onset [ʃn] is included because it is the only pronunciation of the initial onset in *schnapps* and *schnitzel*, even though there are just two such words. On the other hand, the non-initial onset [ʃn] in words like *fashionable* [fæ][ʃnə][bl̩] is not included, because it can be pronounced as [ʃn̩] in such cases (among other options), where the [n] is syllabic.

With this in mind, I divide the onset clusters in CELEX into two groups: (a) those that need to be accounted for and (b) those that need not. I show all the clusters in (31)–(33), where group (a) is labeled as "included" and group (b) is labeled as "excluded." An example for each cluster is given in parentheses.

(31) Onset CCC clusters
Included (all initial): str (*string*), skr (*screen*), skw (*square*), spr (*spring*), spl (*splash*), stj (*studio*), skj (*skew*), spj (*spew*), skl (*sclerosis*), tsw (*Tswana*)
Excluded: klw (*cloisonné* [kloi]), psj (*pseudo* [sj]), gjw (*multilingual* [gj]), frw (*sangfroid* [sɑ̃:][fʁwɑ̃:]), krw (*croissant* [kʁwʌ][sɑ̃:]), trw (*octroi* [ɔk][tʁwɑ:])

(32) Onset CR clusters
Included: bj (*beauty*), bl (*black*), br (*bring*), dj (*duty*), dr (*dry*), dw (*dwell*), ʃr (*shrink*), ʃw (*schwa*), fj (*few*), fl (*fly*), fr (*fry*), gj (*argue*), gl (*glad*), gr (*green*), gw (*penguin*), hj (*huge*), kj (*cute*), kl (*class*), kr (*cry*), kw (*quick*), lj (*volume*), mj (*music*), mw (*moiré*), nj (*news*), nw (*peignoir*), pj (*pure*), pl (*plot*), pr (*price*), pw (*puissance*), sj (*suit*), sl (*sleep*), sr (*Sri Lanka*), sw (*swim*), tj (*tube*), tr (*try*), tw (*twin*), vj (*view*), vw (*reservoir*), zl (*zloty*), zj (*presume*), ʒw (*bourgeois*), θj (*enthuse*), θr (*three*), θw (*thwart*)
Excluded: ʤj (*plagiarize* [ʤ]), ðr (*brethren* [brɛð][rən]), ʃj (*negotiable* [ʃə] / [ʃɪə]), rj (*marijuana* [rɪ]), ʧr (*natural* [ʧʊ][rəl]), zw (*Venezuela* [nɛz][weɪ]), ʒj (*Eurasia* [ʒə], [ʃə], or [ʒɪə]), vr (*ivory* [vər])

(33) Other onset CC clusters
Included: st (*stop*), sp (*spot*), sk (*sky*), sn (*snake*), sm (*smack*), sf (*sphere*), ʃm (*schmaltz*), ʃn (*schnitzel*), km (*Khmer*), kn (*Knesset*), kv (*kvass*), sv (*svelte*)
Excluded: ʧn (*fortunate* [ʧn̩]), ʃt (*gestalt* [gəʃ][tælt]), ʃp (*spiel* [sp]), pf (*pfennig* [f]), ph (*pooh* [p]), ps (*psoriasis* [s]), pʃ (*pshaw*, interjection), ts (*zeitgeist* [z])

For each excluded cluster, the relevant part of the alternative pronunciation is also give in brackets, indicating why the cluster is not needed. For example, [ʤj] is found in *plagiarize* [pleɪ][ʤjə][raɪz], which can be pronounced as [pleɪ][ʤɪə][raɪz] or [pleɪ][ʤə][raɪz], which does not use the cluster [ʤj]. The cluster [ʃt] is found in *gestalt* [gə][ʃtælt], which can be syllabified as [gəʃ][tælt], which does not use the onset cluster [ʃt]. The cluster [pʃ] is excluded because it is only found in *pshaw*, which is an interjection. The clusters [frw], [krw], and [trw] are excluded because they only occur in French accented pronunciation, where the transcribed [r] is pronounced as [ʁ], which is not an English sound (as discussed earlier, the English [r] is already rounded [r^w], and so the English [fr, kr, tr] are the same as [fr^w, kr^w, tr^w] or [frw, krw, trw]).

The decision on which clusters to include is not meant to be absolute. For example, *Tswana* and *bourgeois* may have alternative pronunciations in other dictionaries and so [tsw] and [ʒw] may be moved to the excluded list. On the other hand, *Buenos Aires*, which is absent in CELEX, may be found in other dictionaries and so [bw] may need to be added to the included list. However, as will be seen below, the criteria for which clusters to include are not always a crucial factor in the conclusions to be reached.

8.5.2. **The sonority-based analysis**

Most analyses of onset clusters employ the notion of sonority, which is roughly the loudness of a sound (Jespersen 1904). The basic idea is that each syllable has a sonority peak, which is the main vowel, and the sounds before the vowel should have increasing sonority. I review two such proposals: Kenstowicz (1994) and Gouskova (2004).

According to Kenstowicz (1994), English consonants and glides are divided into four degrees of sonority, shown in (34).

(34) Sonority scale for English (Kenstowicz 1994)

Sound class	Example	Sonority scale
Glides	[j, w]	4
Liquids	[l, r]	3
Nasals	[m, n]	2
Obstruents	[f, t, k, d]	1

The Minimal Sonority Distance (MSD) in an English onset cluster is two, which means that the sonority of the second sound should be at least two degrees higher than that of the first. Some examples are shown in (35).

(35) English onset clusters (Minimal Sonority Distance = 2)

Example	Sonority distance	Prediction
[fn]	1 < MSD	bad
[fl]	2 = MSD	good
[kw]	3 > MSD	good

The sonority of [f] is 1 and that of [n] is 2, so the sonority distance of [fn] is 2 – 1 = 1, which is below the MSD, so the cluster is bad (that is, no English words start with [fn]). In [fl], the sonority distance is

3 – 1 = 2, which meets the MSD, so the cluster is good. In [kw], the sonority distance is 4 – 1 = 3, which exceeds the MSD, so the cluster is also good. However, there are two well-known problems for the sonority analysis, which are shown in (36).

(36) Two problems for the sonority analysis
  a. Occurring forms that do not satisfy the MSD: e.g. [st, sn]
  b. Missing forms that satisfy MSD: e.g. [tl, dl]

A common response is to add two more assumptions, shown in (37).

(37) a. An initial [s] can be added
  b. Place dissimilation: e.g. *Alveolar-Alveolar

First, one can assume that [s] is an exceptional sound that can combine freely with other sounds, even if the sonority rise is bad, as in [st]. One might also grant [ʃ] a special status, since [ʃn] and [ʃm] occur in words like *schnapps*, *schnauzer*, and *schmaltz*, even though they violate the MSD. Second, one can assume additional restrictions on the places of the sounds in a cluster, besides the sonority requirement. For example, one can propose that the two sounds cannot both be alveolar; this will rule out [tl] and [dl]. The assumption is not without problems, though. For example, place dissimilation does not apply to [st]. In addition, alveolar pairs like [nt] (as in *ant*), [nd] (as in *and*), [tn] (as in *lateness*), [dn] (as in *sadness*), [tl] (as in *lately*), and [dl] (as in *gladly*) do not cause any dissimilation problems, presumably because they can form an alveolar geminate, in which a single alveolar place feature (or gesture) is linked to both sounds. If so, why can't [tl] and [dl] be resolved in the same way?

Leaving aside questions about place dissimilation, let us see to what extent the sonority theory makes correct predictions. Ideally, all clusters predicted to be good are found and all clusters predicted to be bad are absent. Let us begin with CR clusters. There are twenty-one choices for C (consonant) and four choices for R (approximant). Therefore there are eighty-four possible CR clusters in all, which are shown in (38), where the relevant sonority distance (SD) is indicated.

(38) Onset CR clusters (84 in all)
  Predicted to be good and found (42 cases)
    SD = 2: bl, br, dr, ʃr, fl, fr, gl, gr, kl, kr, mj, mw, nj, nw, pl, pr, sl, sr, tr, θr
    SD = 3: bj, dj, dw, ʃw, fj, gj, gw, hj, kj, kw, pj, pw, sj, sw, tj, tw, vj, vw, zj, ʒw, θj, θw

Predicted to be good but not found (27 cases)
SD = 2: ðl, ðr, ʤl, ʤr, ʃl, hl, hr, ʧl, ʧr, vl, vr, ʒl, ʒr, θl, zr;
SD = 3: bw, ðj, ðw, ʤj, ʤw, ʃi, fw, hw, ʧj, ʧw, zw, ʒj
Predicted to be bad and not found (13 cases)
SD = 0: ll, lr, rl, rr
SD = 1: lw, rj, rw, ml, mr, nl, nr
*Alveolar-Alveolar: dl, tl
Predicted to be bad but found (2 cases)
SD = 1: lj
*Alveolar-Alveolar: zl

There are two kinds of wrong predictions. First, there is over-prediction: many clusters that are predicted to be good are not found. To exclude them, one could add more restrictions. For example, one might say that affricates cannot occur with another sound, which will rule out eight clusters. However, other clusters are more difficult to account for. For example, why is [zr] missing, if we have [sr, tr, dr]? Second, there is under-prediction: there are clusters that are predicted to be bad but are found. We can keep [zl] by changing the constraint *Alveolar-Alveolar to *Alveolar[+stop]-Alveolar (an alveolar stop cannot be followed by an alveolar), but then the constraint becomes less general and less motivated phonetically. More troublesome is [lj], whose sonority distance is 1, which is below the required MSD of 2—yet the cluster is found. If we lower the MSD to 1, then we would make over-prediction even worse.

Next we examine onset CC clusters that are not CR. Twelve such clusters occur in CELEX, which are repeated in (39).

(39) Onset CC that are not CR (12 found):
sp, st, sk, sn, sm, sf, sv, ʃm, ʃn, km, kn, kv

In all the cases the sonority rise is below the MSD of 2. Therefore, such clusters must be accounted for in terms of the special status of the first C. There are three problems, though. First, although most of them are [s]+C, not all [s]+C are found. Second, it is hard to explain why [ʃ] and [k], but not others, can also appear as the first C. Third, it is hard to explain why [ʃ] and [k] do not combine with other consonants. Specifically, excluding [l, r], there are nineteen consonants that can occur in the onset. Therefore, there should be 19 [s]+C clusters, 19 [ʃ]+C clusters, and 19 [k]+C clusters. However, as shown in (40), most of them are missing.

(40) Onset CC clusters that are not CR:
[s]+C found (7 cases): sp, st, sk, sn, sm, sf, sv

[s]+C missing (12 cases): sb, sd, sg, sθ, sð, ss, sz, sʃ, sʒ, sh, stʃ, sʤ
[ʃ]+C found (2 cases): ʃm, ʃn
[ʃ]+C missing (17 cases): ʃp, ʃt, ʃk, ʃf, ʃv, ʃb, ʃd, ʃg, ʃθ, ʃð, ʃs, ʃz, ʃʃ, ʃʒ, ʃh, ʃʧ, ʃʤ
[k]+C found (3 cases): km, kn, kv
[k]+C missing (16 cases): kp, kt, kk, kf, kb, kd, kg, kθ, kð, ks, kz, kʃ, kʒ, kh, kʧ, kʤ

Some statements can be made about various subsets of the clusters. For example, [s] does not occur with voiced stops, and the affricates [ʧ] and [ʤ] do not occur in any cluster. However, there does not seem to be a simple generalization about the occurrence patterns.

Finally, we examine CCC clusters. All except [tsw] start with [s]. There does not seem to a good reason for the occurrence of [tsw]. If it is made of three sounds, why are there no other CCC clusters? If [ts] is an affricate, why do other affricates not occur with [w]? For example, one might expect [dzw] to be a simpler form than [tsw] since [dzw] has a single value for [voice] while [tsw] has two. For [s]+CC forms, there are missing forms, too, as shown in (41).

(41) Onset [s]+CC clusters:
Found: spj, spl, spr, stj, str, skj, skl, skr, skw
Missing: spw, stw

Even if we limit the CC to occurring CR and the C to voiceless stops [p, t, k], there are still two missing forms [spw] and [stw] (since we have [pw] in *puissance* and [tw] in *twin*).

In summary, the sonority-based analysis of Kenstowicz (1994) requires two additional assumptions: (a) the special status of [s] (and [ʃ]) and (b) further restrictions on sound sequences (phonotactics), such as the prohibition against two alveolar sounds. Even so, there are missing clusters that are predicted to be good and occurring clusters that are predicted to be bad.

A different version of sonority-based analysis is offered by Gouskova (2004), who uses a finer-grained sonority scale, adapted from that of Jespersen (1904). According to Gouskova, there is a universal sonority scale between adjacent sounds (not including vowels), shown in (42), where W is any glide, R any [r]-like sound, L any lateral, N any nasal, Z any voiced fricative, S any voiceless fricative, D any voiced stop, and T any voiceless stop. Clusters in the same column have the same sonority rise shown on the top row.

(42) Universal scale of sonority rise (Gouskova 2004)

0	+1	+2	+3	+4	+5	+6	+7
WW	RW	LW	NW	ZW	DW	SW	TW
RR	LR	NR	ZR	DR	SR	TR	
LL	NL	ZL	DL	SL	TL		
NN	ZN	DN	SN	TN			
ZZ	DZ	SZ	TZ				
DD	SD	TD					
SS	TS						
TT							

Although Gouskova does not discuss English, it is obvious that the MSD must be set at 3, in order to exclude the lack of [nr, mr, bm, bn, tn, tm, . . .], even though some clusters whose sonority rise is 2 or lower are found, such as [lj] and [st] (to which I return shortly).

Like Kenstowicz (1994), to account for the special status of [s] and the lack of [tl], Gouskova must make two additional assumptions, repeated in (43).

(43) a. An initial [s] can be added.
 b. Place dissimilation: e.g. *Alveolar-Alveolar

Let us consider the predictions of the theory. I begin with CR clusters. The eighty-four actual CR clusters are shown in (44), where their sonority distance (SD) is indicated. Since Gouskova does not indicate the sonority rise between an affricate and an approximant such as [ʤl], I have listed such clusters separately. These clusters are predicted to be good because their sonority rise exceeds the MSD, whether an affricate is treated as a stop or a fricative.

(44) Onset CR clusters (84 in all)

Predicted to be good and found (MSD ≥ 3; 42 cases)

SD = 3 (NW, ZR, DL)	mj, mw, nj, nw; bl, gl
SD = 4 (ZW, DR, SL)	vj, vw, zj, ʒw; br, dr, gr; fl, sl
SD = 5 (DW, SR, TL)	bj, dj, dw, gj, gw; fr, θr, sr, ʃr; pl, kl
SD = 6 (SW, TR)	fj, hj, sj, θj, θw, sw, ʃw; pr, tr, kr
SD = 7 (TW)	pj, pw, tw, tj, kj, kw

Predicted to be good but not found (24 cases)

SD = 3 (NW, ZR, DL)	ðr, vr, ʒr, zr
SD = 4 (ZW, DR, SL)	ðj, ðw, zw, ʒj; θl, ʃl, hl
SD = 5 (DW, SR, TL)	bw; hr
SD = 6 (SW, TR)	ʃj, fw, hw
Others	ʤl, ʤr, ʧl, ʧr, ʤj, ʤw, ʧj, ʧw

Predicted to be bad and not found (16 cases)

SD < 0 (RL)	rl
SD = 0 (LL, RR)	ll, rr
SD = 1 (RW, LR, NL)	rj, rw; lr; ml, nl
SD = 2 (LW, NR, ZL)	lw; mr, nr; ðl, vl, ʒl
*Alveolar-Alveolar	dl, tl

Predicted to be bad but found (2 cases)

SD = 2	lj
*Alveolar-Alveolar	zl

As in the analysis of Kenstowicz (1994), there are missing clusters that are predicted to be good. In addition, the presence of [lj] presents a problem if the MSD = 3.

Next we examine onset CC clusters that are not CR. As shown in (45), some CC clusters have a good sonority rise, but not all such clusters occur. In addition, some occurring clusters do not have a good sonority rise.

(45) Sonority analysis of onset CC that are not CR:

Predicted good and occurring: SD ≥ 3 (SN, TZ, TN)

sn, sm, ʃm, ʃn; kv; km, kn

Predicted good but missing: SD ≥ 3 (SN, TZ, TN)

fn, fm, θn, θm, hn, hm; pv, tv, pð, tð, kð, pʒ, tʒ, kʒ, pz, tz kz; pn, pm, tn, tm

Predicted bad but occurring (unless [s] is special):

SD = 2	sv
SD = 0	sf
SD = −1	sp, st, sk

Once again, many clusters are predicted to be good but missing. Even if we require the first C to be [s, ʃ, k], we will still miss [kð, kʒ, kz]. Besides, [sv, sf, sp, st, sp] occur but violate the MSD. If [s] is free of the sonority requirement, there is another problem that not all [s]+C clusters are found. In particular, there should be nineteen [s]+C clusters, but only seven occur, as shown in (46).

(46) Occurring [s] + C (7 cases): sp, st, sk, sn, sm, sf, sv

Missing [s] + C (12 cases): sb, sd, sg, sθ, sð, ss, sz, sʃ, sʒ, sh, stʃ, sdʒ

While some statements can be made about the missing clusters, there does not seem to be a simple generalization. Instead, it seems that some clusters just occur and some do not. It can also be shown that for CCC clusters Gouskova (2004) faces the same problem as Kenstowicz (1994).

In summary, sonority-based analyses require two additional assumptions: (a) restrictions on sound sequences (phonotactics) besides sonority, such as the prohibition against two alveolar sounds, and (b) the special status of [s] (and [ʃ]). Even so, there are missing clusters that are predicted to be good and occurring clusters that are predicted to be bad.

8.5.3. **The complex-sound analysis**

The basic claim of the complex-sound analysis is that, if we treat the initial C separately, as other analyses do, then all CC onsets can form a complex sound. This prediction is verifiable: we can check all occurring CC clusters and see if that is the case. The analysis does not claim that all possible complex sounds will be used in English (a point to which I return later).

As discussed in Chapter 2, the most important constraint that governs possible and impossible complex sounds is the No Contour Principle (Duanmu 1994), repeated in (47).

(47) No Contour Principle
An articulator cannot make the same feature (F) twice within one sound.

*Articulator	*Articulator	*Articulator	*Articulator
/\	/\	/\	/\
[+F][−F]	[−F][+F]	[+F][+F]	[+F][+F]

I first show that all CC clusters which do not involve a special initial C can be represented as a complex sound, and that those CC clusters which cannot be represented as a complex sound are not found word-medially. I begin with occurring CR clusters, of which there are forty-three, shown in (48).

(48) Occurring onset CR clusters (44 in all)

Different articulators (25)	bj, bl, br, dw, ʃw, fj, fl, fr, gl, gr, gw, hj, kl, kr, kw, mj, nw, pj, pl, pr, sw, tw, vj, ʒw, θw
Same articulators (19)	mw, pw, vw, dj, gj, kj, lj, nj, sj, tj, θj, sl, sr, ʃr, zl, zj, dr, tr, θr

As discussed in Chapter 2, [j] has both Coronal and Dorsal. Twenty-five of the occurring clusters have different articulators for C and R and can be represented as complex sounds. For illustration, the analysis of [bl] and [hj] are shown in (49), where only relevant features are shown.

(49) Complex sounds involving different articulators

[b]	Labial-[+stop]
[l]	Coronal-[+lateral]
[bl]	Labial-[+stop], Coronal-[+lateral]
[h]	Vocal-cords
[j]	Coronal, Dorsal-[−back]
[hj]	Vocal-cords, Coronal, Dorsal-[−back]

Since [b] is Labial and [l] is Coronal, [bl] is Labial-Coronal. Similarly, since [h] is Vocal-cords (glottal) and [j] is Coronal-Dorsal, [hj] is Vocal-cords-Coronal-Dorsal.

Next I examine the nineteen occurring clusters that involve the same articulator. All except five can be represented as complex sounds. They are shown in (50)–(53).

(50) [dj, gj, kj, lj, nj, sj, tj, zj, θj]

Example:	[s]	Coronal-[+fricative]
	[j]	Coronal, Dorsal-[−back]
	[sj]	Dorsal-[−back], Coronal-[+fricative]

(51) [mw, pw, vw]

Example:	[p]	Labial-[+stop]
	[w]	Labial-[+round]
	[pw]	Labial-[+stop, +round]

(52) [dr, tr] (affricates)

Example: [dr^w] Coronal-[−anterior, +stop, +fricative], Labial-[+round]

(53) Unaccounted for: [sl, sr, ʃr, zl, θr]

Example: [z] is Coronal-[−lateral], [l] is Coronal-[+lateral]
[θ] is Coronal-[+anterior], [r] is Coronal-[−anterior]

In (50) and (51), the sounds that share the same articulator do not have conflicting features. Therefore, they can be represented with one articulator that carries both features. For example, [p] and [w] both use Labial, but since their features under Labial are different, we can represent [pw] with one Labial articulator that carries both features.

Many people have proposed that [dr] and [tr] are affricates (Jones 1950, Abercrombie 1967, Gimson 1970, Wells 1990, Lawrence 2000). Therefore, they are represented as affricates in (52). One might wonder how a stop and an approximant can make an affricate, since affricates are usually made of a stop and a fricative (such as [t] + [s] → [ts]). However, there are well-known cases of the former. For example, [t] and [j] can make [ʧ], as in *get you* [gɛʧu] and [d] and [j] can make [ʤ], as in *did you* [dɪʤu].

There are five remaining clusters, [sl, sr, ʃr, zl, θr], that do not seem to be possible complex sounds. If they are not complex sounds, then the first C must be a special case. This is true for [s], as previous analyses agree, and possibly for [ʃ] as well. In addition, [zl] occurs in a single word, *zloty*, and so it is a special case anyway. The remaining case is [θr]. In sonority-based analyses, [θr] is a good onset, but in the present analysis it is not. The two approaches make different predictions. If [θr] is a good onset, we expect it to occur both initially and medially. If [θr] is not a good onset, we expect it to occur only initially. Indeed, in the complex sound analysis, none of [sl, sr, ʃr, zl, θr] should occur word-medially. To find out, I examined the distribution of these clusters in CELEX. The results are shown in (54).

(54) Distribution of non-initial onset clusters [sl, sr, ʃr, zl, θr]

[zl]	0	
[sr]	1	alternative form (*nursery* [nɜ:][sə][rɪ] or [nɜ:][srɪ])
[ʃr]	5	prefixed (*enshrine*)
[sl]	35	compounds without space (*oversleep*)
	7	prefixes (*asleep*)
	3	suffixed (*bracelets*)
	5	after short vowel (*Muslim* [mʊ][slɪm] → [mʊs][lɪm])
[θr]	15	compounds (*overthrow*)
	20	prefixed (*misanthropic*)
	8	suffixed (*anthropologist*)
	2	after VN (*anthrax* [æn][θræks] → [æ̃θ][ræks])
	1	*arthritis* [ɑ:][θraɪ][tɪs] (American: [ɑɚ][θraɪ][tɪs])

There is no occurrence of [zl] in non-initial positions. There is one case of non-initial [sr], in the word *nursery*. The word has two pronunciations, where [sr] is an alternative form, which need not be included (the word could also be analyzed as [nɜs][rɪ] or [nɝs][ri]). There are five cases of [ʃr], all of which are prefixed words, where [ʃr] is word-initial without the prefix (the word *enshrine* could also be analyzed as [ĩʃ][rain]). The cluster [sl] occurs in compounds and prefixed words; it also occurs in three suffixed words, where it lies across morpheme boundaries. In the remaining five cases [sl] lies after a short vowel, and therefore the [s] can be syllabified as the coda of the short vowel. Excluding compounds, the cluster [θr] only occurs in three words (and their related forms): *anthrax*, *anthropology*, and *arthritis*. The first two can be analyzed with a nasalized vowel ([æn][θr . . .] → [æ̃θ][r . . .]), following Malécot (1960), Bailey (1978), and Fujimura (1979). The only remaining case, then, is *arthritis*. The

British pronunciation [ɑː][θraɪ][tɪs] is not a problem, because it can be analyzed as [ɑθ][raɪ][tɪs], where there is no [θr] onset. The analysis of the American pronunciation is less obvious. In feature terms, [ɑɚ] can be represented as [ɑ˞], as shown in (55).

(55) [ɑ] Labial-[−round], Dorsal-[+back, +low]
[ɚ] Coronal-[+anterior], Dorsal-[+back]
[ɑ˞] Labial-[−round], Coronal-[+anterior], Dorsal-[+back, +low]

The retroflex vowel [ɚ] is usually treated as a mid vowel, which ought to be [−high, −low], which ought to conflict with the [+low] feature of [ɑ]. However, because [ɚ] is the only retroflex vowel, it does not need to be specified for height (nor does it need to be specified for [round]). Therefore, there need not be a feature conflict. The result is [ɑ˞], a retroflex [ɑ]. Whether this is the correct analysis I shall leave open.

In summary, most onset CR clusters can be represented as a single complex sound. Those that cannot are found in word-initial position only, where the first C can be treated as lying outside the onset. There is no compelling evidence, then, of true CR clusters that cannot be represented as a complex sound.

Finally, I discuss CCC clusters and CC clusters that are not CR, repeated in (56) and (57), all of which occur in word-initial position only.

(56) Onset CCC clusters (10, initial only): str, skr, skw, spr, spl, stj, skj, spj, skl, tsw
Without [s]: tr, kr, kw, pr, pl, kj, pj, kl
[tsw]: Coronal-[+stop, +fricative], Labial-[+round]

(57) Other onset CC (12, initial only): st, sp, sk, sn, sm, sf, sv, ʃm, ʃn, km, kn, kv
*[sm] [s] is [−nasal]
[m] is [+nasal]

Nine of the CCC clusters start with [s]. If we exclude [s], we get nine CR clusters, all of which have been analyzed. The cluster [tsw] can be analyzed as two sounds, an affricate [ts] and [w]. Since the sounds use different articulators, they can be represented as a complex sound [ts^w]. Most CC clusters in (57) cannot be represented as a complex sound, as exemplified with [sm]. However, the first C need not be in the onset. This is possible because these clusters only occur in word-initial position, where an extra C can be tolerated. In English the extra C is usually [s] and sometimes [ʃ] (*schnauzer*, *schmooze*, *schlock*), but other consonants are occasionally possible in borrowed words (some not included in CELEX), such as [km] in *Khmer*, [kn] in *Knesset*, [kv] in *kvass*, [vl] in *Vladimir*, and [vr] in *vroom*.

We have seen that the sonority-based analysis makes over-predictions, in the sense that it predicts more good clusters than are found. For example, the analysis of Kenstowicz (1994) predicts that [ðl, ðr, hl, hr, vl, vr, ʒl, ʒr, θl, zr, . . .] are good, but they are not found. Similarly, the analysis of Gouskova (2004) predicts that [ðr, vr, ʒr, zr, . . .] are good, which are not found. The complex-sound analysis also over-predicts, in that not all possible complex sounds are found. For example, [hr], [hw], [zw], [mw], and [dzw] are possible complex sounds, but they are not found in CELEX or used widely. For example, [hw] is only used by some American speakers in words like *what* and *white* (others use [w]), [mw] is only found in borrowed French words (such as *moiré*), and [zw] is only used in words like *zwieback* and *Venezuela*, where alternative pronunciation or syllabification is possible (e.g. [swi:][bæk] and [vɛ][nɪz][weɪ][lə]). In (58) I show a comparison between the sonority-based analysis and the complex-sound analysis.

(58) Comparison between the sonority analysis and the complex-sound analysis

	Sonority	Complex-C	Example
Over-prediction	yes	yes	[hr], [hw], [zw], [zr], [dzw]
Under-prediction	yes	no	[lj]
Special C	anywhere (?)	initial	[st], [sf], [sl], [ʃm], [ʃn], [θr]
Place dissimilation	yes (?)	no	*[ʃj], *[tl]

While both approaches over-predict, there are three differences. First, the sonority-based analysis also under-predicts, in the sense that a cluster that is predicted to be bad can still occur, such as [lj]. In contrast, the complex-sound analysis does not have this problem. Second, in the sonority-based analysis, the special C is often thought to be part of the onset. If so, one expects the special C to occur in non-initial onset clusters, too, which it does not. In contrast, in the complex-sound analysis, the special C can only occur word-initially, which is true. Third, the sonority-based analysis assumes place dissimilation in order to eliminate some clusters. In contrast, in the complex-sound analysis, there is no need for the assumption, because there is just one slot in the onset, which cannot hold two separate sounds—unless the two sounds can merge into one, in which case there is no sequence of identical place features (articulators).

