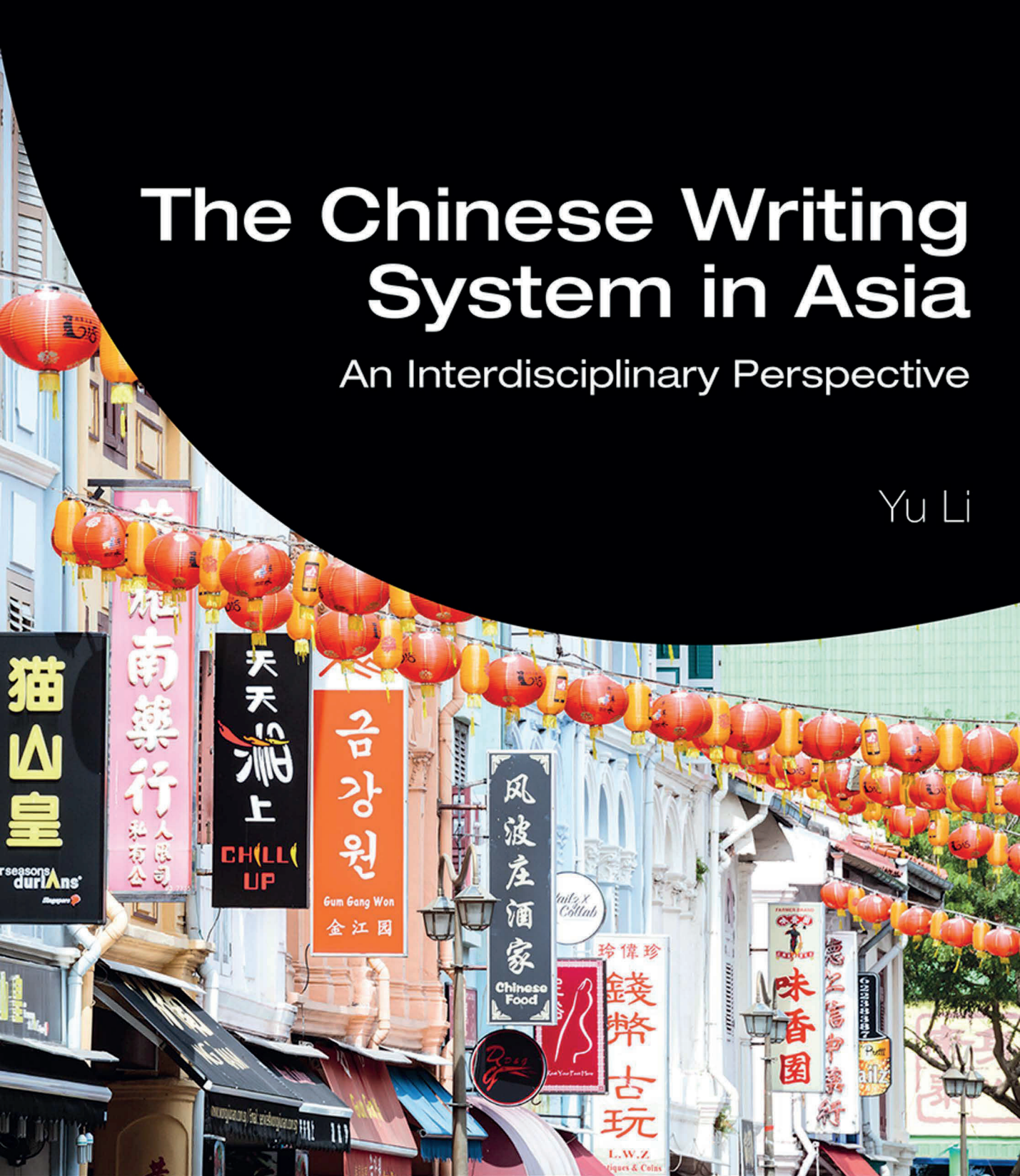


The Chinese Writing System in Asia

An Interdisciplinary Perspective

Yu Li



The Chinese Writing System in Asia

The Chinese Writing System in Asia: An Interdisciplinary Perspective integrates a diverse range of disciplinary approaches in examining how the Chinese script represents and actively shapes personal and social identities in and beyond Asia. It is an ideal read for students and scholars interested in a broad and culturally rich introduction to research on the Chinese writing system. It can also serve as the main text of an undergraduate course on the subject.

Key features of this volume include:

- Insights from studies of the Chinese writing system in linguistics, script reform and technology, gender, identity, literature, and the visual arts;
- Examples embedded in inquiries of the cultural history and contemporary society of Asia;
- Rigorous yet accessible discussions of complex concepts and phenomena that assume no prior knowledge of Asian languages or linguistics;
- Supplementary multimedia materials and resources, including instructional support, available online.

Yu Li is Assistant Professor of Chinese in the Department of Modern Languages and Literatures at Loyola Marymount University, USA. Her research interests include Chinese linguistics, Chinese language pedagogy, East Asian calligraphy, and the Chinese linguistic landscape in the diaspora.

“This is a comprehensive introduction to the Chinese writing system from a wide range of disciplinary perspectives. The author writes beautifully on the linguistic, socio-cultural, anthropological, literary and artistic aspects of Chinese characters.”

— **Hongyuan Dong**, *Assistant Professor of Chinese Language and Linguistics,
The George Washington University in Washington DC, USA*

“This book is a welcome new introduction to the Chinese writing system. I see the strengths of the book in three areas: its wide scope in the topics it covers, its thoroughness in the treatment of the topics, and its balance between scholarly discussions of the topics and its attention to popular interest in the topics.”

— **Guohe Zheng**, *Professor of Japanese,
Ball State University, USA*

The Chinese Writing System in Asia

An Interdisciplinary Perspective

Yu Li

First published 2020
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2020 Yu Li

The right of Yu Li to be identified as author of this work has been asserted by her in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Li, Yu, 1978– author.

