



Introductory Linguistics for Speech and Language Therapy Practice

JAN McALLISTER AND JIM MILLER

WILEY Blackwell

Introductory Linguistics for Speech and Language Therapy Practice

Jan McAllister

Senior Lecturer, University of East Anglia

Jim Miller

Emeritus Professor, University of Edinburgh



WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

This edition first published 2013

© 2013 by John Wiley & Sons, Ltd

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester,
West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
350 Main Street, Malden, MA 02148-5020, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

The contents of this work are intended to further general scientific research, understanding, and discussion only and are not intended and should not be relied upon as recommending or promoting a specific method, diagnosis, or treatment by health science practitioners for any particular patient. The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of fitness for a particular purpose. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of medicines, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each medicine, equipment, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. Readers should consult with a specialist where appropriate. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising herefrom.

Library of Congress Cataloging-in-Publication Data

McAllister, Jan.

Introductory Linguistics for Speech and Language Therapy Practice / Jan McAllister, Senior Lecturer, University of East Anglia ; Jim Miller, Emeritus Professor, University of Edinburgh.
pages cm

ISBN 978-0-470-67110-8 (pbk. : alk. paper) – ISBN 978-1-118-57193-4 (emob) –
ISBN 978-1-118-57194-1 (epub) – ISBN 978-1-118-57195-8 (epdf) 1. Speech therapy–Practice.
2. Speech therapists. 3. Linguistics–Study and teaching. 4. Language disorders–Treatment.

I. Title.

RC428.5.M42 2013

616.85'50068–dc23

2012047723

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

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover image: © iStockphoto.com/piccerella

Cover design by Andy Meaden

Set in 10/12pt Sabon by Aptara® Inc., New Delhi, India



Contents

1	Introduction	1
1.0	Why do speech and language therapy students need to study linguistics?	1
1.1	Why do speech and language therapy students need this book?	2
1.2	Aims of this book, and what this book will <i>not</i> aim to do	3
1.3	Some preliminaries	4
1.3.1	Levels of description in language	5
1.4	How this book is organised	8
1.5	Exercises	9
	Exercises using clinical assessments	9
	Further reading	9
2	Words and Non-words	10
2.0	Introduction	10
2.1	Why do SLTs need this knowledge?	11
2.2	Learning objectives	12
2.3	Words, word-forms and lexemes	12
2.4	Testing word processing and related abilities	14
2.5	Principles of selection of items in clinical resources	15
2.6	Words and non-words	22
	Chapter summary	23
	Exercises using clinical resources	24
	Further reading	24
3	Word Meaning	25
3.0	Introduction	25
3.1	Why do SLTs need this knowledge?	26
3.2	Learning objectives	26
3.3	Reference and sense	27
3.4	Lexical semantics	28
3.4.1	Lexemes, categories and concepts	28
3.4.2	Decomposition of word meaning	29
3.4.3	Basic concepts	31
3.4.4	Semantic fields	32
3.4.5	Sense relations	34
3.4.6	Lexical ambiguity	37

Chapter summary	38
Exercises using clinical resources	39
Further reading	40
4 Sentence Meaning	41
4.0 Introduction	41
4.1 Why do SLTs need this knowledge?	41
4.2 Learning objectives	42
4.3 Sentence-level meaning	42
4.3.1 Semantic roles	44
4.3.2 Reversibility	47
4.3.3 Contradiction	49
4.3.4 Syntactic ambiguity	50
Chapter summary	51
Exercises using clinical resources	51
Further reading	52
5 Parts of Speech	53
5.0 Introduction	53
5.1 Why do SLTs need this knowledge?	54
5.2 Learning objectives	54
5.3 Identifying parts of speech	55
5.3.1 Content words and function words	55
5.3.2 Why is it sometimes difficult to identify parts of speech in English?	55
5.3.3 Traditional, meaning-based criteria	57
5.3.4 Form-based criteria	57
5.3.5 Tests for parts of speech	61
Chapter summary	83
Exercises using clinical resources	83
Further reading	84
6 Word Structure	85
6.0 Introduction	85
6.1 Why do SLTs need this knowledge?	86
6.2 Learning objectives	86
6.3 Words and morphemes	87
6.4 Free and bound forms	89
6.5 Allomorphs	90
6.6 Common mechanisms of word formation in English	92
6.6.1 Compounding	92
6.6.2 Affixation	94
6.7 Mean length of utterance in morphemes	104
Chapter summary	104
Exercises using clinical resources	105
Further reading	106

