

ROUTLEDGE  
ROUTLEDGE  
HANDBOOKS



# The Routledge Handbook of the English Writing System

Edited by Vivian Cook and Des Ryan

# The Routledge Handbook of the English Writing System

*The Routledge Handbook of the English Writing System* provides a comprehensive account of the English writing system, both in its current iteration and highlighting the developing trends that will influence its future. Twenty-nine chapters written by specialists from around the world cover core linguistic and psychological aspects, and also include areas from other disciplines such as typography and computer-mediated communication.

Divided into five parts, the volume encompasses a wide range of approaches and addresses issues in the following areas:

- theory and the English writing system, discussing the effects of etymology and phonology;
- the history of the English writing system from its earliest development, including spelling, pronunciation and typography;
- the acquisition and teaching of writing, with discussions of literacy issues and dyslexia;
- English writing in use around the world, both in the UK and America, and also across Europe and Japan;
- computer-mediated communication and developments in writing online and on social media.

*The Routledge Handbook of the English Writing System* is essential reading for researchers and postgraduate students working in this area.

**Vivian Cook** is Emeritus Professor of Applied Linguistics at Newcastle University, UK, and founding co-editor of the journal *Writing Systems Research*.

**Des Ryan** is a doctoral student at Trinity College Dublin, Ireland.

## **Routledge Handbooks in Linguistics**

*Routledge Handbooks in Linguistics* provide overviews of a whole subject area or sub-discipline in linguistics, and survey the state of the discipline including emerging and cutting edge areas. Edited by leading scholars, these volumes include contributions from key academics from around the world and are essential reading for both advanced undergraduate and postgraduate students.

A full list of titles in this series is available at: [www.routledge.com/series/RHIL](http://www.routledge.com/series/RHIL). Recently published titles:

### **The Routledge Handbook of Syntax**

*Edited by Andrew Carnie, Yosuke Sato and Daniel Siddiqi*

### **The Routledge Handbook of Historical Linguistics**

*Edited by Claire Bowerman and Bethwyn Evans*

### **The Routledge Handbook of Language and Culture**

*Edited by Farzad Sharifian*

### **The Routledge Handbook of Semantics**

*Edited by Nick Riemer*

### **The Routledge Handbook of Morphology**

*Edited by Francis Katamba*

### **The Routledge Handbook of Linguistics**

*Edited by Keith Allan*

### **The Routledge Handbook of Language and Media**

*Edited by Daniel Perrin and Colleen Cotter*

### **The Routledge Handbook of Phonological Theory**

*Edited by S. J. Hannahs and Anna Bosch*

### **The Routledge Handbook of Theoretical and Experimental Sign Language Research**

*Edited by Annika Hermann, Roland Pfau and Josep Quer*

### **The Routledge Handbook of Linguistic Anthropology**

*Edited by Nancy Bonvillian*

### **The Routledge Handbook of the English Writing System**

*Edited by Vivian Cook and Des Ryan*

---

# The Routledge Handbook of the English Writing System

*Edited by Vivian Cook and Des Ryan*

First published 2016  
by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge  
711 Third Avenue, New York, NY 10017

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

© 2016 selection and editorial matter, Vivian Cook and Des Ryan; individual chapters, the contributors

The right of the editors to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted 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: Cook, Vivian, 1940– editor. | Ryan, Des (Linguist) editor.

Title: The Routledge Handbook of the English writing system / edited by Vivian Cook, Des Ryan.

Description: New York : Routledge, [2016] |

Series: Routledge Handbooks in Linguistics | Includes bibliographical references and index.

Identifiers: LCCN 2015045837 | ISBN 9780415715973 (hardback) | ISBN 9781315670003 (ebook)

Subjects: LCSH: English language – Writing – Handbooks, manuals, etc. | English language – Written English – Handbooks, manuals, etc. | BISAC: LANGUAGE ARTS & DISCIPLINES / General. | LANGUAGE ARTS & DISCIPLINES / Linguistics / General. | LANGUAGE ARTS & DISCIPLINES / Alphabets & Writing Systems.

Classification: LCC PE1074.7 .R68 2016 | DDC 421/.1–dc23

LC record available at <http://lcn.loc.gov/2015045837>

ISBN: 978-0-415-71597-3 (hbk)

ISBN: 978-1-315-67000-3 (ebk)

Typeset in Times New Roman  
by HWA Text and Data Management, London

# Contents

---

<i>List of figures</i>	<i>ix</i>
<i>List of tables</i>	<i>xiii</i>
<i>List of contributors</i>	<i>xv</i>
<i>Acknowledgements</i>	<i>xix</i>
1 Introduction <i>Des Ryan</i>	1
2 Background to the English writing system <i>Vivian Cook</i>	5
<b>PART I</b>	
<b>Theory and the English writing system</b>	<b>25</b>
3 English among the writing systems of the world <i>Richard Sproat</i>	27
4 Linguists' descriptions of the English writing system <i>Des Ryan</i>	41
5 Phonology and English spelling <i>Iggy Roca</i>	65
6 Punctuation <i>Frank Kirchhoff and Beatrice Primus</i>	93

<b>PART II</b>		
<b>The history and development of the English writing system</b>		<b>111</b>
7	The etymological inputs into English spelling <i>Simon Horobin</i>	113
8	Changing functions: English spelling before 1600 <i>Merja Stenroos and Jeremy J. Smith</i>	125
9	Modernization and standardization since the seventeenth century <i>Phil Scholfield</i>	143
10	Spelling variation as documented in historical dictionaries: The <i>OED</i> as a test case <i>Philip Durkin</i>	163
11	Accent variation reflected in the standard writing system of English <i>Jesper Kruse</i>	175
<b>PART III</b>		
<b>Learning and teaching English</b>		<b>189</b>
12	Children's first language acquisition of the English writing system <i>Nenagh Kemp</i>	191
13	Teaching literacy to English children: Policy and practice <i>Terezinha Nunes</i>	205
14	Dyslexia and the English writing system <i>Liory Fern-Pollak and Jackie Masterson</i>	223
15	The structure of literacy teaching: A case study from England <i>Rhona Stainthorp</i>	235
<b>PART IV</b>		
<b>Society and the English writing system in the world</b>		<b>259</b>
16	Sociolinguistics and the English writing system <i>Florian Coulmas</i>	261
17	The evolution of British and American spelling <i>D. W. Cummings</i>	275

18	The spelling of Scots: Tradition, adaptation and reform <i>Jennifer Bann and John Corbett</i>	293
19	Irish English and the English writing system <i>Raymond Hickey</i>	315
20	Eye dialect and pronunciation respelling in the USA <i>Michael D. Picone</i>	331
21	The orthography of English-lexicon pidgins and creoles <i>Mark Sebba</i>	347
22	The English writing system in the linguistic landscapes of the world <i>Jeffrey L. Kallen</i>	365
23	Japanese uses of the English writing system: A case study <i>Takeshi Okada</i>	397
24	Spelling reform <i>Valerie Yule and Ishi Yasuko</i>	413
	<b>PART V</b>	
	<b>Processing the written symbol</b>	<b>429</b>
25	Typography and the printed English text <i>Will Hill</i>	431
26	Processing the written word <i>Melvin J. Yap and Susan J. Rickard Liow</i>	453
27	Computer-mediated communication and the English writing system <i>Lauren Squires</i>	471
28	Texting and other messaging: Written system in digitally mediated vernaculars <i>Tim Shortis</i>	487
29	Spellcheckers <i>Roger Mitton</i>	517
	<i>Index</i>	531

*This page intentionally left blank*

# Figures

---

2.1	Modern inscription in Newcastle upon Tyne	10
3.1	Taxonomy of writing systems according to Sampson (1985)	28
3.2	Taxonomy of writing systems according to DeFrancis (1989)	30
3.3	Planar taxonomy, after Sproat (2000)	31
3.4	Left: pípá 枇杷 ‘loquat’. Right: Pípá 琵琶 ‘pipa, lute’	37
4.1	The graphematic foot (left) and phonological foot (right) for the word <i>late</i>	53
5.1	Levels of language structure	68
5.2	Structuralism and generativism	69
5.3	The structure of the syllable	75
5.4	GVS historical changes	83
8.1	An example of handwriting from a fifteenth-century English document	134
9.1	A printed text of 1683	144
9.2	A late seventeenth century handwritten text from a commonplace book	144
9.3	Blackletter or Gothic typeface in the King James Bible (1611)	148
9.4	Round hand or copperplate handwriting guide	155
15.1	Data on average class size, pupil–teacher ratio and pupil–adult ratio over the last forty years	238
15.2	Percentage of pupils achieving level 4 in reading SATS since 1997	250
15.3	The top 12 and the bottom 12 countries relative to the PIRLS centre point of 500 in 2011	253
16.1	A simplest model of the dialectologist conception of language	262
16.2	A more complex model of language variation, including writing	263
16.3	Variation in written English: fuzzy edges of national varieties	269
17.1	<i>humour/humor</i> in British English	280
17.2	<i>humor/humour</i> in American English	280
17.3	<i>connection/connexion</i> in American English	283
17.4	<i>characterize/characterise</i> in British English	285
17.5	<i>medallist/medalist</i> in British English	287
17.6	<i>medalist/medallist</i> in American English	287

17.7	Typicality of British and American English	290
18.1	Map of Scots dialect areas reproduced from the <i>Dictionary of the Scots Language</i>	294
18.2	Cluster analysis for Aitken's Vowel 3 in five Scottish novels	307
19.1	Later insular Irish script, based on Latin (half-)uncial models, later used for texts in Old English; uppercase forms are identical with lowercase forms	315
21.1	Extract from blog <i>Jamaica Woman Tongue</i> by Carolyn Cooper	360
21.2	<i>Notis I go long</i>	361
22.1	Directions to the <Centre> (Newry, Co. Down, 2005)	368
22.2	American <Center> in two fonts (New York, 2007)	368
22.3	<JEWELRY> (Annapolis, 2014) and <JEWELLERY> (Dublin, 2014)	369
22.4	<HARBOUR> spelling in American English (Annapolis, 2014)	371
22.5	Anomalous <JEWELERY> and <LICENSE> (Dublin, 2014)	371
22.6	<Center> spelling in Amsterdam (2014)	371
22.7	<our> endings in Amsterdam English (2014)	372
22.8	<Parlour> and <color> spellings (Hong Kong, 2014)	372
22.9	<THRU> (Washington, D.C., 2014); <Tech>, <Sec> (Philadelphia, 2014)	373
22.10	<KORNER> (St. Johns, NL, 1999); <Rite> (Arlington, VA, 2014)	375
22.11	<Kwality> (Dublin, 2014); <Kool Cutz> (Warrenpoint, Co. Down, 2012)	375
22.12	<Eclectix> and <LAIDSF LIFECYCLE> (San Francisco, 2002)	375
22.13	Missing an <a> in <bckspace> (Amsterdam, 2014)	376
22.14	Casual <Ya> and <Sea> referencing <i>sea</i> and <i>see</i> (Annapolis, 2014)	376
22.15	<Full of> becomes <Fulla> (Warrenpoint, 2012)	377
22.16	<Rare Oul Times> in Monaghan town (2012)	377
22.17	<LOBSTAH>, <Chowda> (Annapolis); <PLANE> (Philadelphia) (2014)	378
22.18	Polysemy in <InDeep'n'Dance> (Amsterdam, 2014)	378
22.19	Billboard textspeak (Dublin, 2006); <@> a CMC zone (Dublin, 2014)	379
22.20	<WARE> (Newry, Co. Down, 2005); <DELLY> (Annapolis, 2014)	380
22.21	Using <BOOTH> for a car boot sale (Dublin, 2014)	381
22.22	Non-standard spellings: Dublin, Longford Town, and Amsterdam (2014)	381
22.23	Internal variation using <'s> (Dublin, 2014)	382
22.24	Plural <SOFA'S> (Dublin) and <Bike's> (Amsterdam) (2014)	383
22.25	<i>A different kettle of fish altogether</i> as one word (Dublin, 2014)	383
22.26	Separating morphemes onto banners (San Francisco, 2002)	383
22.27	Layouts: <i>upstairs</i> (Galway, 2005); <i>Solo</i> (Dublin, 2014)	384
22.28	<i>Cider</i> referring to north-siders and south-siders (Dublin, 2014)	385
22.29	Referencing French with <beaudevin> (Charlotte, NC, 2014)	385
22.30	The <i>beef</i> + <i>bistro</i> hybrid (Dublin, 2014)	386
22.31	Referencing Irish: <i>bó</i> and Irish typography (Dublin, 2014)	386
22.32	<you me town> is also 'dream town' (Fukuoka, 2007)	387
22.33	Orientations for English in Hong Kong (2014) and Fukuoka (2007)	388
22.34	THE DON in multiple guises (Fukuoka, 2007)	389
22.35	The star replaces an apostrophe (Arlington, VA, 2014)	390
22.36	Orthographic reversal and polysemy in <i>Evolution</i> (Amsterdam, 2014)	390
22.37	Pictures for <o> in <i>Cornish</i> and <ok> in <i>Wok</i> (Dublin, 2014)	391
22.38	Virtues of Ohio indexed with an ice cream cone (Arlington, VA, 2014)	393
22.39	Replacing <A> with pictures to index Japanese cuisine (Dublin, 2014)	393
23.1	Japanese keyboard (for Windows® PC)	400

23.2	Station sign written with different romaji styles	404
23.3	Newspaper article written vertically	407
23.4	The driver's door of Hikari Taxi cab	409
23.5	The right-hand side of the cab	409
25.1	a) Punch and matrix from Devinne; b) sort from Diderot	433
25.2	Venetian 301 e	434
25.3	Day double pica	436
25.4	Fell 3-line pica italic	438
25.5	Adobe Caslon ct ligature and Adobe Caslon st ligature	439
25.6	Th ligature	439
25.7	Baskerville BT italic ampersand	440
25.8	'Ye' sign from 1734 tombstone	442
26.1	McClelland and Rumelhart's (1981) interactive activation model of letter recognition	454
26.2	Coltheart <i>et al.</i> 's (2001) DRC model of visual word recognition and reading aloud	461
26.3	Seidenberg and McClelland's (1989) parallel distributed processing model	462
28.1	Private message to microblogging followers later broadcast in news story (2010)	490
28.2	Excerpt from a mass-mediatised homework assignment written in SMS style (2003)	492
28.3	Imagined 'grassroots literacy' as typeset in a printed novel (Dickens 1861)	494
28.4	Actual 'grassroots literacy' in township text message (Blommaert 2013: 452)	496
28.5	Early mobile phone pad with key to 'multi-tap' letter entry (2003)	497
28.6	Elaborated and demotic hetero-graphic styles in later text message (2011)	499
28.7	Sequence of earlier SMS interaction between two adolescents (2003)	500
28.8	Facsimile images of a smartphone and SMS interaction by older adults (2009)	503
28.9	Facsimile images of contrasting types of smartphone SMS interaction (2009)	503
28.10	Rationale for a choice of 'Greeklish' in 'Greeklish' (Androutopoulos 2012)	504
28.11	Excerpt from Sebba 2007: 11, recast in a familiar English 'hetero-graphy'	507