Title: The Chinese writing system in Asia : an interdisciplinary
perspective / Yu Li.

Description: Abingdon, Oxon : New York, NY : Routledge, 2020. |

Includes bibliographical references and index.

Identifiers: LCCN 2019025053 (print) | LCCN 2019025054 (ebook) |
ISBN 9781138907317 (hardback) | ISBN 9781138907324 (paperback) |
ISBN 9780429345333 (ebook) | ISBN 9781000698343 (adobe pdf) |
ISBN 9781000698701 (mobi) | ISBN 9781000699067 (epub)

Subjects: LCSH: Chinese characters—Asia.

Classification: LCC PL1171 .L387 2020 (print) | LCC PL1171 (ebook) |
DDC 495.11/1—dc23

LC record available at <https://lcn.loc.gov/2019025053>

LC ebook record available at <https://lcn.loc.gov/2019025054>

ISBN: 978-1-138-90731-7 (hbk)

ISBN: 978-1-138-90732-4 (pbk)

ISBN: 978-0-429-34533-3 (ebk)

Typeset in Minion Pro

by Apex CoVantage, LLC

Visit the eResources: www.routledge.com/9781138907324

Contents

Preface ix

PART I LINGUISTIC PRELIMINARIES 1

CHAPTER 1 Foundational concepts 2

Phonetics and phonology 2

Emes and *Allos* 3

Notational conventions 5

CHAPTER 2 What is writing? 6

The definition of writing 6

Language is not writing 6

Visual representations of non-languages are not writing 7

Writing is systematic 8

Writing is secondary to speech 9

CHAPTER 3 What kinds of writing systems are there? 12

Classification of writing systems 12

The linguistic universal of writing systems 15

CHAPTER 4 *Pīnyīn* tutorial 17

Simple finals 18

Initials 19

Compound finals 21

Spelling rules 21

Tones 23

PART II WRITING CHINESE 29

CHAPTER 5 The Chinese speech 30

Chinese languages and dialects 30

Modern Standard Chinese 32

| | |
|-------------------|---|
| CHAPTER 6 | Written Chinese 40 |
| | Classical Chinese as a written standard 40 |
| | The impact of Classical Chinese on Modern Standard Chinese 42 |
| CHAPTER 7 | The Chinese writing system 45 |
| | Nature of the Chinese script 45 |
| | Formation of Chinese characters 46 |
| | The character inventory 50 |
| | Changes in script style 50 |
| | Structure of the Chinese writing system 51 |
| CHAPTER 8 | Demythifying the Chinese script 56 |
| | The ideographic myth 56 |
| | The universality myth 59 |
| PART III | BORROWING THE CHINESE WRITING SYSTEM 63 |
| CHAPTER 9 | Chinese characters in Asia: An overview 64 |
| | From shared borders to common vocabularies 64 |
| | Why borrow? 68 |
| | General borrowing strategies 69 |
| CHAPTER 10 | Writing Korean 77 |
| | <i>Idu</i> ‘clerical reading’ 78 |
| | <i>Hyangchal</i> ‘vernacular letters’ 79 |
| | <i>Gugyeol</i> ‘oral formulae’ 80 |
| | <i>Hangeul</i> 81 |
| CHAPTER 11 | Writing Japanese 87 |
| | <i>Kanbun</i> 88 |
| | <i>Kanji</i> 89 |
| | <i>Man’yōgana</i> 93 |
| | <i>Hiragana</i> and <i>Katakana</i> 94 |
| | A mixed script 97 |
| CHAPTER 12 | Writing Vietnamese 101 |
| | <i>Chữ nho</i> ‘Confucian script’ 101 |

Chữ nôm ‘southern script’ 102

Quốc ngữ ‘national script’ 105

PART IV REFORMING THE CHINESE SCRIPT 111

CHAPTER 13 Phonetic writing before *pīnyīn* 112

Early attempts at Romanization (1580s–1620s) 112

Later proposals of phonetic scripts (1890s–1930s) 115

CHAPTER 14 *Pīnyīn* 127

What is *pīnyīn*? 127

History of *pīnyīn* 129

Current uses of *pīnyīn* 130

Spelling rules 131

CHAPTER 15 Simplification of Chinese characters 136

Early attempts at simplification (1900–1950s) 136

Script simplification by the PRC (1950s–1980s) 139

International impact 146

CHAPTER 16 Writing and technology in modern China 149

Ordering information 149

Computer input methods 152

PART V IDENTITY AND GENDER IN WRITING CHINESE 157

CHAPTER 17 Handwriting and personhood 158

Handwriting and moral character 158

Handwriting and personality traits 160

Parallels between calligraphy techniques and moral preferences 161

CHAPTER 18 Sexism in the Chinese writing system 163

The status of women in China 163

Sexism in Chinese characters 166

CHAPTER 19 *Nǚshū*: Women’s script 169

Who, where, and when? 169

| | |
|-------------------|--|
| | What does <i>nǚshū</i> look like? 171 |
| | What kind of writing system is <i>nǚshū</i> ? 173 |
| | What is <i>nǚshū</i> used to write and why? 174 |
| CHAPTER 20 | Script choice in writing Japanese 179 |
| | Linguistic constraints 179 |
| | Extra-linguistic factors 181 |
| PART VI | CHINESE CHARACTERS IN ART AND LITERATURE 187 |
| CHAPTER 21 | Chinese calligraphy 188 |
| | The Chinese script as foundation for calligraphy 188 |
| | Calligraphic evolution of the Chinese script 189 |
| CHAPTER 22 | A calligraphy workshop 200 |
| | Getting ready 200 |
| | Warm-up 203 |
| | Basic strokes 204 |
| | Stroke order 208 |
| | Composition 209 |
| | Writing exercise 210 |
| CHAPTER 23 | Modern calligraphy in China 212 |
| | The '85 New Wave 212 |
| | The Modernists and their works 214 |
| CHAPTER 24 | Chinese characters in Avant-garde art 219 |
| | Background 219 |
| | Artworks 220 |
| CHAPTER 25 | Chinese characters and Western Modernist poetry 226 |
| | Imagism and Modernist poetry in the English language 226 |
| | Chinese characters and Imagist poetry 227 |
| | <i>References 233</i> |
| | <i>Index 238</i> |

Preface

The idea of this book came about in the second year of my teaching career. After having taught a few Chinese language courses, I was eager to create something new to offer my students, something that would spark their curiosity, challenge their thinking, and inspire their creativity. I wanted them to see, as I did at the time and still do now, that learning Chinese was not just about being able to communicate in the language; it could be and should be, in no small measure, a journey of intellectual discovery and growth. And so was my experience researching for and teaching the course *Chinese Writing Systems in Asia* at Emory for nearly ten years since then.