7	Sentence Structure 1: Phrases and Clauses	107
7.0	Introduction	107
7.1	Why do SLTs need this knowledge?	108
7.2	Learning objectives	109
7.3	Syntax, morphology and the lexicon	109
7.4	Phrases	110
7.4.1	Criteria for phrases	111
7.4.2	Types of phrases	112
7.4.3	Heads, modifiers and phrase types	113
7.4.4	Phrases: noun phrases	114
7.4.5	Phrases: prepositional phrases	115
7.4.6	Phrases: adjective phrases	116
7.4.7	Phrases: adverb phrases	116
7.4.8	Phrases: verb phrases	117
7.5	Clauses	118
7.5.1	Elements of the clause	119
7.5.2	Elements of the clause: verb	121
7.5.3	Elements of the clause: subject	121
7.5.4	Elements of the clause: object	123
7.5.5	Elements of the clause: complement	127
7.5.6	Elements of the clause: adverbial	129
7.5.7	Phrases and grammatical functions	131
7.5.8	Notation for clauses	131
	Chapter summary	133
	Exercises using clinical resources	133
	Further reading	134
8	Sentence Structure 2: Constructions and Main Clauses	135
8.0	Introduction	135
8.1	Why do SLTs need this knowledge?	136
8.2	Learning objectives	137
8.3	Declarative, interrogative and imperative clauses	137
8.3.1	Basic constructions: declarative clauses	138
8.3.2	Imperative clauses	139
8.3.3	Interrogative clauses: yes–no interrogatives	140
8.3.4	Interrogative clauses: wh interrogatives	141
8.3.5	Interrogative clauses: tag questions	141
8.4	The active, passive and middle constructions	143
8.4.1	Active clauses/constructions	143
8.4.2	Passive clauses/constructions	144
8.4.3	Middle clauses/constructions	145
8.5	Negation	147
	Chapter summary	151
	Exercises using clinical resources	152
	Further reading	152

9	Sentence Structure 3: Subordinate Clauses and Sentences	153
9.0	Introduction	153
9.1	Why do SLTs need this knowledge?	153
9.2	Learning objectives	154
9.3	Main and subordinate clauses	154
9.4	Recognising different types of subordinate clause	155
9.4.1	Relative clauses	155
9.4.2	Adverbial clauses	156
9.4.3	Complement clauses	157
9.5	Clauses that modify nouns	159
9.6	Optional extra on relative clauses	161
9.7	Sentences	165
	Chapter summary	166
	Exercises using clinical resources	168
	Further reading	168
10	Sentence Structure 4: Non-finite Clauses	169
10.0	Introduction	169
10.1	Why do SLTs need this knowledge?	170
10.2	Learning outcomes	170
10.3	Infinitives and (Type 1) gerunds	170
10.3.1	Understood subject and control	172
10.4	Type 2 gerunds	173
10.5	Bare-verb clauses	175
10.6	Free participle clauses	175
10.7	Reduced clauses	176
	Chapter summary	179
	Exercises using clinical resources	180
	Further reading	180
11	Language in Use 1: Deixis and Reference	181
11.0	Introduction	181
11.1	Why do SLTs need this knowledge?	181
11.2	Learning objectives	182
11.3	Reference and deixis	182
11.3.1	Deixis: introduction	183
11.3.2	Deixis: personal pronouns	184
11.3.3	Deixis: demonstratives	186
11.3.4	Deixis: spatial expressions	188
11.3.5	Deixis: verbs of movement	189
11.3.6	Deixis: temporal expressions	191
11.3.7	Deixis: tense	193
11.3.8	Deixis: extended uses	194
11.4	Reference	197
	Chapter summary	202
	Exercises using clinical resources	202
	Further reading	203