*This page intentionally left blank*

# Tables

---

4.1	Consonant alternations with constant spelling	46
4.2	Vowel alternations with invariant spelling	46
5.1	Distinctive features of English consonants	67
5.2	The twelve RP simplex vowels	71
5.3	Old English five-vowel system	73
5.4	Vowel sounds and spellings	82
9.1	Examples of sound changes in Southern British English 1600–1720 not matched by spelling changes	146
9.2	Examples of non-naturalised spellings of loanwords predominantly arriving after 1600 and their effect on the writing system	147
9.3	Establishment of < i j u v > as separate letters with their own values	149
9.4	Main areas of optional letter variation still existing in the seventeenth century	150
9.5	Common uses of the apostrophe for optional letter omission in the seventeenth century	152
9.6	Lower case letter sequences recognised as ligatures in Johnson’s dictionary	154
11.1	Wells’s Standard Lexical Sets (Wells 1982: 123) updated to the transcription system used in the <i>Longman Pronunciation Dictionary</i> , 3rd edn.	177
17.1	The 14 British/American pairs agreed upon by all four dictionaries	275
17.2	The 31 <our, or> pairs	276
17.3	The eight <our, or> pairs agreed upon by all four dictionaries	276
17.4	The ten <our, or> pairs agreed upon by just three of the four dictionaries	276
17.5	British/American contrasts and pairs	277
17.6	Sixteen ostensible British/American pairs	288
17.7	The five of the 16 ostensible pairs agreed upon by all four dictionaries	288
17.8	The five of the 16 ostensible pairs agreed upon by W2 and W3	289
17.9	The six of the 16 ostensible pairs listed by W3 and not by W2	289
18.1	Older and Modern Scots Chronology (cf. Smith, 2012: 6)	297
18.2	Older Scots consonants (Bann and Corbett, 2015)	298
18.3	Older Scots Vowels	301

List of tables

18.4	Modern Scots consonants: diagnostic variants (Bann and Corbett, 2015)	303
18.5	Modern Scots vowels: diagnostic variants (Bann and Corbett, 2015)	304
19.1	Consonantal features in the language of the <i>Kildare Poems</i>	319
19.2	Phonological processes in the language of the <i>Kildare Poems</i>	320
19.3	Historical features of Irish English pronunciation	323
19.4	Attestations of historical features of Irish English	326
23.1	Examples of Japanese truncation	401
23.2	Japanese-made ‘English-like’ words	402
23.3	Japanese-peculiar English-like words and funny word choices	410
27.1	Nine features of digital writing, after Danet (2001)	477

# Contributors

---

**Jennifer Bann** has a PhD in Victorian ghost stories and spiritualism from the University of Stirling, UK. Until recently she worked at the University of Glasgow, UK, where her research interests included Scots language, digital humanities, medical humanities and corpus linguistics. She now works for the civil service.

**Vivian Cook** is Emeritus Professor of Applied Linguistics at Newcastle University, UK. His main interests are how people learn second languages and the written language of street signs. He is a founder of the European Second Language Association and co-editor of the journal *Writing Systems Research*.

**John Corbett** is Head of the Department of English at the University of Macau, China, and a senior research fellow at the University of Glasgow, UK. His research interests include Scots language and literature, digital humanities and intercultural language education.

**Florian Coulmas** is Professor of Japanese Society and Sociolinguistics at the Institute of East Asian Studies (IN-EAST) of the University of Duisburg-Essen, Germany. His scholarly publications are in the sociology of language and Japanese studies. He is currently working on a project about language policy.

**D. W. Cummings** is Emeritus Professor of English at Central Washington University, USA, whose major works in English orthography are *American English Spelling* (Johns Hopkins University Press, 1988) and the lexical databases and other material at [dwcummings.com](http://dwcummings.com).

**Philip Durkin** is Deputy Chief Editor of the *Oxford English Dictionary*, where he has led the dictionary's specialist team of etymologists since the late 1990s. He trained as a mediaevalist and historian of the English language at the University of Oxford, UK, where he completed a doctorate on previously unedited Middle English prose texts. His research focuses on etymology and language contact.

**Liory Fern-Pollak** is a lecturer in psychology in the School of Psychology, Social Work and Human Sciences at the University of West London, UK. Her research interests include neurolinguistics, dyslexia, biophysics and education.

**Raymond Hickey** is Professor of General Linguistics and Varieties of English at the Department of Anglophone Studies in the University of Duisberg-Essen, Germany. His main research interests are sociolinguistics, varieties of English, the eighteenth century, the Irish language as well as general questions of language contact, shift and change.

**Will Hill** is Deputy Head of Cambridge School of Art, UK. His work includes vernacular lettering, the design of experimental display typefaces and research on revival and historic reference in type design. It also reflects an ongoing interest in visual poetics, concrete poetry and the use of language in public art.

**Simon Horobin** is Professor of English Language and Literature and tutorial fellow at Magdalen College, Oxford, UK. His research interests are in the history of the English language, especially in the medieval period, most recently focusing on English spelling: its history, its role in society and its future.

**Jeffrey L. Kallen** is Associate Professor of Linguistics and Phonetics in Trinity College Dublin, Ireland. His main areas of research include the English language in Ireland, linguistic landscape, topics in semiotics, discourse analysis, dialectology, and corpus-based approaches to language variation.

**Nenagh Kemp** is a senior lecturer in psychology at the University of Tasmania, Australia. Her research focuses on spelling development in children and spelling strategies in adults, as well as on the written language of digital communication.

**Frank Kirchoff** is a PhD student of linguistics at the University of Cologne, Germany. His research focuses on punctuation systems, syntax and intonation.

**Jesper Kruse** teaches English language and literature at the University of Copenhagen, Denmark where he is a research assistant in the Department of English, Germanic and Romance Studies. His research interests include global accents of English, poetic metre, and typography in verse texts.

**Jackie Masterson** is Professor of Psychology of Reading in the Department of Psychology and Human Development at the UCL Institute of Education, London, UK. Her research interests are in literacy development and difficulties, cross-linguistic investigations of reading and spelling, and object and action naming.

**Roger Mitton** is a visiting research fellow in the Department of Computer Science and Information Systems at Birkbeck, University of London, UK. His research has included electronic dictionaries, text corpora and spellchecking.

**Terezinha Nunes** is Professor of Educational Studies in the Department of Education at Oxford University, UK. Her work spans the domains of children's literacy and numeracy, with

a special interest in educational applications. Her literacy research focuses on the connections between morphological awareness, spelling and vocabulary growth.

**Takeshi Okada** is Professor of Applied Linguistics at Graduate School of International Cultural Studies, Tohoku University, Japan. After working on a corpus-based analysis of Japanese-specific English spelling errors at Birkbeck College, University of London, UK, he is now leading a research group of blended e-learning system for Japanese EFL learners.

**Michael D. Picone** is Professor of French and Linguistics at the University of Alabama, USA. He is interested in lexicological, phonological and language-contact topics, as well as contemporary and historical profiles of language use in Louisiana, in Alabama and elsewhere.

**Beatrice Primus** is Professor of Linguistics at the University of Cologne, Germany, chief editor of *Written Language and Literacy* and co-editor of the monograph series *Linguistische Arbeiten* published by Mouton de Gruyter. Her major areas of research are syntax, semantics and writing systems.

**Susan J. Rickard Liow** is Director of the Speech and Language Pathology Programme at the National University of Singapore. Her research interests include cross-linguistic influences on normal and abnormal reading and spelling development in bilinguals.

**Iggy Roca** is Emeritus Professor in the Department of Language and Linguistics at the University of Essex. His main research activity lies in the areas of phonology and morphology. He has written two books on phonology, edited a number of collections, and published papers mainly on stress, syllables and grammatical gender.

**Des Ryan** is a final-year PhD student in Trinity College Dublin, Ireland. The title of the thesis is 'Principles of English spelling formation'. He is also interested in word-formation, James Joyce, creative spelling, and the mix of writing, graphic design and, potentially, sound and movement.

**Phil Scholfield** is a visiting fellow at the University of Essex, UK, in the sunset years of a long career in applied linguistics with an interest, amongst many other things, in the lexical history of English, lexicography, and writing systems of the world.

**Mark Sebba** is a reader in sociolinguistics and language contact at Lancaster University, UK. He has done research on pidgin and creole languages and on conversational code switching in bilingual communities. His most recent research has been about written bilingual and multilingual texts and the multilingual linguistic landscape.

**Tim Shortis** is a visiting research associate at the University of Bristol, UK. His PhD thesis examines diachronic change in linguistic and semiotic resources as exemplified by orthographic practices in texting and messaging. His current research focuses on the impact of digital media on literacy and pedagogy, multi-modal approaches to written representation, and poetry recitation

**Jeremy J. Smith** is Head of the School of Critical Studies at the University of Glasgow, UK. His research includes Old and Middle English language and literature, early modern English, book history, and the use of computers in teaching and research.

**Richard Sproat** is a research scientist at Google. His research interests include: accent prediction, articulatory phonetics, Chinese linguistics, computational morphology, computational modelling of language change, corpus-based methods, pronunciation modelling, syntax, text analysis/normalization, text-to-scene conversion, text-to-speech synthesis and writing systems.

**Lauren Squires** is an assistant professor in the Department of English at The Ohio State University, USA. She conducts research on language variation, computer-mediated communication, mass media, and sociolinguistic perception, with particular interest in notions of ‘standard English’.

**Rhona Stainthorp** is Professor in the Institute of Education at the University of Reading, UK. Her interests include the development of reading and writing, the effective teaching of reading and writing, individual differences in reading and writing, and children with learning difficulties.

**Merja Stenroos** is Professor of English Linguistics at the University of Stavanger, Norway. Her research work focuses on Middle English linguistic variation and literacy practices, and includes the compilation of the Middle English Grammar Corpus (MEG-C), 2011 (version 2011.0), and the Middle English Local Documents Corpus (MELD).

**Melvin J. Yap** obtained his PhD from Washington University, USA, in 2007 and since then has been based at the National University of Singapore. His research interests include visual word recognition and the influence of individual differences and semantics on lexical processing.

**Ishi Yasuko** received a doctoral degree in applied linguistics from the University of Essex, UK, in 2011 for a dissertation entitled ‘English silent letters and how English spelling should be: Modified spelling and people’s perception of this’. She currently lives in Japan.

**Valerie Yule** is a researcher in literacy and imagination, clinical child psychologist, academic, school psychologist and teacher working in disadvantaged schools and hospitals. She is concerned with research to make literacy easier by removing barriers, studies of children’s imagination and applied imagination, and social innovations.

# Acknowledgements

---

We would like to thank our amazing contributors who have taught us much by their writings and have worked so hard on their chapters.

Vivian Cook would also like to express his appreciation to generations of former Essex and Newcastle students, to the co-editors of *Writing Systems Research*, Jyotsna Vaid and Benedetta Bassetti, who have contributed greatly to his understanding of writing, and to his fellow editor Des Ryan, whose enthusiasm kept this massive project going. As always, the book would never have seen the light of day without the musical accompaniment of the Brotherhood of Breath, Keith Jarrett and Dr Michael White.

Des Ryan would like to thank the Irish Research Council for funding his doctoral thesis and allowing him to work on this project simultaneously. Thanks to Sharifa for her love, support and patience throughout. Thanks to Sean Farrell and Elske Rahill for their generous accommodation for the summer of 2013, and thanks to Aislinn Ryan (no relation) and others at Irlen Ireland, as he wouldn't have been able to do this kind of work without his Irlen glasses. Thanks to the backroom staff in Trinity College Dublin, especially Kieran Lewis and Annemarie Naughton, for helping him learn how to juggle several workloads while keeping sane. Thanks to Vivian himself for having the idea for this book and asking me to work with him. He has been a great role model and very supportive all along.

*This page intentionally left blank*

# Introduction

*Des Ryan*

---

Interest in the English writing system can be viewed as a reaction to the presence or absence of change. The recent burst of interest in the system has undoubtedly been a response to the changes, over the last generation, in how and where printed writing is produced. No longer is it regulated almost exclusively by the house styles of printers and publishers; it now exists in the deregulated written domains of the internet, computer-mediated communication, advertising and elsewhere. In short, the middleman has been removed from much writing, spelling has been freed up and a creative explosion has occurred, documented here by Lauren Squires (Chapter 27) and Tim Shortis (Chapter 28), who also shows how many seemingly new forms have been recycled from the margins of ‘vernacular’ spelling.

This level of change has not been seen since the introduction of print when the reverse was happening and the chaos of the early modern English writing system needed standardising. Phil Scholfield (Chapter 9) examines the process of standardisation which was largely complete when Dr Samuel Johnson’s 1755 dictionary became a standard reference for English spelling. This prescriptive dictionary has since been superseded by the *Oxford English Dictionary*, a historical and descriptive dictionary which is, ironically, often treated as a prescriptive standard reference, for British spelling at least. Philip Durkin (Chapter 10) sketches the enormous historical variation collected by the *OED*. American spelling took longer to standardise and D.W. Cummings (Chapter 17) explains how the variation between the two spelling systems is not so clear cut. Florian Coulmas (Chapter 16) argues that new spelling conventions were only possible in America due to its state of independence, unique among English-speaking countries in the nineteenth century. Merja Stenroos and Jeremy J. Smith (Chapter 8) also explain how social conditions affected the development of the English writing system, covering a massive span from Old English through to the emergence of print in the late mediaeval/early modern period. Technological conditions also play a role and Scholfield in turn examines the effect of printed forms on the writing system. This chapter complements Will Hill’s history of typography and its relation to English writing (Chapter 25), ends with the emergence of screen print and publicly available graphic design tools. Their availability has helped this explosion of creativity, which Jeffrey Kallen views through the lens of contemporary linguistic landscapes, public forms of writing where one

can see hybrid mixtures of written and pictorial forms (Chapter 22). Not since the time of mediaeval manuscripts has there been such cross-modal interaction in English writing.

The stability of English spelling during the time of print has allowed for some very strong traditions to emerge, most notably a tendency for total invariance in the spelling of (almost) every word. This is reflected in the ongoing need for spellcheckers, whose history is outlined by Roger Mitton (Chapter 29). And of course, children must learn how to spell. Nenagh Kemp (Chapter 12) and Terezinha Nunes (Chapter 13) both discuss the acquisition of the English writing system by L1 learners but from two opposing perspectives. Kemp looks at how children develop their ability to spell while Nunes looks at how spelling is taught, comparing the policies and practices of England and New Zealand. Rhona Stainthorp (Chapter 15), on the other hand, looks at the changing contexts in which teaching actually takes places, focusing on classrooms in England, while Liory Fern Pollak and Jackie Masterson (Chapter 14) focus on the challenges of teaching the writing system to people with various kinds of dyslexia. Melvin J. Yap and Susan J. Rickard Liow (Chapter 26) discuss how the written word is visually recognised and processed by the human brain.