I would like to make two further remarks about the issue of over-prediction. First, it is a problem only if all possible syllables are needed. If we do not need so many, as I shall argue in Chapter 9, then some syllables will be left unused, even if the syllables are phonologically well formed. Second, supposing there are more words than possible syllables, we might still not expect all syllables to be used with equal frequencies. Instead, which sounds or syllables a language uses can be influenced by historical accidents, such as from which languages it has imported words. In addition, when new words are created, it is unlikely that the creator will try to fill unused syllables first so as to make the phonological system more balanced. Rather, it is more likely that frequent sounds or syllables would come to the mind of the creator more readily, and as a result the imbalance of usage becomes more skewed.

8.5.4. **Summary of onset analysis**

I have shown that, excluding a word-initial consonant, all English onset clusters can be represented as a complex sound. This result supports the CVX theory, in which there is only one onset slot. The complex-sound analysis is better than the sonority-based analysis in that the former does not have the problem of under-prediction. In addition, the complex-sound analysis does not need to assume place dissimilation. Moreover, the complex-sound analysis makes a better prediction of onset clusters in word-medial positions.

8.6. SUMMARY

On the basis of an exhaustive study of the CELEX lexicon of English, I have shown that the maximal syllable in word-medial positions is CVX (CVV or CVC). This size is smaller than has been conceived in previous literature, including Giegerich (1985), for whom the maximal English syllable is CCVX, Borowsky (1989), who allows occasional rhymes that are larger than VX, Pierrehumbert (1994), who allows VVC, VGC, and VNC rhymes as well as onset clusters, and Lowenstamm (1996) and Scheer (2004), for whom a CVX is CVCV.

In word-initial positions, an extra C may occur, which is usually [s] or [ʃ] but can occasionally be another consonant, such as [θ] in [θr] and [k] in [km, kn]. In word-final positions, there can be up to three extra Cs after VX. The first can be almost any consonant. It is supported by a potential vowel and anti-allomorphy (Chapter 3). The two Cs after VXC must be [s, z, t, d, θ], which are either suffixes or suffix-like and are covered for by the "affix rule" (Chapter 3). Since the extra Cs can be accounted for by morphology, there is no need to assume that they are inside a syllable. Nor is there any evidence that the CVX size needs to be expanded.

Unlike other analyses, which rarely exploit feature theory, the present study uses feature theory to solve two problems. First, some exceptions to the VX limit on rhymes can be explained, such as VNC (now ṼC) and VVN.C (now ṼṼ.C). Second, feature theory offers a simpler analysis of onset clusters, according to which only those onset clusters that can form a complex sound are found (Chapter 2).

The complex-sound analysis also has another advantage. It predicts that in a CR cluster, such as [kl, kr, kj, kw], the R is voiceless, a fact well known to phoneticians (e.g. Ladefoged 1982: 84). This is because CR is a single sound, in which there can only be one voice value. In the sonority-based analysis, a CR cluster is two sounds and a separate rule must be proposed for the voice change in R.

The fact that the maximal syllable size is considerably smaller than it appears at first sight has theoretical implications. First, CVX is a minimal size for a heavy syllable that most phonologists already assume, and if CVX is also the upper limit for languages that are rich in consonant clusters, such as English, there is a good chance that it is the upper limit universally. Any smaller alternative, such as the CV-only theory (Lowenstamm 1996, Scheer 2004), must make excessive use of empty sounds that do not have phonetic realizations (Chapter 3). Second, CVX offers a guide for possible ways of syllabification. For example, Hoard (1971), Bailey (1978), and Wells (1990) propose that the stressed syllable can take as many consonants as possible. Thus, *dolphin* is [dɒlf.ɪn] and *cauldron* is [kɔːldr.ən]. In the CVX analysis, the words are [dɒl.fɪn] and [kɔl.drən], because [dɒlf] and [kɔːldr] are not possible syllables. The same criticism applies to extra-sized syllables in Selkirk (1982) and Hammond (1999), such as *fealty* [fɪlt.i], *bulky* [bʌlk.i], and *alcove* [ælk.kov]. Third, the CVX limit provides strong evidence for the existence of the syllable as a phonological entity, despite the lack of a precise definition of what syllables are, or the lack of speaker intuition for where word-medial syllable boundaries are (but see syllabification based on the Weight-Stress Principle in Chapter 3).

9

English II: syllable inventory and related issues

In this chapter I discuss additional issues on the English syllable, including possible and actual sizes of the syllable inventory, diphthongs, and voicing in coda consonants. I start with the inventory of consonants and vowels.

9.1. SOUND INVENTORY

I focus on General American English, which is based on the pronunciation of the American Midwest. There are some twenty consonants (excluding glides), given in (1). Sounds with limited use are shown in parentheses.

(1) Consonants in American English

Labial	Dental	Alveolar	Alveo-palatal	Retroflex	Velar	Glottal
p, b		t, d			k, g	(ʔ)
f, v	θ, ð	s, z	ʃ, ʒ			h
		tr, dr	ʧ, ʤ			
m		n			ŋ	
		l		r		

Initial only: h, (ʔ)
Final only: ŋ, (ʒ)

The contrast between [p, t, k] and [b, d, g] could be one of [voice], or it could be one of [aspirated] (Iverson and Salmons 1995, Jessen and Ringen 2002, Vaux and Samuels 2005).

Phonetically [ʧ, ʤ, tr, dr] are all affricates, although [ʧ, ʤ] are often listed as single sounds but [tr, dr] are listed as consonant clusters. In the onset the affricates behave like clusters in that they do not combine with another consonant.

Sounds that are only used by some people are not included, such as [x] in [bɑx] *Bach* and [lɒx] *Loch (Ness)*, for which most American speakers use [k].

The glottal stop is used in vowel-initial words, such as [ʔist] *east*, which differs from [jist] *yeast*. It also occurs in some interjections, such as [ʔʌʔo] *uh-oh*. One could argue that interjections are marginal words and that in words like *east* the glottal stop is not an intended sound. This is supported by the fact that when words like *east* follow another word there is no glottal stop, as in *run east* or *travel east*. If so, the contrast between *east* and *yeast* can be [ist] vs. [jist], and there is no need to include [ʔ] as an independent sound. In syllable-coda positions [ʔ] can occur as a variant of a stop for some speakers, such as [bʌʔ] for [bʌt] *but*.

The sound [r] is not a trill but an approximant. In addition, it is [+round] in syllable onset position. A more accurate IPA transcription is [ɹʷ]. For convenience I use the more common symbol, [r] (or [rʷ]). In syllable coda position [rʷ] does not occur; instead, we find [ɚ].

The velar nasal [ŋ] does not occur word-initially. One might argue that it results from a combination of [nk] or [ng], which I do not pursue.

The sound [h] does not occur in syllable-final position. It is therefore in complementary distribution with [ŋ]. However, since [h] and [ŋ] are phonetically quite different, they are not treated as allophones of the same phoneme.

The sound [ʒ] occurs initially in the word *genre* for some speakers. Other speakers may pronounce *genre* with [ʤ] and do not use [ʒ] word initially. Indeed, [ʒ] does not occur very often elsewhere either, although *garage* and *measure* are very common words.

What we see is that not all sounds occur with equal frequency. In fact, there is a large range of difference. Consider the data in (2), which shows the number of occurrences (in parentheses) of some English sounds, calculated by myself from 52,447 uninflected words in the CELEX English "lemma" lexicon (Baayen et al. 1993). Components of affricates and diphthongs are not counted separately. For example, the frequency of [ɪ] does not include that in [eɪ] and the frequency of [t] does not include that in [ʧ].

(2) Least frequent C ʒ (252), ð (606), θ (1,377), ʧ (1,967), ʤ (2,939),
Most frequent C k (17,975), l (22,901), s (23,092), t (25,562), n (25,797)
Most frequent V ɛ (9,929), ə (23,693), ɪ (38,123)

One can see that [ʒ] and [ð] are the least frequent consonants in English, whose occurrences are a small fraction of those of the most common consonants, such as [k, l, s, t, n], or the most common vowels, such as [ɛ, ə, ɪ].

The vowels in American English are shown in (3). I assume the simplest feature system, with two degrees of frontness and three degrees of height. Also, [tense] can be interpreted as [advanced] of Tongue-root (Halle and Stevens 1969).

(3) Vowels in American English

		−back		+back	
		−round	+round	−round	+round
+high	+tense	i *beat*			u *boot*
	−tense	ɪ *bit*			ʊ *book*
−high, −low	+tense	e *bait*		ɝ (ɚ) *Berber*	o *boat*
	−tense	ɛ *bet*		(ə) ʌ *abut*	
+low		æ *bat*		ɑ *spa*	ɒ *law*

Diphthongs

iɚ	eɚ	uɚ	oɚ	aɚ	ai	au	oi	(ou	ei)
beer	*bear*	*tour*	*or*	*are*	*buy*	*how*	*boy*	*(go*	*day)*

In general, vowels are more variable than consonants. For example, some speakers of American English use the same vowel for *spa* and *law*.

In diphthongs that end in [ɚ], such as *beer* [biɚ], *bear* [beɚ], and *or* [oɚ], I follow Clark and Hillenbrand (2003) and assume that the main vowel is tense, rather than lax.

The vowel [ə] is the unstressed version of [ʌ] and [ɚ] is the unstressed version of [ɝ]. One can relate [ɝ] and [ɚ] to the consonant [r] (or [r^{w}]), because when [ɝ] or [ɚ] is followed by a vowel, [r] will appear in between, as in *fur and*... [fɝ rənd . . .] and *beer and*... [biɚ rənd . . .]. Therefore, some linguists use [r] for [ɚ] or for both [ɝ] and [ɚ], as shown in (4).

(4)

	beer	*father*	*fur*	*real*
Present	biɚ	fɑðɚ	fɝ	r^{w}il
Ladefoged (2006)	bɪr	fɑðɚ	fɝ	ril
Hammond (1999)	bir	fɑðr̩	fr̩	ril

The final sound in *beer* is phonetically more similar to that in *father* than to the first sound in *real*. In this regard, the present representation is phonetically more accurate than those of Hammond (1999) and Ladefoged (2006).

The tense vowels [o] and [e] are monophthongs when they are followed by [l] or [ɚ], as in [oɚ] *or*, [gol] *goal*, [eɚ] *air*, and [pel] *pale*. They are often diphthongs in open syllables, as in [gou] *go* and [dei] *day*.

There are different ways of transcribing the diphthongs [ai, au, oi, ei, ou], even if we ignore the difference between a high vowel and a glide (e.g. [ai] vs. [aj]). One question is whether we should use tense or lax vowels for each half of a diphthong. Some common options are shown in (5).

(5)

ai	aɪ		
au	aʊ		
oi	oɪ	ɔɪ	ɔi
ei	eɪ	(ɛɪ)	
ou	oʊ		

For each of the five sets of symbols, there is no contrast among the members of the set. So how can we decide which form is better? One might argue that, since each half of a diphthong has about the same duration as a lax vowel, one ought to use lax vowels for both parts of a diphthong, such as [əʊ] for the vowel in *go* in British English. However, no one uses [ɛɪ] for British or American English, probably because the starting point is higher than [ɛ]. One might also point out that the ending point of a diphthong does not quite reach the height of [i] or [u], and so one might use [ɪ] or [ʊ] for the end point. However, it is possible that the ending target of a diphthong is a tense high vowel, and owing to the lack of time the target is not quite reached. There are two other reasons against using lax vowels. First, a diphthong is like a tense vowel, and so we should represent diphthongs with tense vowels. Second, the lax vowels [ɪ] and [ʊ] do not occur in word-final position in American English, but diphthongs can. Therefore, it is more consistent to use tense vowels to end a diphthong. For now I use tense vowels to represent diphthongs, although I shall argue later that the feature [tense] is not contrastive in diphthongs.

Most non-low vowels show a contrast in the feature [tense]. A common test for [tense] in English is the occurrence in a stressed open syllable, such as a CV word. Tense vowels can occur in a stressed CV word but lax vowels cannot. By this test, the vowel in *law* is a tense vowel. In many studies, such as Kenyon and Knott (1944), Chomsky and Halle (1968), Giegerich (1992), Hammond (1999), and Ladefoged (2006), *law* is transcribed as [lɔ]. In feature analysis (e.g. Chomsky and Halle 1968, Giegerich 1992, Hammond 1999), the vowel in *law* is a low vowel. Since [ɔ] is often used for a mid vowel, I use [ɒ] for the vowel in *law* instead. On a related note, the first part of the diphthong in *boy* seems closer to the vowel in *go* and *or* than to the vowel in *law*, and therefore [oi] is a better transcription than [ɔi]. Like [ɒ], the low vowel [ɑ] is a tense vowel, since it can occur in *spa* and *bra*. The low vowel [æ] is probably a tense vowel, too, since it can occur in a word like *yeah* [jæ], although such words are marginal.

In feature theory diphthongs are made of two vowels each. Thus, [oi] is made of [o] and [i], [ou] of [o] and [u], and [ei] of [e] and [i]. On the other hand, the vowel [a] in the diphthong [ai] and [au] does not occur by itself. I will discuss this point later.

The glides [j] and [w] are not listed as separate phonemes. Instead, they are treated as variants of the high vowels [i] and [u] when they occur in the onset of a syllable. For example, [wai] *why* can be represented as [uai] where [u] is in the onset, [ju] *you* can be represented as [iu] where [i] is in the onset, and [wi] *we* can be represented as [ui] where [u] is in the onset. Similarly, *east* can be represented as [ist] where there is no onset, and *yeast* can be represented as [iist] where the first [i] is in the onset.

9.2. UNUSED SYLLABLES IN ENGLISH

In this section I discuss the number of possible monosyllables in English in comparison with the number of actual monosyllables found in the CELEX lexicon. It is well known that, while there are restrictions within the onset or within the rhyme, there are few restrictions between onsets and rhymes (Kessler and Treiman 1997). Therefore, we can treat the number of possible syllables as the product of the number of possible onsets and the number of possible rhymes. In addition, let us ignore words with syllabic consonants, of which there are six. Finally, let us limit possible rhymes to within VX, which includes V, VV, and VC.

The only V rhyme found is [ə]. VV rhymes include tense vowels and diphthongs. In VC rhymes, C can be any consonant that can occur in the coda, and V can be either lax or tense, assuming that the length of a tense V is flexible (Pike 1947, Giegerich 1985; Chapter 8 above). For ease of calculation I follow CELEX and treat [ʧ, ʤ] as single Cs but [tr, dr] as CC clusters. In addition, a glide before the nuclear vowel is counted as C and the offglide of a diphthong is counted as V. In (6) I list possible V, VV, and VC rhymes. Since [e, o] are often monophthongs before C in American English, I have listed them as monophthongs.

(6) V, VV, and VC rhymes in English
V and VV 1 + 7 + 8 = 16
ə; i, u, e, ɝ, o, ɑ, ɒ; iɚ, eɚ, uɚ, oɚ, aɚ, ai, au, oi
VC (6 + 7) × 21 = 273 (including tense V)
ɪ, ʊ, ɛ, ʌ, æ, ə; i, u, e, o, ɝ, ɑ, ɒ
p, b, t, d, k, g, f, v, θ, ð, s, z, ʃ, ʒ, ŋ, m, n, l, r, ʧ, ʤ

The list of possible onsets is shown in (7). A sample word is given for each CC and CCC clusters, based on discussions in Chapter 8.

(7) Onsets in English
C = 23
p, b, t, d, k, g, f, v, θ, ð, s, z, ʃ, ʒ, h, m, n, l, r, ʧ, ʤ; j, w
CC = 56
bj (*beauty*), bl (*black*), br (*bring*), dj (*duty*), dr (*dry*), dw (*dwell*), ʃm (*schmaltz*), ʃn (*schnitzel*), ʃr (*shrink*), ʃw (*schwa*), fj (*few*), fl (*fly*), fr (*fry*), gj (*argue*), gl (*glad*), gr (*green*), gw (*Gwen*), hj (*huge*), kj (*cute*), kl (*class*), km (*Khmer*), kn (*Knesset*), kr (*cry*), kv (*kvass*), kw (*quick*), lj (*volume*), mj (*mule*), mw (*moiré*), nj (*news*), nw (*peignoir*), pj (*pure*), pl (*plot*), pr (*price*), pw (*puissance*), sf (*sphere*), sj (*suit*), sk (*sky*), sl (*sleep*), sm (*smack*), sn (*snake*), sp (*spot*), sr (*Sri Lanka*), st (*stop*), sv (*svelte*), sw (*swim*), tj (*tube*), tr (*try*), tw (*twin*), vj (*view*), vw (*reservoir*), zj (*presume*), zl (*zloty*), ʒw (*bourgeois*), θj (*enthuse*), θr (*three*), θw (*thwart*)
CCC = 10
str (*string*), skr (*screen*), skw (*square*), spr (*spring*), spl (*splash*), stj (*studio*), skj (*skew*), spj (*spew*), skl (*sclerosis*), tsw (*Tswana*)

If any onset can combine with any rhyme, the total number of possible syllables in English is 24,276, shown in (8). I added one to the number of onsets because a syllable may have no onset.

(8) Possible syllables in English: Onsets × Rhymes = Total
(1 + 23 + 56 + 10) × (16 + 273) = 26,010

The actual number of monosyllables in CELEX is much smaller. The data are shown in (9), where 1,259 VC rhymes contain a tense V.

(9)	All monosyllables	6,760
	No-repeat pronunciations	3,801 (many have final consonant clusters)
	V, VV, and VC rhymes	2,574 (1,259 VC rhymes have tense V)

The actual number of monosyllables with V, VV, and VC rhymes is hardly one tenth of all possible ones. It is interesting to ask why. An examination of onset clusters shows that many of them are unproductive. For example, some onsets only occur in occasional foreign words, such as [tsw] in *Tswana*, and some only combine with a specific vowel, such as Cj, which only precedes [u]. If we exclude such onsets, the number of onsets is reduced by one third, as shown in (10), and the number of possible syllables is also reduced by one third, as shown in (11).

(10) Onsets in English, excluding uncommon ones

C 23 – 1 = 22

Included: p, b, t, d, k, g, f, v, θ, ð, s, z, ʃ, h, m, n, l, r, ʧ, ʤ; j, w

Excluded: ʒ

CC 56 – 26 = 30

Included: bl (*black*), br (*bring*), dr (*dry*), dw (*dwell*), ʃm (*schmaltz*), ʃn (*schnitzel*), ʃr (*shrink*), ʃw (*schwa*), fl (*fly*), fr (*fry*), gl (*glad*), gr (*green*), gw (*Gwen*), kl (*class*), kr (*cry*), kw (*quick*), pl (*plot*), pr (*price*), sf (*sphere*), sk (*sky*), sl (*sleep*), sm (*smack*), sn (*snake*), sp (*spot*), st (*stop*), sw (*swim*), tr (*try*), tw (*twin*), θr (*three*), θw (*thwart*)

Excluded: bj (*beauty*), dj (*duty*), fj (*few*), gj (*argue*), hj (*huge*), kj (*cute*), km (*Khmer*), kn (*Knesset*), kv (*kvass*), lj (*volume*), mj (*music*), mw (*moiré*), nj (*news*), nw (*peignoir*), pj (*pure*), pw (*puissance*), sj (*suit*), sr (*Sri Lanka*), sv (*svelte*), tj (*tube*), vj (*view*), vw (*reservoir*), zj (*presume*), zl (*zloty*), ʒw (*bourgeois*), θj (*enthuse*)

CCC 10 – 4 = 6

Included: str (*string*), skr (*screen*), skw (*square*), spr (*spring*), spl (*splash*), skl (*sclerosis*)

Excluded: stj (*studio*), skj (*skew*), spj (*spew*), tsw (*Tswana*)

(11) Possible syllables in English, excluding uncommon onsets

$(1 + 22 + 30 + 6) \times (16 + 273) = 17{,}051$

The number of possible syllables is still several times the number of actual syllables. If we focus on syllables with VC rhymes where V is lax, we still get a similar ratio. The number of such VC rhymes is calculated in (12). I have excluded [r, ʒ] from the coda position, because they are rare (CELEX is based on British English). The CELEX vowel [O] can be [ɒ] in American English, as in *off* and *law*, or [ɑ], as in *hot* and *bomb*. I have left [O] as it is.

(12) VC 7 × 19 = 133 (excluding [r, ʒ])
ɪ, ʊ, ɛ, ʌ, æ, ə, O (ɑ, ɒ)
p, b, t, d, k, g, f, v, θ, ð, s, z, ʃ, ŋ, m, n, l, ʧ, ʤ

The number of possible syllables with VC rhymes is calculated in (13), along with the actual number of such syllables found.

(13) Syllables with VC rhymes, where V is a lax vowel
Possible number: (1 + 22 + 30 + 6) × 133 = 7,847
Actual number: 1,069

As seen before, the number of possible syllables is several times the number of actual syllables.

In (14) I summarize the results, where [V:C] is a VC rhyme with a tense vowel. The number of "possible" syllables is the product of the number of actual onsets and the number of actual rhymes. An "actual" onset or rhyme is any that is found in CELEX. A "common actual" onset or rhyme is any that is found in CELEX, excluding unproductive ones, such as those that only occur in foreign words (e.g. [km] in *Khmer* or [zl] in *zloty*), and those that only precede a specific vowel (e.g. [ʒ] and all onset clusters that end in [j]).

(14) Possible and actual monosyllables in English:

Onset	Rhyme	Possible syllables	Actual syllables	Ratio
All actual	V, VV, VC, and V:C	26,010	2,574	10.1
Common actual	V, VV, VC, and V:C	17,051	2,574	6.6
Common actual	VC	7,847	1,069	7.3

The large number of unused syllables requires an explanation, which I discuss next.

9.3. ACCOUNTING FOR UNUSED SYLLABLES

Some unused syllables can be described with phonological generalizations. For example, there are constraints on word-initial and word-final consonant clusters (Fudge 1969). It is also clear that some unused syllables are accidental gaps, in the sense that English could have used them and English speakers often consider them to be possible words. For example, neither [bɪk] nor [blɪk] occurs in CELEX, although they both sound fine to English speakers. Similarly, the lack of [θɪg] is probably an accidental gap, since there is [θɪk] *thick*, showing that [θ] can precede [ɪ], and [bɪg] *big*, showing that [ɪ] can precede [g].

For unused syllables with real onsets and real rhymes, it is not always clear whether their absence is due to phonological restrictions or accidental gaps. It seems to me that the latter is often the case. Let us take a closer look at syllables with VC rhymes, where V is a lax vowel. The possible number of VC combinations is shown in (15), along with the actual number found in monosyllables. I have excluded [r, ʒ] from the coda because [r] does not occur in British pronunciation and [ʒ] occurs quite rarely. Also, the CELEX [O] is left as it is, which can be [ɑ] or [ɒ] in American English.

(15)	Possible VC:	7 × 19 = 133
		ɪ, ʊ, ɛ, ʌ, æ, ə, O (ɑ, ɒ)
		p, b, t, d, k, g, f, v, θ, ð, s, z, ʃ, m, n, ŋ, l, ʧ, ʤ
	Actual VC:	101 (excluding [ʌx] *ugh*)

There are 32 missing VC rhymes, because not all codas are used after each lax vowel. The data are shown in (16), where unused codas are shown in parentheses.

(16) Used and unused codas in VC rhymes

	Used	Unused
ɪ	19	0
æ	18	1 (ð)
O	18	1 (ð)
ʌ	17	2 (ð, θ)
ɛ	17	2 (ð, ŋ)
ʊ	8	11 (p, b, g, v, θ, ð, z, m, n, ŋ, ʤ)
ə	4	15 (p, t, d, k, g, f, θ, ð, z, ʃ, m, n, ŋ, ʧ, ʤ)
All	101	32

About half of the missing VC rhymes happen with [ə]. The four VC rhymes found with [ə] are [əl] *'ll*, [əm] *'em*, [kəs] *cos* (reduced form of *because*), and [əv] *'ve*, each occurring in one word. These are word forms that cannot be stressed. For words that can be either stressed or unstressed, such as *of* and *them*, the stressed form is listed as the primary one and with a different vowel. Thus, the lack of VC rhymes with [ə] is not due to any articulatory constraint against [ə] + C, but due to the fact that English does not have many words that cannot be stressed.

A third of the missing VC rhymes happen with [ʊ]. It can be seen that there are no [ʊp, ʊb, ʊf, ʊm] where the coda is labial. In addition, the only labial coda after [ʊ] is found in [wʊf] *woof*. If we exclude *woof*, there is no [ʊ] followed by a labial, which seems to show a dissimilation effect against two labial sounds in the rhyme. However, there are labial pairs [Op, Ob, Of, Ov, Om]. In addition, the dissimilation effect can be avoided by merging the two labial gestures into one long gesture. Therefore, the lack of [ʊ] plus a labial C is still unexplained.

As an alternative analysis, one might note that the missing codas for [ʊ] are all voiced except [p, θ]. In addition, [ʊp] is found in [ʊps] *oops*. One might suggest, therefore, that there is a phonetic constraint against a voiced coda after [ʊ]. It is well known that a vowel is longer before a voiced coda. If so, [ʊ] would be long [ʊ:] before a voiced coda, in which case it would be harder to distinguish from [u:]. In the case of [O], it is the only round low vowel, and so there is no confusion with another vowel even if [O] is lengthened before a voiced coda. However, [ʊ] does occur before [l], as in *bull*, *pull*, *full*, and *wool*. Indeed, there are contrastive pairs of *pull*/*pool* and *full*/*fool*, which show that it is possible for [ʊ] to have a voiced coda (unless one represents *bull*, *pull*, *full*, and *wool* as [bl̩, pl̩, fl̩, wl̩], with a stressed syllabic [l]).

There is little to say about other missing VC rhymes, except that [ð] seems to be the most unproductive coda in VC, which only occurs in the lone word *with*.

We noted earlier that the actual number of monosyllables with VC rhymes is less than one seventh of all possible ones. The unused VC rhymes can only account for a small fraction of all unused syllables. Therefore, most unused syllables are due to the lack of onset-rhyme combinations. In particular, excluding unproductive onsets such as [zl], [vw], [sr], [tsw], Cj, CCj, etc., there are still fifty-nine onsets (twenty-two C onsets, thirty CC onsets, six CCC onsets, and one count for lack of an onset), but even the most productive rhyme hardly occur with half of them. In (17) I list the 101 VC rhymes and the number of onsets they occur with (in parentheses).

(17) VC rhymes and the number of onsets they occur with
[ɪl] (29), [ɪp] (26), [æk] (25), [ɪt] (25), [Ot] (25), [æt] (24), [ɪk] (23), [Op] (23), [æg] (22), [æʃ] (22), [æp] (21), [æm] (20), [ʌm] (20), [ɪn] (19), [Ok] (19), [ʌg] (19), [æd] (18), [æn] (18), [ɛd] (18), [Ob] (18), [Od] (18), [ʌf] (18), [ɛl] (17), [ɛn] (17), [ɛt] (17), [ɪŋ] (17), [Og] (16), [ʌb] (16), [ʌt] (16), [æb] (15), [ɪm] (15), [ɪg] (14), [ɪʧ] (14), [ʌk] (14), [ɛs] (13), [Os] (13), [Oʃ] (13), [ʌn] (13), [ʌʃ] (13), [æŋ] (12), [ɛk] (12), [ɪf] (12), [ɪb] (11), [ɪz] (11), [ʌd] (11), [æʧ] (10), [ɛʤ] (10), [ɪd] (10), [On] (10), [Oŋ] (10), [ʌs] (10), [ʌʤ] (9), [ʌŋ] (9), [ɛg] (8), [ɛʧ] (8), [ɪs] (8), [Of] (8), [Oʧ] (8), [ʊk] (8), [ʌl] (8),

[æs] (7), [ɛf] (7), [ɛm] (7), [Ol] (7), [Om] (7), [ʌʧ] (7), [ɛp] (6), [ɪʃ] (6), [Oθ] (6), [ʊd] (6), [Oʤ] (5), [ʌp] (5), [ʌv] (5), [æl] (4), [ɛb] (4), [ɛʃ] (4), [ɪʤ] (4), [ɪθ] (4), [ɪv] (4), [ʊl] (4), [ʊʃ] (4), [ʊt] (3), [ʌz] (3), [æʤ] (2), [æf] (2), [æv] (2), [æz] (2), [ɛθ] (2), [ʊʧ] (2), [æθ] (1), [əl] (1), [əm] (1), [əs] (1), [əv] (1), [ɛv] (1), [ɛz] (1), [ɪð] (1), [Ov] (1), [Oz] (1), [ʊf] (1), [ʊs] (1)

Total VC rhymes	101
Total syllables with VC rhymes	1,069
Possible onsets per rhyme	59
Average onsets per VC rhyme	10.6

Although for each rhyme there are fifty-nine possible onsets, the average VC rhyme occurs with just 10.6 of them. Let us take a close look at what syllables are missing. The most productive VC rhyme is [ɪl], shown in (18), and one of the least productive rhymes is [ɛθ], shown in (19).