12	Language in Use 2: Frames and Scripts, Metaphor and Irony, Implicature, Explicature	204
	12.0 Introduction	204
	12.1 Why do SLTs need this knowledge?	205
	12.2 Learning outcomes	206
	12.3 Frames and scripts	206
	12.4 Metaphor, irony, humour	209
	12.5 Implicature	211
	12.6 Explicature	216
	12.7 Presupposition	219
	Chapter summary	221
	Exercises using clinical resources	222
	Further reading	222
13	Language in Use 3: Speech Acts, Conversation	223
	13.0 Introduction	223
	13.1 Why do SLTs need this knowledge?	224
	13.2 Learning objectives	224
	13.3 Frames, scripts and norms	224
	13.4 Speech acts	226
	13.5 Conversation: scripts and routines	233
	13.5.1 Turn taking	233
	13.5.2 Adjacency pairs	234
	13.5.3 Expansions	235
	13.5.4 Preference organisation	236
	13.5.5 Opening a conversation	237
	13.5.6 Politeness	238
	13.5.7 Topic management	238
	13.5.8 Repair	241
	13.6 Paralinguistic signals	244
	Chapter summary	246
	Exercises using clinical resources	246
	Further reading	247
14	Narrative 1: Introduction – Coherence and Cohesion	248
	14.0 Introduction	248
	14.1 Why do SLTs need this knowledge?	249
	14.2 Learning objectives	249
	14.3 Tasks for speakers and writers	250
	14.4 Concepts for analysing narrative	250
	14.5 Spoken text	252
	14.6 Coherence and cohesion	254
	14.6.1 Coherence	254
	14.6.2 Cohesion	255
	Chapter summary	259
	Exercises using clinical resources	260
	Further reading	260

15	Narrative 2: Given and New, Theme, Focus	261
15.0	Introduction	261
15.1	Why do SLTs need this knowledge?	261
15.2	Learning objectives	262
15.3	Given and new information	262
15.4	Theme	264
15.5	Focus	270
15.5.1	Focus: tonic accent	270
15.5.2	Focus: IT clefts	272
15.5.3	Focus: WH clefts	274
15.5.4	Focus: TH clefts	275
15.6	Conclusion	276
	Chapter summary	277
	Exercises using clinical resources	278
	Further reading	278
16	Narrative 3: Syntax and the Organisation of Text	279
16.0	Introduction	279
16.1	Why do SLTs need this knowledge?	279
16.2	Learning objectives	280
16.3	Tense and aspect in narrative	280
16.3.1	Tense and aspect: reminder of the main points	280
16.3.2	Simple Past versus Progressive in narrative	281
16.3.3	Perfect	281
16.3.4	Tense and aspect: problems of usage	282
16.4	Analysing clinical test narratives	283
16.4.1	The runaway tractor story	283
16.4.2	The safari park story	286
16.4.3	The sports day story	286
16.4.4	The squirrel story	287
16.5	Some loose ends	288
16.5.1	End weight	288
16.5.2	End weight: extraposition	288
16.5.3	Passives in narrative text	290
16.5.4	Non-finite clauses	292
	Chapter summary	294
	Exercises using clinical resources	294
17	Conclusion	297
	Appendix A: Islands of Reliability for Determining Parts of Speech	299
	Appendix B: SLT Resources Referenced in This Text	302
	Answers to Exercises	306
	Annotated Bibliography	343
	Index	345

1

Introduction

This book is a practical introduction to the aspects of linguistics that generalist speech and language therapists (SLTs) need to understand in order to be able to use the published tools that are available for analysing clients' language abilities. Linguistics is the study of the organising principles of language. It is concerned with language in general, not with specific languages, although from a practical point of view we may wish to apply the principles to a particular language (here, English).

If you are reading this book you are probably interested in Speech and Language Therapy. Perhaps you are studying on a degree course leading to registration as an SLT – if so, one would certainly hope that you are interested in Speech and Language Therapy! Or perhaps you are considering enrolling on such a degree, or are a qualified SLT already but want to brush up on your language analysis skills. Whatever your interest, read on!

1.0 Why do speech and language therapy students need to study linguistics?

We said in the previous section that this book was about linguistics, but some students seem to wonder why, when they have enrolled on a degree in Speech and Language Therapy, they need to spend time studying linguistics, which is, after all, a separate academic discipline which can be studied to degree level in its own right. A short answer to this question is that professional bodies say that students need this knowledge. The Royal College of Speech and Language Therapists (RCSLT), which is the professional body that provides leadership and sets professional standards for SLTs in the United Kingdom, produces curriculum guidelines that specify the knowledge bases that speech and language therapists must acquire before registration, and one academic discipline that they stipulate is general linguistics. This is in addition to the related subjects of phonetics, psycholinguistics, sociolinguistics, bilingualism and language development.