Stability gives way to inertia, and the lack of internal consistency across the writing system has prompted several waves of people to try and reform the English writing system. While this agenda is less fashionable nowadays, the need for change is still there, argues long-time advocate Valerie Yule, alongside Ishi Yasuko (Chapter 24). A partial solution to the need for reform is that people may in fact change their speech to fit the spelling, slowly, over generations. An implicit counter-argument comes from Jesper Kruse (Chapter 11) who discusses the vast range of phonological changes which have occurred throughout English since its spelling stabilised. Specifically, his focus is on cases where the spelling represents obsolete phonological contrasts that only survive in a few accents (e.g. *wait* and *weight* etc). While these curiosities are fascinating, their obsolescence suggests that changing phonology remains a more powerful force on people's speech than the conservativeness of spelling. And of course the more the phonology changes, the more need there is for spelling reform...

Iggy Roca in Chapter 5 analyses phonological information which is not included in spelling and must be supplied by the reader. My own chapter (Chapter 4) is its counterpart, outlining the different kinds of information that are included in English spellings. Not only are letters matched to sounds, as one might expect, but they also try to keep the spelling of related words constant, as much as possible. This often includes keeping the spelling of borrowed words and these are examined in greater detail by Simon Horobin (Chapter 7). In Chapter 6, Frank Kirchoff and Beatrice Primus complete the theoretical section of the book, providing a new approach to the core of English punctuation. This chapter also provides a useful starting point for those interested in the different approaches to writing systems taken in Germany in recent decades.

Finally, the relative stability of English spelling in recent centuries has had a series of parallel and alternative histories in the spelling of dialects. Foremost among these is the spelling of Scots whose strong orthographic history is viewed here, by Jennifer Bann and John Corbett in Chapter 18, as an ongoing dialogue between its own traditions and the need to adapt and reform while under the dominating influence of standard English. Raymond Hickey (Chapter 19) presents the stunted history of the spelling of Irish English, first as a mediaeval dialect and later in the staged portrayal of Irish accents in theatre and fiction. Local speech has also developed orthographic traditions throughout North America, and Michael Picone (Chapter 26) focuses on the comic traditions of Louisiana. He discusses many of the problems of dialect representation but also the benefit to linguists in triangulating dialect spelling against putative historical phonetics.

A particularly revealing chapter about the nature of the English writing system is Mark Sebba's discussion (Chapter 21) of the emergent spelling of Jamaican creole. Like English in the late Middle Ages, Jamaican creole is spelt as a compromise between representing local pronunciation and representing the English origins of words. This tension between sound and meaning is at the heart of all writing systems, as can be seen from the typology of writing systems presented by Richard Sproat (Chapter 3) who allows us to view how English writing compares to other languages. One example of this can be seen in Takeshi Okada's chapter on the relationships between Japanese and English writing in Japan (Chapter 23).

For those readers new to some, if not all, areas of the English writing system, Vivian Cook (Chapter 2) provides an introductory overview of the central ideas and themes that contributors assume that they are familiar with. Writing systems research is a growing field emerging from the shadow of linguistics, a discipline which has concentrated on speech for a whole century. There is plenty more work to be done.

*This page intentionally left blank*

# Background to the English writing system

Vivian Cook

---

This book brings together contributions on many different aspects of the English writing system, not only the state of the art but also growth points for the future. The overall scope of the English writing system is vast; the 27 contributions here inevitably represent only a selection of the many aspects it encompasses. The book demonstrates the variety and heterogeneity of the English writing system, going from Old English to Twitter, from children's acquisition of spelling to dialect spelling in Louisiana and from theoretical phonology to Dublin street signs. Consequently each chapter may cater for very different readers ranging from psychologists to linguists to teachers.

This chapter introduces some of the background issues in the study of the English writing system. It is intended as a brief overview, looking at some of the basic ideas that contributors assume readers are familiar with. In addition it draws attention to some aspects which have not been treated at length in our contributions, such as the material nature of writing.

## General issues of writing systems and English

### *The relationship of spoken and written language*

Linguists have typically maintained that written language is a representation of spoken language, seen in such classic statements as de Saussure's 'Language and writing are two distinct systems of signs; the second exists for the sole purpose of representing the first' (de Saussure 1916, trans. Baskin 1960: 23). If writing is completely dependent on speech, there is little reason for it to receive attention in its own right.

From its origins, written language has, however, been far more than speech written down. According to Samuel Butler (1908):

The written symbol extends infinitely, as regards time and space, the range within which one mind can communicate with another; it gives the writer's mind a life limited by the duration of ink, paper, and readers, as against that of his flesh and blood body.

While many linguists emphasise the primacy of speech (Daniels 1996), others today claim that speech and writing exist largely in their own right rather than one being subordinated to the other; ‘The sound system and the writing system are the two modes of *expression* by which the lexicogrammar of a language is represented’ (Halliday and Mattheisen 2013: 7). Speech can indeed be written down; writing can be read aloud. But speech is only written down for specialised purposes, like court transcripts; writing is read aloud in a few limited situations, like newsreading from teleprompters. The writing system is an alternative form of representation for the language, not the language itself, and is not subservient to the phonology, though related to it. The important relationship is between writing and language, not between writing and speech.

Crucial differences between speech and writing arise from their dissimilar engagement with time. Unless recorded, speech is gone the moment after it is said; writing is available effectively until it physically disintegrates or disappears. Among other things, writing is an external physical memory system for recording and for planning, to avoid the vagaries and limitations of our internal memory systems, exemplified par excellence in diaries and desktop calendars. As St Augustine (397, II, 5) put it, ‘because words pass away as soon as they strike upon the air, and last no longer than their sound, men have by means of letters formed signs of words’.

### ***Types of writing system***

According to Perfetti (1999: 168), a writing system ‘determines in a general way how written units connect with units of language’. The same language can be represented by more than one writing system: English can be represented in Braille, shorthand or written Morse code. The complementary terms to ‘writing system’ are ‘script’ and ‘orthography’. A script is the actual physical symbols of the writing system, for instance Roman or Cyrillic alphabets; an orthography is the rules for using a script in a particular writing system, that is to say how the symbols spell out words etc. However, these terms vary considerably in meaning between writers. Sproat (2000: 25) for example uses ‘the terms “orthography” and “writing system” interchangeably’.

The classification of writing systems has a long and chequered history; for recent accounts see Borgwaldt and Joyce (2013) and Rogers (2005). The main issue is how to reconcile the two levels of language that written symbols correspond to, on the one hand items in the lexicon, whether words or morphemes, called morphographic or meaning-based systems, on the other hand sounds in the phonology, whether syllables (Japanese kana), phonemes (Italian) or consonant phonemes only (Arabic), called phonographic or sound-based systems.

Lexical correspondences for individual symbols occur only on the fringes of the English system. On a standard keyboard, < £ \$ % & @ # > all have particular meanings corresponding to English words or phrases, *pound, dollar, per cent, ampersand, at, hash*. (Conventionally angle brackets < > are used to enclose written forms, just as strokes // and square brackets [ ] enclose spoken forms, though this usage is by no means standard to all writing system researchers and does not suit all the chapters in this volume.) None of these symbols correspond directly to the pronunciation: you either know that < & > corresponds to *ampersand* /æmpəsænd/ or you don’t. As we will see, English also involves lexical correspondences for individual words such as *yacht* and *of*, which cannot be handled as straightforward phonological correspondences.

English text messaging provides some examples of phonological correspondence at the syllable level: in *C u 4 t* <C>, <u>, <4> and <t> correspond to spoken syllables rather than individual sounds. However most phonological correspondences in English connect letters or combinations of letters to distinctive sounds of the language, i.e. phonemes.

In the seminal paper by Katz and Frost (1992: 71), a writing system in which each symbol corresponds to a particular sound of the language, and, vice versa, each sound corresponds to a symbol, is called ‘transparent’ or ‘shallow’. In transparent systems such as Turkish, contrasted with English in literacy education in Chapter 15, every letter can be read aloud and every sound can be spelled according to a set of one-to-one correspondences between letters and sounds. Transparency is relative rather than absolute – Turkish is more transparent than English – and transparency can be estimated in various ways (Neef and Balestra 2013).

Full transparency implies that the relationship between sounds and letters is isomorphic: one letter one sound, one sound one letter. English is a long way from meeting this ideal; as William Salesbury already pointed out in 1547, ‘You cannot fail to know that in English they do not read and pronounce every word literally and fully as it is written’. Two letters may correspond to a single sound, <sh> to /ʃ/ in *sharp*; three letters to one sound, <pph> to /f/ in *sapphire*; or a single letter to two sounds, <x> to /ks/ in *tax*. Some so-called ‘silent’ letters have no direct link to sounds in particular words, say the <h> in *hour* or the <l> in *salmon* compared to those in *hour* /hʊər/ and *salmonella* /sælməneɪl/.

Conversely a writing system ‘in which the letter-phoneme relation is substantially equivocal’ is called ‘opaque’ or ‘deep’ (Katz and Frost 1992: 71). It would be impossible to work out the pronunciation of the following words through correspondence rules: *hour* /aʊəl/, *Leicester* /lestəl/, *lieutenant* (in British English /leftənənt/), *ptarmigan* /tɑ:mɪgən/, *colonel* /kənəl/, *veille* /rɪvæl/ and *hiccough* /hɪkʌp/ (though, since 1950, more likely to be spelled *hiccup*, according to Ngram Viewer).

Rather than one-to-one correspondences, English has many alternatives in both directions. The single vowel letter <a> for example corresponds to at least five different English sounds: /æ/ *bad*, /ɑ:/ *bath*, /ə/ *about*, /e/ *many*, /ɒ/ *cauliflower*. In reverse the spoken diphthong /eɪ/ corresponds to twelve different spellings: <a> *lake*, <au> *gauge*, <et> *ballet*, <ai> *aid*, <er> *foyer*, <ay> *stay*, <é> *café*, <ea> *steak*, <eigh> *weigh*, <ée> *matinée*, <ae> *sundae*, <ey> *they*.

While the English writing system is mainly phonological, it is far from transparent. If the goal of a writing system is to represent the sheer sounds of the language as faithfully as possible then the English system is highly inefficient. But, if written language represents language rather than speech, English may be representing other levels of language than the sounds. The <g>s in *sign* and *malign* for instance seem redundant, unnecessary ‘silent’ letters – if we are looking for a direct sound correspondence. But these <g>s connect to families of words in which the <g> is not silent, such as *signify* and *malignant*. Without the <g>, these underlying links in our minds would not have been activated, supporting the concept of spelling representing something deeper than surface phonology, one of the threads that runs through the chapters of this book.

### ***Dual routes in reading***

So far the English writing system has been treated in terms of language as ‘an abstract external entity’, as described in grammars and dictionaries (Cook 2010): ‘the English language’. But the writing system is also part of language in the senses of ‘the possession of a community’ and of ‘the knowledge in the mind of an individual’, in which the writing system is not an external entity but an internal system in the mind of the user. Using the writing system involves processing written and spoken information and relying on memory processes to retain the information for shorter or longer periods.

The continuum between lexical and phonological writing systems outlined above parallels two ways of processing the text, visually for meaning and phonologically for sound:

Reading theorists have reached unanimity concerning the existence in the human reading system of two separate procedures for reading aloud – that is, dual routes from print to speech.

(Coltheart 2005: 23)

Different models have now been developed within this broad dual-route architecture, including connectionist models such as Seidenberg and McLelland (1989) and the Dual Route Cascaded Model (Coltheart *et al.* 2001).

Following the lexical route means seeing the word *blossom*, recognising it visually as a whole word <blossom>, finding its pronunciation /blɒsəm/ in a mental lexicon along with its meaning and then saying it aloud. Following the phonological route means seeing a word such as *blossom*, recognising the letters <b l o s s o m>, converting them into the corresponding phonemes /blɒsəm/, and then saying them aloud: the meaning is available by matching in the mental lexicon but is not necessary for reading aloud.

The two routes thus involve alternative ways of processing text. Following the lexical route a reader can recognise the word *through*, and look it up in their mental dictionary as a whole to retrieve the meaning without knowing its pronunciation. The phonological route, however, involves ‘assembling phonology from a word’s component letters’ (Katz and Frost 1992: 71). A reader using the phonological route can recognise the letters of *salad* as <salad> and apply the correspondence rules to get /sæləd/, without knowing its meaning.

Both routes are used by readers of English, some words being processed entirely through the phonological route, some through the lexical. The advantage of the phonological route is that it can apply to unknown words. English newsreaders demonstrate this when they read foreign politicians’ names aloud, say Ukraine’s energy minister, *Volodymyr Demchyshyn*, China’s president, *Xi Jinping*, or the Zulu king, *Goodwill Zwelithini kaBhekuzulu*. Any reader has the ability to deal with non-words, whether conforming to English spelling, like *broave* or *shart*, or not, like *\*qish* or *\*rawh* (asterisk indicates impossible spelling in English). Companies are confident that people can pronounce their non-English names, *Aviva*, *Skandia*, *AXA*, *Ansva* and *Kwelm*, to take just insurance groups.

Some English words like *colonel* and *Wednesday* are necessarily processed through the lexical route since their spoken correspondences are virtually unique, namely /kə:nəl/ and /wenzdi/. The spellings of these words have to be learned individually, not through letter to sound correspondence rules. Seidenberg (1992) claims that the lexical route is used for the most frequent 200 words of English, dense with function words like *the*, *where* and *of* – the only word in which <f> corresponds to /v/.

Those who believe that the only valid route is phonological exclaim at the sheer number of exceptions to correspondence rules. Those who believe that English readers use both routes regard these exceptions as separate lexical entries, not dauntingly large in number compared to the 1,945 characters Japanese children have to learn at primary school or the 40,000 or so in a dictionary of ‘traditional’ Chinese. Methods for teaching reading tend to favour one route or the other, whether the ‘look and say’ method’s emphasis on the lexical route or the ‘phonics’ method’s concentration on the phonological route. Chapters 15 and 25 both draw attention to the need for children to be taught both routes. Different forms of dyslexia are indeed associated with one or the other route; some children who have initially mastered the phonological route find it impossible to go on to the lexical route, vital for efficient silent reading (Frith 1985). Aphasia too can affect either route (Funnell 1983).

## ***The phoneme and the writing system***

Most writing research has discussed phonology in terms of phonemes – the minimum sound units that distinguish one word from another, say /tent/ *tent* distinguished from /dent/ *dent* by the /t~d/ contrast, or *red* /red/ from *rod* /rɒd/ by the /e~ɒ/ contrast. In this view, speech consists of strings of discrete contrasting phonemes, rather than being a continuous stream of sound: /skim/ is a temporal sequence of four phonemes /s/, /k/, /ɪ/ and /m/, parallel to the visual sequence of four letters in <skim>: both speech and writing can be chopped up into discrete contrasting segments that occur one after the other, whether phonemes or letters.