(18) Unused syllables with [ɪl] (59 – 29 = 30):
vɪl, θɪl, ðɪl, zɪl, ʃɪl, lɪl, jɪl; blɪl, dwɪl, ʃmɪl, ʃnɪl, ʃwɪl, flɪl, glɪl, gwɪl, klɪl, krɪl, plɪl, prɪl, sfɪl, slɪl, smɪl, snɪl, θwɪl; strɪl, skrɪl, skwɪl, sprɪl, splɪl, sklɪl

(19) Unused syllables with [ɛθ] (59 – 2 = 57):
ɛθ, bɛθ, ʤɛθ, fɛθ, gɛθ, hɛθ, kɛθ, mɛθ, nɛθ, pɛθ, rɛθ, sɛθ, tɛθ, ʧɛθ, wɛθ, vɛθ, θɛθ, ðɛθ, zɛθ, ʃɛθ, lɛθ, jɛθ; drɛθ, frɛθ, grɛθ, kwɛθ, skɛθ, spɛθ, ʃrɛθ, stɛθ, swɛθ, trɛθ, θrɛθ, twɛθ, blɛθ, dwɛθ, ʃmɛθ, ʃnɛθ, ʃwɛθ, flɛθ, glɛθ, gwɛθ, klɛθ, krɛθ, plɛθ, prɛθ, sfɛθ, slɛθ, smɛθ, snɛθ, θwɛθ; strɛθ, skrɛθ, skwɛθ, sprɛθ, splɛθ, sklɛθ

I do not see any good phonological reason to rule out such syllables, and English speakers might find them to be possible potential words. In fact, some of them occur in words not collected by CELEX, such as [mɛθ] *meth*, [bɛθ] *Beth*, and [sɛθ] *Seth*.

What is the reason, then, for the large number of unused syllables in English? The answer, it seems to me, is that English simply does not need so many words. Let us consider how many words are used in English in comparison with how many possible syllables there are.

The CELEX lexicon contains 160,595 entries, of which 52,447 are uninflected (called "lemmas"). The uninflected list does include derived word forms. For example, *catch* (verb), *catch* (noun), *catching* (adjective), *catcher* (noun), *catchy* (adjective), and *catchily* (adverb) are listed as separate entries. If we count the number of morphemes, or morphologically simple words, the lexicon is a lot smaller. CELEX labels each entry with one of eight categories grouped under the heading "morphological status." We can divide the words into two groups: morphologically "simple" and morphologically "derived." This is shown in (20).

(20) Simple and derived words in CELEX (total = 52,447)

Status	Example	Simple	Derived
C (complex)	*aback*, *sandbank*		29,117
F (fused)	*I've*		173
I (irrelevant morphology)	*A/C*, *Almighty*		2,128
M (morpheme)	*camel*	7,401	
O (obscure morphology)	*anarchy*, *vegetable*	3,571	
R (root possible)	*behind*, *chapter*	2,066	
U (unclear morphology)	*alpenstock*, *interland*	725	
Z (zero conversion)	*abandon* (noun)		7,266
Sum		13,763	38,684

Under morphological status, cases C, F, and I are morphologically complex. Case Z repeats words already counted elsewhere, from which the current ones can be derived. For example, the noun *abandon* can be derived from the verb *abandon* (with a zero suffix). Cases O, R, and U seem to contain both simple and derived words. For example, pairs like *anarchy*-*monarchy*, *vegetable*-*vegetation*, *behind*-*before*, and *interland*-*intercity* seem to suggest that these words can be decomposed. However, for the sake of argument I group O, R, and U with "simple" words. Even so, just one quarter of the 52,447 words in CELEX are morphologically simple, from which others can be derived.

It is also relevant to note that some of the entries are uncommon words. It is well known that most of the time the average English speaker uses just a few thousand words. In fact, the *Longman Handy Learner's Dictionary* (Urbom 1999) uses merely 2,000 words for all definitions. A simple way to define uncommon words is by the frequency of occurrence. Based on a corpus of 18 million word tokens, CELEX has provided for each word its frequency of occurrence. Let us assume, quite conservatively, that frequent words are those that cover 99.9% of the tokens in the corpus. By this definition, words whose frequency is 5 or more are frequent and all others infrequent. Among the 13,763 morphologically simple words, 10,565 are frequent.

English has both polysyllabic and monosyllabic words. Their numbers are shown in (21), counted in four ways, where "simple" means morphologically simple and "frequent" means occurring five times or more. The data show that the number of morphemes used in English is of the order of 10,000, of which some 4,000 are monosyllabic.

(21)	Vocabulary type	All	Polysyllabic	Monosyllabic
	All "lemmas"	52,447	45,687	6,760
	Non-compounds	42,089	35,329	6,760
	Simple	13,763	9,842	3,921
	Simple and frequent	10,565	7,078	3,487

Next we consider the number of possible monosyllables in English. As discussed in Chapter 8, English allows up to three consonants at the beginning of a monosyllable. In addition, English allows an extra C after a VX rhyme at the end of a word, such as [p] in *help* or [k] in *task*. Additional consonants beyond the extra C are usually suffixes, which should not be many in morphologically simple words. The total number of possible syllables, therefore, should be the number of possible onsets times the number of VX rhymes times the number of choices for the extra final C. This is shown in (21), where 0 denotes the lack of an onset or the final C.

(22) Possible monosyllables in English
Total = Onsets × VX × C = 59 × 298 × 10 = 170,510

Onsets (common):	1 + 22 + 30 + 6 = 59 (0, C, CC, CCC)
VX:	16 + 273 = 289 (VV, VC, where V in VC can be tense)
C:	10 (0, p, t, k, f, θ, s, ʃ, ʧ, (ŋ, m, n, l, r))

The number of common onsets was discussed earlier, as was the number of VX rhymes. For the final C, I use just one series of voicing, because it usually agrees in voicing with the preceding C. In addition, because a sonorant C does not occur after an obstruent C, I counted all sonorant Cs as one sound, indicated by parentheses.

The data show that the number of possible monosyllables is seventeen times the number of all morphologically simple words in English, and over forty times the number of morphologically simple monosyllabic words. Not surprisingly, most possible syllables are unused, not because they are ruled out by phonetic or phonological constraints, but because English simply does not need so many words.

9.4. SYLLABLES IN POLYSYLLABIC WORDS

Let us now consider polysyllabic words. In particular, let us consider whether polysyllabic words contain more or new syllable types.

If a polysyllabic word is made of a string of syllables, and if each of the syllables can occur in a monosyllable, we would not expect there to be a big difference between syllable types in polysyllabic words and those in monosyllabic words. Of course, there might be occasional clusters that only occur in some polysyllabic words, such as [tswa] in *Tswana*, but we do not expect many such cases. Surprisingly, however, there turn out to be 1,453 more syllable types in polysyllables than in monosyllables. The data are shown in (23), extracted from the CELEX "lemma" lexicon, where "compounds" refers to entries that contain an internal space or hyphen.

(23)

	No. of words	No. of syllable types
Monosyllables	6,760	3,801
Polysyllables	36,012	5,254
Compounds	10,358	
Total "lemmas"	52,447	

There are three major sources of extra syllable types in polysyllables, shown in (24).

(24) Major sources of new syllable types in polysyllables
  a. Unstressed syllables
  b. Open lax vowels
  c. Suffixes

The first source is due to the fact that most monosyllables are stressed, whereas many syllables in polysyllabic words are unstressed. The lack of stress can give rise to vowel-less syllables and those whose main vowel is [ə]. This can be seen in the data in (25).

(25)

	Vowel-less	[ə] as main V
Monosyllables	6	6
Polysyllables	86	552

There are just six vowel-less syllables in monosyllabic word forms (*ahem*, *'d*, *hem*, *psst*, *'s*, and *shh*), but there are eighty-six in polysyllabic words. Similarly, only six monosyllabic word forms use an unstressed [ə] in its primary pronunciation (*o'*, *'ll*, *'em*, *'re*, *cos*, and *'ve*), but 552 syllables in polysyllabic words do so.

The second source of extra syllables comes from the method of syllabification. CELEX always syllabifies CVCV as CV.CV. For example, *happy* is [hæ][pɪ] and *belly* is [bɛ][lɪ], even though [hæ] and [bɛ] do not occur in monosyllables. The number of open lax vowels in polysyllables is shown in (26), where [ɔ] is a lax vowel in British pronunciation.

(26)

	æ]	ɛ]	ɪ]	ʊ]	ɔ]	ʌ]	All
Monosyllables	0	0	0	1	0	1	2
Polysyllables	49	50	57	42	51	44	293

The six open lax vowels account for about 300 extra syllables in polysyllabic words. However, as discussed in Chapter 3, CVCV should be syllabified as CVC.V when the first V is short and stressed. If so, new syllables with open lax vowels can be excluded.

The third source of extra syllables is affixes in polysyllabic words. For example, consider syllables with [ɪə] and [ʊə], whose counts are shown in (27).

(27)

	Monosyllables	Polysyllables
Syllables with [ɪə]	42	156 (*-ial*, *-ian*, *-ious*, etc.)
Syllables with [ʊə]	13	99 (*-ual*, *-uary*, etc.)

Many polysyllables with [ɪə] have a suffix, such as *-ial* [ɪəl], *-ian* [ɪən], and *-ious* [ɪəs]. Similarly, many polysyllables with [ʊə] have a suffix, such as *-ual* [ʊəl] and *-uary* [ʊə.rɪ]. For example, from *concept* we get *conceptual*, where [ʧʊəl] is not found in monosyllables. Similarly, from *sense* we get *sensual*, where [sjʊəl] is not found in monosyllables either.

It is worth noting that neither monosyllables nor polysyllables contain many new rhymes the other set does not have. In (28) I show the rhymes that are unique to one set, along with the counts of their occurrences, where [əC] is [ə] plus any C. Following the discussion in Chapter 8, word-final Cs beyond VX are ignored (e.g. [ʊps] counts as [ʊp] and [ʊnd] counts as [ʊn]).

(28)

Monosyllables	Polysyllables	Counts	Example
[ʌx], [ʊp], [ɛv]		3	*ugh*, *oops*, *rev*
	[ʊz], [ʊŋ], [ʊn]	4	*gooseberry*, *Weltanschauung*, *Burundi*
	[ɔ̃:], [ɑ̃:], [æ̃:], [æ̃]	35	(foreign words only)
	[æ], [ɔ], [ɛ], [ɪ]	207	
	[ə], [əC]	331	

The three rhymes unique to monosyllables occur once each in a marginal word. Most rhymes unique to polysyllables involve an unstressed [ə] or an open lax vowel, which we have discussed. Of the remaining rhymes, [ʊz] occurs once in *gooseberry*, [ʊŋ] occurs once in *Weltanschauung*, [ʊn] occurs once each in *Burundi* and *dachshund*, and [ɔ̃:], [ɑ̃:], [æ̃:], and [æ̃] occur in foreign-accented words only. Therefore, besides rhymes with [ə] or an open lax vowel, monosyllables and polysyllables have similar rhyme inventories.

In summary, there are more syllable types in polysyllabic words than in monosyllabic words. Most of the new syllable types can be accounted for by three facts: (a) there are more unstressed syllables in polysyllabic words; (b) the method of syllabification in CELEX creates open lax vowels; and (c) some vowel-initial suffixes create additional rhymes in polysyllables.

9.5. MORPHEME INVENTORIES IN ENGLISH AND CHINESE

It is interesting to compare the size of morpheme inventory in English with that in Chinese. The two turn out to be quite similar.

For simplicity and consistency, I take the number of morphemes in Chinese to be roughly the same as the number of characters. This method ignores homographs. For example, the character 干 can mean 'dry' or 'to do' but it is counted as one morpheme. The analysis also over-counts disyllabic morphemes, although there are not many in Chinese. For example, 蜻蜓 'dragonfly' and 玛瑙 'amber' are each one morpheme, but they are counted as two morphemes each, even though the parts have no meaning by themselves.

For English, I take the number of morphemes to be roughly the same as the number of words that are labeled as single morphemes in CELEX, i.e. those whose morphological status is M. This method will count homographs because CELEX lists them separately. For example, *bank* (of money) and *bank* (of river) are listed separately and will be counted as two morphemes. However, the analysis excludes bound morphemes, such as *bio-*, *pre-*, *-ology*, *-er*, and *-ly*. The undercount of bound morphemes in English is compensated by the inclusion of homographs, which are excluded in Chinese. Therefore, the overall effects of the counting method probably balance out for the two languages.

In both languages, zero derivations (i.e. a change of word category without an overt affix) are excluded. For example, in English *dry* (adjective) is included but *dry* (verb) is not. Similarly, in Chinese 干 is counted once, although it can be a verb 'to dry', a verb 'to do', an adjective 'dry', or a noun 'dried food'.

I use two electronic corpora for the comparison, Da (2004) for Chinese and CELEX for English. The basic information of the corpora is given in (29).

(29)	Chinese	English
Corpus	Da (2004)	CELEX
Size	259 million characters	18 million words
Morphemes	12,041 character types	7,401 monomorphemic words

The English corpus has fewer morphemes because it covers modern English only, while the Chinese corpus covers both classic and modern texts. In addition, many characters in the Chinese corpus are rarely used. If we ignore uncommon morphemes, the similarity between the languages becomes more evident. Let us consider the coverage of character or word tokens in each corpus. The data are shown in (30), up to the 7,000th most frequent morpheme. The Chinese calculation is made by Da (2004). The English calculation is made by me.

(30) Cumulative corpus coverage by the number of most frequent morphemes

Most frequent	Chinese coverage	English coverage
1,000	86.1740%	87.3571% (*wise*, 723)
2,000	95.5529%	94.2505% (*liquid*, 204)
3,000	98.3248%	97.2358% (*leap*, 78)
4,000	99.3046%	98.6762% (*loom*, 35)
5,000	99.7321%	99.4708% (*tankard*, 16)
6,000	99.9268%	99.8682% (*clunk*, 5)
7,000	99.9802%	100.0000% (*gull* (verb), 0)

The first 1,000 most frequent Chinese morphemes cover 86% of the corpus and the first 1,000 most frequent English morphemes, which ends at *wise* (occurring 723 times), cover 87% of the corpus. In both languages, the first 4,000 most frequent morphemes cover 99% of all occurring tokens, and the first 6,000 most frequent morphemes cover 99.9% of all occurring tokens. The bottom 454 morphemes in English, such as *asp* (noun), *barm* (noun), and *gull* (verb), do not occur in the frequency corpus (the frequency corpus was one of the sources of CELEX), but it is reasonable to assume that they are infrequent. In any case, in both Chinese and English, the first 6,000 most frequent morphemes cover 99.9% of all occurring tokens.

It is unclear how many morphemes are used in other languages. Given the fact that English is used worldwide and has borrowed many words from other languages, and the fact that the Chinese corpus covers not only modern usage but a large amount of texts from classic literature, it is reasonable to assume that morpheme inventories in other languages are unlikely to be much larger than those in English and Chinese.

9.6. DIPHTHONGS

A diphthong is made of two vowels. It can end in a retroflex vowel, such as [iɚ] *ear* and [eɚ] *air*, or start with a high vowel, such as [ia] and [ua], or end with a high vowel, such as [ai] and [au]. In the second kind, the high vowel is in the onset and can be transcribed as a glide, e.g. [ja] for [ia] and [wa] for [ua]. There does not seem to be any restriction on such diphthongs. For example, [i] and [u] can be followed by any vowel in American English, as seen in (31).

(31) [wi] *we*, [wɪt] *wit*, [wu] *woo*, [wʊd] *wood*, [wei] *way*, [wɛt] *wet*, [woɚ] *war*, [wʌt] *what*, [wæk] *wack*, [wɒk] *walk*, and [gwɑm] *Guam*
[jist] *east*, [jɪp] *Yip*, [ju] *you*, [jʊd] *you'd*, [jeits] *Yates*, [jɛt] *yet*, [joɚ] *your*, [jʌp] *yup*, [jæk] *yak*, [jɒn] *yawn*, and [jɑɚd] *yard*

The only restriction is that there are just two starting high vowels, [i, u], instead of four, [i, ɪ, u, ʊ]. What we can say then is that high vowels (or glides) cannot contrast in [tense] in the onset.

Let us now focus on diphthongs that end in a high vowel, which are subject to many restrictions. If any vowel can combine with any high vowels, there should be many diphthongs. For example, with eleven vowels [i, ɪ, u, ʊ, e, ɛ, o, ʌ, æ, ɑ, ɒ] and four high vowels [i, ɪ, u, ʊ] in American English, there are forty-four possible combinations, shown in (32).

(32) Diphthong combinations that end in a high vowel
ii, ɪi, ui, ʊi, ei, ɛi, oi, ʌi, æi, ɑi, ɒi
iɪ, ɪɪ, uɪ, ʊɪ, eɪ, ɛɪ, oɪ, ʌɪ, æɪ, ɑɪ, ɒɪ
iu, ɪu, uu, ʊu, eu, ɛu, ou, ʌu, æu, ɑu, ɒu
iʊ, ɪʊ, uʊ, ʊʊ, eʊ, ɛʊ, oʊ, ʌʊ, æʊ, ɑʊ, ɒʊ

However, English uses just five of them, which raises two questions. First, why are most diphthongs not used? Second, which of the diphthongs are used and how should we transcribe them? For example, should the diphthongs be [ai, au, oi, ou, ei] or [aɪ, aʊ, ɔɪ, oʊ, eɪ]? Should we use [ai, au] or [ɑi, ɑu]? In what follows, I address such questions by examining possible constraints that reduce the number of diphthongs.

First, we can see that there are no diphthongs that contrast in [tense], such as [ei]-[eɪ], [ei]-[ɛi], or [ai]-[aɪ]. In addition, there are no diphthongs that contrast in [back], such as [ɛi]-[ʌi], [ɛu]-[ʌu], [æi]-[ɑi], or [æu]-[ɑu]. Therefore, I propose the constraints in (33) and (34).

(33) No-[tense]
The feature [tense] is not contrastive in diphthongs.

(34) No-non-high-[back]
The feature [back] is not contrastive for non-high vowels in diphthongs.

Given No-[tense], there is no contrast between [i] and [ɪ], or [u] and [ʊ], and we can only have two of the vowels. Let us transcribe them as [I, U], which are unspecified for [tense]. Similarly, I use [ə] for the mid vowel unspecified [tense] and [back] (merging [e, ɛ, ʌ]), and [A] for the low vowel unspecified [tense] and [back] (merging [æ, ɑ]). There are now six vowels [I, U, ə, o, A, ɒ] (instead of eleven) and two high vowels [I, U] (instead of four). The list of diphthongs, therefore, is reduced from forty-four to twelve, shown in (35).

(35) Diphthong combinations for English
II, UI, əI, oI, AI, ɒI
IU, UU, əU, oU, AU, ɒU

[II] and [UU] can be excluded by Merge, introduced in Chapter 5, so that [II] is the same as [i:] and [UU] is the same as [u:]. This leaves a list of fourteen diphthongs.

To further reduce the list, let us consider feature compatibility in the diphthongs. First, let us consider the features of the six vowels, shown in (36).

(36) Vowel features in diphthongs

[+high]	[I]: [−back]
	[U]: [+back]
[−high, −low]	[ə]: [−round]
	[o]: [+round]
[+low]	[A]: [−round]
	[ɒ]: [+round]

Following a common practice, I assume that only contrastive features need be specified. The high vowels [I] and [U] can differ in [back] or [round], but only one is needed, and I choose [back]. For mid and low vowels, I choose the feature [round] to distinguish each pair.

Given the above features, we can use Rhyme-Harmony, introduced in Chapter 5 and repeated in (37), to reduce two more diphthongs, shown in parentheses in (38).

(37) Rhyme-Harmony
Vowels in a rhyme cannot have opposite values in [back] or [round].
That is: *[+back][−back], *[−back][+back]
*[+round][−round], *[−round][+round]

(38) Diphthong reduced by Rhyme-Harmony

Differ in [back]	(UI, IU)
Not differ in [back]	əI, oI, oU, ɒI, ɒU, əU, AI, AU

Since [o] and [ɒ] are unspecified for back, they can each combine with [I] and [U]. This leaves eight diphthongs.

The eight diphthongs are still more than what are found in English. I suggest a further constraint, described in (39); a similar constraint was proposed for Shanghai in Chapter 6, where GV and VX do not contrast in the height of the nuclear vowel. The effect of the constraint is that only one diphthong in each vertical column in (40) is used. The unused diphthong is shown in parentheses.

(39) No-Height
Diphthongs do not contrast in height.

(40)

High	(II)	(IU)	(UI)	(UU)
Mid	əI	(əU)	oI	oU
Low	AI	AU	(ɒI)	(ɒU)

The only exception is [əI, AI], where we predict just one of them to occur. It is interesting to note that [əI] is often [e] in American English, in which case No-Height is true for all diphthongs.

In summary, I have proposed four constraints, repeated in (41), that reduce forty-four possible diphthongs to five that actually occur in English.

(41) Constraints on diphthongs in American English
No-[tense]: The feature [tense] is not used in diphthongs.
No-non-high-[back]: The feature [back] is not used for non-high vowels in diphthongs.
Merge: Two tokens of the same feature merge into one long feature.
Rhyme-Harmony: Vowels in a diphthong cannot have opposite values in [back].
No-Height: Diphthongs do not contrast in height.

The analysis is not the only one possible. For example, if we use a different way of feature specification, such as using [round] for [I] and [U], we might need to revise the analysis but achieve similar results. It is also unclear at this point which constraints have a more general nature or hold at the surface level. For example, if [oi] is fully specified at the surface level, it would violate both back harmony and round harmony. Similarly, in Australian English (Harrington et al. 1997), [ɑe] (as in [hɑe] *high*) seems to violate back harmony and [æɔ] (as in [hæɔ] *how*) seems to violate both back harmony and round harmony.

However, regardless of how such cases are analyzed, it seems clear that diphthongs use fewer features than monophthongs, probably because each part of a diphthong is too short to make use of too many features contrastively.

Finally, let us consider the frequencies of diphthongs. In (42) I show the data for English, where "all" means out of the entire CELEX lemma lexicon and M means out of those words that are labeled as monomorphemes. For convenience I represent the diphthongs with regular symbols, without using upper-case letters.

(42) Frequencies of English diphthongs

	ai	au	oi	ou	ei
All	3,903	1,142	436	4,032	5,168
M	181	19	11	195	129

The least frequent diphthong is [oi], where the two parts differ in [round] (if fully specified). Slightly more frequent is [au], where the two parts again differ in [round] (if fully specified). The remaining three are many times more frequent, and their parts agree in [round].

It is unclear whether the relation between feature similarity within a diphthong and its frequency is accidental or motivated. It would be interesting, for future research, to find out whether similar frequency relations are found in other languages.

9.7. VOICING IN CODA CONSONANTS, SYLLABIFICATION, AND VOWEL LENGTH

There is a common preference for obstruent codas to be voiceless. For example, in German, Polish, Russian, and Turkish, some or all obstruent codas are unvoiced. Similarly, in Chinese, Thai, Korean, and Vietnamese, obstruent codas are limited to [p, t, k, ʔ]. Moreover, there is to my knowledge no language in which obstruent codas are voiced only.

In English, both voiceless and voiced obstruents can occur in the coda. For example, we find [ɪk] in *picnic* and [ɪg] in *signature*. However, English also has a preference for voiceless obstruent codas in the sense that they are more frequent than voiceless ones. For illustration, let us consider obstruent codas in nonfinal syllables of monomorphemic words, which are shown in (43), where X][indicates a coda X before another syllable.

(43) Frequencies of obstruent codas in non-final syllables of monomorphemic words

	p][	t][	k][	ʧ][	f][	s][	ʃ][	θ][	All
[−voice]	19	21	76	1	9	34	3	2	165
[+voice]	6	12	20	0	1	18	0	1	58
	b][	d][	g][	ʤ][	v][	z][	ʒ][	ð][	All

In every case the voiceless coda is more frequent than its voiced counterpart, and on average there are three times as many voiceless obstruent codas as voiced ones.

It is interesting to compare coda frequencies with onset frequencies. For accuracy, let us avoid word-initial onsets, which can be skewed by certain clusters that only occur there (e.g. sC clusters). Similarly, let us avoid cases where syllabification is ambiguous. For example, CELEX syllabifies VCV as V.CV, although it is likely to be VC.V if the first V is short and stressed (Chapter 3). To avoid such complications, and for ease of calculation, let us consider the frequencies of obstruent onsets after a long vowel [V:] in monomorphemic words, which are shown in (44). The onset counts include those followed by [r, l, w, j], which should not affect the result in significant ways, since there are far more simple onsets than clusters and since [r, l, w, j] occur with both voiceless and voiced obstruents.

(44) Frequencies of obstruent onsets after [V:] in monomorphemic words

	:][p	:][t	:][k	:][f	:][s	:][ʧ	:][ʃ	:][θ	All
[−voice]	36	81	41	8	36	7	8	2	219
[+voice]	53	59	45	27	23	15	2	3	227
	:][b	:][d	:][g	:][v	:][z	:][ʤ	:][ʒ	:][ð	All

The data show no obvious preference for voiced or voiceless obstruents in the onset.

The preference for voiceless obstruent codas may help explain a well-known property of nasal duration. It has been observed that a nasal consonant tends to lack its own duration when it is followed by a voiceless stop, but not when it is followed by a voiced stop (Malécot 1960, Bailey 1978, Fujimura 1979). Some examples are shown in (45).

(45)				
	N before voiceless C	[tɛ̃t] *tent*	[æ̃pl̩] *ample*	[æ̃kɚ] *anchor*
	N before voiced C:	[tɛ̃nd] *tend*	[æ̃mbl̩] *amble*	[æ̃ŋgɚ] *anger*

The data can be analyzed in (46), where a stressed syllable is heavy, onsets are optional (Chapter 3), the maximal rhyme is VX (Chapter 8), and voiced codas [b, d, g] are avoided.

(46)			
	[tɛ̃t] *tent*	[æ̃p][l̩] *ample*	[æ̃k][ɚ] *anchor*
	[tɛ̃n][d] *tend*	[æ̃m][bl̩] *amble*	[æ̃ŋ][gɚ] *anger*

The analysis also predicts that the codas [mp], [nt], and [ŋk] should be common both word-finally and word-medially, but the codas [mb], [nd], and [ŋg] should not. The prediction is correct, as the data in (47) show. The data are based on the 7,401 monomorphemic words in CELEX, where # indicates a word boundary.

(47) NC clusters in monomorphemic words

N + [−voice]		N + [+voice]	
mp#	55	1	mb#
nt#	122	86	nd#
ŋk#	79	0	ŋg#
mp][	4	0	mb][
nt][	1	0	nd][
ŋk][	3	0	ŋg][
m][p	67	64	m][b
n][t	86	104	n][d
ŋ][k	40	51	ŋ][g

While [mp], [nt], and [ŋk] codas are found both word-finally and word-medially, there is no [mb], [nd], or [ŋg] coda word-medially. Word-finally there is no [ŋg] either, and there is just one [mb], which occurs in *iamb*, which is pronounced without [b] by some speakers. We do find many words that end in [nd], as well as many that end in [nt], but this is probably due to the fact that [t, d] are are used as suffixes in English and they both should be tolerated. Of relevance is the fact that across syllable boundaries [mb, nd, ŋg] are just as common as [mp, nt, ŋk]. Therefore, the lack of [mb, nd, ŋg] in the coda supports the view that the rhyme is limited to VX and voiced obstruent codas are avoided.