Now this book almost feels like a gift to myself: the next time when I teach this course, I will be able to use it as a main text. I have had to rely on a collection of scholarly writings that was challenging to present as a coherent body of text in an introductory course. Having a systematic and accessible set of readings on hand will be a welcome change. I hope it will make it easier for interested colleagues in the field to offer similar courses as well. Besides what is in the print edition and the electronic version of the book, I have also made available additional materials and tips for teaching on the companion website, including exercises and activities that can be integrated into classwork or assigned as homework.

Although intended as a textbook for a college course, this volume is by no means restricted to a classroom audience. General-interest readers curious or enthusiastic about the Chinese script will also find the broad-ranging topics interesting and accessible. This book is designed to be an interdisciplinary introduction to the Chinese writing system. The chapters integrate studies of the Chinese language, writing, and linguistics into inquiries regarding script borrowing, writing reform, technology, gender, identity, visual art, and literature, and they assume no prior knowledge about any of these topics or about Chinese language or linguistics in general. Parts I and II are foundational and will be best read first, but the remaining ones can be more flexibly sequenced.

It is also with the development of my pedagogical field in mind that I have written this book. As strongly as ever, I believe that there is much more to be said and done about teaching Chinese than the worthy endeavor of raising students' communicative proficiency. Language and writing are at the core of humanistic flourishing – as we see in this volume, the Chinese writing system has been and will continue to be part and parcel of the forging and reshaping of social, cultural, and certainly personal identities – of those who've embraced it and those who've abandoned it. Language and culture instructors of Chinese are obligated to bring our students closer to such understandings, and I hope this book is a step in that direction.

A small step, that is. The subject of Chinese characters is endlessly fascinating, and I have had to limit the topics to the selected few that there is room to accommodate and worry about their uneven treatment. Indeed, among the numerous decisions made during the writing process, I suspect not all were wise, and any true wisdom that does come through is probably rubbing off the many works I have consulted and to which I am greatly indebted.

I also owe a great deal to the many individuals who have helped with this project and wish to thank them here: my Emory colleagues and friends Juliette Apkarian, Julia Bullock, Bumyong Choi, Cheryl Crowley, Seth Goss, Sun-chul Kim, Aya McDaniel, Maria Sibau, and Amanda Wright, who gave me invaluable feedback on various parts of this project; my editor Kelly Besecke, who patiently read through every word and offered key insights on making them fit better together; my research assistant Tianqi Wang, who created most of the illustrations, helped obtain permissions to use copyrighted images, and organized materials for the companion website; and my dear friends Ron Janssen, Jin Liu, and Qi Wang as well as my family in China, for inspiring me and keeping me going. It is to them that I dedicate this book.

PART I

Linguistic preliminaries

1

Foundational concepts

THE KEY TO UNDERSTANDING how a writing system works is to understand how its symbols systematically represent the units of the language it writes. This is often an intricate matter that involves linguistic concepts of which we may not be intuitively aware. It is therefore necessary to familiarize ourselves with these concepts first, so that we will be able to discuss the working mechanisms of writing systems with more clarity and efficiency. In what follows, we will start with defining and distinguishing between a pair of terms that provide the disciplinary setting in which we will discuss the other terms: phonetics and phonology. From there we will look at three analogous pairs of concepts: phoneme and allophone, morpheme and allomorph, and grapheme and allograph.

■ Phonetics and phonology

Phonetics and phonology are both branches of linguistics that study the sounds of human languages. They are different yet closely related. **Phonetics** deals with speech sounds as concrete physical entities, and it is primarily interested in their articulatory and acoustic aspects, for example: How are speech sounds made? What are the gestures and movements involved? What are the articulatory features that distinguish one speech sound from another? What are the physical properties that differentiate them from each other? How are speech sounds represented, analyzed, and read on a computer? As you see, phonetics treats speech sounds as concrete and physical elements that have stable inherent qualities. However, it is important to understand that speech sounds do not function as isolated entities in fulfilling their linguistic roles. They interact with each other, often causing phonetic – that is, articulatory and acoustic – changes, and they tend to behave in ways that form predictable patterns. This is where phonology comes in.

Phonology is primarily concerned with the patterns in which speech sounds behave as members of a linguistic system. In investigating the regularities and patterns in sound distribution, phonology views speech sounds as abstract linguistic units. Such abstraction is necessary, because – for example – a given speech sound may pattern differently with other sounds in different languages, so that the same phonetic entity may have a different phonological status in a different language. One example is the consonants /θ/ and /s/ in English vs. French. θ is an International Phonetic Alphabet (IPA) symbol we use to represent the final sound in the English word *myth*, and s is the IPA symbol for the final sound in *miss*.¹ In English phonology, /θ/ and /s/ are regarded as two distinctive sound categories, because changing from one to the other would change the meaning of the word (i.e., from “myth” to “miss,” or vice versa). In other words, /θ/ and /s/ **contrast** with each other (i.e., they are used to differentiate word

meaning) in the English phonology. In French, however, the [θ] sound is normally not used, so no two words are different from each other solely on the basis of [θ] vs. [s]. If a speaker for some reason substitutes [θ] for a [s] in a French word, the result may sound odd to a native French person, but it would not be interpreted as a different word. Thus, in the subconscious mind of the French speaker, [θ] and [s] are two possible physical realizations of the same sound category: they are phonetically different, but phonologically the same. By contrast, /θ/ and /s/ are phonologically distinctive in English. This example tells us that in order to be able to describe the distribution of speech sounds in different languages, we would not only want to talk about them as concrete entities in the physical world (i.e., phonetics), but would also need to understand at an abstract level how they relate to one another in a particular language (i.e., phonology).