Treating speech as a string of phonemes does not, however, account for the discontinuous elements that occur in speech or writing, called ‘split digraphs’ in Brooks (2015: 6). The difference in pronunciation between <note> /nəʊt/ and <not> /nɒt/ is shown by the final <e>, which has no direct sound correspondence; <o\_e> can be analysed as a single unit split by an intervening consonant <t>, with the correspondence /eɪ/, as Albrow (1972) argues, and the same applies to <a\_e> <hat/hate>, <e\_e> <gen/gene>, <i\_e> <writ/write>, and <u\_e> <cut/cute>, to which Brooks (2015: 432) adds <y\_e> <byte>. This is sometimes called ‘The fairy e rule’ – ‘Fairy E waves her wand and makes the vowel in front say its name’, an often used rule of thumb in primary schools – or, as Hart (1569: 33) perhaps first put it, ‘for the quantitie of the preceeding vowel’.

The differences between phonemes come down to the distinctive features that make them up, such as voice, which distinguishes voiced /d/ *dent* from unvoiced /t/ *tent*, or continuant, which inter alia distinguishes plosive /t/ *tin* from fricative /θ/ *thin*, or lateral, which distinguishes /l/ *lip* from /r/ *rip*. ‘These distinctive features occur in lumps or bundles, each of which we call a phoneme’ (Bloomfield 1933: 79); a table of distinctive features for English consonants is given in Chapter 5. Writing systems research cannot confine itself to the phoneme, indispensable as the term may be as an overall label.

## **The properties of the English writing system**

Any writing system has many elements. While most discussion of the English writing system concentrates on the letter–sound correspondences, these are only part of the whole system.

### ***Direction***

English text is normally read from left-to-right in rows from top-to-bottom of the page, unlike Arabic (right-to-left) or traditional Japanese (columns). In some circumstances, English is written in columns top-to-bottom, or sideways top-to-bottom. English books are read by turning pages from right-to-left, while the reverse happens in Japanese, traditionally written vertically; the sequence of reading speech balloons in English comics is left-to-right, in Japanese the opposite. English letters also face in a particular direction (Treiman and Kessler 2013): <b> is not <p> (top-bottom inversion) and <d> is not <b> (left-right inversion), a problem for some dyslexics.

### ***Script and letters***

English now uses a Roman alphabet of 26 letters, in lowercase and capitals, plus italic lowercase and capitals, called by Gill (1931) different alphabets. The differences between these forms of the alphabet form a useful resource for the English writing system: a capital letter may mark



Figure 2.1 Modern inscription in Newcastle upon Tyne

a grammatical difference between a proper name and a common noun <Hall> versus <hall>; italics may show emphasis <He's *the* expert on glottochronology>; and so on. Historically the forms of the alphabet were used separately, i.e. for complete texts rather than combined for contrastive purposes, as described in Tschichold (1928: 79). The range of forms has been amplified comparatively recently by the addition of bold and small caps (Bringhurst 2005), all now available through word processing programs rather than restricted to typesetters. Chapter 25 discusses how letters are used in printed texts.

Capital letters are written within virtual squares, easily discernible in say <NEWCASTLE> in the Times New Roman typeface or indeed <NEWCASTLE> in the Keystroke typeface. Printed English since Victorian times has been heavily influenced by the classical Roman letters carved in stone on inscriptions such as Trajan's Column in AD 113 rather than other forms of Roman letter (Gray 1960). The modern inscription in Figure 2.1 brings out their use of serifs: 'The broadening of triangular forms at the terminals of letters' (Hill 2010: 186) – seen at the tips of the strokes at the bottom and top strokes of <A> and <T> etc – and the varying width of the line, as seen in the <S> and <N>.

To many the capital letter is somehow the prototypical form of a letter: official forms demand to be 'printed' in capitals or 'block' capitals; amateur handwritten notices tend to be all capitals; modern text art chiefly uses capitals, as seen in *Art and Text* (Beech, Harrison and Hill 2009), as do balloons in strip cartoons. Yet, if anything, capitals are harder to read than lowercase as their square shapes make them more difficult to differentiate. The letters for UK motorway signs for instance were based on research that demonstrated the most legible signs from a speeding car combined capitals and lowercase (Kinneir 1980).

The lowercase or 'minuscule' letters on the other hand have ascenders above the line as in <d> and <b> and descenders below the line as in <p> and <y>. They are derived from cursive letters written quickly by hand with brushes or quills rather than laboriously carved on stone with chisels. While capitals are sometimes called 'big' and lowercase 'little', size is relevant only to the few lowercase letters that have similar shapes to their capital versions, say <c/C j/J o/O s/S x/X>. The shapes of most capital/lowercase pairs are quite distinct, as in <a/A h/H q/Q t/T>.

Italic letters have a distinctive slant < *a l p v f g k n* >; the italic letter < *a* > has a closed form compared to regular open <a> in most typefaces, although not all. Modern uses of italics are for emphasis <I do *not* believe in ectoplasm>, for stage directions in plays <*Exit, pursued by a bear*>, for citing book titles <*The History of The Decline and Fall of the Roman Empire*>, and so on.

At the start of printing in Europe in the fifteenth century, printed letters were derived from handwritten forms; Chapter 25 describes the complex relationship between the two ever since. Printed texts are produced by machine in as many copies as are needed – they are reproducible; lettering is produced by the individual's hand, usually as a single copy. According to the type designer Fred Smeijers (2011: 19), 'There are just three kinds of letters: written, drawn or lettered'. Writing produced by a brush or pen is individually done by hand and is effectively



## BOX 2.1 ENGLISH CORRESPONDENCE UNITS FOR WRITTEN CONSONANTS

### *Relational units for English consonants (Venezky 1970: 54)*

#### *Major units*

b c ch ck d dg f g gh h j k l m n p ph q r rh s sh t tch th u v x w wh xs y z

#### *Minor units*

kh sch gn

### *English consonant graphemes (Brooks 2015: 255–257)*

#### *Main system*

b bb c ce ch ci ck d dd dg dge ed ew f ff g ge gg h i j k l le ll m mm n ng nn p ph pp q  
r rr s se sh si ss ssi t tch th ti tt v ve u ue w wh x y z z z

#### *Others*

bh bd bp bt bu bv + 189 more

in turn subdivides into a nucleus /æ/ and a coda /g/. Children are believed to acquire the spelling of onset and rime separately, and then to separate the letters corresponding to the rime gradually: they first learn say *bag* as <b> and <ag> and then separate <ag> into <a> and <g> (Goswami 1999). Children often acquire a consonant cluster as a unit rather than as separate phonemes, spelling *street* as *set* and *screams* as *sceem* (Treiman and Kessler 2014).

As a fairly opaque writing system, English has complex rules for linking speech to writing and vice versa, as much of the rest of this book demonstrates. Carney (1994) for example details thirteen correspondence rules for the letter <g> with /g/ *guide*, /dʒ/ *contagious*, etc, 41 correspondence rules for the letter <o> with /əʊ/ *goat*, /ɔ:/ *floor*, etc. As the English language has never had a body to dictate what the forms of the language should be, unlike say l'Académie française for French, English spelling 'rules' are for linguists descriptions of how words behave in English. They emerge out of our writing because that is how English writing works, not because some authority tells us what to do. This does not prevent popular discussion taking the rules of spelling to be commandments engraved on stone by an unnameable, unchallengeable authority.

### *The principle of invariance*

One implicit assumption about the modern English writing system is that a word is always spelled in the same way, regardless of its sound correspondence: *scissors* has to be spelled <scissors> not <sizerz>, even if the latter corresponds more accurately to its pronunciation /sɪzəz/. A written word is seen as fixed and unchanging. A limited dispensation from invariance is afforded to proper-names, as in *Vivian*, *Jyvyan*, *Vivien*, and *Vivienne* (with a gender difference

between the first two and last two in British English); the possessor of a name can insist on how it is spelled or said, say *Keynes* /keɪnz/ for the economist or *Menzies* /mɪŋɪs/ or *C.J. Cherryh* with final silent <h> for the novelist.

This insistence on invariance is comparatively new in English, and is often at odds with consistent letter to sound correspondence rules. When a language is spoken with multiple accents, the spelling cannot both reflect how a word is said and always be the same. Middle English is famous for its variable spellings, not only across dialects but also within the writing of the same individual; much the same was true of Older Scots, as discussed in Chapter 18. Looking at citations for *scissors* in the *Oxford English Dictionary* (2015), in the fifteenth century the spellings included *cysars*, *cysurs*, *cysour*, *cisours*, *sesours*, *sisours*, *sisoures*, *scissors*, and *sysowre*; there were around ten variant spellings per century until 1700; even the twentieth century had five: *cissers*, *cithors*, *scissors*, *sisors*, *sizzers*. The modern spelling <scissors>, though first found in 1484, was only one of the variants for many centuries; between about 1750 and 1820 *scissars* was the most popular form according to Google Ngram Viewer. The <sc> spelling is one of the many examples of English erroneously adopting a spelling based on Latin, in this case treating it as coming from *scissor* ‘a cutter of cloth’ rather than from *cisoria* ‘cutting instrument’. Other historically inaccurate re-spellings include the <s> of *island* added to Middle English *iland/éaland* on the belief that it was derived from French *île* (isle) rather than from an Old English word for ‘water’ *eag* still seen in the <ey> of *Anglesey* and *ait*.

At some point then English spelling fixed on the spelling of individual words, rather than relying on general sound/letter correspondences, probably through the word-based attempts by the great dictionaries of Johnson (1755), Webster (1828) and others to lay down a fixed form of the language. Once the spelling of a word is set, any deviation is a mistake and a solecism, even when the spoken form is obvious. Most popular discussion of spelling concerns invariance: using anything but the accepted spelling of a word is a sign of lack of education and carelessness and a betrayal of the English language, according to many highly literate English people: ‘Spelling is one of the outward and visible marks of a disciplined mind’ (Kilpatrick 1988). Competitions like the annual Scripps National Spelling Bee in the USA concentrate on the invariant spelling of infrequent words like *cypseline*, *pyrrhuloxia* and *scherenschnitte* (words in the 2015 competition, absent from the 100 million running words of the British National Corpus). It is very dangerous to your social prestige and employment prospects to spell *paid* as <payed>, to confuse *compliment* and *complement*, or *their*, *there* and *they’re*, or to forget how many <c>s and <m>s there are in *accommodate*. British newspapers attacked the then prime minister Gordon Brown in 2009 for sending a handwritten letter of condolence to a dead soldier’s mother that she felt was disrespectful because he misspelled *Mrs Janes* as <Mrs James>, *greatest* as <greatst>, *your* as <you> and *colleagues* as <colleagus>, possibly more due to his poor eyesight and illegible handwriting than lack of respect. The extent of the problem that English-speaking children have with invariant spelling can be gauged from Peters (1970) test of 967 ten-year-olds’ spelling of *saucer*; only 47 per cent were right; the most popular mistake was *sauser*, followed by *sorser*, *suacer* and *sacer*, and so on down to 126 one-off spellings such as *scorceri* and *suarser*; Chapter 29 discusses children’s misspellings of *scissors*.

The importance of spelling for teaching children to read and write is emphasised in Chapter 12. Spelling words correctly is a vital part of literacy education; children will be marked down at examinations and when submitting applications for college or jobs if they make many spelling mistakes. Indeed in 2016 a British schoolboy was investigated by the police when he wrote in his school homework that he lived in a <terrorist house> rather than a <terraced house>. According to Kreiner et al (2002), a writer who makes more than two per

cent of spelling errors seems poor or unintelligent. Nevertheless using alternative spellings to the usual letter/sound correspondence rules is very much a feature of English, as discussed in Chapter 4.

### **Orthographic regularities**

One of the principles of English spelling proposed by Venezky (1999: 6) is ‘Letter distribution is capriciously limited’. Orthographic units are not free to occur anywhere in the word or syllable. For instance <k> and <ck> both correspond to the sound /k/ yet <ck> only occurs at the end of syllables as in *back* and *tick*, never at the beginning; there are no English words \**ckab* or \**ckit*, though these are perfectly pronounceable as /kæb/ *cab* and /kit/ *kit*.

Such rules are called variously ‘orthographic constraints’ (Treiman 1993) and ‘orthographic regularities’ (Cook 2004a). Correspondence units that start words but do not end them include: <wr> *write*, <wh> *whom*, <j> *jug* and <rh> *rhesus*. Those that finish words but do not start them include <tch> *match* and <ng> *ring*. Double consonants occur freely both within words, *officer*, *adder* and at the end of words, *gruff*, *odd*, but cannot occur at the beginning, \**ffame*, \**ddont*.

An efficient user of the English writing system must know not only the standard spelling correspondences and the particular spellings of many individual words but also the orthographic regularities about where letters may occur. Treiman (1993) found that the eight-year-olds she tested had already substantially mastered these regularities, being able to tell that *beff* is a possible English word but \**ffeb* is not, though both are equally acceptable phonologically. It is interesting just how odd the words look that break these rules: \**ckall* is unambiguously /kɔ:l/ *call*, \**dgell* /dʒel/ *gel*, \**farh* /fa:/ *far*, yet they look completely strange and unEnglish.

### **The spelling of inflectional morphemes**

The past tense ending in regular English verbs is nowadays typically spelled <ed> but has three spoken forms: /t/ *looked*, /d/ *opened* and /ɪd/ *waded*. The inflection <ed> conveys the meaning of ‘past’ but does not correspond to the actual pronunciation, which is predictable from its phonetic context. The exact sound/letter correspondences are neglected in favour of a morpheme correspondence.

So the <ed> in content words is different from the <ed> in verbal inflections; the adjectives *learned* /lɜ:nɪd/ and *blessed* /blesɪd/ for example contrast with the past tense verbs *learned* /lɜ:nɪt/ or /lɜ:nd/ and *blessed* /blest/. Before the eighteenth century, <ed> tended to be used for stressed adjectives, as in the historically related form <learnèd>. Since about 1700 the spelling <ed> in inflections has not corresponded unambiguously to a single pronunciation (see Cook 2004b for an account of the different past tense rules for speech and writing). These morpheme-related aspects of spelling support the argument that English graphemes connect to the English language at different levels of language.

The possessive inflection <s> shows another relationship in that the apostrophe has traditionally distinguished between plural <s> and possessive <’s>, as in *tigers/tiger’s*, and between singular and plural possession through position, *tigers’/tiger’s*. However in many public texts today <’s> is a common way of showing plural, known as the greengrocer’s apostrophe. A street sign advertises tours of *Colchester’s Church’s*; a university notice directs one to the *LECTURE THEATRE’S*; a newspaper headline proclaims *DRIVER’S SMASHED UP MY FLAT* – to the horror of those who regard the apostrophe as the last line of defence against the barbarians at the gate. Chapter 9 discusses the history of the apostrophe.