Pater (2004) proposes a universal constraint against a nasal followed by a voiceless C (*NC̥). The proposal could be used to explain why [tɛnt] *tent* is realized as [tɛ̃t], while [tɛnd] *tend* is realized as it is. On the other hand, Pater cannot explain the lack of [mb, nd, ŋg] codas, unless he also assumes that the maximal rhyme is VX and that voiced obstruent codas are avoided.

It is also well known in English that vowels are shorter before a voiceless consonant than before a voiced one (Jones 1950, Peterson and Lehiste 1960). For example, Jones (1950) notes that tense vowels are "half-long" before a voiceless obstruent coda, exemplified in (48).

(48)

Before voiceless C	[sit] *seat*	[lus] *loose*	[rut] *root*
Before voiced C	[si:d] *seed*	[lu:z] *lose*	[ru:d] *rude*

One is tempted to offer a similar analysis, shown in (49), where voiced obstruents are excluded from the coda.

(49) [sit] *seat* [lus] *loose* [rut] *root*
[siː][d] *seed* [luː][z] *lose* [ruː][d] *rude*

A problem for the above analysis is that the length variation is also found in lax vowels. For example, [ɪ] is longer in *hid* than in *hit*, and [ɛ] is longer in *led* than in *let*. If we follow the same analysis, we are led to the syllabification in (50).

(50) [hɪt] *hit* [lɛt] *let*
[hɪː][d] *hid* [lɛː][d] *led*

However, there are two problems with syllables like [hɪː] and [lɛː]. First, they contain a stressed open lax vowel, which does not occur in monosyllables. Second, a "long" lax vowel, such as that in *hid*, is much shorter than a long tense vowel, such as that in *seed*. Therefore, it is questionable whether the lax vowels in *hid* and *led* are indeed phonologically long.

Instead of representing all length variations in phonology, or excluding all of them, we might consider represent some length variations in phonology, such as those that are large. In this regard, it is relevant to note that, according to Jones (1950), tense vowels are "half-long" before voiceless codas, whereas there is only "a limited extent" of shortening in lax vowels before voiceless codas. Therefore, one might consider representing (some) length variations in tense vowels phonologically but not those in lax vowels.

9.8. SUMMARY

If we freely combine actual onset clusters and actual rhymes, English has some 20,000 possible syllables. If we allow an extra C that commonly occurs after a VX rhyme (such as [p] in *help* and [k] is *ask*), English has some 200,000 possible syllables. However, the actual number of syllables used in English is just over 5,000, including word-final consonant clusters, and if we limit the rhyme to VX, there are fewer than 3,000 actual syllables. The huge gap between the number of possible syllables and the number of used syllables requires an explanation.

Unlike previous analyses, which attempt to rule out unused syllables by phonological rules, I have argued that the main reason for not using them is not phonological. Instead, English simply does not need so many syllables. Specifically, similar to Chinese, English has some

10,000 morphemes, and the first 6,000 frequent ones cover 99.9% of all usage. In addition, of the 10,000 morphemes about 4,000 are monosyllabic, and of the 6,000 frequent morphemes about half are monosyllabic. Therefore, English only needs a few thousand syllables to make every morpheme distinct, which is far below the number of possible syllables English allows.

I have also argued that diphthongs use fewer features than monophthongs. In particular, English diphthongs do not use the feature [tense], nor the feature [back] for non-high vowels. Finally, I have shown that there is a preference in English for obstruent codas to be voiceless. The preference can account for the fact that [Vmp, Vnt, Vŋk] rhymes are common, because they can be realized as [Ṽp, Ṽt, Ṽk], whereas [Vmb, Vnd, Vŋg] rhymes are rare, because the realizations [Ṽb, Ṽd, Ṽg] have a voiced coda and are avoided.

10

German

The German syllable has been studied at length in several works (e.g. Giegerich 1985, 1989, Hall 1992, 2002b, Wiese 1996). As in English, a German word can start with up to three Cs, such as [ʃpl] in *Splitt* 'gravel', where the first C is limited to [ʃ] or [s]. A common analysis is that the onset is limited to CC and [ʃ] and [s] are treated in some special ways.

A German word can end with up to five Cs, such as [rpsts] in *(des) Herbsts* 'autumn (genitive)' (Moulton 1956). But as in English, if we ignore word-final coronals, the German rhyme is generally limited to VX (i.e. VV or VC) in non-final positions and VXC in final positions. Three analyses have been proposed. According to Giegerich (1985, 1989), if we exclude the final C, the German rhyme is limited to VX. The question for this analysis is why the final C can survive when it is not part of a rhyme. Alternatively, according to Wiese (1996), the German rhyme is maximally VXC. The question for this analysis is why VXC rhymes are rare in non-final positions. Finally, according to Hall (2002b), the German rhyme is maximally VX in non-final positions but VXC in final positions. The question for this analysis is why the final rhyme can be larger, especially in view of the fact that final rhymes seem to have less ability to attract stress than non-final rhymes. For example, with regard to stress assignment, a final VC is equal to a non-final V and a final VXC is equal to a non-final VX (Giegerich 1985).

In this chapter I propose that the maximal syllable in German is CVX, similar to that in English. To support the proposal, I provide the arguments in (1).

(1) Arguments for CVX in German
   a. CCC onsets are limited to initial position only.
   b. Medial CC onsets are possible complex sounds.
   c. Non-final rhymes are limited to VX.

d. The C in the final VXC is a “potential onset” (supported by a potential V and anti-allomorphy).
e. Consonants beyond the final VXC are suffixes or suffix-like.

I have discussed “complex sounds” in Chapter 2 and “potential V” and “anti-allomorphy” in Chapter 3. In addition, the “affix rule” (Chapter 3) can account for the presence of suffix or suffix-like consonants beyond VXC. The analysis of the rhyme is similar to that of Giegerich (1985, 1989). The analysis of the onset differs from previous proposals.

The arguments in (1) are based on an exhaustive examination of the CELEX German lexicon (Baayen et al. 1993), which contains 365,530 entries, of which 51,728 are uninflected. Of the uninflected entries there are 6,531 that are labeled as “monomorphemic” words. Since other entries are mostly made of elements from the morphologically simple words, I shall examine the syllables in the 6,531 simple words only. I start with consonants and vowels, followed by rhymes and then onsets.

10.1. SOUND INVENTORY

CELEX lists the consonants in (2) for German. The sounds [ç, x] are often allophones of the same phoneme and [r, ʀ] are alternative transcriptions.

(2) German consonants
[p] ***P****akt*, [b] ***B****ad*, [t] ***T****ag*, [d] ***dann***, [k] ***k****alt*, [g] ***G****ast*
[f] ***f****alsch*, [v] ***W****elt*, [s] *Gla****s***, [z] ***S****uppe*, [ʃ] ***Sch****iff*, [ʒ] ***G****enie*, [ç, x] *Ba****ch****/i****ch***,
[h] ***H****and*
[pf] ***Pf****erd*, [ts] ***Z****ahl*, [tʃ] *Ma****tsch***, [dʒ] ***G****in*
[m] ***M****aß*, [n] ***N****aht*, [ŋ] *Kla****ng***
[l] ***L****ast*, [r, ʀ] ***R****atte*, [j] ***J****acke*, [w] ***w****aterproof*

Kohler (1999) transcribes [x] and [ʀ] as [χ] and [ʁ], and Wiese (1996) transcribes them as [x] and [r]. In addition, neither Kohler nor Wiese includes [w], which occurs in foreign words only.

According to CELEX, Wiese (1996), and Kohler (1999), German has the vowels in (3), classified according to Wiese (1996). The vowel [ə] occurs in unstressed syllables only.

(3) German vowels

		−back		+back	
		−round	+round	−round	+round
+high	+tense	i: *Lied*	y: *für*		u: *Hut*
	−tense	ɪ *Mitte*	ʏ *Pfütze*		ʊ *Pult*
−high, −low	+tense	e: *mehl*	ø: *Möbel*		o: *boot*
	−tense	ɛ *Bet* ɛ: *Käse*	œ *Götter*	ə *Beginn*	ɔ *Glocke*
+low		a *hat* a: *klar*			

Diphthongs: [ai] *weit*, [au] *Haut*, [ɔy] *freut*

The tense vowels [i, y, u, e, ø, o] are long in stressed syllables and short in unstressed syllables (Giegerich 1985), but there is no contrast in length alone, such as [i:] vs. [i]. However, the pair [a, a:] and [ɛ, ɛ:] do differ in length, and I shall return to them.

Wiese (1996) transcribes the diphthongs as [aɪ, aʊ, ɔʏ] and Kohler (1999) transcribes them as [aɪ, aʊ, ɔɪ]. In addition, Wiese (1996) uses [ɐ] for what CELEX transcribes as [ər], as seen in the second syllable of *Lehrer*. CELEX also uses a number of other vowels (different from those above) that occur in foreign words. They are shown in (4).

(4) [ɑ:] *Adv**a**ntage,* [ɔ:] ***A**llroundman,* [ɜ:] *Teamw**or**k*, [eɪ] *N**a**tive,* [aɪ] *Sh**y**lock,* [ɔɪ] *Playb**oy**,* [aʊ] *Allr**ou**ndsportler,* [æ] *R**a**gtime,* [ɑ] *K**a**levala,* [ʌ] *Pl**u**mpudding,* [œ̃:] *Parf**um**,* [æ̃] ***im**promptu,* [ɑ̃:] *Dét**en**te,* [æ̃:] *Bass**in**,* [ɒ̃:] *Affr**on**t*

These vowels are not included in Wiese (1996) or Kohler (1999). Among them [ɑ, ɑ:] seem to contrast in length, but they do not occur in the same environments. Also, [ɔ:] seems to contrast with [ɔ] in length, but there are no minimal word pairs either. Therefore, the vowels in (4) do not present a serious problem for our discussion on syllable size.

10.2. RHYME SIZE IN NON-FINAL POSITIONS

The basic information of the CELEX German lexicon is shown in (5), which has 6,531 morphologically simple words.

(5)

All entries	365,530
Uninflected entries	51,728
Simple words	6,531
Syllables in simple words	13,445

CELEX provides for each word both a phonetic transcription and a CV transcription, both being syllabified. For example, *reimen* 'to rhyme' is [rai][mən] in the phonetic transcription and [CVV][CVC] in the CV transcription, where brackets indicate syllable boundaries.

In the CV transcription, each diphthong or long vowel is VV. In addition, VVV is found once in the word *royal* [ro:a][ja:l] [CVVV][CVVC], along with an alternate transcription [ro:][a][ja:l] [CVV][V][CVVC]. Therefore, we can ignore VVV. There is little doubt that German has many VX rhymes. The question is whether German has non-final rhymes that are longer than VX. By searching the CV transcription for relevant sequences, we can locate all rhymes that are longer than VX. The results are shown in (6).

(6) Words with non-final rhymes larger than VX

Simple words	6,531
CCC][	0
VCC][	69
VVC][	234
All VX+ rhymes	303
Words with VX+	296

Among the 6,531 simple words, there are 303 non-final rhymes that exceed VX. They occur in 296 words because seven of the words contain two rhymes each that exceed VX.

The 296 words can be divided into five groups according to the nature of the extra-large rhyme and whether the word is morphologically complex. The groups are listed in (7)–(11). If a word can be grouped in more than one way, it is grouped just once. For example, *Branntwein* [brant][vain] 'burnt-wine (brandy)' can be grouped under [VNC] rhymes or under morphologically complex words (*Brannt-wein*), and it is grouped with the former only.

(7) [VNC] or [VVN] rhymes (48 words)

Advantage, allenthalben, Ammonshorn, Antlitz, antworten, Askorbinsaeure, bedienstet, Blindgaenger, Branntwein, brenzlich, brenzlig, Dienstag, entbloeden, entgegen, entgegen, entlang, entlang, entlegen, Entstalinisierung, entweder, entwegt, entwischen, entwoehnen, Gemeinde, grundfesten, Handlanger, Hundsfott, Lambda, landlaeufig, Lindwurm, Meintat, Neutestamentler, Pendule, Plankton, punkto, Roentgen, Rosenmontag, Sintflut, Suendflut, Tausendsasa, Tausendsassa, Tinktion, Tinktur, Transport, transversal, Traumwandler, ueberwendlings, Unktion

(8) [V:C] rhymes where the vowel is [i:, y:, e:, ø:, u:, o:] (74 words)
Behoerde, Buchstabe, derselbe, Diebstahl, Dipolantenne, duester, Erde, etwelche, gehorsam, Griesgram, Herde, Hochverraeter, Hygroskop, jedweder, jeglicher, Kehraus, Kloster, Knoblauch, Kohlrabi, Kuerbis, Langspielplatte, Libertaet, Linolschnitt, Louisdor, Meltau, Menostase, monochrom, Muesli, mutmassen, Niednagel, Notdurft, notwendig, Notzucht, Nutrition, nutritiv, Ostern, Phonometrie, plustern, Prostitution, Rebhuhn, Reklame, ruchlos, Ruebsen, Schuster, Sekretaer, Spierling, Stegreif, Tetrachlorkohlenstoff, Thermostat, Tournee, Truthahn, Truthenne, Truthuhn, urbar, Urheber, Urkunde, Urlaub, Uroskopie, Ursache, Ursprung, urtuemlich, Vorfahr, Vorgelege, vorlaeufig, Vormund, vornehm, Vorrang, Vorrat, Vorschein, Vorschmack, Wermut, Werwolf, zappenduster, Zitrin

(9) Morphologically complex words (149 words)
Absorption, absorptiv, aeufnen, ahnden, Althaendler, anheimgeben, anheimstellen, Argwohn, atmen, aufbaenken, aufbaumen, aufbuerden, aufdonnern, aufdunsen, aufflackern, aufkuenden, aufmoebeln, aufmuepfig, aufpeppen, aufpropfen, aufrauhen, auserkoren, ausfitten, ausgefeimt, ausgefuchst, ausixen, auskolken, ausmaeren, ausmerzen, ausmugeln, auspichen, ausrasten, ausrotten, Aussaat, Aussatz, Ausschuss, ausstatten, Ausstich, Austenit, Auster, auswaertig, ausweiden, Bauernfang, behaupten, beichten, beschlagnahmen, bewerkstelligen, biedermeisterlich, Bifokalglas, bodmen, daemlich, dermaleinst, durchtrieben, durchwachsen, einbrechen, einbuchten, einbuergern, Einfalt, einfinden, einfloessen, Einfuhr, eingedenk, eingefleischt, Eingeweide, einhaendigen, Einhalt, einhellig, einlegen, einmuenden, Einoede, einrichten, einruecken, einruesten, Einsaat, einschaerfen, einschieben, einschmelzen, einschreiten, einsiedeln, einsprengen, einstehen, Eintracht, eintreiben, einverstanden, Einwaage, einwecken, einwenden, erbeingesessen, erbreisten, ernten, erste, explosiv, Faehnrich, fahnden, feilschen, Feldwebel, gebaerden, Gedaechtnis, graetschen, Heimkunft, jauchzen, Karfreitag, Karsamstag, Karwoche, Kleinasien, knautschen, kostenaufwendig, kuenftig, Labskaus, Leichnam, leisten, leuchten, Maedchen, Maerchen, Mastdarm, Meineid, meistens, Nachricht, Pausback, pusten, quietschen, raeuspern, raunzen, Reinfall, Reinmachefrau, Reissaus, ruelpsen, scheusslich, Schienbein, Schornstein, seufzen, surfen, Teilkasko, Thunfisch, Tonbank, unvordenklich, uralaltaisch, verleumden, verwahrlosen, vierte, vierzehnte, Wahnkante, Wahnwitz, wahrscheinlich, Weissbinder, Weissgerber, werden, Yamswurzel, Zirkumpolarstern

(10) [V:C] rhymes where the vowel is [a:, ɛ:] (9 words)
Bratsche, Garten, haetscheln, Hafner, Hahnrei, Laerche, Latsche, Party, watscheln

(11) Other words with VX+ rhymes (16 words)
Arkturus, Deichsel, Erbse, Halfter, Halfter, Halfter, Juxta, Kleister, Leiste, Meister, Partner, Peitsche, Sexta, Teiste, Veilchen, Weichsel

As discussed in Chapters 2 and 3, [VNC] and [VVN] rhymes can be analyzed as [ṼC] and [ṼṼ]. This is illustrated in (12).

(12)

CELEX	Analysis	Example
[VNC]	[ṼC]	*Antlitz* [ant][lɪts] → [ãt][lɪts]
[VVN]	[ṼṼ]	*Meintat* [main][ta:t] → [mãĩ][ta:t]

Next, [V:C] rhymes that contain the tense vowel [i:, y:, e:, ø:, u:, o:] can be analyzed as [VC], where the tense vowel is short. Two examples are shown in (13).

(13)

CELEX	Analysis	Example
[y:s]	[ys]	*duester* [dy:s][tər] → [dys][tər]
[e:r]	[er]	*Erde* [e:r][də] → [er][də]

Finally, words with complex morphology can be excluded, because what appears to be medial in these words may not be. For example, in *auserkoren* [aus][ɛr][ko:][rən] the first rhyme exceeds VX, but because there is a morpheme boundary *aus-erkoren*, [aus] is final and not a true exception. The same is true for *aus-statten*, *auf-baumen*, etc.

The words in (10) and (11) are still to be accounted for. Those in (10) contain [a:C] and [ɛ:C] rhymes; they cannot be analyzed as [aC] and [ɛC], if there is a length contrast between [a:] and [a] and between [ɛ:] and [ɛ]. However, according to Moulton (1956: 373), [a] and [a:] have "a qualitative as well as a quantitative difference" and the qualitative difference is "both auditorily and visually perceptible." For this reason, he transcribes the pair as [ᴀ] and [a:]. If Moulton is correct, then [a:C] can be represented as [aC], which remains different from [ᴀC]. This accounts for seven of the words in (10). There are sixteen words left, therefore, two from (10) and fourteen from (11) (excluding two repetitions of *Halfter* in (11)). A closer look at them suggests that further analysis is possible, which I show in (14), where [kˢ, pˢ] are affricates. Some words have alternative solutions, which are separated by a semicolon.

(14) Remaining 16 words with non-final VX+ rhymes

Word	CELEX	Analysis
haetscheln	[hɛ:t][ʃəln]	[het][ʃəln]?
Laerche	[lɛ:r][xə]	[ler][xə]?
Arkturus	[ark][tu:][rʊs]	perceived compound *Ark-turus*?
Deichsel	[daik][səl]	[dai][kˢəl]; perceived suffix -*sel*
Erbse	[ɛrp][sə]	[ɛr][pˢə]

Halfter	[half][tər]	perceived suffix -*er*
Juxta	[jʊks][ta]	affricate [k^s]
Kleister	[klais][tər]	perceived suffix -*er*
Leiste	[lais][tə]	perceived suffix -*e*
Meister	[mais][tər]	perceived suffix -*er*
Partner	[part][nər]	perceived suffix -*er*
Peitsche	[pait][ʃə]	perceived suffix -*e*; [pai][tʃə] with affricate [tʃ]
Sexta	[zɛks][ta:]	affricate [k^s]
Teiste	[tais][tə]	perceived suffix -*e*
Veilchen	[fail][xən]	perceived suffix -*chen*; syllabic [l]
Weichsel	[vaik][zəl]	perceived suffix -*sel*

First, most of the words contain a suffix-like ending, which could be treated as a "perceived suffix" (Chapter 3). For example, *Kleister* may have a perceived suffix -*er*. If we exclude -*er*, the root is *Kleist*, and if we exclude the final coronals [s, t], the rhyme is [ai], which is within VX. Second, if [ks], [tʃ], and [ps] are pronounced as affricates, we can account for a few more words. Finally, the vowel [ɛ:] is marginal for some speakers (Moulton 1956), and the words *haetscheln* and *Laerche* can probably be accounted for, too. In summary, non-final rhymes in German seem to be limited to VX, in agreement with previously proposals (Giegerich 1989, Hall 2002b).

10.3. FINAL RHYMES

Word-final rhymes in the list of morphologically simple words are shown in (15).

(15) Word-final rhymes

Type	Count	Example
VC	2,757	*antworten* [ant][vɔr][tən]
V	1,101	*erste* [e:r][stə]
VVC	1,058	*Urlaub* [u:r][laup]
VCC	964	*Ostern* [o:s][tərn]
VV	533	*Kohlrabi* [ko:l][ra:][bi:]
VVCC	59	*entwegt* [ɛnt][ve:kt]
VCCC	45	*Angst* [aŋst]
VVCCC	10	*Dienst* [di:nst]
VCCCC	4	*Herbst* [hɛrpst] (*Ernst, selbst, selbst*)
All	6,531	

A German word can end with up to five consonants, although the last is always an inflection (Moulton 1956, Giegerich 1989). Without inflection the maximal final consonant cluster is CCCC after a short vowel and CCC after a long vowel, or VXCCC, which is confirmed in (15). Moulton (1956) also suggests that the first C after VX can be one of many consonants except voiced obstruents, whereas the final CC are always [t, s] or to a limited extent [ʃ], although he does not provide quantitative data. Let us check Moulton's generalization by examining all the words that end in VCCC, VCCCC, VVCC, and VVCCC. The results are shown in (16)–(19), where each cluster is followed by its count. The affricate [ts] in [rts] (*Erz*, *Harz*, *Quarz*) and [rtst] (*Arzt*) corresponds to orthographic 'z' and is treated as one C.

(16) Word-final VCCC (45 words)
Last C is [s, t] [rst] 8, [nst] 7, [nft] 5, [ŋks] 5, [kst] 4, [lst] 3, [ŋkt] 3, [ŋst] 3, [rft] 2, [rkt] 2, [nʃt] 1, [pst] 1, [rxt] 1

(17) Word-final VCCCC (4 words)
Last CC are [s, t] [lpst] 2, [rnst] 1, [rpst] 1

(18) Word-final VVCC (59 words)
Last C is [s, t] [rt] 17, [st] 11, [xt] 7, [nt] 4, [ks] 3, [rts] 3, [kt] 2, [ns] 2, [ft] 1, [mt] 1, [ps] 1, [pt] 1, [rk] 1, [sk] 1, [ʃt] 1
Exceptions [lʃ] 1 (*Koelsch* [kø:lʃ]), [pʃ] 1 (*huebsch* [hy:pʃ]), [rʃ] 1 (*Barsch* [ba:rʃ])

(19) Word-final VVCCC (10 words)
Last CC are [s, t] [nst] 3, [pst] 3, [rst] 2, [rtst] 1
Exception [mpt] 1 (*prompt* [pro:mpt])

The data confirm that the first C after VX is fairly unrestricted and the Cs after VXC are limited to [s, t]. The four exceptional words involve a long tense vowel. If tense vowels can be short, then these words are not true exceptions. The analysis is shown in (20), which agrees with the intuition of an anonymous reviewer that the V is short.

(20)

Word	Analysis
Koelsch [kø:lʃ]	[kølʃ] ([ʃ] within VXC)
huebsch [hy:pʃ]	[hypʃ] ([ʃ] within VXC)
Barsch [ba:rʃ]	[barʃ] ([ʃ] within VXC)
prompt [pro:mpt]	[prompt] or [prõ:pt] ([p] within VXC)

In summary, as in English, final consonants after VXC in German are either [s] or [t], which are used as suffixes. Therefore, they are covered by the "affix rule" (Chapter 3). In addition, the C after VX is supported by a "potential vowel" (Chapter 3), which is available in vowel-initial suffixes. Therefore, the final rhyme in German is limited to VX. The conclusion is similar to that of Giegerich (1989), although he allows extra consonants after VX to be appended to a syllable in a later stage, whereas I see no need for the appendix.

10.4. INITIAL ONSETS

The 6,531 initial onsets in the list of morphologically simple words are shown in (21) and (22).

(21)

	Types	Tokens
Vowel-initial	1	876
C	23	4,312
CC	38	1,271
CCC	8	72
All	70	6,531

(22) C onsets: [k] 473, [m] 464, [b] 331, [h] 297, [l] 293, [v] 278, [f] 275, [r] 267, [p] 255, [t] 233, [g] 215, [z] 213, [ʃ] 171, [n] 156, [d] 143, [ts] 118, [j] 77, [pf] 23, [ʒ] 15, [ʤ] 9, [x] 3, [tʃ] 2, [s] 1

CC onsets: [ʃt] 136, [tr] 101, [kl] 85, [kr] 83, [ʃp] 76, [br] 69, [gr] 66, [ʃl] 64, [fl] 53, [pl] 52, [pr] 49, [ʃv] 48, [kn] 46, [bl] 44, [ʃm] 36, [fr] 35, [ʃn] 33, [kv] 30, [ʃr] 29, [tsv] 29, [gl] 28, [dr] 27, [sk] 11, [pfl] 9, [vr] 5, [ps] 4, [gn] 3, [pfr] 3, [sl] 3, [sm] 3, [pn] 2, [sp] 2, [sts] 2, [fj] 1, [ks] 1, [sf] 1, [sn] 1, [sv] 1

CCC onsets: [ʃtr] 44, [ʃpr] 17, [ʃpl] 5, [skr] 2, [skl] 1, [skv] 1, [spl] 1, [str] 1

As previously known, the first C in CCC is [s] or [ʃ]. Also, CC and CCC constitute 21% of all initial onsets. This percentage is higher than that of non-initial onsets, to be seen below.

10.5. NON-INITIAL ONSETS

There are 13,445 syllables in the 6,531 morphologically simple words. Their distributions in the word are shown in (23), ranging from the first syllable (S1) to the sixth (S6).

(23)

Syllable position	S1	S2	S3	S4	S5	S6	All
Count	6,531	5,081	1,398	357	65	13	13,445

If we exclude S1, there are 6,914 non-initial syllables. Their onsets range from zero to three-consonant, as shown in (24). The notation [C] is an "ambisyllabic" C in CELEX. For example, *bellen* is transcribed as [bɛ[l]ən], where [l] is ambisyllabic.

(24)

Onset	[C]	[C]C	[	[C	[CC	[CCC	All
Count	1,681	65	402	4,495	264	7	6,914

The vast majority of onsets do not involve consonant clusters. In particular, if we exclude the ambisyllabic [C], there are just 271 onset clusters, or 4%, which is far below the 21% of word-initial clusters. In (25) and (26) I list the types and frequencies of non-initial onsets, where the ambisyllabic [C] is ignored.

(25)

	Types	Tokens
[C] or [	1	2,083
[C]C or [C	23	4,560
[CC	29	264
[CCC	3	7
All	56	6,914

(26) C onsets: [t] 695, [d] 416, [r] 379, [n] 324, [b] 308, [l] 305, [g] 304, [z] 292, [k] 285, [m] 257, [p] 176, [f] 172, [ts] 138, [v] 130, [s] 102, [x] 74, [ʃ] 72, [h] 68, [j] 27, [ʒ] 17, [pf] 15, [tʃ] 2, [ʤ] 2
CC onsets: [tr] 48, [ʃt] 34, [st] 24, [gr] 24, [fl] 15, [br] 14, [kr] 12, [dr] 11, [pl] 9, [fr] 9, [ʃv] 7, [pr] 7, [ʃp] 6, [ʃm] 5, [gn] 5, [ʃl] 4, [gl] 4, [dl] 4, [bl] 4, [ʃr] 3, [kv] 3, [ʃn] 2, [sk] 2, [ks] 2, [kl] 2, [sts] 1, [pn] 1, [gv] 1, [dn] 1
CCC onsets: [ʃtr] 3, [str] 2, [ʃpr] 2

A comparison between initial and non-initial onsets is given in the next section. Onsets listed in other studies but not found in CELEX will also be discussed later.

10.6. ANALYSIS OF ONSETS

There are more initial onset clusters than medial onset clusters, both by onset types and by their frequencies. A comparison is shown in (27).