THE STORY BEYOND

Speaking with a French accent

English distinguishes [θ] (*myth*) vs. [s] (*miss*), or their voiced counterparts [ð] (*breathe*) vs. [z] (*breeze*) phonologically, while French does not normally use [θ] or [ð]. In fact, the sounds [θ] and [ð] are rare in human languages, while [s] and [z] are very common. Someone speaking English with a French accent would tend to substitute [s] for [θ] and [z] for [ð]. For example, *thank you very much for these* would become something like *sank you very much for zese*. *I think of you all the time* would sound like *I sink of you all ze time*.

Emes and Allos

Speech sounds may have different relationships in the same language as well. Take English as an example: some sounds contrast with each other while others complement each other. What do we mean by “contrast” or “complement” in this context? When speech sounds can occur in exactly the same phonological environment, and switching from one to another creates differences in meaning, we say that they **contrast** with each other, or that they are **in contrastive distribution**. For instance, the initial consonants of the English words *mac* /mæk/, *tack* /tæk/, and *lack* /læk/ are /m/, /t/, and /l/, respectively. They can all occur at the beginning of a word and in front of the vowel /æ/. In fact, they constitute the only differences in the words given, and changing from one to another alters the meaning of the word. In this case, these three sounds are contrasting with one another, and we say that they belong to distinctive phonological categories, or phonemes, in English. **Phonemes** are abstract sound categories in the speaker’s mind that contrast with each other in a given language. *Phone* here means “speech sound.” It comes from the Greek word *phōné* ‘sound, voice.’ The English suffix *-eme* indicates a significant contrastive unit in linguistics at the level the stem of the word suggests. In this case, a *phon-eme* is a smallest contrastive *sound* unit. We will soon learn about *morph-eme* and *graph-eme*.

A phoneme may sound different depending on the context, and the concrete phonetic realizations of a phoneme are called **allophones** of this phoneme. *Allo-* means “other.” Its origin is the Greek form *állos* ‘other.’ In linguistics, *allo-* is used to indicate non-contrastive,

alternative forms of *-emes*. Besides *allo*-phones, we will also learn about *allo*-morphs and *allo*-graphs in this chapter. As an example for understanding the concept of allophones, think about how the phoneme /t/ in English may be pronounced differently based on the context. When it occurs at the beginning of a word, most speakers pronounce it as an aspirated consonant [t^h], as in *tack* [t^hæk]. **Aspiration** refers to the strong puff of air that accompanies the production of the consonant, and it is marked as a raised *h* ([^h]) in IPA. When /t/ occurs after the sound [s] in the same word, by contrast, most English speakers use the unaspirated version of the sound, [t], as in *stack* [stæk]. If you place your hand close in front of your mouth when saying *tack* and *stack*, you should be able to feel a stronger puff of air in *tack*. Thus, [t^h] and [t] are phonetically different sounds. In fact, native speakers of English are fully able to hear the difference between them.

Allophones are context-dependent, so that the phonological environments in which they occur do not overlap with each other. In other words, they are **in complementary distribution**. In the example of *tack* and *stack*, [t^h] is always at the beginning of a word and never occurs right after [s] in the same word, while [t] always occurs immediately after [s] and never at the beginning of a word. Thus, as allophones of the same phoneme /t/ in English, [t^h] and [t] complement each other in distribution. For this reason, an allophone is also a **predictable** phonetic realization of a phoneme as determined by the phonological context: at the beginning of a word, the English /t/ is always [t^h], while immediately after [s] within a word, it is always [t].

Companion website

Exercise 1.1 Phonemes or allophones

Now that we understand the terms phoneme and allophone as defined in phonology, we will learn two additional pairs of concepts analogous to them, both of which will be very useful to us in the discussion of writing systems: morpheme and allomorph, grapheme and allograph. A **morpheme** is the smallest combined unit of sound and linguistic meaning or function in a given language. It is a word-like unit but more rigorously defined than word. In fact, “word” is notoriously difficult to define in linguistic terms: Is *hot dog* one word or two words? What about *bookcases*? And *fun-filled*? This is one reason why we need the concept of a morpheme. It allows us to say for sure how many basic units there are: *hot dog* contains two morphemes, *hot* and *dog*; *bookcases* consists of three morphemes, *book*, *case*, and the plural marker *-s*; and *fun-filled* also has three morphemes, *fun*, *fill*, and the past participle suffix *-ed*. Each of these morphemes is a smallest meaningful or functional unit and cannot be further divided without losing its meaning or function. Indeed, different from phonemes, morphemes not only involve sounds but also come with meanings or grammatical functions. In this sense, morphemes are at a linguistic level higher or more complex than phonemes.

Like phonemes, however, a morpheme is also an abstract representation in the speaker’s mind, and its actual phonetic realizations, or **allomorphs**, depend on phonological contexts. For example, the plural morpheme in English is /-z/. It is realized as the voiceless [s] when following a voiceless consonant. Thus, we have /-s/ in *cats* [kæts], where [t] is a voiceless consonant. The plural morpheme /-z/ becomes [-əz] with a short vowel inserted in front when it comes right after the sounds [s] (as is the final sound in *kiss*), [z] (*buzz*), [ʃ] (*wish*), [ʒ] (*mirage*), [tʃ] (*church*),

and [dʒ] (*judge*). Therefore, we have *kisses* [-səz], *buzzes* [-zəz], *wishes* [-ʃəz], *mirages* [-ʒəz], *churches* [-tʃəz], and *judges* [-dʒəz]. All elsewhere, the plural morpheme /-z/ is realized simply as [-z]. The allomorphs [-s], [-əz], and [-z] are context-dependent, they are in complementary distribution, and we can also say that they do not contrast with each other.