## The accents used as correspondence norms in the English writing system

A system that combines sound/letter correspondences with invariant spellings has to select a single reference point out of the many different English accents across the globe. British English has traditionally been described in terms of Received Pronunciation (RP): ‘the regionless upper-class and upper-middle-class accent of British – mainly English – English’ (Trudgill 2003: 114). Cruttenden (2014) has adopted General British (GB) to include other countries than England, allowing for variation under the heading of Regional General British. Kruse in Chapter 11 here uses Southern British Standard (SBS) and North American Standard (NAS). The norm for discussing letter/sound correspondences in spelling research in England has mostly been taken to be SBS/RP/GB, as in say Carney (1994) and Brooks (2015), not very different from Puttenham’s (1589: 144) choice of ‘the vsuall speach of the Court, and that of London and the shires lying about London within lx miles, and not much aboute’. For the USA, Venezky (1970) relies on the General American (GenAm) accent of Kenyon and Knott (1951), which ‘corresponds to the layman’s perception of an American accent without marked regional characteristics’ (Wells 1982: 470).

Choosing one accent as the correspondence standard creates the problem that any chosen accent is not used by all speakers. Trudgill claims that approximately three per cent of British speakers speak RP (Trudgill 2001), a tiny minority in the UK, compared say to the Northern accents used by half the population of England (Wells 1982). RP is furthermore a status accent, the one to which the speech of politicians and newsreaders tends to gravitate. In a survey, the accent called Queen’s English, a lay equivalent to RP, was evaluated highest of 34 accents for prestige (Coupland and Bishop 2007).

The choice of a correspondence standard is then a social decision, as Coulmas points out in Chapter 16, rather than one based on considerations about how many speakers use it, its general comprehensibility or ease of learning etc. Choosing an accent such as RP or GenAm ignores the accents of most native users of English, whether Brooklyn, Geordie or Toronto, and indeed the multifarious accents of second-language users. The present book is rich with varieties of English such as Scots (Chapter 18) and Irish (Chapter 19), as well as English used by speakers of other languages (Chapters 21 and 23).

One kind of variation in English spelling is captured as American versus British, say – American style first – *feces/faeces*, *plow/plough* and *traveler/traveller*. Cummings (1988: 26) points out that ‘the differences between American and, say, British English spelling are quite modest’, amounting to a few hundred words. His Chapter 17 here shows how American and British spelling differ not just from one another but from dictionary to dictionary. These two main varieties broadly extend to the rest of the English-using world. Take the word *labour* in on-line English language newspapers around the world. Hardly surprisingly, British-style *labour* is found in Canada, Thailand, New Zealand, India and Nigeria; American-style *labor* in Israel, Korea, Singapore, Japan and the Philippines. Australia distinguishes *labour* from the *Labor Party*; the *Australian Government Style Manual* generally recommends ‘what is often thought of as British rather than American practice’ (Peters and Delbridge 1989: 129). Despite this, *color* appears in Australian newspapers two and a half times as often as *colour* (Peters and Delbridge 1989).

Some writers try to convey regional dialect accent through spelling, illustrated in the box below; people assert their local identity in print. Showing dialect through non-standard spelling is nevertheless a double-edged weapon. Readers with other accents may struggle to get through a few lines of Geordie poetry or *Uncle Remus*. Dialect speakers may be pleased

## BOX 2.2 SOME DIALECT SPELLINGS

*Edinburgh*: One ay the things thit concerned us maist wis the fact thit ye couldnae really relax in his company, especially if he'd hud a bevvy.

*Irvine Welsh*

*South Florida*: You ain't been used tuh knockin' round and doin' fuh yo'self, Mis' Starks. You been well taken keer of, you needs a man.

*Zora Neale Hurston*

*Geordie (Newcastle upon Tyne)*:

Me nyem it's Billy Oliver,  
Iv Benwell Town aw dwell;  
And aw's a cliver chep, aw's shure,  
Tho' aw de say'd mysel.

*Bill Oliver's Ramble, 1842*

to see their accent reflected in speech: in *Nippers*, a series of readers for children (1968–1974), Leila Berg tried to cover all the variations of the child's word for 'mother', *mum*, *mummy*, *mom*, *mam*, etc. Or they may feel stereotyped as yokels who cannot spell properly. Jaffe (2000) shows the delicate balance in the representation of black American English where the more accurate the portrayal of the accent in writing, the more it stigmatises the speaker as uneducated.

Showing people's actual pronunciation through non-standard spelling needs to be distinguished from the traditional literary convention through which novelists can indicate a dialect accent by using non-standard spellings that correspond to standard speech, called 'eye-dialect', as argued in Chapter 20. To take some eye-dialect examples: unstressed vowels can be shown with <er> for /ə/, as in <fer> for <for>, <ter> for <to> and <yer> for <you>; alternative non-standard spellings can be given for the standard pronunciation as in <wot> for <what>, /wɒt/ in both cases, <luv> for <love> /lʌv/, <mister> for <Mr> /mɪstə/. These reflect the typical RP pronunciation, but not the accepted spelling. A novelist can label a character as dialectal, uneducated or uncouth through eye-dialect without showing any actual difference in accent from the standard.

The invariance of word spelling creates a dilemma for those spelling reformers who want to make spelling better reflect pronunciation. If this means choosing one 'standard' out of all the accents available, much of the spelling will be opaque for many readers. If it means having different spellings for each dialect, a reader could only read texts with ease that were written in their own dialect. And, if it means adhering to the pronunciation at one moment of time, Dr Johnson (1755) pointed out:

some have endeavoured to accommodate orthography better to the pronunciation, without considering that this is to measure by a shadow, to take that for a model or standard which is changing while they apply it.

Thus spelling would have to keep up with changes in pronunciation. In both cases many people would be disadvantaged. The argument for spelling reform in Chapter 24 in fact favours simplification rather than adaptation to pronunciation.

The accent chosen also has implications for the teaching of reading. Perhaps the majority of English-speaking children are taught letter/sound correspondences that are not based on their own accent, nor always those of their teachers. Correspondences based on RP force them in effect to learn the phonology of another dialect. Essex children have spellings such as <wo> for <wall> and <fevr> for <feather>, revealing their local realisation of final /l/ as a vowel /u:/ and of /ð/ as /v/ (Bromley 2002). Children who speak the Hoosier dialect in Indiana spell <when> as <win> and <pen> as <pin>, showing their local /i/ pronunciation (Treiman 1993). In RP, words like *muffin* and *rocket* have /ɪ/ in the second syllable, in Australian English /ə/; Australian children aged 6–8 spell <muffin> as <muffen> ten times as often as UK children, while UK children spell <rocket> as <rockit> three times as often as Australian children (Kemp 2009), developed further in Chapter 12.

Children who do not speak the target accent can have specific problems with some spellings. This shows up in children's difficulties with the presenters' accents in the British TV spelling competition *Hardspell* and with the presenters' perception of the children's accents in the Scripps National Spelling Bee (McMenamin and Kerr 2014). In England a school might well encourage children to use an RP 'standard' accent because of the broader life opportunities it affords, but this is a separate issue from teaching children how to read and write, even if most teaching of spelling has prejudged it by using RP. The need for children to be aware of another accent when learning English spelling puts a burden on those who do not speak the 'standard' accent – the majority in many classes.

## Punctuation

English makes use of a fairly standard set of 'Western' punctuation marks (Nunberg 1990). The actual forms used in English differ from those found in continental Europe chiefly over quotation marks. English uses single and double quotation marks at the top of the line <“ ” ‘ ’ > rather than the up and down marks <„ „ > used for German and in many East European languages, or the goose feet <« » > found in French and Russian texts, the reverse goosefeet found in Switzerland <» « > (Cook 2004a) or the long dashes used in Spanish <— — >.

Punctuation originated as a way of providing hints for poor readers on how to read manuscripts aloud (Parkes 1992), reflected in advice such as:

A Comma Stops the Voice while we may privately tell one, a Semi Colon two; a Colon three: and a Period four.

(Mason 1748)

This seems little different from that available to University of Hull students today:

Where you think a reader should make a major pause (draw breath), use a full stop. Where you think a reader should make a smaller pause, use a comma.

(University of Hull 2007)

Note that these punctuation marks do not represent actual pauses in speech, which seldom occur at grammatical boundaries, but are guides to potential pauses in reading aloud.

Punctuation also helps the reader to understand the grammatical structure of the sentence, the focus of Chapter 6. Using the scheme in Halliday (1985) as a starting point, a paragraph can be shown by indent or leading (pronounced /ledɪŋ/); a lexical sentence by <! ? . >; a word by space, a morpheme by <' - >; and so on; line breaks also function as punctuation in street signs

(Cook 2013). The point where the phonological and grammatical functions of punctuation coincide is the overlap between the grammatical clause and the phonological tone group: ‘other things being equal, each clause is spoken as one tone group’ (Halliday 1985: 36). There is often tension between the two systems; some writers punctuate more by structure, some more by pauses; editing a book such as this reveals the wide individual differences in punctuation, particularly in the use of commas.

An easily overlooked feature of English is using spaces to separate words. Since spaces are essentially invisible to the reader, they are scarcely perceived as punctuation. Yet word spaces are not necessary to a script; letter-based writing systems such as Vietnamese and Thai do not have them. Historically, spaces only became standardised in European writing about the seventh century AD, originating from Irish scribes (Saenger 1997). Harris (1986) regards the invention of the word space as comparable in importance to the invention of zero in mathematics. In particular it facilitated silent reading; Saenger (1997) claims this had profound effects on intellectual life through the privacy it afforded compared with the public nature of reading aloud.

### ***The materiality of writing***

Writing and speech take different physical forms, whether material texts or sound waves. Writing is above all making symbols on a surface:

For most of the five thousand years of writing history, all our techniques and technologies have been aimed at making visible marks stick to surfaces.

(Levy 2001: 34)

A writing system reflects the strengths and limitations of the material on which texts are written and the material that the letters are made of: ‘texts are material objects’ (Kress and van Leeuwen 1996: 231). A writing system is tied to the technology available at a particular moment in time. Clay tablets require a different kind of writing from printed books, blackboards a different kind of writing from computer monitors. The materiality of writing has mostly been considered by typographers, as in Chapter 25, and calligraphers like Clayton (2014).

Kress and van Leeuwen (1996: 232) distinguish three material elements of writing:

- the *surface* on which marks are made, such as paper pages, blackboards, stone
- the *substance* they are made of, such as ink, paint, pixels
- the *means* by which they are made, whether pens, printers, brushes, chisels.

Let us take some examples of how material has affected letters.

### ***Runes***

The Old English letters <þ> (thorn) and <ƿ> (wynn) were taken from the early runic futhork, relics of which are scattered sparsely across England and Scotland. Runes were made by carving with a knife or sharp object on something solid like stone or bone. A glance at the bone inscription <𐺱𐺦𐺢𐺢𐺢> *raihan* meaning ‘roe deer’, found at Caistor-by-Norwich and dated about 400 AD, shows that the runes are largely made of straight lines, because of the difficulty in cutting curves, like Ogham described in Chapter 8. Letter shapes are a consequence of the materials and tools available to the writer (Jackson 1981).

### ***Serif Roman capital letters***

The technique of making Roman inscriptions was to draw the outline of the letter with a brush before cutting it out with a chisel. The serifs on Roman letters show the chisel following the flourish made by the brush at the end of the stroke (Catic 1968). Serif letters predominated until the introduction of sans serif fonts in the early nineteenth century, which became the very sign of modernity for the twentieth century (Tschichold 1928). Serif and sans serif are now familiar to every PC user through Times New Roman and Arial respectively. A quirk of Roman lettering technique has become a staple of our lives.

### ***Reading on a screen***

Letters on a computer monitor appear quite differently from those on a printed page, essentially lit from the back like a stained glass window compared to lit from the front like a painting: ‘The screen mimics the sky, not the earth’ (Bringhurst 2005: 193). The orientation of reading is usually different, typically a screen being vertical in front of the reader, a book horizontal. These demands led to a generation of fonts specifically designed to be legible on screen, such as Verdana and Georgia, designed by Matthew Carter (Re 2003), discussed in Chapter 25.

### ***Pen and paper***

The forms of letters depend upon the instrument used to make them, particularly affecting the thickness of the line. The development of the minuscule letter in England in the tenth century depended upon square-cut quill pens made from goose feathers (Clayton 2014), rediscovered by Johnston (1906). Most modern biros and felt-tips are pointed and so have no variation in line width. The development of pen technology goes hand-in-hand with advances in papermaking technology, which also affect the history of print (Müller 2014). Serif screen fonts still mimic the effects of the pen both in the varying width of line and the ‘stress’ showing the angle of pen-hold; sans serif fonts tend to have uniform thickness of line.

### ***Materials and socio-semiotics***

The conventional choice of materials goes with the socio-semiotic meaning of the text. Scollon and Scollon (2003: 135) describe three aspects of material; permanence/durability, temporality/newness and quality:

- *Permanence*: a written text can exist for seconds or millennia. Cook (2014) distinguishes functional permanence in which permanence is dictated by the intended use, say manhole covers and street-name signs, from asserted permanence in which the endurance and respectability of the sign and its owner is proclaimed through brass-plates, metal engravings, letters carved in stone and the like. Permanence is shown by the choice of material – stone or brass versus paper or plastic – and is often associated with serif Roman capital letters on stone; indeed a twentieth-century stone sign in Colchester with relief letters proclaims <LAW COVRTS>.
- *Temporality/newness*. Many written texts have a short life, typically handwritten in ink on paper. They are disposable not only physically but also indexically in that they also imply a limited time period whether <Closed for repairs>, <Special clearance> or job offers.

- *Quality*. In part quality is shown by the means through which writing is produced. Quality is a function of the skill involved in production: carving letters on slate, engraving letters on brass and painting a sign are skilled and expensive activities compared to scribbling with a felt-tip or printing out on a PC printer. To go back to the condolence letter from Gordon Brown, writing it by hand gave it quality – if not legibility.

Materials and the forms of letters thus express an identity, whether the impersonal identity of the permanent sign in quality materials or the individualism of many painted signs. The material is the message, or at least part of it.

## Conclusion

This background chapter has tried to weave together some of the threads from the other chapters. They suggest how diverse, rich and interesting the tapestry of the English writing system can be and how important it is both to the study of English and to the study of language in general. A recent book showed how writing enabled one individual, Ewan Clayton, to span the artistic community of Eric Gill, the world of monastic calligraphy, and the cutting edge technological community of PARC (Palo Alto Research Centre) (Clayton 2014). Research into the English writing system is not arcane academic description; writing permeates every aspect of our lives.