(27)

	Types (count)		Tokens (count)		Tokens (%)	
	Initial	Non-initial	Initial	Non-initial	Initial	Non-initial
[C] or [	1	1	876	2,083	13.4	30.1
[C]C or [C	23	23	4,312	4,560	66.0	66.0
[CC	38	29	1,271	264	19.5	3.8
[CCC	8	3	72	7	1.1	0.1
All	70	56	6,531	6,914	100	100

CC onsets occur 20% of the times initially but just 4% non-initially. In addition, CCC onsets occur seventy-two times initially but just seven times non-initially, even though the total number of initial onsets is roughly the same as that of non-initial onsets. In (28)–(30) I compare the types of onsets in initial and non-initial positions.

(28) Onsets found both initially and non-initially
C-onsets: [b], [d], [ʤ], [f], [g], [h], [j], [k], [l], [m], [n], [p], [pf], [r], [s], [ʃ], [t], [ts], [tʃ], [v], [x], [z], [ʒ]
CC-onsets: [bl], [br], [dr], [fl], [fr], [gl], [gn], [gr], [kl], [kr], [ks], [kv], [pl], [pn], [pr], [sk], [ʃl], [ʃm], [ʃn], [ʃp], [ʃr], [ʃt], [sts], [ʃv], [tr]
CCC-onsets: [ʃpr], [str], [ʃtr]

(29) Onsets found initially only (with counts)
CC-onsets: [fj] 1, [kn] 46, [pfl] 9, [pfr] 3, [ps] 4, [sf] 1, [sl] 3, [sm] 3, [sn] 1, [sp] 2, [sv] 1, [tsv] 29, [vr] 5
CCC-onsets: [skl] 1, [skr] 2, [skv] 1, [spl] 1, [ʃpl] 5

(30) Onsets found non-initially only (with counts): [dl] 4, [dn] 1, [gv] 1, [st] 24

As reported before, CCC onsets are made of an initial [s] or [ʃ] plus a CC cluster that can occur by itself. In previous analyses, such as Giegerich (1985, 1989), Hall (1992), and Wiese (1996), the German onset can hold two consonants, and the initial [s] or [ʃ] is either excluded or attached by a special rule. Such analyses assume that a CC onset is governed by sonority. For example, Giegerich (1989) proposes that a CC onset should be [−son][+son], which can be obstruent + sonorant or obstruent + nasal. Hall (1992) proposes a slightly different analysis, in which a CC onset can be obstruent + sonorant or [k, g] + nasal. For CC clusters that have the proper sonority sequence but do not occur, additional rules are proposed to rule them out.

My analysis differs from previous ones in two ways. The proposal is stated in (31).

(31) Present analysis of onset clusters
a. There is an extra C slot in word-initial position.
b. There is one onset slot (rather than two), which can contain a complex sound.

The idea in (31a) is as previously proposed, but I shall argue that it is not limited to [s] and [ʃ]. Also, unlike previous analyses, which assume that the German onset has two consonant positions, I argue that it has just one, although it can contain a complex sound.

My analysis predicts that only those CC clusters that can be represented as a complex sound are found medially, and CC clusters that cannot be represented as a complex sound do not occur medially. For example, both Giegerich (1989) and Hall (1992) assume that [kn] and [gn] are good onsets and expect them to occur both initially and medially. In contrast, in my analysis, [kn] and [gn] are not possible complex sounds and therefore cannot occur in medial onsets. I shall show that the present prediction is correct.

Let us begin with initial onsets. Initial C onsets need no discussion. Initial CCC onsets need no discussion either, since the CC part after removing the initial [s] or [ʃ] is found in CC onsets. So let us focus on CC onsets. In (32) I divide them into two groups: those that are possible complex sounds and those that are not (see Chapter 2).

(32) Initial CC onsets
  a. Possible complex sounds
    [tr] 101, [kl] 85, [kr] 83, [br] 69, [gr] 66, [fl] 53, [pl] 52, [pr] 49, [bl] 44, [fr] 35, [kv] 30, [gl] 28, [dr] 27, [pfl] 9, [vr] 5, [ps] 4, [pfr] 3, [fj] 1, [ks] 1
  b. True clusters (not possible complex sounds)
    [ʃt] 136, [ʃp] 76, [ʃl] 64, [ʃv] 48, [kn] 46, [ʃm] 36, [ʃn] 33, [ʃr] 29, [tsv] 29, [sk] 11, [gn] 3, [sl] 3, [sm] 3, [pn] 2, [sp] 2, [sts] 2, [sf] 1, [sn] 1, [sv] 1

The present analysis predicts that (i) the onsets in (32a) can occur medially and (ii) those in (32b) cannot. Prediction (i) is largely true, except that [pfl, vr, ps, pfr, fj] are not found medially. It can be seen, though, that these clusters have the lowest frequencies in (32a), and so their absence medially is understandable. Prediction (ii) does not appear correct at first sight, because many of the clusters in (32b) seem to occur medially. Consider the data in (33). There is also a cluster [gv], which is only found medially (just once, in *Ingwer*), but since it is a possible complex sound, it is not included in (33).

(33) True onset clusters (not possible complex sounds)
  a. CC-onsets found initially but not medially (with initial frequency)
    [kn] 46, [sf] 1, [sl] 3, [sm] 3, [sn] 1, [sp] 2, [sv] 1, [tsv] 29
  b. CC-onsets found initially and medially (with medial frequency)
    [ʃt] 34, [ʃv] 7, [ʃp] 6, [ʃm] 5, [gn] 5, [ʃl] 4, [ʃr] 3, [ʃn] 2, [sk] 2, [sts] 1, [pn] 1
  c. CC-onsets found medially only (with medial frequency)
    [st] 24, [dl] 4, [dn] 1
  d. CCC-onsets found medially (with medial frequency)
    [ʃtr] 3, [ʃpr] 2, [str] 2

The result in (33a) supports the present analysis, but the results in (33b–d) do not. Let us take a close look at the exceptions.

The clusters in (33b–d) consists of seventeen types and 106 tokens, which occur in 104 words (two words have two such clusters each). They can be grouped into several different cases, shown in (34)–(39). If a word can go with more than one group, it is only listed in one.

(34) Offending onset clusters after a tense V (e.g. *niedlich* [ni:][dlɪx] → [nid][lɪx]) (6 words)
Adler, *niedlich*, *Horoskop*, *Physiognomie*, *prognostizieren*, *Zustand*

(35) Offending onset clusters after a nasal (e.g. *Instanz* [ɪn][stants] → [ĩs][tants]) (9 words)
anstrengen, *instaendig*, *instant*, *Instanz*, *Instinkt*, *Institut*, *Instrument*, *konstatieren*, *Minstrel*

(36) Offending onset clusters before a suffix (-*e*, -*en*, -*er*, -*ig*) (24 words)
Borste, *Buerste*, *erste*, *Gerste*, *bersten*, *duensten*, *dunsten*, *leugnen*, *ordnen*, *Pfingsten*, *regnen*, *segnen*, *Althaendler*, *Elster*, *Fenster*, *finster*, *Hamster*, *Klempner*, *Kuerschner*, *Monster*, *Muenster*, *Polster*, *Traumwandler*, *garstig*

(37) [p^{s}] as an affricate (2 words)
obskur [ɔp][sku:r] → [ɔp^{s}][ku:r]
obszoen [ɔp][stsø:n] → [ɔp^{s}][t^{s}ø:n]

(38) Offending onset clusters after a short V (and a prefix) (22 words)
abgeschmackt, *beschlagnahmen*, *beschwichtigen*, *beschwipst*, *bespickt*, *bestaetigen*, *Besteck*, *bestehen*, *bestellen*, *bestimmen*, *Geschlecht*, *Geschmeide*, *Geschwader*, *geschwind*, *Geschwulst*, *Gespenst*, *Gestade*, *gestatten*, *gestehen*, *Gestell*, *ungeschlacht*, *ungestuem*

(39) Offending onset clusters at a morpheme boundary (41 words)
abspenstig, *Abstand*, *anheimstellen*, *Anstalt*, *ausstatten*, *Ausstich*, *bedienstet*, *Beispiel*, *bewerkstelligen*, *Buchstabe*, *Diebstahl*, *einschmelzen*, *einschreiten*, *einsprengen*, *einstehen*, *einverstanden*, *Entstalinisierung*, *erstunken*, *Gegenstand*, *Gestruepp*, *imstande*, *Langspielplatte*, *Linolschnitt*, *Pappenstiel*, *Schornstein*, *substantial*, *Tetrachlorkohlenstoff*, *Ueberschwang*, *umstritten*, *Unschlitt*, *unumschraenkt*, *Ursprung*, *verschmitzt*, *verschroben*, *verschwaegern*, *verstockt*, *verunstalten*, *vollstaendig*, *Vorschmack*, *widerspenstig*, *Zirkumpolarstern*

For the words in (34), the first C of the offending onset cluster can be moved into the preceding coda, because a tense vowel need not be long. For the words in (35), the first C of the offending onset cluster can be moved into the preceding coda, where the nasal has moved into the vowel. For the words in (36), if we remove the suffix, the offending cluster is no longer a medial onset. For the words in (37), the [s] of the offending onset cluster can move into the preceding coda and form an affricate [p^s]. For the words in (38), either the offending onset cluster is not genuinely medial, or the first C of the offending onset cluster can be moved into the preceding coda. Finally, for the words in (39), the offending onset clusters are not medial, because they lie at a morpheme or compound boundary. Thus, none of the 104 words contains a true onset cluster that is larger than a complex sound.

In conclusion, medial onset clusters are limited to those that can form a complex sound. An exhaustive examination of the CELEX German lexicon has yielded no counterexamples. This result, along with the finding that the German rhyme is limited to VX, supports the proposal that the maximal German syllable is CVX.

10.7. ONSET CLUSTERS LISTED IN OTHER STUDIES

In (40) and (41) I compare the onset clusters found in CELEX with those listed in Giegerich (1989), Hall (1992), and Wiese (1996).

(40) Onset clusters found in CELEX but not listed by Giegerich, Hall, or Wiese

Onset	Count	Example
[dl]	4	*Adler* [a:][dlər], *Althaendler* [alt][hɛn][dlər], *niedlich* [ni:][dlɪx], *Traumwandler* [traum][van][dlər]
[dn]	1	*ordnen* [ɔr][dnən]
[fj]	1	*Fjord* [fjɔrt]
[gv]	1	*Ingwer* [ɪŋ][gvər]
[sv]	1	*Swing* [svɪŋ]
[skv]	1	*Squaw* [skvo:]
[spl]	1	*Splen* [sple:n]
[str]	3	*Stratus* [stra:][tʊs], *Instrument* [ɪn][stru:][mɛnt], *Minstrel* [mɪn][strəl]

(41) Onset clusters not found in CELEX but listed by others

Onset	Hall (1992)	Wiese (1996)
[gm]	*Magma*	*Gmünd*
[km]		*Khmer*

[kt]	*ktenoid*	
[pʃ]	*Pschorr*	*Pschorr*
[pt]	*Ptolomäus*	
[ʃk]		*Schkopau*
[sr]		*Sri Lanka*
[tv]		*Twist*
[st]	*Stil*	*Stil, Stoiker*

The clusters in (40) have already been discussed. The cluster [st] in (41) is found medially in CELEX but not initially; the word *Stil* is transcribed as [sti:l] by Hall and Wiese but [ʃti:l] by CELEX. In any case, the clusters in (41) all occur word-initially, except [gm] in *Magma*. There are quite a few such words in CELEX, such as *Pigment*, *Magnet*, *Dogma*, etc. Hall syllabifies them as [CV][gmV . . .], but CELEX syllabifies them as [CV[g]mV . . .]. According to the Weight-Stress Principle, such words should be [CVg][mV . . .]. It is worth noting that [gm] is not found after a diphthong, i.e. there is no [CVV][gmV . . .]. Therefore, [gm] is not a compelling medial onset cluster. The conclusion based on the onsets in CELEX is still valid: i.e. there is just one onset slot in German.

10.8. ANOTHER LOOK AT WORD-FINAL CLUSTERS

We have seen that word-final rhymes in German can be up to VXCCC without inflection (e.g. *Herbst* [hɛrpst]) and VCCCCC with inflection (e.g. *(des) Herbsts* [hɛrpsts]). However, some sounds in the final cluster can form a complex sound. It is interesting, therefore, to ask how large the word-final cluster really is if we maximize complex sounds. First, let us consider VCCCCC, for which *(des) Herbsts* [hɛrpsts] is the only example, analyzed in (42).

(42) VCCCCC (1 word)

Reported	Analysis	Notes
[ɛrpsts]	[ɛ, r, p^s, t^s] VCCC	[p^s] and [t^s] as affricates

If [p^s] and [t^s] are affricates, then the actual pronunciation of VCCCCC is VCCC. Next consider VCCCC, analyzed in (43), where V is a short vowel.

(43) Word-final VCCCC (4 words)

Reported	Analysis	Notes
[Vlpst] 2	[V, l, p^s, t] VCCC	[p^s] as an affricate
[Vrnst] 1	[V, r̃, s, t] VCCC	nasalized [r]
[Vrpst] 1	[V, r, p^s, t] VCCC	[p^s] as an affricate

If [r] can be nasalized, then the clusters in (43) are again VCCC. Finally, we consider VVCCC, analyzed in (44), where VV is a diphthong or a long vowel and the rhymes [V:pst] and [V:mpt] have two possible analyses.

(44) Word-final VVCCC (10 words)

Reported	Analysis	Example
[VVnst] 3	[VṼ, s, t] VVCC	*einst* [ainst]
[V:pst] 3	[VV, p^s, t] VVCC	*nebst* [ne:pst]
	[V, p^s, t] VCC	
[V:rst] 2	[V, r, s, t] VCCC	*erst* [e:rst]
[V: rtst] 1	[V, r, t^s, t] VCCC	*Arzt* [a:rtst]
[V:mpt] 1	[Ṽ:, p, t] VVCC	*prompt* [pro:mpt]
	[Ṽ, p, t] VCC	

If a tense vowel can be short and [VN] can be [Ṽ], then VVCCC rhymes need not be longer than VXCC.

In conclusion, if complex sounds are used, the word-final rhyme in German is maximally VXCC, instead of VXCCCC.

10.9. SUMMARY

A German word can start with up to three Cs and end with up to five Cs; therefore, the maximal length of a monosyllabic word is in principle CCCVXCCCC. However, the string can be analyzed as (C)CVX(C)M, where (C) and M can be explained by morphology. M is one or more suffix or suffix-like consonants ([s] or [t] in German) and is covered by the "affix rule" (Chapter 3). The final (C), which can be almost any consonant, is supported by a "potential vowel" that comes from a vowel-initial suffix (Chapter 3). Similarly, the initial (C), which can be [ʃ, s] and sometimes [k, ts, g, p], is supported by a "potential vowel" that comes from a vowel-final prefix. When a potential vowel is not present, the initial or final (C) is supported by "anti-allomorphy" (Chapter 3). The remaining CVX is a maximal syllable, where C, V, and X can each be a complex sound (Chapter 2). In word-medial positions, no syllable exceeds CVX. Therefore, the maximal syllable in German is CVX, as it is in English.

The CVX analysis differs from previous ones in two ways: apparent non-final VXC rhymes are treated within VX (e.g. [VVN] → [ṼṼ], [V:C] → [VC], and [VNC] → [ṼC]), and onset CC clusters are analyzed as a complex sound (e.g. [kw] → [k^w], [br] → [b^r]). Both innovations make use of articulator-based feature theory and a strict definition of what a complex sound is (see Chapter 2).

11

Jiarong (rGyalrong)

Jiarong (rGyalrong) is a Tibeto-Burman language spoken in northern Sichuan, China. The speaker population is estimated to be between 150,000 (Qu 1990) and 200,000 (Lin 1993, Sun 2000). The speakers of Jiarong are Tibetans culturally, but the linguistic affiliation of Jiarong is not completely clear. Qu (1990) considers it to be a Tibetan language, while Sun (2000) considers it to be a Qiangic language, which is a sister of Tibetan.

Most words in Jiarong have a monosyllabic root, which often takes one or more CV prefixes (Lin 1993, Yan 2004). Jiarong is of interest because it reportedly has many onset clusters and coda clusters. For example, CCC onsets are common, and some Jiarong dialects also have CCCC onsets. Some examples are shown in (1).

(1) CCCC onset clusters in Jiarong (Lin 1993: 56)

Ergangli dialect		Caodeng dialect	
nspreɤ	'to scratch an itch'	ʃmbru	'boat'
nspʰjə pʰjə	'to exchange'	ʒmbri la ju	'willow tree'
nzbleɤ	'to steam'	ʒŋgri	'stars'
mfstɕə fstɕə	'mighty'	ʂŋɢri	'bamboo harmonica'

The numbers of onset clusters in Jiarong are rather large; in some dialects there are 300 or more (Lin 1993). Coda clusters in Jiarong can contain up to three consonants. Some examples are shown in (2).

(2) CCC coda clusters in the Caodeng dialect of Jiarong (Lin 1993: 594–5)

[-vndʒ]	ndʒə-ɟji tə-cçʰovndʒ	'You two will break(it)'
[-rndʒ]	ndʒə-ɟji tə-qurndʒ	'You two will help'
[-mndʒ]	ndʒə-ɟji tə-rtsomndʒ	'You two will compile'
[-ŋndʒ]	ndʒə-ɟji tə-fsroŋndʒ	'You two will protect'

Given such data, we want to ask whether Jiarong is a problem for the CVX theory. I shall argue that it is not. In particular, I shall show that extra coda consonants, such as [-nʤ], only occur word-finally and are either suffixes or suffix-like. In addition, onset clusters only occur word-initially or root-initially and they can be explained in two ways. First, a root-initial C can be accounted for in terms of a potential V and anti-allomorphy (Chapter 3), where the potential V comes from a CV prefix. Second, a nasal or [r] in the middle of a cluster may be syllabic; for example, *zndai* 'wall' can be analyzed as [zn̩][dai].

11.1. ONSET CLUSTERS IN JIARONG

Lin (1993: 526) divides Jiarong into three dialect areas: the eastern dialects are represented by Zuokeji, the northern by Caodeng and Ribu, and the western by Ergangli. The number of onsets in these dialects are shown in (3), where the eastern dialect Zhuokeji (ZKJ) was surveyed twice, first in the 1950s and then in the 1980s. Zhuokeji in the 1980s has more onset clusters than Zhuokeji in the 1950s, which shows that a dialect with more onset clusters is not necessarily older (a point to which I shall return).

(3) Onset counts in Jiarong dialects

	ZKJ 50s (East)	ZKJ 80s (East)	Caodeng (North)	Ribu (North)	Ergangli (West)
C	32	34	41	40	41
Clusters	200	227	299	282	325
All	232	261	340	322	366

Except Ergangli, all the dialects in (3) are based on Lin's own fieldwork. Among them Caodeng has the most number of onset clusters, and I shall focus on this dialect here.

The onsets in Caodeng are listed in (4)–(7). Lin (1993) roughly divides CC-onsets into those that end in an approximant and those that do not; similarly, CCC-onsets are divided into those that end in an approximant and those that do not. I have followed his lists, although there are some inconsistencies. In particular, while [lj, wr, wl, jr, jl] are listed with those that end in an approximant, [rw, rl, rj, lw, jl, jw] are not. The cluster [vl] was listed twice erroneously; I have deleted one and adjusted the total number of clusters from 300 to 299.

(4) C-onsets in Caodeng Jiarong (41 in all)

p, p^h, b	t, t^h, d				k, k^h, g	q, q^h, ɢ	
	ts, ts^h, dz	ʈʂ, ʈʂ^h, ɖʐ	ʧ, ʧ^h, ʤ	cç, cç^h, ɟj			
v	s, z, ɬ	ʂ	ʃ, ʒ		x	ʁ	h
	l	r					
m	n			ȵ	ŋ		
w							

(5) CC-onsets in Caodeng Jiarong (229 in all)
pr, pw, pj, p^{h}r, p^{h}w, p^{h}j, tw, tj, t^{h}w, t^{h}j, kr, kl, k^{h}r, gr, qr, ql, qj, q^{h}r, tsw, tsr, ʧw, ʧ^{h}w, cç^{h}w, vr, vl, vj, sr, sl, sw, sj, zr, zl, zw, ʃl, ʃw, ʒr, ʁl, ʁj, lj, wr, wl, jr, jl p^{h}ʃ, k^{h}ʃ, q^{h}ʃ, vt, vt^h, vd, vk, vg, vq, vts, vts^h, vʧ, vʧ^h, vcç, vɟj, vs, vz, vʃ, vʒ, sp, st, st^h, sk, sk^h, sq, scɕ, scɕ^h, sm, sn, sȵ, sŋ, zd, zg, zɟj, zv, zʁ, ʃp, ʃp^h, ʃt, ʃt^h, ʃk, ʃk^h, ʃq, ʃq^h, ʃʈʂ, ʃʧ, ʃm, ʃn, ʃŋ, ʒp, ʒg, xp, xp^h, xt, xts, xʈʂ, xʧ, xʧ^h, xcç, xcç^h, xɟj, xs, xʃ, ʁd, ʁg, ʁv, ʁz, ʁʒ, ʁm, ʁn, ʁȵ, rp, rb, rt, rd, rk, rg, rq, rq^h, rts, rts^h, rdz, rʧ, rʧ^h, rʤ, rcç, rcç^h, rɟj, rv, rs, rz, rʒ, rʁ, rl, rm, rn, rȵ, rŋ, rw, rj, lp, lt, lt^h, ld, lk, lʧ, lʤ, lɟj, lv, lʁ, ln, lŋ, lw, mp, mp^h, mb, mt, mt^h, md, mk, mk^h, mq, mts, mts^h, mdz, mʧ, mʧ^h, mʤ, mcç, mcç^h, mɟj, mʃ, mn, mȵ, mŋ, np, np^h, nt, nt^h, nd, nk, nk^h, ng, nq, nts, nts^h, ndz, nɖʐ, nʧ, nʧ^h, nʤ, ncç^h, nɟj, nm, nŋ, ŋk^h, ŋg, ŋq, ŋɢ, wp, wt, wd, wts, wts^h, wʧ, wʧ^h, wɟj, wv, ws, wz, wʃ, wm, wn, wȵ, jp, jt, jd, jq, jts, jʧ, jʁ, jl, jm, jŋ, jw

(6) CCC-onsets in Caodeng Jiarong (65 in all)
mpj, mp^{h}r, mp^{h}j, mbr, mbl, mbw, mbj, mk^{h}r, mgr, mgl, mq^{h}r, mɢl, mtsr, nbw, ntsw, ndzr, ndzw, nʧ^{h}w, nʤw, ŋk^{h}r, ŋgl, ŋgr, vkr, vʧw, vsr, vʃw, vsj, spr, skr, sqr, sq^{h}j, spj, zgr, ʃkr, ʃqr, ʒgr, xpj, ʁgr, rqj, rtsw, rts^{h}w, rcɕw, lbw, wpr mp^{h}ʃ, vrd, vzɟj, znd, ʒŋg, ʁnd, ʁnɟj, rmb, rnd, rŋg, rnʤ, rnɟj, lndz, wmb, wnd, wnʤ, wnɟj, wrŋ, jmb, jnd, jŋg

(7) CCCC-onsets in Caodeng Jiarong (6 in all)
ʃmbr, ʒmbr, ʒngr, ʒŋɢr, rmbj, jndw

Many of the clusters pose a problem for the sonority analysis. For example, [pr, kr, kl], which have rising sonority, and [rp, rk, lk], which have falling sonority, are both found. I review two previous analyses before presenting my own.

11.1.1. **Previous analyses**

Lin (1993: 66–7) observes that previous studies of Jiarong rarely offer arguments for syllable divisions: the reader is often left wondering whether the proposed syllable divisions are correct and why Jiarong seems so different from other languages.

Lin (1993) offers two arguments for his analysis. First, he relies on his intuition as a Jiarong speaker. Second, he uses evidence from morphology. When the two are in conflict, the second takes priority. For example, consider the root [rbo] 'drum' in (8).

(8) tə-rbo 'drum'
no nə-rbo 'your drum'

Lin (1993: 35) points out that in [tə-rbo] 'drum' the [r] sounds like the coda of the prefix, but when a different prefix is added, [rb] stays with the root. Therefore, he concludes that [rb] is always syllabified with the root.

Lin seems to assume that a morphological boundary should coincide with a syllable boundary. However, no evidence is offered. In fact, few phonologists hold the assumption. For example, most phonologists would assume the analysis in (9), where syllable boundaries and morpheme boundaries do not coincide.

(9) Morphology *help-er*
Syllables [hɛl][pɚ]

Lin (1993: 66) is aware that his analysis of syllables in Jiarong is rather unusual. His reply is that Jiarong is simply a unique language that has unusual syllables.

Hsieh (1999) offers an analysis of the Zuokeji dialect. Following a sonority-based analysis, Hsieh proposes that Jiarong has a special sonority requirement, rephrased in (10).

(10) Sonority requirement for onset clusters in Zuokeji Jiarong (Hsieh 1999)
The sonority cannot rise twice before the nuclear vowel.

Hsieh assumes the sonority scale in (11), with examples in (12), where V is a "vowel," R an "approximant," N a "nasal," Z a "voiced fricative," S a "voiceless continuant," DZ a "voiced affricate," TS a "voiceless affricate," D a "voiced stop," and T a "voiceless stop."

(11) Sonority scale (9 = highest; 1 = lowest)

V	R	N	Z	S	DZ	TS	D	T
9	8	7	6	5	4	3	2	1

(12) Sonority analysis of Jiarong clusters

Cluster	Sonority	Comment
sl + V	5-8-9	one rise; good
st + V	5-1-9	one rise; good
spr + V	5-1-8-9	one rise; good
slŋ + V	5-8-7-9	two rises; bad

There are two problems in Hsieh's proposal. First, like Lin (1993), Hsieh assumes that Jiarong is unusual, where sonority can fall in the onset; in contrast, in other languages sonority cannot fall in the onset. Second, Hsieh's analysis cannot extend to other dialects. For example, the Caodeng dialect has CCC onsets that would lead to two sonority rises, as shown in (13).

	Cluster	Sonority	Comment
(13)	vrd + V	6-8-2-9	two rises
	ʒŋg + V	6-7-2-9	two rises
	znd + V	6-7-2-9	two rises

In addition, it can be seen that four of the six CCCC onsets in Caodeng contain two sonority rises, which also pose a problem for Hsieh's analysis.

In summary, previous analyses essentially assume that Jiarong has unusual syllable structures, which are not found in other languages.

11.1.2. **Present analysis**

A relevant fact of Jiarong is that it has many lexical prefixes, which are mostly CV. For example, Suomo Jiarong has the lexical prefixes shown in (14), where position 1 is closest to the root and position 3 is furthest away.

(14) Lexical prefixes in Suomo Jiarong (Yan 2004: 120–1)

Mostly position 1	tɑ-, tə-, kɑ-, kə-, ɑ-
Mostly position 2	mɑ-, sɑ-, wə-
Mostly position 2 or 3	nɑ-, nə-, wɑ-, ʃɑ-, sɑ-, rɑ-, rə-, mə-, ŋɑ-

Lexical prefixes are used when a root occurs alone, but can be absent when a root combines with other roots or words. The meanings of lexical prefixes are not always obvious (Sun 1998: 112). Yan (2004: 120) suggests that in Suomo Jiarong, [tɑ-] and [tə-] are usually prefixes to nouns, [kɑ-] to verbs, [kə-] to adjectives, and [ɑ-] to locatives. A root in Suomo Jiarong can take up to three prefixes. Some examples are shown in (15), from Yan (2004: 121).

(15)	One prefix	tɑ-rpɑm 'ice'	kə-mɑk 'deep'
	Two prefixes	kɑ-sɑ-jɑ 'begin'	kə-mɑ-ʃɛə 'rich'
	Three prefixes	tə-kɑ-ŋɑ-ro 'all'	kə-sɑ-wɑ-ri 'funny'

As discussed in Chapter 3, V and CV prefixes provide a "potential V," which can take an extra root-initial C as its coda. When a prefix is absent, the extra root-initial C is supported by "anti-allomorphy," which aims to keep a morpheme in the same shape whether it has an affix or not (Chapter 3). An example is shown in (16) and (17), from Lin (1993: 36).