The concepts phoneme, allophone, morpheme, and allomorph apply to speech; in writing, we may use the terms “grapheme” and “allograph” to think about the roles graphic symbols play. **Graphemes** are abstract representations of contrastive graphic categories in the writer’s mind. In writing English, for example, <m> and <t> are two graphemes that can distinguish word meaning: <smart> is a different word than <start>. **Allographs** are actual realizations of a given grapheme that are determined by context. They are written differently in different contexts, but we still recognize them as the same letter or character. In the printed text below, the grapheme <m> has two allographs, each in a different typeface:

smart [Times New Roman]
 smart [Helvetica]

The Times New Roman {m} and the Helvetica {m} are each consistent with their neighboring letters in terms of typeface, as is usually the case in printing practice. Here is another example: the uppercase {M} and the lowercase {m} are also allographs of the same grapheme <m>. Based on the convention for writing English, for example, {M} is used at the beginning of a sentence, while {m} is used elsewhere, unless it is the first letter of a proper name or part of an acronym, where it should be capitalized:

May you have a good vacation.
 You may have a good vacation.
 May is my favorite vacation month.
 My favorite vacation month is May.

■ Notational conventions

Following the notational conventions in linguistics, in this book, phonemes and morphemes are placed between slashes (e.g., /p/). Allophones and allomorphs use square brackets ([p]). We use angle brackets (<p>) for graphemes and braces ({p}) for allographs.

Note

- 1 Transcriptions of speech sounds in the rest of the book are also in IPA. In IPA, each symbol represents a unique speech sound. The use of IPA allows us to specify and communicate with clarity what sounds we are referring to. IPA is also designed to be able to transcribe all speech sounds in human languages, which makes it an indispensable tool for documenting languages not yet written down. We will not discuss IPA in detail in this book. If you are interested in learning how to transcribe English in IPA, many resources are available on the Internet. The companion website has links to such resources.

2

What is writing?

■ The definition of writing

In our day-to-day use of the English language, the word “writing” may mean a number of different things. It can be the act of composing an essay or a poem. It can be a noun, such as “the writings of Edgar Allen Poe.” It can refer to the rhetorical style in which the essays are written, as in “Hemingway’s writing is terse and succinct.” When handwriting is employed, “writing” can also refer to the penmanship of the writer.

Which of the above is the writing with which we are concerned? Well, none of them. For our purpose, **writing** is defined as graphic representation of language in a specific and systematic manner. For the purpose of this course, we will not be concerned with the connotative, or figurative, uses of the word “writing” – that is, the action or outcome of writing, the style of putting words together to express ideas, or the aesthetic appearance of marks made on paper. We will solely use the term in its denotative, or literal, meaning of the visual representation of linguistic utterances with graphic symbols.

Companion website

Exercise 2.1 The definition of writing

To fully understand what constitutes writing, let us think further about what is *not* writing.

■ Language is not writing

The distinction between “writing” and “language” is a necessary and important one in our discussion of writing systems. By defining writing as graphic representation of language, we mandate that writing be seen as separate from language: **language** refers to the speech one produces or hears, while writing refers to the visual representation of speech using graphic marks. Writing is not one kind of language, and the concept of language does not incorporate writing.

Yet, perhaps because many of us today live in a highly literate society, confusion between the two, especially in using “language” to refer to “writing,” is pervasive. You may have heard statements like these:

Chinese has tens of thousands of characters.

Hebrew has no vowels.¹

Nǚshū is a secret language invented by women.

Such statements are linguistically problematic. “Chinese” or “Hebrew” each refers to a language.² The Chinese language is not composed of characters but speech sounds – one does not need to write or read any characters in order to speak or understand it. Hebrew is clearly not lacking in vowels, though it is true that vowels are generally not represented in writing the language. Thus, the first two statements are in fact comments on how the languages are written rather than on the languages themselves. The third statement may require a little more explanation. *Nǚshū* (女书/女書 ‘Women’s Script’) is the name of a script created by and once used among women in rural Jiangyong county, Hunan province in China. It is not a language but a writing system designed for writing the local Chinese dialect. We will learn about *nǚshū* in detail in Chapter 19 of this book.

Statements like these may reveal a lack of understanding about what constitutes language. In cognitive terms, **language** is a complex system residing in our brain that allows us to produce and interpret utterances. For our purpose, since we are primarily concerned with graphic rendering of linguistic utterances, language can be understood specifically as the **speech** produced or potentially produced by this system.

Visual representations of non-languages are not writing

Writing is visual representation of language, and language *only*. Although language conveys information or ideas, not all graphic renditions of ideas or information are writing. A good example to illustrate this point would be **pictography**,³ or so-called picture writing, that is, the use of usually highly stylized or stereotyped pictures to communicate information or serve as memory aids. International airports are where modern pictography abounds: a figure in a skirt or dress indicates the ladies’ room, a person seated in a wheelchair signifies accessibility, and a set of a fork and knife suggests places to eat. Because these symbols are not tied to any particular language, they, to some extent, transcend linguistic barriers and function effectively in assisting travelers navigating the airports. However, precisely because they are not tied to any particular language and do not correspond to any specific linguistic utterances, they cannot be considered writing. They are merely visual cues employed to suggest information with which the viewers are already familiar.

The rule of thumb is that writing represents *specific* utterances of a language in a *systematic* way, so that proficient readers are able to reliably reproduce what is written and consistently render it back into speech intelligible to the community of speakers of that particular language.

Companion website

Exercise 2.2 Examples of writing

Indeed, systematicity is one of the fundamental features of writing, and this is why linguists use the term **writing system** to refer to what we commonly call **writing** or **script**. In the remainder of this book, we will use these three terms interchangeably.