## A note on frequency information

As print is not the best way of accessing frequency information about the English writing system, we have decided not to present frequency data here but to refer readers to the following sources:

### *English frequency counts:*

Norvig, P.: <http://norvig.com/mayzner.html>

Gaines, H.F (1939) *Cryptanalysis*, New York: Dover

Cook, V. J. <http://www.viviancook.uk/Punctuation/PunctFigs.htm>

### *Grapheme representation counts:*

Fan, F. and Altmann, G. (2008) ‘Graphemic representation of English phonemes’, in Altmann, G. and Fengxiang, F. (eds.) (2008) *Analyses of Scripts: Properties of Characters and Writing Systems*, Berlin: Mouton de Gruyter, 25–61

### *Common words:*

D.W. Cummings: [dwcummings.com/cwdb/default.asp](http://dwcummings.com/cwdb/default.asp)

### *Correspondences:*

Gontijo, P.F.D., Gontijo, I. and Shillcock R. (2003) ‘Grapheme–phoneme probabilities in British English’, *Behavior Research Methods, Instruments, and Computers*, 35(1), 136–157

**Spelling data:**

<http://www.viviancook.uk/EnglishSpellingSystem/>

**General follow-up reading**

Since this field is seldom presented as a whole, readers will find follow-up reading suggested for the specific areas in the chapters. Otherwise the following provide some help to new readers.

- Clayton, E. (2014) *The Golden Thread*, London: Aldine Books.  
 Cook, V.J. (2004) *The English Writing System*, London: Edward Arnold.  
 Coulmas, F. (1996) *The Blackwell Encyclopedia of Writing Systems*, Oxford: Blackwell.  
 Hill, W. (2010) *The Complete Typographer*, London: Thames and Hudson.  
 Treiman, R. and Kessler, B. (2014) *How Children Learn to Write Words*, Oxford: Oxford University Press.

**References**

- Albrow, K.H. (1972) *The English Writing System: Notes towards a Description*, London: Longman for the Schools Council.  
 Altmann, G. (2008) 'Towards a theory of script', in G. Altmann and F. Fengxiang (eds) *Analyses of Scripts: Properties of Characters and Writing Systems*, Berlin: Mouton de Gruyter, 149–164.  
 Beech, D., Harrison, C. and Hill, W. (2009) *Art and Text*, London: Black Dog Publishing.  
 Berg, L. (1968–1974) *Nippers*, London: Macmillan Education.  
 Bloomfield, L. (1933) *Language*, New York: Holt.  
 Borgwaldt, S.R. and Joyce, T. (eds) (2013) *Typology of Writing Systems*, Amsterdam: John Benjamins.  
 Bringhurst, R. (2005) *The Elements of Typographic Style*, Version 3.1. Vancouver: Hartley and Marks.  
 Bromley, T. (2002) 'Dialect spelling: a study of the effects of a child's dialect on their spelling', MA dissertation, University of Essex.  
 Brooks, G. (2015) *Dictionary of the British English Spelling System*, Cambridge: Openbook.  
 Butler, S. (1908) *Characters and Passages from Note-Books*, Cambridge: Cambridge University Press.  
 Carney, E. (1994) *A Survey of English Spelling*, London: Routledge.  
 Catich, E.M. (1968) *The Origin of the Serif: Brush Writing and Roman Letters*, Davenport, IA: The Catfish Press.  
 Clayton, E. (2014) *The Golden Thread*, London: Aldine Books.  
 Coltheart, M. (2005) 'Modelling reading: the dual-route approach', in M. Snowling and C. Hulme (eds) *The Science of Reading: A Handbook*, Oxford: Blackwell, 6–23.  
 Coltheart, M., Rastle, K., Perry, C., Langdon, R. and Ziegler, J. (2001) 'DRC: a dual route cascaded model of visual word recognition and reading aloud', *Psychological Review*, 108, 1, 204–256.  
 Cook, V.J. (2004a) *The English Writing System*, London: Edward Arnold.  
 Cook, V.J. (2004b) 'The spelling of the regular past tense in English: implications for lexical spelling and dual process models', in G. Bergh, J. Herriman and M. Mabärg (eds) *An International Master of Syntax and Semantics: papers presented to Aimo Seppänen on the occasion of his 75th birthday*, Göteborg, Sweden: Acta Universitatis Gothoburgensis, 59–68.  
 Cook, V.J. (2010) 'Prolegomena to second language learning', in P. Seedhouse, S. Walsh and C. Jenks (eds.) *Conceptualising Language Learning*, Basingstoke: Palgrave Macmillan, 6–22.  
 Cook, V.J. (2013) 'Standard punctuation and the punctuation of the street', in M. Pawlak and L. Aronin (eds.) *Essential Topics in Applied Linguistics and Multilingualism*, Switzerland: Springer International Publishing, 267–290.  
 Cook, V.J. (2015) 'Meaning and material in the language of the street', *Social Semiotics*, 25, 1, 81–109.  
 Coulmas, F. (1996) *The Blackwell Encyclopedia of Writing Systems*, Oxford: Blackwell.

- Coupland, N. and Bishop, H. (2007) 'Ideologised values for British accents', *Journal of Sociolinguistics*, 11(1), 74–93.
- Cruttenden, A. (2014) *Gimson's Pronunciation of English*, 8th edition. London: Routledge.
- Cummings, D.W. (1988) *American English Spelling: An Informal Description*, Baltimore, MD: Johns Hopkins University Press.
- Daniels, P.T (1996) 'The study of writing systems', in P.T. Daniels and W. Bright (eds) *The World's Writing Systems*, New York: Oxford University Press.
- de Saussure, F. (1916) *Cours de linguistique générale*, in C. Bally, A. Sechehaye and A. Reidlinger (eds). Paris: Payot. Trans W. Baskin (1959) *Course in General Linguistics*, London: Peter Owen.
- Frith, U. (1985) 'Beneath the surface of developmental dyslexia', in K.E. Patterson, J.C. Marshall and M. Coltheart (eds.) *Surface Dyslexia*, Hove: Lawrence Erlbaum, 301–330.
- Funnell, E. (1983) 'Phonological processes in reading: new evidence from acquired dyslexia', *British Journal of Psychology* 74, 159–180.
- Gill, E. (1931) *An Essay on Typography*, London: Lund Humphries.
- Goswami, U. (1999) 'Integrating orthographic and phonological knowledge as reading develops: onsets, rimes, and analogies in children's reading', in P. McMullen and R. Klein (eds.) *Converging Methods for Understanding Reading and Dyslexia*, Hillsdale, NJ: Erlbaum, 57–85.
- Gray, N. (1960) *Lettering on Buildings*, London: The Architectural Press.
- Halliday, M.A.K. (1985) *Spoken and Written Language*, Oxford: Oxford University Press.
- Halliday, M.A.K. and Mattheisen, M.I.M. (2013) *Halliday's Introduction to Functional Grammar*, 4th edition, Abingdon: Routledge.
- Harris, R. (1986) *The Origin of Writing*, London: Duckworth.
- Hart, J. (1569) *An Orthographie*, London: William Seres.
- Hill, W. (2010) *The Complete Typographer*, London: Thames and Hudson.
- Jackson, D. (1981) *The Story of Writing*, London: Studio Vista.
- Jaffe, A. (2000) 'Non-standard orthography and non-standard speech', *Journal of Sociolinguistics*, 4(4), 497–514.
- Johnson, S. (1755) *A Dictionary of the English Language*, London: W. Strahan.
- Johnston, E. (1906) *Writing and Illuminating and Lettering*, London: Pitman.
- Katz, L. and Frost, R. (1992) 'The reading process is different for different orthographies: the orthographic depth hypothesis', in L. Katz and R. Frost (eds.) *Orthography, Phonology, Morphology and Meaning*, Amsterdam: Elsevier, 67–84.
- Kemp, N. (2009) 'The spelling of vowels is influenced by Australian and British English dialect differences', *Scientific Studies of Reading*, 13(1), 53–72.
- Kenyon, J.S. and Knott, T.A. (1951) *A Pronouncing Dictionary of American English*, Springfield, MA: Merriam.
- Kilpatrick, J. (1988) 'Financial planner should invest in spelling lessons', *Rome-News-Tribune*, October 27 3c.
- Kinneir, J. (1980) *Words and Buildings*, London: The Architectural Press.
- Kreiner, D.S., Schnakenberg, S.D., Green, A.G., Costello, M.J. and McClin, A.F. (2002) 'Effects of spelling errors on the perception of writers', *Journal of General Psychology*, 129, 5–17.
- Kress, G. and van Leeuwen, T. (1996) *Reading Images: The Grammar of Visual Design*, London: Routledge.
- Levy, D.M. (2001) *Scrolling Forward: Making Sense of Documents in the Digital Age*, New York: Arcade Publishing.
- Lewery, A.J. (1989) *Signwritten Art*, Newton Abbot: David and Charles.
- Mason, J. (1748) *Essay on Elocution; or, Pronunciation. Intended chiefly for the Assistance of those who Instruct Others in the Art of Reading. And of Those who are often Called to Speak in Publick*, London: R. Hett.
- McMenamin, G.R. and Kerr, L.N. (2014) 'Jasmine and the Bee: spelling the initial [tʰ] in English', *Language*, 90, 4, 149–164.
- Müller, L. (2014) *White Magic*, Cambridge: Polity.

- Neef, M. and Balestra, M. (2013) 'Measuring graphematic transparency: German and Italian compared', in S.R. Borgwaldt and T. Joyce (eds) *Typology of Writing Systems*, Amsterdam: John Benjamins, 113–146.
- Nunberg, G. (1990) *Linguistics of Punctuation*, Stanford, CA: Center for the Study of Language and Information.
- Oxford English Dictionary* (OED) (2015) online www.oed.com. Oxford: Oxford University Press.
- Parkes, M.B. (1992) *Pause and Effect: An Introduction to the History of Punctuation in the West*, Aldershot: Scolar Press.
- Perfetti, C.A. (1999) 'Comprehending written language; a blueprint of the reader', in C. Brown and P. Hagoort (eds) *The Neurocognition of Language*, Oxford: Oxford University Press, 167–210.
- Peters, M.L. (1970) *Success in Spelling: A Study of the Factors Affecting Improvement in Spelling in the Junior School*, Cambridge: Cambridge Institute of Education.
- Peters, P. and Delbridge, A. (1989) 'Standardisation in Australian English', in P. Collins. and D. Blair (eds) *Australian English*, Brisbane: University of Queensland Press, 127–137.
- Puttenham, T. (1589) *The Arte of English Poesie*, London: The Black-Friers.
- Re, M. (2003) *Typographically Speaking: the Art of Matthew Carter*, New York: Princeton Architectural Press.
- Roberts, J. (2005) *Guide to Scripts used in English Writings up to 1500*, London: British Library.
- Rogers, H. (2005) *Writing Systems: A Linguistic Approach*, Oxford: Blackwell.
- Saenger, P. (1997) *Space Between Words: The Origins of Silent Reading*, Redwood City, CA: Stanford University Press.
- Salesbury, W. (1547) *A Dictionary in Englyshe and Welshe*, London: John Waley.
- Scollon, R. and Scollon, S. (2003) *Discourses in Place: Language in the Material World*, Abingdon: Routledge.
- Seidenberg, M.S. (1992) 'Beyond orthographic depth in reading: equitable division of labour', in R. Frost and L. Katz (eds.) *Orthography, Phonology, Morphology and Meaning*, Amsterdam: Elsevier, 85–118.
- Seidenberg, M.S. and McClelland, J.L. (1989) 'A distributed, developmental model of word recognition and naming', *Psychological Review*, 96, 523–568.
- Smeijers, F. (2011) *Counterpunch*, 2nd edition, London: Hyphen Press.
- Sproat, R. (2000) *A Computational Theory of Writing Systems*, Cambridge: Cambridge University Press.
- St Augustine (397) *On Christian Doctrine*, <http://www.ccel.org/ccel/augustine/doctrine.txt>
- Treiman, R. (1993) *Beginning to Spell: A Study of First-Grade Children*, Oxford: Oxford University Press.
- Treiman, R. and Kessler, B. (2013) 'Similarities among the shapes of writing and their effects on learning', in S.R. Borgwaldt and T. Joyce (eds) *Typology of Writing Systems*, Amsterdam: John Benjamins, 41–61.
- Treiman, R. and Kessler, B. (2014) *How Children Learn to Write Words*, Oxford: Oxford University Press.
- Trudgill, P. (2001) *Sociolinguistic Variation and Change*, Edinburgh: Edinburgh University Press.
- Trudgill, P. (2003) *A Glossary of Sociolinguistics*, Edinburgh: Edinburgh University Press.
- Tschichold, J. (1928) *The New Typography*, 1998 edition, Oakland, CA: University of California Press.
- University of Hull (2007) 'Punctuation: Elementary Guidance', <http://www2.hull.ac.uk/student/pdf/Punctuation%20Elementary%20Guidance2.pdf>
- Venezky, R.L. (1970) *The Structure of English Orthography*, The Hague: Mouton.
- Venezky, R.L. (1999) *The American Way of Spelling*, New York: Guilford Press.
- Webster, N. (1828 [1970]) *An American Dictionary of the English Language*, New York: Johnson Reprint Corp.
- Wells, J.C. (1982) *Accents of English*, 3 volumes, Cambridge: Cambridge University Press.

*This page intentionally left blank*

**Part I**

# Theory and the English writing system

---

*This page intentionally left blank*

# English among the writing systems of the world<sup>1</sup>

Richard Sproat

---

I am reminded at this point of a fellow I used to know whose name was Henry, only to give you an idea of what an individualist he was he spelt it <HEN3RY>. The 3 was silent, you see.  
Tom Lehrer

## Introduction

Tom Lehrer was of course being silly when he used the introduction quoted above to his song ‘We will all go together when we go’, yet at the same time the joke would hardly have been possible were it not for the vagaries of English orthography. While no standard English word (such as one would find in a dictionary) includes numerals, there are many cases of words where a ‘silent’ letter has been inserted for one reason or another. A famous case is *debt*, originally spelled <dette> or <dete>, where <b> was introduced for etymological reasons (Latin *debitum*) certainly by the sixteenth century.<sup>2</sup> After this the spelling with the <b> became standard, and English orthography happily accommodated this purely etymological ‘silent’ addition. Another example is *island*, from Old English *īegland*, with an <s> introduced on false analogy with *isle* (from Latin *insula*), with which *island* is completely unrelated. Johnson ensured the survival of <debt> by including it in his *Dictionary of the English Language* (1755). (Interestingly, he incorrectly gives the etymology of *island* as coming from Latin *insula*.)

These are some of the many aspects of English orthography that conspire to make it one of the most complicated modern writing systems. How does one classify such a system among the writing systems of the world?

In this chapter we will examine some of the proposals for where English falls in the typology of writing systems. But in order to do that, we first need to examine some proposals for the taxonomy of writing systems in general, and it is to that topic that we first turn.