(16) tə-ʒba [təʒ][ba] 'face'

(17) ʒba ntɕʰok ʒ[ban][tɕʰok] 'face dip (dimple)'

In (16), the root-initial cluster [ʒb] is split between two syllables. In (17) 'face' has dropped its prefix, but the initial [ʒ] is kept by anti-allomorphy. Since anti-allomorphy can explain the [ʒ] in (17), there is no problem for the CVX theory.

The combination of a potential V and anti-allomorphy can account for all CC clusters. In addition, they can account for those CCC clusters whose last two CC can form a "complex sound" (Chapter 2), such as [mpj, mpʰr, mpʰʃ, mpʰj, mbr, mbl, mbw, mbj, mkʰr, . . .]. The remaining CCC clusters in Caodeng are shown in (18), along with the six CCCC clusters. They total 29, or 10% of all the clusters.

(18) vrd, znd, ʒŋg, ʁnd, ʁnɟj, rmb, rnd, rŋg, rnʤ, rnɟj, lndz, wmb, wnd, wnʤ, wnɟj, wrŋ, jmb, jnd, jŋg, mtsr, ndzr, vsr, vzɟj
ʃmbr, ʒmbr, ʒngr, ʒŋɢr, rmbj, jndw

Most of these clusters contain a medial [r] or nasal. If sonorant consonants can be syllabic, as they can in English, then these clusters do not contain extra consonants, as shown in (19).

(19) zndai [zn̩][dai] 'wall'
ʁnɟju [ʁn̩][ɟju] 'window'
ʃmbru [ʃm][bru] 'boat'
ʒngri [ʒn̩][gri] 'stars'
ʒŋɢri [ʒŋ̍][ɢri] 'bamboo harmonica'

The idea that sonorant consonants in a cluster can be syllabic has also been proposed for Georgian. Some examples are shown in (20), from Butskhrikidze (2002: 88).

(20) Syllabic consonants in Georgian
naɣm-s [naɣm̩s] 'mine'
ipn-s [ipn̩s] 'ash tree'
saxl-s [saxl̩s] 'house'
tetr-s [tetr̩s] 'white'
k'lde [k'l̩de] 'rock'
trtvili [tr̩tvili] 'hoar-frost'

There are four onset clusters still to be accounted for, which are shown in (21), with a sample word each.

(21)

Cluster	Word	Gloss
mtsr	k^{h}ɐ-mtsrof	‘suck’
ndzr	kɐ-ndzri	‘twist’
vsr	kɐ-vsroŋ	‘protect’
vzɟj	kɐ-vzɟjər	‘change’

A possible solution is provided by Lin (1993), who points out that the consonants in a cluster are often separated by [ə]. Some examples are shown in (22), from the Ribu dialect.

(22) Optional [ə] in consonant clusters in Ribu Jiarong (Lin 1993: 448)

With [ə]	Without [ə]	
tə-ʁɛ	tʁɛ	‘musk deer’
kə-mə-r̥kə	kə-mr̥kə	‘thief’
kə-nə-mʧ^{h}ə	kə-nmʧ^{h}ə	‘early’
kə-mə-ɢɛ	kə-mɢɛ	‘thief’

Given such cases, the Caodeng clusters in (21) can be analyzed in (23).

(23)

k^{h}ɐ-mtsrof	[k^{h}ɐ][məts][rof]	‘suck’
kɐ-ndzri	[kɐ][nədz][ri]	‘twist’
kɐ-vsroŋ	[kɐ][vəs][roŋ]	‘protect’
kɐ-vzɟjər	[kɐ][vəz][ɟjər]	‘change’

In this analysis, no syllable exceeds CVX. Indeed, it is possible that the clusters in (19) have an optional [ə], too. For example, ‘wall’ could be either [zn̩][dai] or [zən][dai].

11.1.3. **Summary**

Given CV prefixes, the possibility that nasals and [r] can be syllabic, and the fact that [ə] is often inserted into a consonant cluster, all Jiarong onset clusters can be accounted for within the CVX theory. In (24) I list the typical cases and their analysis, where R is an approximant that can combine with the preceding C to form a complex sound C^R, N is a syllabic consonant ([r] or a nasal), and (CV) is a potential prefix.

(24)

Cluster	Analysis
CRV	[(CV)C][RV], [(CV)][C^{R}V], [C^{R}V]
CCV	[(CV)C][CV], [Cə][CV], C[CV]
CCRV	[(CV)C][C^{R}V], [Cə][C^{R}V], C[C^{R}V]
CNCV	[CN̩][CV], [CəN][CV]
CCCV	[CəC][CV], [(CV)C][Cə][CV]
CNCRV	[CN̩][C^{R}V], [CəN][C^{R}V]

Each cluster may have two or more analyses, depending in part on whether there is a prefix. For example, CC may contain an inserted [ə] when there is no prefix (i.e. [Cə][CV]), but not when there is a prefix (i.e. [(CV)C][CV]). Such predictions remain to be verified.

11.2. CODA CLUSTERS

Excluding suffixes, the maximal Jiarong rhyme is either VC or VG, where V is a short vowel and VG is a diphthong (Lin 1993: 598). The only exception is the rhyme [ols], which occurs in the word [qʰols] 'scripture bag' of the Ergangli dialect. I shall return to this case.

Jiarong words can take a vowel suffix, which can be [i], [u], or [ɑ] (Lin 1993: 581–3). The suffix can create what Lin calls a "triphthong," such as [ɛii], [ɐiɑ], and [ouɑ]. An example from Caodeng is shown in (25), where the verb root is [sjou] 'finish', followed by two suffixes.

(25) ŋi sjou-ɑ-ŋ 'I shall finish'

Jiarong also has consonant suffixes, which can be C or CC. The suffixes can vary from dialect to dialect. Some examples are shown in (26).

(26) Consonant suffixes in Jiarong (Lin 1993: 588–97)

Caodeng dialect	[n, ʧʰ, nʤ]
Ergangli dialect	[ŋ, n, ɲ̥, s, ɣ]

Since an unsuffixed word can end in VC, a consonant suffix can create a coda cluster of CC or CCC. Two examples in Caodeng are shown in (27), where the verb root is [qur] 'help'.

(27) Coda clusters in the Caodeng dialect of Jiarong (Lin 1993: 589, 594)

[r-n]	nə-ɟji rə-qurn	'You will help'
[r-nʤ]	nʤə-ɟji tə-qurnʤ	'You two will help'

Having presented the basic data, I review two previous analyses, Lin (1993) and Hsieh (1999). Then I offer an analysis in the CVX theory.

11.2.1. **Previous analyses**

Lin (1993) assumes no limit to how many vowels or consonant a rhyme can contain. For example, he calls the vowel sequences [ɛi-i], [ɐi-ɑ], and [ouɑ] "triphthongs," discussed in the section on "rhymes." Similarly, CC and CCC clusters created by suffixing are considered to be in the coda. Thus, [sjouɑŋ] in (25) is a single syllable, and so is [qurnʤ] in (27).

Lin offers no evidence that [sjouɑŋ] constitutes a single syllable, instead of two syllables [sʲou][ɑŋ]. The latter is similar to the standard analysis in English, such as *going* [gou][ɪŋ] and *buying* [bai][ɪŋ]. Lin does not explain why Jiarong is different from other languages.

Hsieh (1999) takes a similar view and considers consonant suffixes to be in the coda. In particular, although he otherwise assumes that the maximal rhyme is VX, he assumes another constraint, shown in (28), which can override the VX limit on rhyme size.

(28) PARSE-(Morpheme)
Morpheme M in the input must be parsed in the output.

Because of (28), a suffixed word like [qur-nʤ] is thought to form an extra-large syllable [qurnʤ], where the rhyme is VCCC.

Hsieh offers no evidence for the syllabification assumed. In addition, there is some redundancy in the analysis: if a suffix C can be kept by (28), there is no need to assume that it must be inside a syllable as well. Hsieh seems to follow the "all-in" assumption that every sound must belong to a syllable, an assumption I have argued against in Chapter 3. If we give up the assumption, we can account for suffix Cs without expanding the VX rhyme size.

11.2.2. **Present analysis**

I assume that a vowel suffix can create a separate syllable in Jiarong, as it does in English. For example, in *ŋi sjou-a-ŋ* [ŋi][sjou][ɑŋ] 'I shall finish', the suffixes form a syllable [ɑŋ], similar to *see-ing* [si:][ɪŋ] in English. Therefore, in what follows I focus on consonant suffixes only.

The analysis of coda clusters in Jiarong is quite simple. It is shown in (29), where M is a suffix consonant.

(29) Coda clusters in Jiarong

Maximal string	VVCC, VGCC, VCCC
General	[VX]MM

In the analysis the rhyme is VX. Additional consonants MM can occur because they are suffixes (or suffix-like). Therefore, there is no need to assume that MM are part of the preceding syllable. M or MM may or may not form a separate syllable, depending on the nature of the sound(s). For example, when the suffix contains a nasal, it may form a separate syllable. Two examples from Zuokeji Jiarong are shown in (30), from Lin (1993: 588, 593), where I use brackets to indicate possible syllabification.

	Transcription	Possible analysis	Gloss
(30)	ŋɑ top-ŋ	[ŋɑ][top][ŋ]	'I shall hit'
	nʤo tə-wɐt-nʧʰ	[n][ʤo][tə][wɐt][n]ʧʰ	'you two will wear'

When M is an obstruent, there is no evidence that it forms a separate syllable. Therefore, I assume that it is unsyllabified. Two examples from Zuokeji Jiarong are shown on (31), from Lin (1993: 590, 592), where [cçʰ] is an affricate.

	Transcription	Possible analysis	Gloss
(31)	ŋə nʤɛ-cçʰop-ʧʰ	[ŋən][ʤɛ][cçʰop]ʧʰ	'we two shall break (it)'
	ŋə nʤɛ-kɐk-ʧʰ	[ŋən][ʤɛ][kɐk]ʧʰ	'we two shall peel'

The final [ʧʰ] is kept not because it is part of the preceding syllable but because it is a suffix.

In the Jiarong dialects examined by Lin (1993), there is one unsuffixed word that has a VCC rhyme. It occurs in the Ergangli dialect and is shown in (32).

(32) Unsuffixed VCC rhyme in Ergangli Jiarong (Lin 1993: 598, 600)
qʰols 'scripture bag'

If the rhyme is limited to VX, we must explain why the final [s] is able to stay, even though it is not a suffix. The answer lies in the "affix rule" (Chapter 3), which allows not only suffix consonants but also suffix-like consonants to occur without being in a syllable. In other words, the final [s] in [qʰols] can be explained if [s] is a suffix elsewhere in Ergangli Jiarong. This is indeed the case. Consider the suffixes in the three dialects of Jiarong, shown in (33).

(33) Consonant suffixes in three Jiarong dialects (Lin 1993: 588–97)

Zuokeji (east)	-ŋ, -ɲ, -ʧʰ, -nʧʰ
Caodeng (north)	-ŋ, -n, -ʧʰ, -nʤ
Ergangli (west)	-ŋ, -n, -ɲ, -s, -ɣ

While Ergangli indeed has the suffix *-s*, the other two dialects do not. This is consistent with the fact that no extra final [s] occurs in Zuokeji or Caodeng.

11.2.3. **Summary**

I have shown that the maximal non-final rhyme in Jiarong is VX. Extra consonants beyond VX can occur at word edges, but they can be explained by morphology: they are either suffixes or suffix-like, and so they are covered by the "affix rule" (Chapter 3). Therefore, there is no need to assume any rhyme larger than VX.

11.3. A HISTORICAL PERSPECTIVE

In this section I discuss two questions. First, how did consonant clusters arise in Jiarong? Second, what is likely to happen to the consonant clusters?

For the first question, we know that consonant clusters can come from vowel deletion. Some examples in English, both historical and synchronic, are shown in (34).

(34) Consonant clusters from vowel deletion in English
Historical (Gimson 1970: 237–8)
Initial *state* [est], *scholar* [esk]
Medial *Gloucester* [st], *Leicester* [st], *evening* [vn], *lightening* [tn]
Synchronic (in casual speech)
Initial *potato* [pt], *police* [pl]
Medial *flavoring* [vr], *thickening* [kn]

The same process seems to be going on in Jiarong. Some examples are shown in (35), where CVCVCV has changed to CVCCV.

(35) Vowel deletion in Jiarong dialects (Lin 1993: 9)

Suomo	Zuokeji	
kə-mə-ŋam	kə-mŋam	'pain'
ka-mə-sam	ka-msam	'hear'
ka-mə-ʧur̥	ka-mʧir̥	'make circles'
ka-mə-tɕa	ka-mtɕa	'many'

Although the examples do not tell us whether the change is vowel deletion (from Suomo to Zuokeji) or vowel insertion (from Zuokeji to Suomo), further evidence does. According to Yan (2004), vowel deletion reflects the simplification of lexical prefixes, with the eastern dialects having the most prefixes and western dialects having the fewest. In prefix simplification, vowel deletion takes place first, creating a consonant cluster, which is then simplified. Some examples are shown in (36) and (37).

(36) Prefix simplification in Jiarong dialects (Lin 1993: 9, 634, 665)

Suomo (east)	Zuokeji (east)	Ergangli (west)	
ʧə tə	ʃtə	tɕə	'this'
kə-mə-ŋam	kə-mŋam	ŋam	'pain'

(37) More examples of prefix simplification in Jiarong dialects (Yan 2004: 124)

Suomo (east)	Caodeng (north)	Zhongzhai (west)	
kɑ-ʃə-rŋɑ	kɑ-sə-rŋɑ	sŋe	'lend to'
kɑ-nɑ jo	kɑ nɑ nɟɑ	nɑnjo	'wait for'
kə-rɑm	kə-rɑm	ʁru	'dried food'
kɐ-jɐs	tɑk-jɛs	ʁjɐ	'soot'
kə-tsu	ʁzə	zə	'monkey'

The data show that older forms of Jiarong, such as the Suomo dialect, in fact do not have many large consonant clusters. When vowel deletion occurs, many large consonant clusters are created. However, such clusters are likely to be simplified soon, as seen in some western dialects. Similarly, consider some cognate words in Old Chinese, written Tibetan, and Trung, shown in (38), from Mei (2006: 600).

(38)

Old Chinese	Written Tibetan	Trung	
snjit	–	sɯ-ɲɪt	'seven'
sljiak	zla-ba	sɯ-la	'evening/moon'
sljəp	slob-pa	sɯ-lap	'practice/study/teach'

The initial consonant clusters in Old Chinese and written Tibetan did not stay for long. Modern Chinese has no initial consonant clusters. In written Tibetan, there are many initial consonant clusters, which likely reflect the pronunciation of the seventh century, when writing was first introduced to Tibetan. However, in modern Lhasa Tibetan, there are hardly any initial consonant clusters. It is likely, therefore, that Jiarong is moving in the same direction.

11.4. SUMMARY

Jiarong has many large consonant clusters, but the syllable size need not be large. In some cases, the first consonant of a root-initial cluster is supported by a CV prefix, serving as the coda of the prefix vowel (e.g. [tɑ-rpɑm] → [tɑr][pɑm] 'ice'). In other cases, the first two consonants of a root-initial cluster are supported by an optional [ə] (e.g. [zndai] → [zən][dai] 'wall'). Word-final clusters mostly come from consonant suffixes; otherwise, the rhyme is VX. Therefore, there is no need to assume that the Jiarong syllable is larger than CVX.

Consonant clusters in Jiarong are mostly the result of vowel deletion in CV prefixes (Yan 2004). But owing to their lack of semantic or syntactic functions, many prefixes are dropping out, as seen in the western dialects. The next step appears to be the simplification of onset clusters, which is happening in some western dialects. It is likely that Jiarong is moving in the direction of modern Lhasa Tibetan, which has lost most onset clusters from classic Tibetan.

12

Theoretical implications

12.1. THE CVX THEORY AND ITS PREDICTIONS

I have discussed several languages and proposed that their maximal syllable is CVX. The proposal is called the CVX theory and is summarized in (1).

(1) The CVX theory
  a. The word structure is C_mCSCC_m, where C_m is one or more affix or affix-like consonants, C is a potential coda of a prefix-final V or a potential onset of a suffix-initial V, and S is one or more syllables. Since C_m and C can be accounted for by morphology, they need not be part of S.
  b. The maximal size of a syllable is CVX (CVC or CVV), where V can be filled by a vowel or a syllabic consonant. C can be a complex sound (such as [k^w] and [ts]), and so can V (such as the nasalized vowel in [t^hɛ̃t] *tent*).
  c. Syllabification is determined by the Weight-Stress Principle, according to which the rhyme of a stressed syllable is VX and that of an unstressed syllable is V.

Extra consonants at word edges can be predicted from morphology as follows:

(2) Predicting word-edge consonants from morphology
  a. If a language has vowel-initial suffixes, there can be an extra C in root-final position.
  b. If a language has vowel-final prefixes, there can be an extra C in root-initial position.
  c. If a language has consonant prefixes or suffixes, they can appear at word edges as additional extra Cs.
  d. The maximal syllable is CVX otherwise.

Although I have only discussed several languages, there are reasons to believe that the CVX theory has a more general nature. First, English and German have large consonant clusters and have long been thought to have very large syllables. If their syllables are in fact no larger than CVX, one would like to re-examine other languages that are thought to have very large syllables. Second, there is evidence for the CVX theory from major language families. In the Indo-European family, I have discussed English and German, and others have discussed Polish (Bethin 1992), Georgian (Butskhrikidze 2002), and Hindi (Kumar 2005). In the Sino-Tibetan family, I have discussed Chinese and Jiarong. In the Amerindian family, there is similar evidence from Bella Coola (Bagemihl 1991) and Spokane Salish (Bates and Carlson 1992). Although I have not discussed African languages, they are known to have fairly simple syllable structures. Thus, evidence for the CVX theory goes beyond a few random languages.

A skeptic may point out that even if the CVX theory is true for the languages I have mentioned, there could be others that allow larger syllables. Indeed, even if CVX is the maximal syllable size in all languages we have ever known, past or present, it does not necessarily mean that future languages cannot use larger syllables. Therefore, the CVX theory could be an artifact of languages that happen to have existed, rather than a principle that limits possible syllables.

There are three criticisms of the skeptic's position, which we can call the "super-syllable" theory. First, if there is no evidence for super-syllables, there is no reason to assume them. Second, there is no way to falsify it. Consider some predictions of the CVX theory and "super-syllable" theory against possible findings, shown in (3).

(3)	Findings	CVX theory	"Super-syllable" theory
	"Super-syllables" not found	Prediction correct	Prediction correct
	"Super-syllables" found	Prediction wrong	Prediction correct

Regardless of empirical findings, the "super-syllable" theory is always correct, and so it has no predictive value. Third, it is hard to distinguish one super-syllable theory from another. For example, one super-syllable theory can claim that the maximal syllable is CCVXCC, and another can claim that the maximal syllable is CCCVXCCC. Neither claim is falsifiable (because there are probably no syllables larger than CVX).

A further shortcoming of the super-syllable theory is that it offers no guidance for research. Consider the scenarios in (4).

(4)

Scenarios	CVX theory	"Super-syllable" theory
"Super-syllables" reported	Study needed	Study not needed
"Super-syllables" disproved	Result important	Result trivial
"Super-syllables" confirmed	Result important	Result trivial

When a language appears to have super-syllables, the CVX theory calls for a close study, because super-syllables are unexpected. In contrast, the super-syllable theory calls for no further study, because super-syllables are expected. When reported super-syllables are disproved, the CVX theory considers the result to be important, because a problem is solved and the theory is confirmed. For the super-syllable theory, the result is trivial, because both small syllables and super-syllables are expected. When super-syllables are confirmed, the CVX theory considers the result to be important again, because a change in theory is needed. In contrast, for the super-syllable theory the result is trivial again, because both small syllables and super-syllables are expected and no new insight is gained. Therefore, from a methodological point of view, the CVX theory is better, because it is both simpler and more likely to yield research progress, as I have shown in the analysis of English, German, and Jiarong.

12.2. THE CVX THEORY AND THEORIES OF GRAMMAR

Let us now consider the CVX theory in a broader perspective, in relation to theories of grammar.

12.2.1. **Tabula rasa**

A common view of many linguists is that children are born with no grammar. Instead, they learn the grammar that a speech community has arbitrarily constructed. Let us call this view the tabula rasa (blank slate) theory. Of course, just as no one can draw without at least something, such as chalk and a slate, no one disputes the fact that humans have some physical endowment, such as ears and the vocal tract. What the tabula rasa theory claims is that the child is born without specific linguistic knowledge. For example, Abercrombie (1967: 70) says:

> The full range of possible human phonetic performance is very wide . . . Only a selection, however, from this full range is put to use by the speakers of any single language—a selection, moreover, which is not only limited, but also different in very many (one might almost say nearly all) respects from the selection used by speakers of every other language.

Similarly, Port and Leary (2005) argue that there are no universal phonological categories. Instead, each language makes up its own categories, which differ from those in other languages. Similarly, Goldsmith (2007) argues that, given a body of data, we can construct a grammar with a probabilistic model without inherent linguistic knowledge.

If each language makes up its own syllable structure, we expect syllables to differ from language to language, probably dramatically. For example, Abercrombie (1967: 73–4) suggests that the maximal syllable is CCCVCCCC in English, CVC in Cantonese, and CV in Japanese. However, I have argued that the maximal syllable is CVX in English, and it can be shown that it is CVX in Japanese too. If the CVX theory is correct, there is a question for the tabula rasa theory: Why is there so little variation in syllable structure?

One could construct a proposal that derives syllable structure from perceptual, functional, or other pragmatic reasons. One might even imagine a functional or pragmatic explanation of why the maximal syllable ends up the same in all languages, even though they start out with nothing. However, such proposals remain to be seen and validated.

Indeed, even for those who view grammar as a pure probabilistic model, such as Goldsmith (2007), the computation would be easier if there are inherent constraints. For example, there is no need to consider dog barking or bird chirping (not human imitations of them) as possible human speech. Similarly, if the CVX theory is correct, there is no need to consider many other conceivable syllable sizes. Such a grammar would be not only simpler but also a better model of human behavior.

12.2.2. **Substantive and formal universals**

Chomsky (1965) argues that part of our knowledge of language is innate, which is "universal grammar." In addition, Chomsky and Halle (1968: 4) state:

> It is useful to divide linguistic universals roughly into two categories. There are, first of all, certain "formal universals" that determine the structure of grammars and the form and organization of rules. In addition, there are "substantive universals" that determine the set of elements that may figure in particular grammars. For example, ... general linguistic theory might propose, as substantive universals, that the lexical items of any language are assigned to fixed categories such as noun, verb, and adjective, and that phonetic transcriptions must make use of a particular, fixed set of phonetic features.

Chomsky and Halle (1968) offer no discussion of syllable structure. But if the maximal syllable is CVX in all languages, we could add it to the list of formal or substantive universals.

12.2.3. **Principles and parameters**

Chomsky (1981) proposes that universal grammar consists of a set of inviolable constraints, or "principles," and a set of "parameters," whose values are set by each given language. Parameters for the maximal syllable have been proposed (e.g. Clements and Keyser 1983, Blevins 1995). Some parameters of Blevins (1995: 219) are shown in (5), each having a binary choice.

(5) Can the onset contain two sounds?
Can the nucleus contain two sounds?
Is the coda allowed?
Can the coda contain two sounds?
Can extra C occur initially?
Can extra C occur finally?

If the CVX theory is correct, there are no parameters for the maximal syllable. Instead, we can add a new principle, which is the CVX limit on syllable size. In this regard, the CVX theory simplifies the principles-and-parameters model by eliminating parameters from at least one area of grammar.

12.2.4. **Optimality Theory and inviolable constraints**

Prince and Smolensky (1993) point out that, while certain phonology tendencies seem to be prevalent, truly universal rules or constraints are hard to find. As a solution, Optimality Theory is proposed, according to which linguistic constraints can be in conflict and must be ranked. In addition, different languages can rank constraints differently, and each ranking yields a grammar. As Prince and Smolensky (1993: 208) put it, "Within Optimality Theory, all constraints have exactly the same status. The theory does not recognize, for example, a difference between 'violable' and 'inviolable' constraints. All constraints are *potentially* violable."

With regard to syllable structure, Prince and Smolensky (1993) assume that it can vary from language to language. Consider three violable constraints they proposed, shown in (6).

(6) *Complex-Onset: No more than one C is allowed in the onset.
*Complex-Coda: No more than one C is allowed in the coda.
Parse: Underlying segments must be parsed into syllable structure.

If there are N constraints, there are N! (N factorial) ways to rank them (ignoring equal ranking of two or more constraints), and so there are N! possible languages. This is known as "factorial typology." With the three constraints in (6), there are six possible rankings. If we input /CCVCC/, the typology yields four possible outputs, which are shown in (7) and in evaluation tables (tableaus) in (8), where syllabified sounds are shown in brackets. In each case, the losing candidates have either violated a higher-ranked constraint, or violated a constraint more often, than the winning candidate (the output). When a constraint is violated twice, two * signs are shown in a table cell.

(7)

Ranking	Output for /CCVCC/
*Complex-Onset >> *Complex-Coda >> Parse	C[CVC]C
*Complex-Coda >> *Complex-Onset >> Parse	C[CVC]C
*Complex-Onset >> Parse >> *Complex-Coda	C[CVCC]
*Complex-Coda >> Parse >> *Complex-Onset	[CCVC]C
Parse >> *Complex-Onset >> *Complex-Coda	[CCVCC]
Parse >> *Complex-Coda >> *Complex-Onset	[CCVCC]

(8) Illustration: '√' = optimal output; '*' = constraint violation

/CCVCC/	*Complex-Onset	*Complex-Coda	Parse
[CCVCC]	*	*	
[CCVC]C	*		*
C[CVCC]		*	*
√ C[CVC]C			**

/CCVCC/	*Complex-Coda	*Complex-Onset	Parse
[CCVCC]	*	*	
[CCVC]C		*	*
C[CVCC]	*		*
√C[CVC]C			**

/CCVCC/	*Complex-Onset	Parse	*Complex-Coda
[CCVCC]	*		*
[CCVC]C	*	*	
√C[CVCC]		*	*
C[CVC]C		**	

/CCVCC/	*Complex-Coda	Parse	*Complex-Onset
[CCVCC]	*		*
√[CCVC]C		*	*
C[CVCC]	*	*	
C[CVC]C		**	

/CCVCC/	Parse	*Complex-Onset	*Complex-Coda
√[CCVCC]		*	*
[CCVC]C	*	*	
C[CVCC]	*		*
C[CVC]C	**		

/CCVCC/	Parse	*Complex-Coda	*Complex-Onset
√[CCVCC]		*	*
[CCVC]C	*		*
C[CVCC]	*	*	
C[CVC]C	**		

However, if the CVX theory is correct, there is no typology in the maximal syllable. In particular, there are no CC onsets or CC codas. This means that *Complex-Onset and *Complex-Coda are never violated and must be top-ranked universally.

Prince and Smolensky (1993) recognize the fact that certain structural properties are universal. For example, a syllable node can dominate an onset node or a mora node but not vice versa. To ensure such universal properties, Prince and Smolensky (1993: 4) propose that the "generator" component of a grammar (GEN), where all surface candidates for a given input are produced, is "a fixed part of Universal Grammar" and "contains information about the representational primitives and their universally irrevocable relations." In other words, inviolable constraints have a special status and are placed in a different part of the grammar, which is GEN. In addition, there are no conflicts among inviolable constraints. On this view, the relation between inviolable constraints on one hand and violable constraints and their ranking on the other is similar to that between "principles" and "parameters" (Chomsky 1981).

Most discussions in Optimality Theory have focused on violable constraints. The CVX theory shows that inviolable constraints are just as important, and that there are more inviolable constraints and fewer violable ones than currently conceived.

12.3. PREFERENCE CONSTRAINTS AND INVENTORY SELECTION

Even if the maximal syllable size is CVX, languages can still differ in many ways, because not every language uses all possible sounds or syllables. For example, consider coda consonants in VC rhymes in English, Cantonese, Standard Chinese (SC), and Shanghai, shown in (9), where S = fricative, TS = affricate, T = stop, and N = nasal.