Now let us further examine the ways in which writing is systematic.

■ Writing is systematic

Writing, first of all, has a systematic relationship to language. A script generally consists of a set of visual symbols (graphemes) linearly or otherwise arranged to represent the speech units of a language. For this to work efficiently, a high degree of consistency is required. Each grapheme in the script usually represents a constant unit of the language, such as a single phoneme or a single morpheme. For example, the letters <a k i l m> in writing English each represent a phoneme in the most ordinary cases. In writing Chinese, a character, such as <人走文>, corresponds to a morpheme: *rén* ‘person,’ *zǒu* ‘to walk,’ or *wén* ‘writing.’ Of course, the relationships between a grapheme and a speech unit as exemplified by the writing of English and Chinese are not the only possible relationships between script and language. In writing Cherokee or *Yi*⁴ (彝), for example, an individual symbol represents a meaningless syllable. As another example, the graphemes of the Japanese *hiragana* (平仮名) and *katakana* (片仮名) scripts represent speech units that may be slightly smaller than syllables. The nature of the relationship between language and script serves as the basis on which linguists categorize writing systems. For example, Cherokee and *Yi* scripts are both syllabic writing systems. We will look at the classification of writing systems in more detail in Chapter 3 of this book. For now, it is sufficient to understand that the corresponding relationship between a grapheme and a speech unit is generally constant in that it applies to the entirety of a writing system.

Writing systems in general also have internal organizations of their own. For instance, English is usually written horizontally from left to right, while Arabic writing is normally right to left. English letters are arranged sequentially. In writing Korean, however, letters are combined into blocks of syllables, often with some letters stacked on top of others. In both English and Korean, words are separated by spaces, but in Chinese they are strung together with no spaces in between. In English, the first letter of each sentence is usually capitalized, while in writing German, the first letter of each noun is capitalized regardless of its position in the sentence.

The internal structure of a writing system has little to do with the language being written and is primarily a matter of convention. For this reason, it may change over history. For example, some ancient Greek texts of the 6th century BCE were written in alternate lines of opposing directions. That is, the text flowed from left to right in one line, and then right to left in the line below it, and then left to right again below that, and so on. This pattern was likened to how an ox turned in ploughing a field and was given the name *boustrophedon*, a word Greek

in origin meaning “to turn like oxen (in plowing).” Chinese text is another example. Today Chinese is usually written in horizontal lines from left to right and the lines are arranged from top to bottom on a page, just like in writing English. Before the 20th century, however, most Chinese texts used a very different arrangement: writing proceeded from the upper right corner of the page in columns arranged from right to left with each column starting at the top. This change was the result of Western influence on China.

■ Writing is secondary to speech

To say that writing is secondary is to make the point that speech is primary. The primacy of speech has been one of the cardinal principles of modern linguistics – it is one of the features that distinguishes this modern scientific discipline from its earlier state in the European tradition prior to the 20th century.⁵ Evidence for this principle comes from historical, cultural, and developmental sources. The history of writing is much briefer than that of human language or speech. Speech is a naturally evolved, defining characteristic of the human species. Writing, on the other hand, is a cultural invention that emerged much later in history. We know from archaeological findings that the history of writing goes back only about 5,000 years, while scholars have estimated that human speech emerged between 200,000 and 60,000 years ago.⁶

That writing is secondary to speech is also evident when we consider that, although all languages are spoken, not all languages have been written down. In literate societies today, it is not uncommon for bilingual or multilingual communities to use one language purely in spoken form, usually for ordinary conversations such as in the home setting, and another language in writing (and/or speech) for more formal purposes. For example, a student native to the city of *Shànghǎi* may speak the *Shànghǎi* dialect, which is not mutually intelligible with Mandarin, with her friends and family, but converse and write in Mandarin for schoolwork. In linguistics, such a phenomenon is referred to as **diglossia**. The local speech variety in a diglossic situation, in this case the *Shànghǎi* dialect, is often not officially written.

Another set of evidence comes from language acquisition. All people, who are physically able, acquire the ability to speak as children. Only some learn to write, while others may never do so. The acquisition of speech happens naturally for children, regardless of the particular language involved. The learning of writing, on the other hand, requires a conscious effort.

Further insight for the primacy of speech comes from the relationship between language and writing. The same language may be written using a variety of writing systems. In theory, in fact, a given language can be written in any kind of writing system. Which system is in use is, to a great extent, a matter of cultural convention rather than determined by the nature of the language involved. Furthermore, writing systems can be and also have been artificially designed or reformed. In this sense, writing is very much a cultural phenomenon.

Last but not least, the primacy of speech is reflected in our definition of language and writing. Given that language refers to speech, and writing is visual representation of language, it logically follows that speech comes before writing and serves as the basis for writing. A pair of terms that we often hear is “spoken language” and “written language,” as if speech and writing were equal alternative forms in which language could be rendered. However, these terms may cause confusion about the relationship between language, speech, and writing. The term

“spoken language” in itself is redundant – it is equivalent to saying “spoken speech,” since by our definition language refers to speech⁷ and all languages are spoken.

The term “written language” may cause considerable confusion as well, for it has been found to mean a range of different concepts. Some use it to refer to a language that has been out of use (or has never been in use) for oral communication and is now primarily used in the written form, such as Latin or Classical Chinese. Others use it to mean the writing system with which a language is recorded or represented. Still others find the term applicable to anything that is written down. To avoid confusion, we will restrict the meaning of the term to the first sense. That is, a **written language** is a language whose current use is limited to writing, and it may or may not have been used in speech in the past. Latin and Classical Chinese were both spoken at an earlier point in history. Today, they may be used to compose text, and such text may be read aloud by someone on appropriate occasions, but the languages are no longer employed in daily oral communication as they once were. In essence, a written language is a language that is usually represented in the written form.