## Taxonomies of writing systems

One issue that it would be useful for the reader to bear in mind at the outset is the fact that a writing system for a language is designed or adapted, usually by native, but in any case

competent speakers of a language, in order to be used by native or competent speakers of that language. A writing system is not a system of phonetic transcription: rather it is a way to represent the words of the language. In this much, Chomsky and Halle (1968 – see below) were correct. Always this will involve *some* representation of the pronunciation, but that representation is not required to be regular, though of course regularity can often make the task of learning the system easier. But no matter how English orthography ends up being classified, it is no less ‘natural’ than other more seemingly regular systems.

For centuries scholars have been trying to make sense of the differences between how writing systems encode information. Bishop Wilkins in his *Essay Towards a Real Character, and a Philosophical Language* (1668) discusses ‘Aethiopic’, which he notes contains 26 consonants and seven vowels, the latter of which are applied ‘to every one of their 26 Consonants’, which along with 20 additional ‘aspirated syllables’ makes for an ‘alphabet’ of 202 units. The system, according to Wilkins, is ‘exceedingly complicated and perplexed’ (p. 14).<sup>3</sup> Antoine Court de Gébélin in his *Monde Primitif* (1775) declared on the basis of an extensive survey of writing systems known at the time that there were three types of writing in the world, namely Chinese, Egyptian Hieroglyph<sup>4</sup> and alphabetic (p. 399).

It wasn’t until the twentieth century, however, that systematic taxonomies of writing systems were developed. Ignace Gelb, often dubbed the ‘father of grammatology’, presented (1963) a teleological view of writing that saw all writing systems as being on a linear path from pictographic and ‘limited systems’ (into which category he classified Mayan, now known to be full-fledged writing), through ‘word syllabic systems’ (Chinese, Sumerian), through syllabaries and ultimately segmental alphabets. Gelb’s system forced some odd classifications: thus West Semitic abjads (writing systems where the basic characters of the script denote consonants) such as Phoenician were classified as syllabaries largely because he did not have any other category into which to shoehorn them.

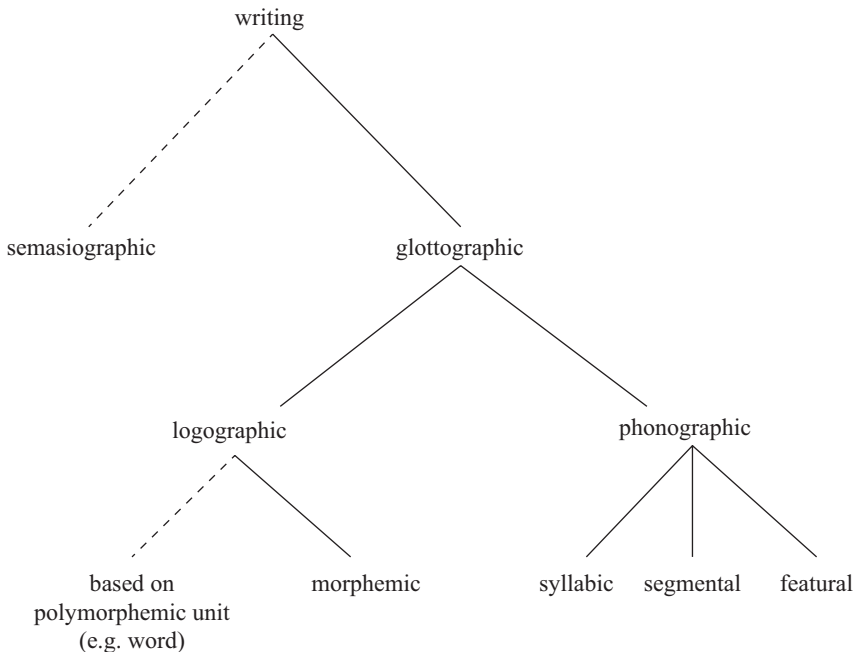


Figure 3.1 Taxonomy of writing systems according to Sampson (1985)

Scholars of writing systems since Gelb have tended to take less teleological views, presenting instead classifications that put the various types of writing more on a par. Sampson (1985) and DeFrancis (1989) provide arboreal (or branching) classifications according to the function of the core units of each system.

Thus Sampson, for example, proposed the taxonomy depicted in Figure 3.1. At the top level of the taxonomy systems divide into *semasiographic* (systems based purely on meaning) and *glottographic*, those encoding specifically *linguistic* information, such as representing phonemes, syllables or morphemes.

Sampson's semasiographic category has always been controversial: DeFrancis (1989) effectively demolishes Sampson's 'Yukaghir love letter' as an example of writing of any kind, showing that it was instead a form of parlor game. But there have been genuine attempts at developing semasiographic systems, the most famous of these being Charles Bliss's system that has come to be known as *Blissymbolics*. Rogers (2004) discusses this system, as does Sproat (2009), who notes that brave though the attempt was, Blissymbolics fails a crucial test for being a full writing system: it is impossible to express in Blissymbolics all of the nuances expressible in spoken language, and which are automatically expressible in any glottographic system.<sup>5</sup> Today the main function of Blissymbolics is as a communication aid for children with severe communications deficits associated with intellectual disability or autism (see, again, Sproat 2009 for discussion).

With glottographic systems Sampson is on more solid ground, and here he divides the world into *logographic* systems, where each symbol represents a morpheme, and *phonographic* systems where the symbols of the system represent one or another phonological unit ranging from syllables down to phonetic features.

DeFrancis (1989) not only disagrees with Sampson on the existence of semasiographic writing. He also takes issue with the whole notion of *logographic* systems at one end of the scale (see also DeFrancis and Unger 1994), as well as featural systems at the other. His taxonomy is presented in Figure 3.2.

For DeFrancis all *full* writing systems are essentially phonological, with the main distinction being between syllabic and segmental systems, and within the latter, consonantal and alphabetic. Consonantal systems (e.g. the writing systems for Arabic or Hebrew) are those systems that are 'defective' in that they represent only consonants, or perhaps consonants with a limited amount of vocalic information. But within each basic category there is another distinction: between systems that are 'pure' and those that also encode morphological information. For Chinese, for example, many characters are a composite of two pieces, one piece that gives a hint about the pronunciation, and another that gives a hint about the meaning. Thus 橡 *xiàng* (in its modern Mandarin form) 'oak' is composed of two pieces, 木 *mù* 'wood, tree', which is typically used for characters denoting names of trees; and 象 *xiàng* 'elephant', which is used purely for its phonetic value. In this particular instance the semantic and phonetic hints are rather good, but this is by no means representative, and quite often the phonetic hint is quite poor. The reasons for this are several, including the huge phonetic changes in Chinese over the roughly 3500 year history of the script. Nonetheless the fact that over 95 percent of the Chinese characters ever developed are based on this semantic-phonetic construction – as well as the fact that this method for creating new characters was also borrowed into the Chữ Nôm system for Vietnamese (see Sproat 2000) – means that this was apparently considered the normal way to make new characters, which in turn is a powerful argument for DeFrancis' theory.

Treating Egyptian as 'morphoconsonantal' makes a great deal of sense since, like Chinese, there are components of the writing system that clearly relate directly to the meaning of the morphemes being written. On the other hand, classifying English (and Korean) as

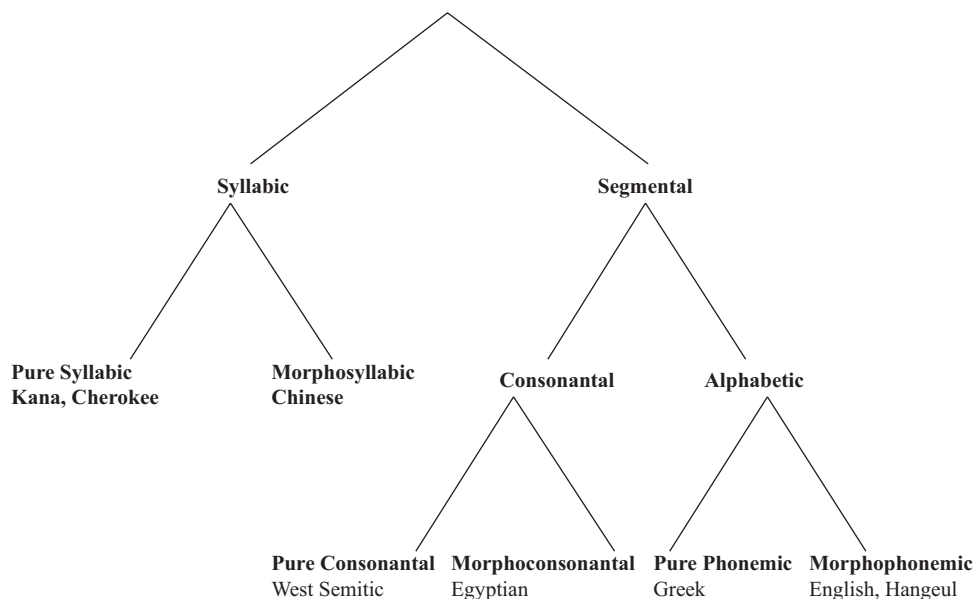


Figure 3.2 Taxonomy of writing systems according to DeFrancis (1989)

‘morphosegmental’ is not as obvious, since here there are no components that have a clear function of indicating the meaning. This issue relates of course directly to the topic of this chapter, and we return at length to it below.

DeFrancis’ neat separation of each phonographic type into ‘pure’ and ‘morpho-’ hints at a further refinement that could be made to taxonomies of writing systems. The problem with arboreal representations such as DeFrancis’ is that the type of phonography, and the existence or not of logography, are really two separate, indeed orthogonal, dimensions. Furthermore, while it usually makes sense to classify a system as basically segmental, basically consonantal, and so forth, logography is really a matter of degree. There are some systems that are heavily logographic, like Chinese. There are others, like Japanese, where a large portion of what one finds written on a page is purely phonographic, but there are still large amounts of logographic information conveyed (on Japanese writing, see Okada, this volume). There are some systems, like Finnish, where the amount of logography is essentially nil. Considerations such as these led Sproat (2000) to propose the planar classification system in Figure 3.3. This system was further modified by Rogers (2004), whose major points of disagreement with Sproat (2000) had to do with the precise divisions of the phonographic dimension.<sup>6</sup>

At one end of the phonographic dimension are the (actually quite rare) full syllabic systems where every syllable of the language has one or more separate symbols to represent it. Much more common are what Sproat termed ‘core syllabic’ systems – developing an observation first noted by Poser<sup>7</sup> – where the basic symbols generally do represent basic syllables (often just V or CV), but where for more complex syllables one must use a combination of symbols. Following the syllabic systems there are various segmental systems ranging from alphabetic, where all or nearly all segments of the language are represented in writing, to consonantal systems where the system basically just represents consonants. Egyptian is separated out as a unique polyconsonantal system – polyconsonantal since many of the signs represent sequences of two or more consonants (with no representation of the vowels). In this, and in

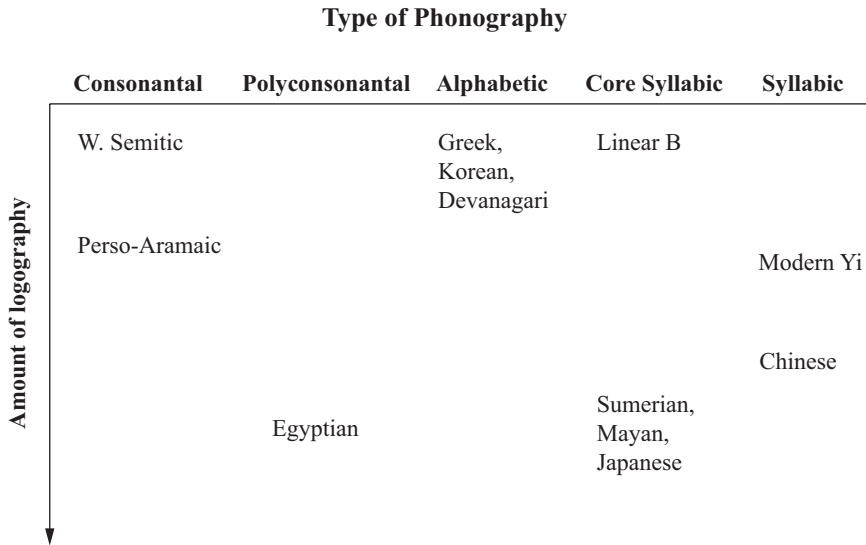


Figure 3.3 Planar taxonomy, after Sproat (2000)

the treatment of *alphasyllabic* systems such as Devanagari, Rogers' system differs. He does not treat 'polyconsonantal' as a separate category, and on the other hand does separate out alphasyllabaries.

As noted above, the logographic dimension is a scale rather than a set of categories. 'Amount of logography' is, as noted by Sproat (2000), very impressionistic – but see Penn and Choma (2006) for some interesting proposals on corpus-based ways to estimate it. Writing systems that have a large degree of logography – Chinese, Sumerian, Egyptian, etc. – are placed higher on the logographic scale than, e.g., most West Semitic writing systems, or Greek. On the other hand, some systems whose basic symbols are purely phonetic nonetheless can behave logographically: a good example is *heterograms* in Perso-Aramaic writing systems, a topic we return to below.

Sproat places English at the low end of the logographic scale – which as we shall see differs from where Sampson would have placed English had he adopted a planar taxonomy like Sproat's.

Having laid out a few different ways to think about writing system classification, we are now ready to address the main question of this chapter: where does English fit?

## Theories of where English falls

### *Preliminaries*

It is no secret that English spelling is quirky and unpredictable. Indeed it probably has the most complicated letter-to-sound correspondence of any language that uses a segmental writing system. In quantitative studies of the complexity of letter-to-sound correspondence among Western European languages, English always tops the list (van den Bosch *et al.* 1994). This quiriness also shows up in another way, in the difficulty that children have in learning the

system, compared to any other alphabetic system (Dehaene 2009; see in particular Figure 5.3, p. 231). Its complexity has also manifested itself in innumerable proposals to replace the system by something based more directly on pronunciation. For various reasons, such proposals have never gained a foothold (for more, see Yule and Ishi, this volume).

But what is one to make of this unpredictability and where therefore English belongs in the taxonomy of writing systems just discussed?

Let us start with the obvious observation that the basic elements of English spelling, the letters of the Roman alphabet, were, as originally designed, purely phonographic symbols. Certainly if one traces the letters back via Latin, through Etruscan, Greek and thence to Semitic scripts ending ultimately with the Proto-Sinaitic scripts of roughly 1800 BCE (Daniels and Bright 1996), one finds that the symbols were originally pictographic, with the *alpha* / *alef* א / <a> representing an ox head, the *beta* / *beth* ב / <b> representing a house, and so forth. But even there the symbols derived their function via the acrophonic principle – ‘*alef*’ (symbol for the glottal stop), being the word for ‘ox’, *beth* (<b>) the word for ‘house’, etc. At no time did the symbols have anything to do semantically with what they depicted. This is in marked contrast to all known independently developed ancient writing systems – Egyptian, Sumerian, Chinese and Mayan – where some of the elements were used for their semantic values.