(9) Possible coda consonants in VC rhymes

	[l]	TS	S	T	N
English	+	+	+	+	+
Cantonese	–	–	–	+	+
SC	–	–	–	–	+
Shanghai	–	–	–	–	–

Such differences can be accounted for either by language-specific restrictions, such as *Coda (No-coda) for Shanghai, or by ranked preferences. For example, nasals seem to be better codas than stops, which in turn seem to be better codas than fricatives and affricates.

Languages can also differ in what sounds they choose to use. For example, consider high vowels in English and Standard Chinese, shown in (10).

(10) High vowels in English: [i, ɪ, u, ʊ]
High vowels in Standard Chinese: [i, y, u]

There is no reason why English cannot use [y], or why Chinese cannot use lax vowels; it just so happens that they do not. Such differences in inventory selection, i.e. which sounds a language chooses to use out of a universal inventory of sounds, are largely arbitrary. There may sometimes be preference rankings based on articulatory ease. For example, if a language uses [y], it is likely to use [i] as well, and if a language uses lax vowels, it is likely to use corresponding tense vowels. However, American English has a mid back unrounded lax vowel [ʌ] but no corresponding mid back unrounded tense vowel [ɤ].

Inventory selection can also happen to syllable structures. A set of universal constraints will define a set of possible structures, which is the universal inventory. Each language can choose a subset of the inventory for its own use. Some possible cases are shown in (11).

(11)

	Heavy	Light
Universal inventory	CVX, VX	CV, V
Language A	CVX, VX	CV, V
Language B	CVX	CV
Language C	VX	V

If the universal inventory has four general syllable types, a language may choose to use all, as English and Standard Chinese do, or some of them, as language B and language C do. An example of language B is Chengdu Chinese, where every syllable has an onset; for example, [ai] 'love', [ən] 'favor', and [əu] 'lotus' in Standard Chinese are [ŋai], [ŋən], and [ŋəu] respectively in Chengdu. Language C is in principle possible, and Arrernte is reported to be an example (Breen and Pensalfini 1999), where no syllable has an onset. However, I shall argue below that Arrernte may still have CVX or CV syllables.

Languages can also have alternative phonological forms for the same morpheme or word, often called "free variation." Examples include the deletion of [t] in certain environments (e.g. *first rate* [fɝs(t) reit] and *facts* [fæk(t)s]), vowel alternations (e.g. *economics* [i/ɛ], *data* [ei/æ], *envelop* [ɛ/a], *Loch Ness* [k/x], etc.), and stress alternations (e.g. primary stress on the first or second syllable in *Detroit*, the second or third syllable in *Caribbean*, the first or or third syllable in *necessarily*). In the present analysis, all forms in a free variation satisfy the CVX theory.

Variations like the above are traditionally accounted for by language-specific rules. In Optimality Theory they are accounted for by violable markedness or preference constraints. Recent analyses of free variation include Boersma and Hayes (2001) and Coetzee (2006). Since such issues are extensively covered in the literature, I do not discuss them further.

12.4. THE WEIGHT-STRESS PRINCIPLE AND THE CV EFFECT

It is often observed that certain syllable types occur more commonly than others. For example, Jakobson (1958: 21) states:

> There are languages lacking syllables with initial vowels and/or syllables with final consonants, but there are no languages devoid of syllables with initial consonants or of syllables with final vowels.

According to Jakobson, CV syllables are found in every language, but other syllables are not. Similarly, Hooper (1976a: 199) considers CV syllables to be "optimal" universally and Steriade (1982: 78) considers them to be "maximally unmarked."

In the phonemic tradition, long vowels and diphthongs are often treated as single vowels. Therefore, CV syllables in Jakobson's generalization include both [CV] (such as [tə] in [tə][dei] *today*) and [CVV] (such as [bi:] *bee* and [bai] *buy*).

I have argued in Chapter 3 that the onset of a syllable is optional. If so, a language may in principle have just two kinds of syllables, VX and V, without any CV syllables. How then would the CVX theory explain the fact that CV syllables are indeed very common?

In fact, while the CV preference is well known, an explanation is lacking. Steriade (1982) accounts for the CV effect by a "CV rule," whereby a CV pair is always built into one syllable. In Zec (1988), syllables are built after mora assignment, but a separate assumption is still needed, similar to the CV rule, so that VCV is syllabified as V.CV. Hooper (1976a) accounts for the CV effect by assuming that the syllable initial must be "stronger" than the syllable final, where the strength scale is obstruents > nasals > liquids > glides > vowels. If VCV is syllabified as VC.V, the final of the first syllable C is stronger than the initial of the second syllable V, which is not the optimal result. In contrast, if VCV is syllabified as V.CV, the final of the first syllable V is weaker than the initial of the second syllable C, which is the optimal result. However, the question for Steriade is why there should be a CV rule, and the question for Hooper is why the syllable initial should be stronger than the syllable final.

In the CVX theory the CV effect can be derived from the Weight-Stress Principle (WSP, see Chapter 3), according to which a stressed rhyme should be heavy (VX) and an unstressed rhyme should be light (V). For illustration, let us focus on disyllabic words. In (12) I show how they can be syllabified under the WSP, where upper-case indicates stressed syllables, lower-case indicate unstressed syllables, and (c) indicates an unsyllabified final consonant. For simplicity, I only consider cases where each word has one stress.

(12)

String	Initial stress	Final stress
CVCV	[CVC][v], [CV:][cv]	[cv][CV:]
CVCVC	[CVC][v](c), [CV:][cv](c)	[cv][CVC], [cv][CV:](c)
VCVC	[VC][v](c), [V:][cv](c)	[v][CVC], [v][CV:](c)
VCV	[VC][v], [V:][cv]	[v][CV:]

If a language has all the strings and both initial stress and final stress, it will have many syllable types, as English does, including [CV:] and [cv].

If a language has no word-final C but lengthens stressed V, the result is shown in (13). The case is similar to Spanish and Italian, where most syllables are [CV:] and [cv].

(13)

String	Initial stress	Final stress
CVCV	[CV:][cv]	[cv][CV:]
VCV	[V:][cv]	[v][CV:]

Next, consider a language without long vowels or word-final vowels, shown in (14).

(14)

String	Initial stress	Final stress
CVCVC	[CVC][v](c)	[cv][CVC]
VCVC	[VC][v](c)	[v][CVC]

If the language has final stress, there will still be [cv] syllables in [cv][CVC] words, owing to the WSP. However, if the language has initial stress only, it is possible that it will lack [CV:] and [cv] syllables. Even in longer words, as long as they have alternating stress starting from the first syllable, [cv] syllables can be avoided. For example, CVCVCVC would be [CVC][v][CVC].

Now consider another case, where a language has no word-initial consonants, word-final vowels, or long vowels. This is shown in (15).

(15)

String	Initial stress	Final stress
VCVC	[VC][v](c)	[v][CVC]

Breen and Pensalfini (1999) argue that Arrernte is such a language, where all words start with V and end with C. However, as Breen and Pensalfini (1999: 2) note, there are two problems. First, 25% of Arrernte words do start with C. Second, a vowel "can often be heard at the end of many Arrente words" when they are pronounced in isolation. For the first problem, Breen and Pensalfini (1999) propose that such words have an underlying initial V that is not always pronounced. For the second problem, they argue that the final V is not there underlyingly.

Suppose that all Arrernte words start with V and end with C. What Breen and Pensalfini (1999) conclude is that no Arrente syllable has an onset. In other words, Arrente lacks both CV and CVC syllables. In the present analysis, their conclusion is unlikely to be correct. In particular, according to Breen and Pensalfini (1999: 3), stress falls on the second syllable in VCVC words and the syllabification is [vc][VC]. However, if the WSP is correct, an unstressed rhyme cannot be [vc], and VCVC should be [v][CVC]. Therefore, Arrente should have CVC syllables, too, even if it lacks word-initial C.

The above discussion shows that, owing to the WSP and the location of stress, a language will often have some CV or CVV syllables, as previously observed. Let us now take another look at languages that may lack CV syllables (i.e. lacking [cv], [CVV], and [CV:]). Such languages should have no diphthongs, tense–lax contrast, or vowel length contrast. To avoid CV syllables, we must assume that stressed vowels cannot be long and VCV is always [VC][v] (not [V:][cv] or [v][CV:]). However, it is unlikely that a language only has short vowels that cannot be long even with stress. Therefore, I propose the "vowel-length rule" in (16).

(16) A stressed V should be long unless it is contrastively short, lax, or is in a VC rhyme.

e.g.	CVCV → [CVC][v]	(if the first V is contrastively short or lax)
	CVCV → [CV:][cv]	(if the first V is not contrastively short or lax)
	CVCCV → [CVC][cv]	(if CC cannot be a complex sound)

Given the rule, it follows that all languages have CV syllables. In particular, if a language has contrastively short or lax vowels, it also has long or tense vowels; in the latter case, VCV will be [VV][cv] or [v][CVV], where there are CV syllables.

In summary, given the WSP, most languages have CV syllables. In addition, if we have the "vowel-length rule" in (16), then all languages have CV syllables ([CV], [CV:], or [CVV]). The present analysis offers an explanation of the CV effect, whereas other analyses do not.

12.5. WHAT IS THE SYLLABLE AND WHY IS IT SO SMALL?

If the syllable is real, we would like to know more about its nature. In particular, according to the Weight-Stress Principle, if we know the stress pattern, we know the syllable boundaries, and vice versa. Given the overlap between syllable structure and metrical structure, it is natural to ask whether they are the same thing.

Kiparsky (1979, 1981) proposes that the structure of a syllable can be seen as a metrical tree. His proposal is shown in (17), where S is a strong node and W is a weak node, following the representation of Liberman and Prince (1977).

(17) Metrical representation of the syllable (Kiparsky 1981: 250)

```
        σ
       / \
      W   S
     /\   /\
    W S  S W
   /\      /\
 …  S      S …
```

In the representation, the strongest terminal node is the nucleus of the syllable, which corresponds to the first terminal node on the right-hand branch. In the onset (left-hand branch), each terminal node is increasingly stronger when we move towards the nucleus, and in the rhyme (right-hand branch), each terminal node is increasingly weaker when we move away from the nucleus. Kiparsky argues that, if we interpret W as having less sonority and S as having more, the metrical representation yields an ideal sonority contour, whereby sonority rises towards the nucleus and falls from the nucleus (Jespersen 1904). On this view, sonority "is simply the intrasyllabic counterpart of stress."

Kiparsky (1981) also suggests that in principle the universal syllable template can expand without limit, although each language can set its own limit to the expansion. In English, the rhyme size is thought to be set at three terminal nodes (VVC or VCC).

Kiparsky's proposal contains an important insight, but there are two problems. First, the proposed metrical tree is ill formed. Consider the prominence interpretation of a hypothetical syllable CCCVCC, shown in (18). The interpretation is based on the requirement that, for two sister nodes S and W, S should be stronger than W (Liberman and Prince 1977).

(18)

```
                  x
            x     x
        x   x     x   x
[[[ W   S ] S ][ S  [ S  W ]]]
```

The structure is metrically bad because it contains four adjacent strong beats, which violates a fundamental property of rhythm, which is the alternation between strong and weak beats. Since metrical structure is a reflection of rhythm (Hayes 1995), adjacent strong beats (known as "stress clash") should be avoided. Therefore, either syllable-internal structure is not a metrical structure, or the syllable structure proposed by Kiparsky is incorrect.

The second problem with Kiparsky's proposal is that it places no limit on the size of a syllable. If the CVX theory is correct, no language has syllables larger than CVX.

We can, however, keep Kiparsky's insight that syllable structure reflects metrical structure, yet avoid the problems in his proposal. The solution is to limit the syllable size to CVX. As shown in (19), any syllable larger than CVX will result in a stress clash.

```
(19)  [C[VX]]        *[[CC][VX]]      *[C[V[XC]]]
                             x                 x
         x              x    x                 x   x
      [ W [ S W ]]  [[ W S ] [ S W ]]  [ W [ S [ S W ]]]
```

If we expand the onset once, we get CCVX, where there is clash between the second onset position and the nucleus. If we expand the coda once, we get CVXC, where there is clash between the nucleus and the first coda position. If we limit the size to CVX, there is no stress clash, the nucleus is the most sonorant, and the structure is metrically well formed.

We have seen, then, that the metrical interpretation of syllable structure explains why the syllable is so small, which is maximally CVX. Metrical structure can also explain additional properties of the syllable. For example, why is the peak in the middle of the syllable (CVX)? Why is it not at the beginning of the syllable (VXC)? Why is it not at the end of the syllable (CXV)? As far as I am aware, there is no explanation in previous literature. In the metrical representation, there is an answer. Let us assume that the simplest rhythm is the alternation between S and W beats, or . . . SWSW . . . , where there is no "clash" SS or "lapse" WW. In addition, let us assume that each syllable has only one peak. Now consider the structures in (20), where V is a strong beat and C a weak one.

(20)

String	Rhythm	Analysis
*CCV	WWS	Bad rhythm: lapse of WW
*VCC	SWW	Bad rhythm: lapse of WW
*VCV	SWS	good rhythm but bad syllable: two peaks
CVC	WSW	good rhythm and good syllable
VC	SW	good rhythm and good syllable
CV	WS	good rhythm and good syllable

The analysis predicts that CCV and VCC are not good syllables (unless CC can form a complex sound), nor is VCV a good syllable, but CVC, VC, and CV are. In addition, V is a good syllable (no clash or lapse), although it is too short to show rhythm by itself, but C is not a good syllable, since it does not have a peak.

The metrical analysis of the syllable I just offered is only a preliminary one. A full analysis would require the specification of moras and foot boundaries, which have to be left for a separate study.

12.6. WHAT IS UNIVERSAL GRAMMAR?

I have argued that there are universal constraints on syllable structure. The conclusion supports the idea that part of our linguistic knowledge is innate and that this innate knowledge is the same in all languages (Chomsky 1986). Less obvious is whether this innate knowledge is autonomous, specific to language alone. Chomsky suggests that the answer is yes: there is a language faculty in the brain, where the innate linguistic knowledge resides. However, while the constraints on syllable size may hold for all languages, they are probably not specific to language alone. Instead, they seem to be part of a mechanism for rhythm in general. If so, we should recognize additional language universals that do not originate from a language faculty.

12.7. SYLLABLE INVENTORY, "HOLES," AND "OUTLIERS"

The CVX theory offers a way to determine possible and impossible syllables. Speakers do not need to learn the set of possible syllables (the universal inventory), because it is determined by innate universal constraints. Rather, speakers learn the list of syllables a language chooses to use, which is a subset of the universal inventory.

The set of syllables a language needs is rather small. For example, English and Chinese each have about 10,000 morphemes, of which only half are frequently used (see section 9.5). If a language uses polysyllables or tolerates some homophones, the needed list of distinct syllables is no more than a few thousand. In English, for example, there are only 615 distinct CVC morphemes (where V is a short vowel)—barely a quarter of possible CVC combinations (see section 3.9).

Because a language only needs a fraction of available syllables, it has considerable freedom to choose which ones to use. The selected syllables may be subject to some markedness or preference constraints, which favor certain sound combinations over others. For example, a Cantonese syllable generally disallows two labial sounds (Yip 1988); the only coda consonants in Standard Chinese are [n] and [ŋ] (Chapter 5); and syllables in Shanghai generally do not have a coda (Chapter 6). Such patterns have led to the view that the phonological structures in a language can be precisely described by a set of rules, as proposed by Halle (1962). A similar assumption is made by Prince and Smolensky (1993).

On the other hand, because a language does not need many syllables, there are often "holes" in the distribution—syllables that could have been used but are not. For example, Standard Chinese uses [məu] 'strategy', [fəu] 'not', and [pʰəu] 'dissect', but not [pəu], which seems to be a hole. Similarly, as discussed in section 9.3, English has at least fifty-nine onsets (twenty-two C onsets, thirty CC onsets, six CCC onsets, and one for lack of an onset), and so we expect there to be at least fifty-nine different monosyllables with the rhyme [ɪl] (the most frequent VC rhyme). However, only twenty-nine of them occur in the CELEX lexicon. The thirty non-occurring ones are shown in (21).

(21) Non-occurring monosyllables with the rhyme [ɪl]
vɪl, θɪl, ðɪl, zɪl, ʃɪl, lɪl, jɪl; blɪl, dwɪl, ʃmɪl, ʃnɪl, ʃwɪl, flɪl, glɪl, gwɪl, klɪl, krɪl, plɪl, prɪl, sfɪl, slɪl, smɪl, snɪl, θwɪl; strɪl, skrɪl, skwɪl, sprɪl, splɪl, sklɪl

A few of the syllables may be used in words that CELEX failed to collect, such as *shill* and *krill*. It has also been proposed that there is a constraint against the string C+[lɪl] (Clements and Keyser 1983: 21, Davis 1988: 25), or against [lɪl] in general (Pierrehumbert 1994: 186), although there is a word *lilt*. Still, there are many others left, which seem to be holes.

Besides holes, there are often "outliers"—those that do not seem to fit the patterns of other syllables. For example, [ts] is rarely used as an onset in English, but it occurs in *Tswana* [tswa:][na] and *scherzo* [skeɚ][tso]. Similarly, [s] does not occur with a fricative in word-initial position except in *svelte*, *sforzano*, *sphagnum*, *spheroid*, *sphincter*, *sphinx*, and *sphere*. Most of these words can probably be labeled as foreign or uncommon, but it is hard to rule out *sphere* this way. In Chinese there are outliers, too. For example, Cantonese generally disallows two labial sounds in a syllable, but it has the word [pɘm] 'pump'. Similarly, in Standard Chinese a palatal onset usually does not go with a diphthong that ends in [i], but then there is a marginal word [jai] 'cliff', which many people pronounce as [ja].

More complicated still, it is not easy to decide whether an occurring form is an outlier or just fine, nor is it easy to decide whether a non-occurring form is a potential word (a hole) or simply ungrammatical. For example, is [sf] in *sphere* an outlier in English, in the sense that it is not part of English phonology? If [sf] is an outlier, we expect [sfɪt] *sfit* and [sfain] *sfine* to be ungrammatical in English (they seem to be as marginal as *Tswana* or *sforzano*). On the other hand, if [sf] is not an outlier, we expect [sfɪt] *sfit* and [sfain] *sfine* to be holes or potential words in English. Similarly, consider occurring and non-occurring monosyllables with VC rhymes in English, shown in (22).

(22) English monosyllables with VC rhymes
Productive onsets: 59
Occurring VC rhymes: 101
Possible monosyllables with VC rhymes: 5,959
Occurring monosyllables with VC rhymes: 1,069

English has fifty-nine productive onsets (excluding those ending in [j]) and 101 occurring VC rhymes. Therefore, there are 5,959 possible monosyllables with VC rhymes. However, only 1,069 occur in the CELEX lexicon. The 101 VC rhymes are shown in (23). The number of occurring onsets for a rhyme is in parentheses and the CELEX [O] can be [ɒ] or [ɑ] in American English.

(23) VC rhymes and the number of onsets they occur with in monosyllables
[ɪl] (29), [ɪp] (26), [æk] (25), [ɪt] (25), [Ot] (25), [æt] (24), [ɪk] (23), [Op] (23), [æg] (22), [æʃ] (22), [æp] (21), [æm] (20), [ʌm] (20), [ɪn] (19), [Ok] (19), [ʌg] (19), [æd] (18), [æn] (18), [ɛd] (18), [Ob] (18), [Od] (18), [ʌf] (18), [ɛl] (17), [ɛn] (17), [ɛt] (17), [ɪŋ] (17), [Og] (16), [ʌb] (16), [ʌt] (16),

[æb] (15), [ɪm] (15), [ɪg] (14), [ɪʧ] (14), [ʌk] (14), [ɛs] (13), [Os] (13), [Oʃ] (13), [ʌn] (13), [ʌʃ] (13), [æŋ] (12), [ɛk] (12), [ɪf] (12), [ɪb] (11), [ɪz] (11), [ʌd] (11), [æʧ] (10), [ɛʤ] (10), [ɪd] (10), [On] (10), [Oŋ] (10), [ʌs] (10), [ʌʤ] (9), [ʌŋ] (9), [ɛg] (8), [ɛʧ] (8), [ɪs] (8), [Of] (8), [Oʧ] (8), [ʊk] (8), [ʌl] (8), [æs] (7), [ɛf] (7), [ɛm] (7), [Ol] (7), [Om] (7), [ʌʧ] (7), [ɛp] (6), [ɪʃ] (6), [Oθ] (6), [ʊd] (6), [Oʤ] (5), [ʌp] (5), [ʌv] (5), [æl] (4), [ɛb] (4), [ɛʃ] (4), [ɪʤ] (4), [ɪθ] (4), [ɪv] (4), [ʊl] (4), [ʊʃ] (4), [ʊt] (3), [ʌz] (3), [æʤ] (2), [æf] (2), [æv] (2), [æz] (2), [ɛθ] (2), [ʊʧ] (2), [æθ] (1), [əl] (1), [əm] (1), [əs] (1), [əv] (1), [ɛv] (1), [ɛz] (1), [ɪð] (1), [Ov] (1), [Oz] (1), [ʊf] (1), [ʊs] (1)

The non-occurring monosyllables with high-frequency rhymes are probably holes, although even the most frequent rhymes hardly occur with half of the onsets. The hard question is what to do with low-frequency rhymes, such as those that only occur with one onset each. For example, [ɪð] occurs in just one monosyllable (alternative pronunciations excluded), which is *with*. Is *with* an outlier? If it is, monosyllables such as [mɪð, nɪð, tɪð, kɪð, . . .] are ungrammatical. If *with* is not an outlier, monosyllables such as [mɪð, nɪð, tɪð, kɪð, . . .] are accidental holes and potential words. Similar problems exist in Chinese. Should [pɔm] 'pump' be an outlier in Cantonese, if *map* and *Pam* are perfect syllables in English? Should [jai] 'cliff' be an outlier in Standard Chinese, if [ȶai] 'release' occurs in other Mandarin dialects, such as Chengdu? Also in Standard Chinese, the medial glide cannot be [ɥ] if the initial C is a sonorant, but there are two exceptions, [nɥe] 'mistreat' and [lɥe] 'abbreviate'. Should these syllables be outliers, or should we say that most syllables with a medial [ɥ] happen to be holes? Clearly, the issue of "holes" and "outliers" makes it difficult to describe the phonology of a language with a set of precise rules.

12.8. SUMMARY

When one looks at the world's languages, it is easy to get the impression that there is a wide range of patterns. Linguists, anthropologists, psychologists, and philosophers have often wondered about the question: How different can human languages be and what are the limits of variation? A theory of language addresses the question by outlining possible and impossible linguistic structures. My study is an effort in this direction. With regard to syllables, I have proposed that, when word-edge effects are excluded, the maximal size is CVX. The set of possible syllables is therefore far smaller than has been proposed in the literature. In addition, CVX has the metrical structure WSW, or simply SW if we ignore the onset. This may be the reason why the syllable is as small as it is. My study shows that in at least some parts of language there may be no structural variation, despite apparent diversity at first sight.

This page intentionally left blank

REFERENCES

ABERCROMBIE, DAVID (1967). *Elements of General Phonetics*. Chicago: Aldine.

ALCANTARA, JONATHAN BRETT (1998). The architecture of the English lexicon. Doctoral dissertation, Cornell University.

AO, BENJAMIN X. P. (1992). The non-uniqueness condition and the segmentation of the Chinese syllable. *Working Papers in Linguistics* 42: 1–25. Ohio State University.

ARCHANGELI, DIANA (1988). Aspects of underspecification theory. *Phonology* 5: 183–207.

BAAYEN, R. HARALD, RICHARD PIEPENBROCK, and L. GULIKERS (1993). The CELEX lexical database. CD-ROM. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

BAGEMIHL, BRUCE (1991). Syllable structure in Bella Coola. *Linguistic Inquiry* 22.4: 589–646.

BAILEY, C.-J. N. (1978). Gradience in English syllabification and a revised concept of unmarked syllabification. Indiana University Linguistics Club, Bloomington.

BAO, ZHIMING (1990). Fanqie languages and reduplication. *Linguistic Inquiry* 21.3: 317–50.

BATES, DAWN, and BARRY CARLSON (1992). Simple syllables in Spokane Salish. *Linguistic Inquiry* 23.4: 653–9.

BAXTER, WILLIAM (1992). *A Handbook of Old Chinese Phonology*. Berlin: Mouton de Gruyter.

BETHIN, CHRISTINA Y. (1992). *Polish Syllables: The Role of Prosody in Phonology and Morphology*. Columbus, OH: Slavica.

BLEVINS, JULIETTE (1995). The syllable in phonological theory. In John Goldsmith (ed.), *The Handbook of Phonological Theory*, 206–44. Oxford: Blackwell.

—— (2003). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

BOERSMA, PAUL, and BRUCE HAYES (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32.1: 45–86.

BOROWSKY, TONI (1986). Topics in the lexical phonology of English. Doctoral dissertation, University of Massachusetts, Amherst.

—— (1989). Structure preservation and the syllable coda in English. *Natural Language and Linguistic Theory* 7: 145–66.

BREEN, GAVAN, and ROB PENSALFINI (1999). Arrernte: a language with no syllable onsets. *Linguistic Inquiry* 30.1: 1–25.

BROMBERGER, SYLVAIN, and MORRIS HALLE (1989). Why phonology is different. *Linguistic Inquiry* 20.1: 51–70.

BROWMAN, CATHERINE P., and LOUIS M. GOLDSTEIN (1989). Articulatory gestures as phonological units. *Phonology* 6: 201–51.

BURZIO, LUIGI (1994). *Principles of English Stress*. Cambridge: Cambridge University Press.

BURZIO, LUIGI (1996). Surface constraints versus underlying representation. In Jacques Durand and Bernard Laks (eds.), *Current Trends in Phonology: Models and Methods*, vol. 1, 123–41. Salford, UK: European Studies Research Institute/University of Salford.

BUTSKHRIKIDZE, MARIKA (2002). The consonant phonotactics of Georgian. Doctoral dissertation, Universiteit Leiden. [Distributed by Netherlands Graduate School of Linguistics, Utrecht.]

BYBEE, JOAN (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

CHAO, YUEN-REN (1933). Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology, Academia Sinica* 4.2: 121–34.

—— (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 4.4: 363–97. Repr. in Martin Joos (ed.), *Readings in Linguistics I* (Chicago: University of Chicago Press, 1957), 38–54.

—— (1948). The voiced velar fricative as an initial in Mandarin. *Le Maître phonétique* 89: 2–3.

—— (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

CHEN, MATTHEW Y. (2000). *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press.

CHEN, ZHONGMIN (2003). *Studies on Dialects in the Shanghai Area: Their Phonological Systems and Historical Developments*. Munich: Lincom Europa.

CHENG, CHIN-CHUAN (1973). *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.

CHENG, ROBERT L. (1966). Mandarin phonological structure. *Journal of Linguistics* 2.2: 135–262.

CHEUNG, KWAN-HIN (1986). The phonology of present-day Cantonese. Doctoral dissertation, University of London.

CHINESE ACADEMY OF SOCIAL SCIENCES INSTITUTE OF LINGUISTICS (1978). *Xiandai Hanyu Cidian* [A dictionary of modern Chinese]. Beijing: Shangwu Yinshuguan.

CHINESE NATIONAL COMMITTEE ON LANGUAGE AFFAIRS BUREAU OF CHINESE CHARACTERS [GUOJIA YUYAN WENZI GONGZUO WEIYUANHUI HANZI CHU] (1988). *Xiandai Hanyu changyong zi biao* [Frequent characters in modern Chinese]. Beijing: Yuwen Chubanshe.

CHINESE NATIONAL COMMITTEE ON LANGUAGE AFFAIRS [GUOJIA YUYAN WENZI GONGZUO WEIYUANHUI] (2004). *Zhongguo yuyan wenzi shiyong qingkuang diaocha* [Survey of spoken and written Chinese]. http://www.china-language.gov.cn/doc/zhongguodiaocha2004-12/shujv.doc.

CHOMSKY, NOAM (1957). *Syntactic Structures*. The Hague: Mouton.

—— (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.

—— (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

—— (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

—— (1986). *Knowledge of Language: Its Nature, Origin, and Use.* New York: Prager.

—— and MORRIS HALLE (1968). *The Sound Pattern of English*. New York: Harper & Row.