Defined as such, a written language is very different from a writing system or script. The term “writing system” may appear rather technical to those not trained in linguistics. In its place, “written language” has been used in non-specialized (and sometimes specialized) publications to refer to script. A written language, however, is not the same as a writing system; in fact, it can be represented using multiple systems of writing. For example, Classical Chinese is usually written in Chinese characters, but we can conceivably write it in *pinyin* (拼音 *lit.* ‘spell sound’), a writing system based on the Roman (Latin) alphabet.

With the above clarification in mind, it is helpful to be aware that we may still encounter loose usage of these terms in writings by linguists or non-linguists alike. In such cases, we will need to pay close attention to the context and determine the exact connotation of the terms to avoid further confusion. In our own writing, of course, it will be beneficial to adhere to the exact definitions.

Notes

- 1 This example is from Rogers (2005, p. 2).
- 2 The term “Chinese language” in fact encompasses a wide variety of speech, some of which are not mutually intelligible. Linguists in China and the West debate over whether these speech varieties should be considered dialects of the same language or separate languages. We will talk more about this in Chapter 5.
- 3 The term “pictography” is not to be confused with “pictographs” (象形字 *xiàngxíng zì*), which is usually the term used to refer to Chinese characters created based on visual resemblance to the objects or phenomena they represent. Pictographs are part of the Chinese writing system, while pictographic symbols, such as the female figure for “ladies’ room” in public places, are not part of any writing system.
- 4 *Yí* is the standard language spoken by the *Yí* people, one of the 55 officially recognized ethnic minority groups in China.
- 5 The focus of linguistic study before the 20th century was predominantly on what was written down. Even among modern linguists, the primacy of speech has not been without controversy. George Sampson (1985), for example, argued for a status of writing on par with that of speech. He considered writing and speech both aspects of language (pp. 11–13). In the context of our discussion, assuming the primacy of speech makes more sense and will help avoid much confusion.

- 6 Berwick and Chomsky (2016).
- 7 It should be acknowledged that a speech-based definition for “language” has its limitations. Sign languages are full languages (more broadly defined) as well, yet they are excluded from this definition. Defining language as speech, however, makes sense in the context of this book, because we are primarily concerned with writing systems used to represent speech, not sign languages.

3

What kinds of writing systems are there?

■ Classification of writing systems

What makes one writing system similar to or different from another? Is it the appearance of the symbols, the number of the letters or characters, or something else? Recall how we have defined writing: writing is a systematic representation of language using graphic marks. This tells us that what fundamentally distinguishes writing systems from each other is in how the written symbols (graphemes) represent the units of speech – or at what linguistic level such representation takes place. In other words, the nature of the script changes depending on whether a writing system's graphemes correspond to phonemes, syllables, morphemes, or some other units of speech. This correspondence is the basis on which we classify writing systems.

Broadly speaking, we can divide writing systems into two categories: those whose graphemes represent speech sounds only and do not involve meaning are called **phonographic** scripts. Many familiar languages, such as English, Spanish, modern Korean, Arabic, Hebrew, and Cherokee, are typically written using phonographic scripts.

Phonographic scripts can be further differentiated from each other based on the specific unit of speech a grapheme represents. In a **phonemic** writing system, a grapheme corresponds to a phoneme. The Roman alphabet is perhaps the most widely used phonemic writing system – and the most widely used writing system in the world – in terms of the variety of languages it conventionally encodes. Aside from languages with European origins such as Irish, Hungarian, Romanian, and Czech, it is also used for non-European languages including modern Vietnamese, Indonesian, Turkish, Afrikaans, and Hawaiian. The current total number of languages primarily written in the Latin script is more than 130. Languages typically written in other scripts also often have official alternatives in Roman letters. For example, the main East Asian languages we discuss in this book, Chinese, Japanese, and Korean, have all been **Romanized**, or represented in the Latin alphabet, to take advantage of its more universal currency.

A second kind of phonographic writing system is the so-called **consonantal** script. Arabic and Hebrew, for example, are usually written without representing vowels. Rather, each grapheme corresponds to a consonant. This works because the use of vowels in these languages is generally predictable based on the consonants, and a proficient reader will be able to fill out the vowels as she reads.

In a phonographic writing system, a grapheme may also correspond to a single syllable in the language. For example, the *Yí* language in China is usually written this way (Table 3.1). Scripts like this are referred to as **syllabic** writing systems, or **syllabaries**.

Table 3.1 Yí symbols representing syllables

| Yí syllabogram | Romanization |
|----------------|--------------|
| 𐏃 | ap |
| 𐏄 | bi |
| 𐏅 | gguot |

It is also possible for a grapheme of a phonographic script to represent a speech unit whose size is between a single speech sound and a syllable. Japanese *hiragana* (平仮名) and *katakana* (片仮名), collectively known as *kana*, are such writing systems. In *kana*, each symbol represents a **mora**, a phonological unit intermediate between a phoneme and a syllable. More specifically in Japanese, a mora can be a single vowel (V), e.g., *a*; a consonant-vowel sequence (CV) within a syllable, e.g., *ka*-; a nasal (N) at the end of a syllable, e.g., *-(ka)n*; or the first segment of a double consonant (C(C)), e.g., *t(t)*. As you see, a mora can be either a single sound (V, N, C(C)) or a syllable (CV, V). Table 3.2 shows a few examples.

Table 3.2 Japanese *kana* symbols representing morae

| Hiragana | Katakana | Romanization | Moraic structure |
|----------|----------|--------------|------------------|
| あ | ア | a | V |
| か | カ | ka | CV |
| ん | ン | -n | N |
| まった | マッタ | (ma)t(ta) | C(C) |

Why do we need the concept “mora” after all, if it is so cumbersome to define? It has to do with the timing and rhythm of the language, in this case Japanese. Apparently, mora is the timing unit on which the Japanese speech is based. That is to say, each mora is given about the same amount of time in speaking Japanese. The concept of mora, therefore, is indispensable for describing and understanding the Japanese sound system. We will talk more about Japanese morae in Chapter 11 of the book. For now, it is sufficient to know that **moraic** writing systems are also a type of phonographic script.