This fact about the original function of the letters would appear to militate against analyzing English as a partly logographic system – as Sampson (1985) argued, see below – and would seem to suggest that the system must be some form of purely phonographic system. If nothing else, English orthography lacks anything corresponding to the semantic radicals that are part of Chinese writing, whereby lexical-semantic properties of morphemes have an explicit (if highly inconsistent) graphical expression. If the system is not logographic, then it surely must be classified as a phonographic system. Yet the original purpose of the symbols that became English letters gives us substantially less insight into the nature of English orthography than one might expect, for two reasons.

First of all, there are logographic systems that involve phonographic symbols in an otherwise phonographic writing system. A clear case of this are so-called *heterograms* in adaptations of the Aramaic script to Persian languages (Skjærvø 1996). In such cases, an Aramaic *spelling* for a word may be borrowed – e.g. <ŠME> ‘name’ – which is nonetheless to be *read* as a Persian word (*nām*). Such a system is clearly logographic since the Aramaic spelling is being used to represent a Persian morpheme that has totally different pronunciation from what the spelling suggests. Furthermore, these heterograms may combine with ‘phonetic complements’ which spell grammatical endings. The situation is, of course, highly reminiscent of the Japanese adaptation of Chinese script, with syllabic phonograms (known as *kana*) used for the Japanese grammatical endings. It is also reminiscent of the Assyrian adaptation of Sumerian script, except that here we are dealing not with logographic or morphosyllabic symbols, but with segmental ones. English orthography, of course, has nothing approaching Perso-Aramaic *heterograms*, but at the same time there is a clear precedent for systems based purely on phonographic symbols to involve some logographic elements.

On the other hand, supposing one wanted to argue that English orthography is purely phonographic. Clearly it cannot be transparently so: the letter–sound correspondence is in no way as regular as in orthographies such as those in Spanish or Finnish. But where does it say that the letter-to-sound correspondence has to be simple? For example, where does it say that the grapheme <a> must always represent the phoneme /a/ (as it does in Spanish) or that the grapheme <p> always /p/?

Indeed there are at least two sources of complexity in phonographic systems. The first relates to what *level* of phonological information is represented by the spelling, something

that corresponds roughly to the notion of *deep* (versus *shallow*) orthography familiar from psycholinguistic work on reading (Katz and Frost 1992), and something that was formalized in Sproat (2000) as the *orthographically relevant level* (ORL) of phonological representation. In that work I argued, for example, that the main difference between Russian and Belarusian orthography relates to the level of phonological representation assumed by the two orthographies, with Russian representing a deeper level than Belarusian. If one takes that approach, English spelling insofar as it does not obviously represent surface pronunciations in anything like a consistent way, must have a relatively deep ORL; indeed English is always characterized as a *deep* orthography in the psycholinguistics literature. But there is a second complication in that one could have a phonographic system that nonetheless admits of some lexical marking. Thus there may be several ways to spell a given sound, which is clear in English with pairs like *seen* and *scene*, and one must simply memorize which words take which spelling, with perhaps one spelling being the default. This seems obvious and simple enough, but crucially as we will see below, there is a trade-off between lexical marking and orthographic depth, insofar as one can assume a shallower depth, perhaps at the cost of more lexical marking. In any case, if one wants to assume that English is purely phonographic, there are options open to one.

With these preliminary issues in mind, we now turn to some previous attempts to categorize English orthography. Anticipating the conclusion, we will see that English is best characterized as being partly phonographic and partly logographic.

## The taxonomy of English orthography

There have been many descriptions of English spelling. Many of them have focused on developing sets of rules that can characterize the relation between spelling and phonology. Perhaps the best-known example of this is the work of Venezky (1970), who develops a system based on a complex set of ordered rewrite rules from spelling to sound. It is worth noting that the small cottage industry of work on letter-to-sound conversion for speech technology applications largely views the problem as one of inducing the appropriate rules from a set of training examples, though the expression of those rules differs widely depending upon the technique used – see Bisani and Ney (2008) for a recent overview.

But there have been approaches that attempt to shift the burden from the letter-to-sound correspondences to the phonology. Thus, Chomsky and Halle in the *Sound Pattern of English* (1968 – henceforth SPE) presented what has been probably the most controversial thesis on English orthography, namely that English spelling is a near optimal orthography for English, in that it represents underlying morphophonemic representation. The arguments largely, though not exclusively, center on stress-induced alternations in Latinate morphology, cases like *assign* versus *assignment*, the former pronounced /ə'sam/ and the latter /,æsig'neɪʃən/, seemingly quite different in their surface forms. Yet if one assumes that the surface pronunciations are derived via phonological rules from something more like /æsign/ and /æsignātyon/, then the spellings make sense. The surface forms are derived by rules that include reduction of unstressed vowels to schwa, vowel shift, and spirantization of /tj/ to /ʃ/.

Or consider the word *nightingale*. Chomsky and Halle argue for a general rule of 'trisyllabic laxing' whereby a vowel in a syllable third from the end of the word laxes. This explains alternations like *opaque* /o'peɪk/ versus *opacity* /o'pæsiti/, where the <a> is pronounced as a lax /æ/ in the second case, but not in the first. Yet *nightingale* has a tense (diphthong) /aɪ/ in the first syllable, which should have by rights become a lax /ɪ/ by the rule. Is something somehow 'saving' the <i> from being pronounced with a lax variant? Chomsky and Halle argue that there is, and more specifically that there is a missing phoneme following the <i> that is somehow

conspiring to prevent the deletion. Conveniently, they argue, that missing phoneme would appear to be a phoneme that has no surface form in standard English, but is still represented in the orthography as <gh>, namely /x/.

Thus two seemingly quirky and idiosyncratic properties of English, among others, can be explained by arguing that English orthography represents not surface, but underlying phonological representation. And this, argued Chomsky and Halle, is how things should be: orthographies are designed, or at least should be designed, with knowledgeable speakers of the language in mind. They are not intended, for example, to make life easier for second-language learners by representing the superficial pronunciation of words. Rather they are intended to show relationships between words that (putatively) share phonological properties, and are arguably morphologically related. Thus *assign* / *assignation* should share the same spelling for the root since they are related, as should *electric* / *electricity* despite the fact that in both cases the root portions have different pronunciations in the two forms.

The SPE hypothesis thus makes a direct claim about what type of writing system English is: it is a (segmental) phonographic system, just like Finnish, or Spanish. The only difference is that for Finnish and Spanish, the orthography largely represents surface phonology, whereas for English it represents a more abstract level of phonological representation.

There are, needless to say, so many problems with the SPE hypothesis, that few scholars take it seriously, and the thesis is at this stage largely forgotten, except perhaps as a classic instance of a *reductio ad absurdum* of an approach. But before examining some of those problems, it is worth noting one reason why the hypothesis is *not* problematic, namely, that there is nothing a priori silly about the idea that an orthography might represent a somewhat abstract level of representation. The notion of orthographically relevant level argued for by Sproat (2000) makes precisely this claim, and as noted above at least some of what counts as ‘deep’ orthographies in psycholinguistic studies of reading also depends on this assumption. Besides this, there are orthographies where this is obviously true, one such case being Korean (King 1996). Korean tends to be quite consistent in its spelling of morphemes across different derived forms, but the actual surface pronunciation of words is complicated by sometimes regular, and sometimes idiosyncratic morphophonemic changes. For example, the combination of *깨* <ggae> /kkæ/ ‘sesame’ and *잎* <ip<sup>h</sup>> /ip/ ‘leaf’ yields *깨잎* ‘sesame leaf, perilla’, written as <kkaesip<sup>h</sup>> but pronounced /kkænnip/. This is an instance of a fairly common nasalization process in compounds, one that is nonetheless hard to predict on the basis of the components of the compound. To take another example, *꽃* <ggoch> /kkot/ ‘flower’ restores the pronunciation of the <ch> in the nominative form *꽃이* <ggochi> /kkoč<sup>h</sup>i/. Under the SPE hypothesis, English would then be just like Korean, except that the relation between the morphophonological forms represented by the spelling, and the surface pronunciation, is more complicated in the case of English.

This said, there are sufficient problems with the hypothesis that it can largely be rejected as a theory of English spelling. As Sampson (1985) notes, while Chomsky and Halle’s principle that roots should retain their spelling in alternative forms, there are many cases where this principle is broken. Thus a pair such as *opaque* / *opacity* respells the <c> as <q> in the first case, for no apparent reason. Perhaps the form \*<opac> is ill-formed for independent reasons, but it is hard to see what they could be, especially given that English words can certainly end in <c>. Or consider *collide* / *collision*, where there is no apparent reason why the <d> should be respelled as <s> in the second word, since *collidion* would surely be a much ‘better’ spelling from SPE’s point of view. Or why is there a change in the spelling of the *identical* vowel in the two related words *speak* and *speech*?

Sampson also discusses experiments by Carol Chomsky (1970), who attempted to find support for the SPE hypothesis as a pedagogical tool. If the hypothesis was on the right track,

then it ought to help children learning to spell English to point out how logical this seemingly illogical system is. Yet as Sampson notes, Carol Chomsky ran into a problem with one seventh-grader who was not able to recognize the relationship between *signature* and *sign*, and thus was not able to make use of her hint that he consider *signature* in understanding how to spell *sign*. This, strictly speaking, is a problem with morphological awareness rather than a direct problem for the hypothesis about English spelling. Still, the fact the two forms are obviously semantically related, coupled with the SPE hypothesis, would a priori suggest that readers *ought* to be aware that the forms are related, at least when prompted, something that Carol Chomsky found not to be the case.

Yet another reason for doubting the utility of the SPE hypothesis comes from statistical analysis. In Sproat (2000) I compared two hypotheses, one where English spelling is a representation of an underlying morphophonemic level, and the other where it is a representation of surface pronunciation. I based the analysis on a set of 1169 Latinate forms – the portion of the vocabulary where the SPE hypothesis is most relevant. Surface pronunciations were taken from an online dictionary of American English. SPE-style underlying pronunciations were also constructed. Obviously since the latter were influenced almost by necessity by the spelling, the experiment contained an inherent bias towards the SPE hypothesis. In the technical terminology of Sproat (2000), there were two hypotheses as to the nature of the ORL, the first that it is surface phonology, the second an underlying phonology. Then, for each of the two hypotheses, I computed the complexity of the rules required to predict the spelling from the phonological representation [sic]. The measure of complexity comprised two components: the number of rewrite rules needed; and the number of lexical markings required to override the application of the otherwise exceptionless rules.

The number of rules needed for the shallow hypothesis was somewhat larger (69 versus 58) and the number of lexical markings is much larger (1452 versus 509 lexical marks needed). However more than 40 percent of the 1452 lexical marks have to do with the spelling of schwa, not needed for the deep ORL; for example in the word *geophagy*, the spelling of the schwa of the second syllable as <a> needs to be marked. This would appear to support a hypothesis that English orthography represents something more abstract than surface pronunciation, but such a hypothesis need not assume such a deep representation as that assumed by SPE: it would suffice merely that vowels be represented in their unreduced forms. Alternations such as the /eɪ/~ /æ/ alternation in *chaste* ~ *chastity* have little consequence since this only means a few additional rules specifying how to spell surface vowel forms. So on balance, I argued, the statistics provide at best weak support for the SPE hypothesis. And once one notes again that the 1169 examples were from that portion of the vocabulary most relevant to the alternations discussed in SPE – the vast majority of English words do not participate in such alternations – the import of the SPE theory as a general model of English spelling is weakened further.

On balance then the statistics do not strongly support English orthography representing anything nearly as abstract as assumed in SPE. At best one can assume a somewhat abstract phonological representation, with a fair amount of marking of lexical exceptions which even SPE's abstract representations cannot avoid. For example there is no way to predict the <ph> in *photograph* from either deep or surface pronunciation, which has an initial /f/, other than marking the word with an ad hoc feature [+Greek].

And it is these lexical markings that are the crux of the matter: English orthography is surely phonographic to a large degree, whichever level of representation one takes it to represent, so the question of where to put it in the taxonomy of the world's writing systems really comes down to what to do with the unpredictability.

The alternative to focusing solely on how exactly English orthography encodes phonographic information is to propose that, while the orthography is obviously phonographic to a significant extent, it is also partly logographic. This was proposed in Bloomfield and Barnhart (1961: 27), where they ask:

Now someone may ask whether the spelling of knit with k does not serve to distinguish this word from nit ‘the egg of a louse’. Of course it does, and this is exactly where our writing lapses from the alphabetic principle back into the older scheme of word writing. Alphabetic writing, which indicates all the significant speech sounds of each word, is just as clear as actual speech, which means that it is clear enough. Word writing, on the other hand, provides a separate character for each and every word ... Our spelling the verb knit with an extra <k> (and the noun nit without this extra <k>) is a step in the direction of word writing.

Sampson (1985: 203) proposes a similar idea when he suggests:

We may see another kind of method in the apparent madness of our spelling, though, if we avoid letting ourselves be obsessed by the phonographic origins of the Roman alphabet and think of English spelling as at least partly logographic.

Irregular or unpredictable spellings thus become simply a way of encoding the fact that two words that have the same pronunciation are really distinct lexical entries. Thus pairs like Bloomfield and Barnhart’s *knit* / *nit*, or *knave* / *nave*, *knight* / *night*, *brake* / *break*, *see* / *sea*, and so forth can be motivated by a general desire to keep different words spelled differently. Of course the system is hardly consistent: <crane> spells both the bird and the construction equipment; <bank> a riverbank or a financial institution; <mole> is an animal, or a chemical unit of measurement. And there are also many distinct words that share the same spelling but have different pronunciations: <bass> is either a musical instrument or range (/beɪs/) or a fish (/bæs/); <mole>, again, is either of the above-mentioned meanings (/mol/) or a Mexican sauce (/mole/). The objection that the mechanical crane is etymologically derivable from the bird name and therefore *should* be spelled the same can be countered by observing that orthographies that have more systematic semasiographic components as part of their logographic system, like Chinese, often will use different semasiographic components in two words even if they are transparently related. Thus in Mandarin, *pípa* is either a loquat or a type of lute that looks like a loquat; see Figure 3.4. Even though the words are clearly related,<sup>8</sup> the fruit / tree is written 枇杷 using the tree radical 木 that we saw above, whereas the instrument is written 琵琶 using the double-jade radical 王王 commonly used for musical instruments. No such pressures exist in English since the logographic aspect of English is by no means as systematic as it is in Chinese.

Granting that there is something to Bloomfield and Barnhart’s and Sampson’s idea, it is worth noting that such an idea is more easily accommodated to a planar taxonomy of writing systems than it is to Sampson’s arboreal system. In the latter, one must shoehorn English under either logographic, or phonographic systems, neither of which it fits into exactly. In a planar system, on the other hand, every writing system is basically one or another kind of phonographic system, and at the same time has a greater or lesser degree of logography. English merely has more logography than many languages that use an alphabetic script.