CLARK, MICHAEL J., and JAMES M. HILLENBRAND (2003). Quality of American English front vowels before /r/. *Journal of the International Phonetic Association* 33.1: 1–16.

CLEMENTS, G. N. (1985). The geometry of phonological features. *Phonology Yearbook* 2: 225–252.

——(1999). Affricates as noncontoured stops. In Osamu Fujimura, Brian D. Joseph, and Bohumil Palek (eds.), *Proceedings of LP '98: Item Order in Language and Speech*, Vol. 1, 271–99. Prague: Karolinum Press (Charles University in Prague).

——and SAMUEL JAY KEYSER (1983). *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT Press.

COETZEE, ANDRIES W. (2006). Variation as accessing 'non-optimal' candidates. *Phonology* 23: 337–85.

COLEMAN, JOHN (1996). Declarative syllabification in Tashlhit Berber. In Jacques Durand and Bernard Laks (eds.),*Current Trends in Phonology: Models and Methods*, Vol. 1, 175–216. Salford: European Studies Research Institute/University of Salford.

——(2001). The phonetics and phonology of Tashlhiyt Berber syllabic consonants. *Transactions of the Philological Society* 99.1: 29–64.

——and JANET PIERREHUMBERT (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*, Association for Computational Linguistics, 49–56. Somerset, NJ: Association for Computational Linguistics.

DA, JUN (2004). Chinese text computing. Murfreesboro, TN: Department of Foreign Languages and Literatures, Middle Tennessee State University. http://lingua.mtsu.edu/chinese-computing/

DAVIS, STUART (1988). *Topics in Syllable Geometry*. New York: Garland.

——(1999). The parallel distribution of aspirated stops and /h/ in American English. In Karen Baertsch and Daniel A. Dinnsen (eds.), *Optimal Green Ideas in Phonology*, 1–10. Bloomington: Indiana University Linguistics Club.

——(2001). The pair of dactyls in *mediterranean*: on the analysis of aspirated weak syllables in English. Paper presented at MCWOP 7, University of Iowa.

DELL, FRANÇOIS, and MOHAMED ELMEDLAOUI (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7: 105–30.

DONG, SHAOWEN (1958). *Yuyin changtan* [Introduction to phonetics]. Beijing: Wenhua Jiaoyu Chubanshe.

DUANMU, SAN (1990). A formal study of syllable, tone, stress and domain in Chinese languages. Doctoral dissertation, Massachusetts Institute of Technology. Distributed by MIT Working Papers in Linguistics.

——(1993). Rime length, stress, and association domains. *Journal of East Asian Linguistics* 2.1: 1–44.

——(1994). Against contour tone units. *Linguistic Inquiry* 25.4: 555–608.

——(1999). Metrical structure and tone: evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics* 8.1: 1–38.

——(2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

DUANMU, SAN (2003). The syllable phonology of Mandarin and Shanghai. In *Proceedings of the Fifteenth North American Conference on Chinese Linguistics*, 86–102. Los Angeles: University of Southern California Press.

—— GREGORY H. WAKEFIELD, YIPING HSU, G. CRISTINA, and SHANPING QIU (1998). Taiwanese Putonghua speech and transcript corpus. CD-ROM. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

EDKINS, JOSEPH (1853). *A Grammar of Colloquial Chinese, as exhibited in the Shanghai dialect*. Shanghai: Presbyterian Mission Press. (2nd edn. 1868.)

EVERETT, DAN, and KAREN EVERETT (1984). On the relevance of syllable onsets to stress placement. *Linguistic Inquiry* 15: 705–11.

FENG, AIZHEN (1998). *Fuzhou Fangyan Cidian* [Dictionary of Fuzhou dialect]. (Xiandai Hanyu Fangyan Da Cidian [Dictionaries of modern Chinese dialects], editor-in-chief Rong Li.) Nanjing: Jiangsu Jiaoyu Chubanshe.

FÉRY, CAROLINE, and RUBEN VAN DE VIJVER (eds.) (2003). *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press.

FIDELHOLTZ, JAMES L. (1975). Word frequency and vowel reduction in English. In Robin E. Grossman, L. James San, and Timothy J. Vance (eds.), *Papers from the Eleventh Regional Meeting, Chicago Linguistic Society*, 200–13. Chicago: Chicago Linguistic Society.

FIRTH, J. R., and B. B. ROGERS (1937). The structure of the Chinese monosyllable in a Hunanese dialect (Changsha). *Bulletin of the School of Oriental Studies* 8.4: 1055–74.

FRISCH, STEFAN A., N. R. LARGE, and DAVID B. PISONI (2000). Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42.4: 481–96.

FU, JINGQI (1990). Labial–labial cooccurrence restrictions and syllabic structure. *Proceedings of the 1st Meeting of the Formal Linguistic Society of Mid America*, 129–44.

FUDGE, ERIK C. (1969). Syllables. *Journal of Linguistics* 5: 253–86.

FUJIMURA, OSAMA (1979). An analysis of English syllables as cores and affixes. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 471–6.

GAO, MINGKAI, and ANSHI SHI (1963). *Yuyanxue gailun* [Introduction to linguistics]. Beijing: Zhonghua Shuju.

GIEGERICH, HEINZ (1985). *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press.

—— (1989). *Syllable Structure and Lexical Derivation in German*. Bloomington: Indiana University Linguistics Club.

—— (1992). *English Phonology*. Cambridge: Cambridge University Press.

GIMSON, A. C. (1970). *An Introduction to the Pronunciation of English.* 2nd edn. London: Arnold.

GOEDEMANS, ROBERTUS WILHELMUS NICOLAAS (1998). *Weightless Segments: A Phonetic and Phonological Study Concerning the Metrical Irrelevance of Syllable Onsets*. The Hague: Holland Academic Graphics.

GOH, YENG-SENG (2000). Beijinghua shi danyinjie yuyan de zhiyi [Is Beijing Mandarin a monosyllabic language?]. *Dangdai Yuyanxue* [Contemporary linguistics] 2.4: 231–47.

GOLDSMITH, JOHN (1976). Autosegmental phonology. Doctoral dissertation, Massachusetts Institute of Technology. Repr. by Indiana University Linguistics Club.

—— (1981). English as a tone language. In Didier L. Goyvaerts (ed.), *Phonology in the 1980's*, 287–308. Ghent: E. Story-Scientia.

—— (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

GORDON, MATTHEW (2005). A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23: 595–653.

GOUSKOVA, MARIA (2004). Relational hierarchies in Optimality Theory: the case of syllable contact. *Phonology* 21.2: 201–50.

GUO, SHAOYU (1938). Zhongguo yuci zhi tanxing zuoyong [The elastic function of Chinese word length]. *Yen Ching Hsueh Pao* 24. Repr. in *Shaoyu Guo* 1963: 1–40.

GUSSMANN, EDMUND (2002). *Phonology: Analysis and Theory*. Cambridge: Cambridge University Press.

HALL, TRACY ALAN (1992). Syllable structure and syllable-related processes in German. Tübingen: Niemeyer.

—— (2001). The distribution of superheavy syllables in Modern English. *Folia Linguistica* 35.3–4: 399–442.

—— (2002a). Against extrasyllabic consonants in German and English. *Phonology* 19.1: 33–75.

—— (2002b). The distribution of superheavy syllables in Standard German. *Linguistic Review* 19.4: 377–420.

HALLE, MORRIS (1962). Phonology in generative grammar. *Word* 18: 54–72.

—— (1972). Theoretical issues in phonology in the 1970s. In André Rignault and René Charbonneau (eds.), *Proceedings of the Seventh International Congress of Phonetic Sciences*, 179–205. The Hague: Mouton.

—— (1992). Phonological features. In William Bright (ed.), *Oxford International Encyclopedia of Linguistics*, 207–12. Oxford: Oxford University Press.

—— (1995). Feature geometry and feature spreading. *Linguistic Inquiry* 26: 1–46.

—— (1997). On stress and accent in Indo-European. *Language* 73.2: 275–313.

—— (1998). The stress of English words: 1968–1998. *Linguistic Inquiry* 29.4: 539–68.

—— (2005). Palatalization/velar softening: what it is and what it tells us about the nature of language. *Linguistic Inquiry* 36.1: 23–41.

—— and G. N. CLEMENTS (1983). *Problem Book in Phonology*. Cambridge, MA: MIT Press.

—— and KENNETH N. STEVENS (1969). On the feature 'Advanced Tongue Root'. *MIT Research Laboratory of Electronics Quarterly Progress Report* 99: 209–15.

—— —— (1971). A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report* 101: 198–213.

—— and JEAN-ROGER VERGNAUD (1987). *An Essay on Stress*. Cambridge, MA: MIT Press.

HAMMOND, MICHAEL (1999). *The Phonology of English: A Prosodic Optimality Theoretic Approach*. Oxford: Oxford University Press.

HANKS, PATRICK (ed.) (1979). *Collins Dictionary of the English Language*. London: Collins.

HARRINGTON, JONATHAN, FELICITY COX, and ZOE EVANS (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics* 17.2: 155–84.

HARRIS, JOHN (1994). *English Sound Structure*. Oxford: Blackwell.

HAYES, BRUCE (1982). Extrametricality and English stress. *Linguistic Inquiry* 13: 227–76.

—— (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

HOARD, JAMES E. (1971). Aspiration, tenseness, and syllabification in English. *Language* 47.1: 133–40.

HOCKETT, CHARLES F. (1947). Peiping phonology. *Journal of the American Oriental Society* 67.4: 253–67.

HOOPER, JOAN BYBEE (1976a). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

—— (1976b). Word frequency in lexical diffusion and the source of morphophonological change. In William M. Christie, Jr. (ed.), *Current Progress in Historical Linguistics*, 96–105. Amsterdam: North-Holland.

HOWIE, JOHN (1976). *An Acoustic Study of Mandarin Tones and Vowels*. Cambridge: Cambridge University Press.

HSIEH, FENG-FAN (1999). Theoretical aspects of Zhuokeji rGyalrong phonology. MA thesis, National Tsing Hua University.

HSUEH, FENGSHENG (1986). *Guoyu Yinxi Jiexi* [An anatomy of the Pekingese sound system]. Taipei: Taiwan Xuesheng Shuju.

HUANG, PARKER PO-FEI (1970). *Cantonese Dictionary: Cantonese–English, English–Cantonese*. New Haven, CT: Yale University Press.

HYMAN, LARRY (1985). *A Theory of Phonological Weight*. Dordrecht: Foris.

INKELAS, SHARON (1994). The consequences of optimization for underspecification. MS, University of California, Berkeley.

ITO, JUNKO (1986). Syllabic theory in prosodic phonology. Doctoral dissertation, University of Massachusetts, Amherst.

IVERSON, GREGORY K., and JOSEPH C. SALMONS (1995). Aspiration and laryngeal representation in Germanic. *Phonology* 12.3: 369–96.

JAKOBSON, ROMAN (1958). Typological studies and their contribution to historical comparative linguistics. In Eva Sivertsen (gen. ed.), *Proceedings of the Eighth International Congress of Linguists*, 17–25. Oslo: Oslo University Press.

—— GUNNAR FANT, and MORRIS HALLE (1952). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. Cambridge, MA: MIT Press.

JENSEN, JOHN (2000). Against ambisyllabicity. *Phonology* 17: 187–235.

JESPERSEN, OTTO (1904). *Lehrbuch der Phonetik*. Leipzig: Teubner.

JESSEN, MICHAEL, and CATHERINE RINGEN (2002). Laryngeal features in German. *Phonology* 19.2: 189–218.

JI, XIANLIN (ed.) (1988). *Zhongguo Dabaikequanshu Yuyan Wenzi* [Chinese encyclopaedia: language and orthography]. Beijing: Zhongguo Dabaikequanshu Chubanshe.

JONES, DANIEL (1950). *The Pronunciation of English*. 3rd edn. Cambridge: Cambridge University Press.

KAGER, RENÉ (1989). *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris.

KAHN, DANIEL (1976). Syllable-based generalizations in English phonology. Doctoral dissertation, Massachusetts Institute of Technology.

KARLGREN, BERNHARD (1949). *The Chinese Language: An Essay on its Nature and History*. New York: Ronald Press.

KAYE, JONATHAN, and JEAN LOWENSTAMM (1984). De la syllabicité. In François Dell, Daniel Hirst, and Jean-Roger Vergnaud (eds.), *Forme sonore du langage*, 123–59. Paris: Hermann.

KEATING, PATRICIA (1988). Underspecification in phonetics. *Phonology* 5: 275–92.

KENSTOWICZ, MICHAEL (1994). *Phonology in Generative Grammar*. Oxford: Blackwell.

KENYON, JOHN SAMUEL, and THOMAS ALBERT KNOTT (1944). *A Pronouncing Dictionary of American English*. Springfield, MA: Merriam.

KESSLER, BRETT, and REBECCA TREIMAN (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37: 295–311.

KEY, HAROLD (1961). Phonotactics of Cayuvava. *International Journal of American Linguistics* 27.2: 143–50.

KEYSER, SAMUEL JAY, and KENNETH N. STEVENS (1994). Feature geometry and the vocal tract. *Phonology* 11.2: 207–36.

KIM, MI-RYOUNG, and SAN DUANMU (2004). Tense and lax stops in Korean. *Journal of East Asian Linguistics* 13: 59–104.

KIPARSKY, PAUL (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry* 10.3: 421–41.

——(1981). Remarks on the metrical structure of the syllable. In Wolfgang U. Dressler, Oskar E. Pfeiffer, and John R. Rennison (eds.), *Phonologica 1980*, 245–56. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

KOHLER, KLAUS (1999). German. In International Phonetic Association, *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, 86–9. Cambridge: Cambridge University Press.

KRAKOW, RENA A. (1989). The articulatory organization of syllables: a kinematic analysis of labial and velar gestures. Doctoral dissertation, Yale University.

KREIDLER, CHARLES W. (2004). *The Pronunciation of English: A Course Book*. Malden, MA: Blackwell.

KUMAR, AMAN (2005). Aspects of Hindi syllable structure. Doctoral dissertation, University of Michigan, Ann Arbor.

LADEFOGED, PETER (1972). Phonetic prerequisites for a distinctive feature theory. In Albert Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, 273–86. The Hague: Mouton.

——(1980). What are linguistic sounds made of? *Language* 56.3: 485–502.

——(1982). *A Course in Phonetics*. 2nd edn. San Diego, CA: Harcourt Brace Jovanovich.

——(1992). The many interfaces between phonetics and phonology. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer, and John R. Rennison (eds.), *Phonologica 1988*, 165–79. Cambridge: Cambridge University Press.

LADEFOGED, PETER (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages*. Malden, MA: Blackwell.

——(2006). *A Course in Phonetics*. 5th edn. Boston, MA: Thomson & Wadsworth.

——and MORRIS HALLE (1988). Some major features of the International Phonetic Alphabet. *Language* 64.3: 577–82.

LAWRENCE, WAYNE P. (2000). /str/ → /ʃtr/: assimilation at a distance? *American Speech* 75.1: 82–7.

LEBEN, WILLIAM (1971). Suprasegmental and segmental representation of tone. *Studies in African Linguistics*, supp. 2: 183–200.

LEE, WAI-SUM, and ERIC ZEE (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33.1: 109–12.

LEVIN, JULIETTE (1985). A metrical theory of syllabicity. Doctoral dissertation, Massachusetts Institute of Technology.

LI, CHARLES N., and SANDRA A. THOMPSON (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

LI, FANG-KUI (1966). The zero initial and the zero syllabic. *Language* 42: 300–302.

LI, RONG (1983). Guanyu fangyan yanjiu de ji dian yijian [Some comments on dialectal studies]. *Fangyan* 1: 1–15.

LIBERMAN, MARK (1975). The intonational system of English. Doctoral dissertation, Massachusetts Institute of Technology.

——and ALAN PRINCE (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8.2: 249–336.

LIN, MAOCAN, and JINGZHU YAN (1988). The characteristic features of the final reduction in the neutral-tone syllable of Beijing Mandarin, *Phonetic Laboratory Annual Report of Phonetic Research*, 37–51. Beijing: Phonetic Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

LIN, XIANGRONG (1993). *Jiarongyu yanjiu* [Studies on the Jiarong (rGyalrong) language]. Chengdu: Sichuan Minzu Chubanshe.

LIN, YEN-HWEI (1989). Autosegmental treatment of segmental processes in Chinese phonology. Doctoral dissertation, University of Texas, Austin.

LOMBARDI, LINDA (1990). The nonlinear organization of the affricate. *Natural Language and Linguistic Theory* 8: 374–425.

LOWENSTAMM, JEAN (1996). CV as the only syllable type. In Jacques Durand and Bernard Laks (eds.), *Current Trends in Phonology, Models and Methods*, 419–43. Salford, UK: European Studies Research Institute/University of Salford.

LÜ, SHUXIANG (1963). *Xiandai Hanyu dan shuang yinjie wenti chu tan* [A preliminary study of the problem of monosyllabism and disyllabism in modern Chinese]. *Zhongguo Yuwen* 1: 11–23.

LUO, CHANG-PEI, and JUN WANG (1957). *Putong yuyinxue gangyao* [Outline of general phonetics]. Beijing: Kexue Chubanshe.

MADDIESON, IAN (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

MALÉCOT, ANDRÉ (1960). Vowel nasality as a distinctive feature in American English. *Language* 36.2: 222–9.

MARTIN, SAMUEL E. (1957). Problems of hierarchy and indeterminacy in Mandarin phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 29: 209–29.

McCarthy, John (1979a). Formal problems in Semitic phonology and morphology. Doctoral dissertation, Massachusetts Institute of Technology. Published by Garland Press, New York, 1985.

—— (1979b). On stress and syllabification. *Linguistic Inquiry* 10.3: 443–65.

—— (1988). Feature geometry and dependency: a review. *Phonetica* 45: 84–108.

Mei, Tsulin (2006). Shanggu Hanyu Dulongyu tongyuan sishi ci [Forty cognate words in Old Chinese and Trung]. In Henry Y. Zhang, Lillian M. Huang, and Dah-an Ho (eds.), *Bai chuan hui hai: Li Rengui xiansheng qi zhi shouqing lunwenji* [Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday], 595–604. Taipei: Institute of Linguistics, Academia Sinica.

Merriam-Webster (2004). *Merriam-Webster Online Dictionary*. http://www.m-w.com/home.htm.

Moulton, William G. (1956). Syllabic nuclei and final consonant clusters in German. In Morris Halle, Horace G. Lunt, Hugh McLean, and Cornelis H. van Schooneveld (eds.), *For Roman Jakobson: Essays on the Occasion of his Sixtieth Birthday*, 372–81. The Hague: Mouton.

Murray, Robert W., and Theo Vennemann (1983). Sound change and syllable structure in Germanic phonology. *Language* 59.3: 514–28.

Myers, James, and Jane Tsay (2005). The processing of phonological acceptability judgments. In *Proceedings of Symposium on 90–92 National Science Council Projects*, 26–45. Taipei, Taiwan: National Science Council.

Newman, Paul (1995). Hausa tonology: complexities in an 'easy' tone language. In John Goldsmith (ed.), *The Handbook of Phonological Theory*, 762–81. Oxford: Blackwell.

Odden, David (2005). *Introducing Phonology*. Cambridge: Cambridge University Press.

Padgett, Jaye (1995). *Stricture in Feature Geometry*. Stanford, CA: Center for the Study of Language and Information, Stanford University.

Pater, Joe (2000). Non-uniformity in English secondary stress: the role of ranked and lexically-specific constraints. *Phonology* 17: 237–74.

—— (2004). Austronesian nasal substitution and other NC̥ effects. In John McCarthy (ed.), *Optimality Theory in Phonology: A Reader*, 271–89. Oxford: Blackwell.

Peterson, Gordon E., and Ilse Lehiste (1960). Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America* 32.6: 693–703.

Pierrehumbert, Janet (1980). The phonetics and phonology of English intonation. Doctoral dissertation, Massachusetts Institute of Technology.

—— (1994). Syllable structure and word structure: a study of triconsonantal clusters in English. In Patricia A. Keating (ed.), *Phonological Structure and Phonetic Form*, 168–88. Cambridge: Cambridge University Press.

—— (2001). Exemplar dynamics: word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.), *Frequency Effects and the Emergence of Linguistic Structure*, 137–57. Amsterdam: Benjamins.

Pike, Kenneth (1947). On the phonemic status of English diphthongs. *Language* 23: 151–9.

Port, Robert F., and Adam P. Leary (2005). Against formal phonology. *Language* 81.4: 927–64.

Prince, Alan (1990). Quantitative consequences of rhythmic organization. In *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, vol. 2: *The Parasession on the Syllable in Phonetics and Phonology*, 355–398. Chicago: Chicago Linguistic Society.

—— and Paul Smolensky (1993). Optimality Theory: constraint interaction in generative grammar. MS, Rutgers University and University of Colorado.

Prinz, Michael, and Richard Wiese (1991). Die Affrikaten des Deutschen und ihre Verschriftung. *Linguistische Berichte* 133: 165–89.

Prokosch, Eduard (1939). *A Comparative Germanic Grammar*. Philadelphia: Linguistic Society of America.

Pulgram, Ernst (1970). *Syllable, Word, Nexus, Cursus.* The Hague: Mouton.

Pulleyblank, Douglas (1986). *Tone in Lexical Phonology*. Dordrecht: Reidel.

Pulleyblank, Edwin G. (1984). Vowelless Chinese? An application of the three-tiered theory of syllable structure. In *Proceedings of the Sixteenth International Conference on Sino-Tibetan Languages and Linguistics*, vol. 2, 568–619. Seattle: University of Washington.

Qian, Nairong (1997). *Shanghaihua yufa* [A grammar of Shanghai]. Shanghai: Shanghai Renmin Chubanshe.

Qu, Aitang (1990). Jiarong de fangyan: Fangyan huafen he yuyan shibie [rGyalrong dialects: Issues in dialect subclassification and language recognition]. *Mimzu Yuwen* 4: 1–8, 5: 37–44.

Ramsey, S. Robert (1987). *The Languages of China*. Princeton, NJ: Princeton University Press.

Redford, Melissa Annette, and Patrick Randall (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33: 27–46.

Rice, Keren (1992). On deriving sonority: a structural account of sonority relationships. *Phonology* 9.1: 61–99.

Sagey, Elizabeth (1986). The representation of features and relations in nonlinear phonology. Doctoral dissertation, Massachusetts Institute of Technology.

Sapir, Edward (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Scheer, Tobias (2004). *A Lateral Theory of Phonology: What Is CVCV, and Why Should It Be?* Berlin: Mouton de Gruyter.

Selkirk, Elisabeth (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry* 11: 563–605.

—— (1982). The syllable. In Harry van der Hulst and Norval Smith (eds.), *The Structure of Phonological Representations*, part II, 337–83. Dordrecht: Foris.

Steriade, Donca (1982). Greek prosodies and the nature of syllabification. Doctoral dissertation, Massachusetts Institute of Technology.

—— (1987). Redundant values. In *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society*, Part 2: *Parasession on Autosegmental and Metrical Phonology*, 339–62. Chicago: Chicago Linguistic Society.

—— (1989). Affricates are stops. Paper presented at Conference on Features and Underspecification Theories, October 7–9, Massachusetts Institute of Technology.

—— (1999). Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph, and Bohumil Palek (eds.), *Proceedings of LP*

'98: Item Order in Language and Speech, Vol. 1, 205–45. Prague: Karolinum Press (Charles University in Prague).

SUN, JACKSON T.-S. (1998). Nominal morphology in Caodeng rGyalrong. *Bulletin of the Institute of History and Philology, Academia Sinica* 69.1: 103–49.

—— (2000). Parallelism in the verb morphology of Sidaba rGyalrong and Lavrung in rGyalrongic. *Language and Linguistics* 1.1: 161–90.

TREIMAN, REBECCA, and CATALINA DANIS (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language* 27.1: 87–104.

TREIMAN, REBECCA, JENNIFER GROSS, and ANNEMARIE CWIKIEL-GLAVIN (1992). The syllabification of /s/ clusters in English. *Journal of Phonetics* 20.3: 383–402.

TURK, ALICE (1994). Articulatory phonetic clues to syllable affiliation: gestural characteristics of bilabial stops. In Patricia A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, 107–35. Cambridge: Cambridge University Press.

URBOM, RUTH (1999). *Longman Handy Learner's Dictionary*. New edn. Harlow, UK: Pearson Education.

VAUX, BERT (2003). Syllabification in Armenian, Universal Grammar, and the lexicon. *Linguistic Inquiry* 34.1: 91–125.

—— (2004). The appendix. Paper presented at the Symposium on Phonological Theory: Representations and Architecture, CUNY, Feb. 20.

—— and BRIDGET SAMUELS (2005). Laryngeal markedness and aspiration. *Phonology* 22.2: 395–436.

VENNEMANN, THEO (1974). Words and syllables in natural generative phonology. In Anthony Bruck, Robert A. Fox, and Michael W. La Galy (eds.), *Papers from the Parasession on Natural Phonology*, 346–74. Chicago: Chicago Linguistic Society.

—— (1988). *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: Mouton de Gruyter.

WANG, JENNY ZHIJIE (1993). The geometry of segmental features in Beijing Mandarin. Doctoral dissertation, University of Delaware, Newark.

WANG, LI (1944). *Zhongguo yufa lilun* [Chinese grammatical theory]. Shanghai: Shangwu Yinshuguan.

—— (1957). *Hanyu Yinyunxue* [Chinese phonology]. Beijing: Zhonghua Shuju.

WEIDE, ROBERT L. (1998). *The Carnegie Mellon Pronouncing Dictionary* [cmudict.0.6]. Carnegie Mellon University: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

WELLS, J. C. (1990). Syllabification and allophony. In Susan Ramsaran (ed.), *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A. C. Gimson*, 76–86. London: Routledge.

WIESE, RICHARD (1996). *The Phonology of German*. Oxford: Clarendon Press.

—— (1997). Underspecification and the description of Chinese vowels. In Wang Jialing and Norval Smith (eds.), *Studies in Chinese Phonology*, 219–49. Berlin: Mouton de Gruyter.

WOO, NANCY (1969). Prosody and phonology. Doctoral dissertation, Massachusetts Institute of Technology.

XU, BAOHUA, ZHENZHU TANG, RUJIE YOU, NAIRONG QIAN, RUJIE SHI, and YAMING SHEN (1988). *Shanghai Shiqü fangyan zhi* [Urban Shanghai dialects]. Shanghai: Shanghai Jiaoyu Chubanshe.

Xu, Baohua, and Huan Tao (1997). *Shanghai Fangyan Cidian* [Dictionary of Shanghai dialect]. (*Xiandai Hanyu Fangyan Da Cidian* [Dictionaries of modern Chinese dialects], editor-in-chief Rong Li.) Nanjing: Jiangsu Jiaoyu Chubanshe.

Xu, Yi (1986). Putonghua yinlian de shengxue yuyin texing [Acoustic properties of syllable junctures in Standard Chinese]. *Zhongguo Yuwen* 1986.5(194): 353–60.

Yan, Muchu (2004). A study of prefixes in the Suomo dialect of the Jiarong language. *Journal of the Central University of National Minorities* (Philosophy and Social Science edn.) 31.6: 120–24.

Yip, Moira (1980). Tonal phonology of Chinese. Doctoral dissertation, Massachusetts Institute of Technology.

—— (1988). The obligatory contour principle and phonological rules: a loss of identity. *Linguistic Inquiry* 19.1: 65–100.

Yue-Hashimoto, Anne O. (1987). Tone sandhi across Chinese dialects. In Chinese Language Society of Hong Kong (ed.), Wang Li memorial volumes: Chinese volume, 445–74. Hong Kong: Joint Publishing Co.

You, Rujie, Nairong Qian, and Zhengxia Gao (1980). Lun Putonghua de yinwei xitong [On the phonemic system of Standard Chinese]. *Zhongguo Yuwen* 1980.5 (158): 328–334.

Zee, Eric (2003). The phonetic characteristics of the sounds in Standard Chinese (Beijing). Paper presented at the 15th North American Conference on Chinese Linguistics, July 11–13, Michigan State University, East Lansing.

Zhu, Xiaonong (1995). Shanghai tonetics. Doctoral dissertation, Australian National University.

AUTHOR INDEX

LANGUAGE INDEX

SUBJECT INDEX