Unlike phonographic scripts, which represent speech sounds only, the character-based writing system for Mandarin Chinese consists of symbols that do involve meaning. A Chinese character, in most cases, represents the smallest combination of sound and meaning – that is, a morpheme – in speech. We refer to such a writing system as a **morphographic** or a morphemic script. More specifically, since a character usually corresponds to a syllable in terms of speech sounds, we can also say that Chinese uses a **morphosyllabic** writing system.

THE STORY BEYOND

Morphographic? Logographic?

You may have heard of another term that is often used to characterize the Chinese script: logographic. Some also refer to Chinese characters as “logographs” or “logograms.” How is “logographic”

different from “morphographic”? Which adjective offers a more accurate description of the Chinese writing system?

The “logo-” part of the word “logographic” derives from the Greek *lógos*, meaning “word.” A logographic script is thus a writing system in which a grapheme typically represents a word or a phrase. This is not exactly the case for Chinese characters, however. A great portion of the words in Modern Chinese contains more than one morpheme (and more than one syllable in pronunciation) and so must be written using more than one character. In other words, a large number of Chinese characters represents morphemes that cannot stand alone as words and must be combined with other morphemes to form words in Modern Chinese. This is why Chinese writing is not entirely logographic, and “morphographic” is a more accurate characterization.

We can summarize the different types of writing systems in Table 3.3.

Table 3.3 Classification of writing systems

| Category | Type | Examples |
|---------------|----------------|--|
| Phonographic | Phonemic | English, (modern) Korean, (modern) Vietnamese |
| | Consonantal | Arabic, Hebrew |
| | Moraic | Japanese (<i>hiragana</i> and <i>katakana</i>), Cherokee, Cree-Inuktitut |
| | Syllabic | <i>Yí</i> (in China) |
| Morphographic | Morphosyllabic | Mandarin (Chinese) |

So far, we have been assuming that the grapheme-speech relationship for a given pair of writing system and language is uniform. In the case of English written in the Roman alphabet, for example, such an assumption would mean that each Roman letter consistently represents a phoneme in English: <p> would usually be a consonant and <a> would normally represent a vowel – even though the specific consonant and vowel might not always be the same ones.¹

In reality, however, this grapheme-speech relationship is almost never completely consistent. English, in particular, is notorious for its pervasive spelling irregularities. One letter may represent multiple speech sounds, and a single speech sound may be written as two or more letters. One often cited example, to push this point to a certain extreme, is the artificial word *ghoti* as an alternative spelling for the English word “fish.” If we extend the grapheme-phoneme correspondences in the words listed below to *ghoti*, we will indeed be pronouncing it as “fish.” Note that in this form, the letter <i>, which normally represents a vowel, is part of a spelling combination for a consonant.

gh, pronounced *f* /f/ as in *tough* /tʌf/
o, pronounced *i* /ɪ/ as in *women* /'wɪmən/
ti, pronounced *sh* /ʃ/ as in *nation* /'neɪʃən/

Despite such discrepancies, the Roman alphabet in writing English is still, in essence, a phonemic writing system. The fundamental relationship of “one letter, one sound” still holds

THE STORY BEYOND

The life of *ghoti*

The spelling *ghoti* for “fish” traces its origin to English text written in no later than the mid-19th century. It has been attributed to a number of sources, the earliest of which appears to be William Ollier Jr. (born 1824) in a letter drafted by his father. The word was later on often cited to support English-language spelling reform.

Ghoti has taken on a life of its own in modern popular culture. In Klingon, the constructed language of the Klingons in *Star Trek*, the word for “fish” is spelled as *ghotl’*. In the 1966 *Batman* TV episode “An Egg Grows in Gotham,” Ghoti Oeufs Caviar Company is the name of the fictional character Egg-head’s caviar business. Ghoti Hook was the name of a Christian punk band active during 1991–2002 based in Fairfax, Virginia, though the band pronounced its own name as “goatee” instead of “fish.”

The speech synthesizer in Mac OS X pronounces *ghoti* as “fish.”²

true. This is especially understandable if we view the subject from a historical perspective. Many of the spelling irregularities arose from phonological changes in the English language over time. For instance, why is it that we do not pronounce the <k> in words like *knight*, *knot*, and *knowledge* yet we spell them with a <k> in writing? In fact, <k> used to be pronounced in Old English before <n>, and it sounded just like the <k> in *kid* today. In modern German, a language genetically related to English, <k> is also still pronounced before <n> in words like *Knecht* (/knɛçt/ ‘servant’) and *Knoten* (/ˈknoːtn/ ‘knot’).

■ The linguistic universal of writing systems

Given the variety of existent writing systems, you may wonder: Can any language be written in any writing system? Or, is a language inherently tied to a particular (kind of) script? For example, must the English language be written in the Latin alphabet? Must it be written in a phonemic writing system, more broadly speaking? Can it be represented using a system like the one for Mandarin Chinese – that is, a morphographic script?

We in fact already touched on this earlier, and, simply put, the answer is “yes” – any language can be written using any writing system, at least theoretically. To understand how this is, we need to take a deeper look into the nature of writing and language. Although writing systems employ a wide range of symbol-speech relationships, fundamentally, they are all visual means to represent speech, and in this sense, speech serves as the foundation for all systems of writing, and visually representing speech is the defining feature or purpose of writing. As John DeFrancis pointed out in *Visible Speech* (1989), “all full systems of writing are based on representation of sounds. . . . Just how a particular writing system goes about making sound visible is a secondary matter.”³ We have seen that speech can be made visible by having a grapheme represent a single speech sound (phonemic, consonantal), a combination of speech sounds without meaning (moraic, syllabic), or a combination of speech sounds with meaning (morphemic). These linguistic units are universal to languages – both English and Chinese have phonemes, syllables, and morphemes, for example. So there is no theoretical reason why