The RCSLT curriculum guidelines are drawing attention to the fact that linguistics is one of the key underpinning disciplines of Speech and Language Therapy, alongside

subjects like anatomy and physiology, psychology and sociology. A newly qualified speech and language therapist who had not mastered these different disciplines to the required level would struggle to carry out their role.¹ Understanding the structure and function of language systems is as fundamental to the job of an SLT as understanding the structure of the human body is to the job of a doctor. It would be strange, not to say worrying, to be treated by a doctor who did not know about the structure or function of the human body, and it would be just as strange and worrying to be treated by an SLT who did not know about the structure and function of language.

1.1 Why do speech and language therapy students need this book?

Learning linguistics in the context of a Speech and Language Therapy degree presents various challenges. Like most students entering higher education today, those studying Speech and Language Therapy typically arrive with little prior knowledge of linguistics or even basic grammar. Many ‘introductory’ texts assume that students already have an accurate understanding of, for example, part-of-speech labels, but for many, this is not the case. Also, though a practical understanding of linguistics is fundamental to Speech and Language Therapy practice, it competes for space in the curriculum with the other knowledge bases that students need to acquire during their degree (life sciences, clinical skills, etc.). Given that they must acquire quite sophisticated linguistic knowledge in a very limited time, students need to focus on just those aspects of linguistics that they will need to master in order to be able to carry out their clinical work, rather than equally interesting but less relevant topics that are covered in many introductory linguistics texts, such as language evolution, historical linguistics or non-human communication.

There are some texts that omit these peripheral subjects and just focus specifically on the core areas of linguistics. These are typically written for students on linguistics degree courses or people with an interest in linguistics *per se*, rather than as one of several disciplines underpinning their main focus of study. As such, they usually provide more detail than is strictly necessary for an introductory text for SLTs, often illustrating them with examples from a wide range of languages. This is not to say that the material covered in these texts is not interesting and important in its own right; this approach is appropriate where the aim is to provide a wide-ranging and in-depth understanding of the subject.

But the needs of students on Speech and Language Therapy courses are different. A comparison of core linguistics texts with assessments and other clinical resources that are routinely used by SLTs shows that the latter focus on only a subset of

¹Admittedly, the remit of specialist SLT posts may involve tightly focused areas that do not require linguistic analysis skills. Specialisms in dysphagia or motor speech disorders are two areas that spring to mind. But SLTs who occupy such posts generally only do so after working for some time as a generalist, using the wider spectrum of SLT skills.

the structures and concepts covered in the former. Most of these resources also focus on the English language; whilst we acknowledge that many SLTs work in multicultural and multilingual settings, the one language that students on UK speech and language therapy courses absolutely must be able to describe is English. Given the limited amount of time that students of speech and language therapy have to devote to the linguistics strand of their degree, they need a text that will focus tightly on the core structures that are examined in the most commonly used clinical resources, and for students graduating in the United Kingdom, the language that they most need to know about is English.

1.2 Aims of this book, and what this book will *not* aim to do

The aim of the book will therefore be to provide the student with a practical introduction to those core linguistic concepts that are most often the subject of clinical resources and to illustrate these concepts with examples from English. As an introductory text with this applied aim, it will avoid reference to formal linguistic models and engagement with current controversies in the field of linguistics. By introducing the concepts and terminology of traditional linguistic description alongside those employed within speech and language therapy (where these differ), it will enable students to explore the subject in more detail using more advanced texts.

The book will focus on the core areas of (to use traditional linguistic terminology) morphology, syntax, semantics, discourse and pragmatics. It will not attempt to cover phonology, which is traditionally considered part of ‘speech’ rather than ‘language’ by SLTs.² Similarly, we will not attempt to provide in-depth coverage of any psycholinguistic or sociolinguistic concepts, although we will refer to concepts from these fields where necessary to elucidate relevant material. It is beyond the scope of the book to analyse languages other than English, and such topics can be dealt with more effectively in more specialised texts.

The focus of this book is the techniques of core linguistics that are needed to carry out analysis of disordered language, rather than the products of such analysis. The latter is the province of clinical linguistics, and the interested reader is referred to the further reading list at the end of this chapter. This decision has motivated the choice of language data that we analyse in this book. For example, occasionally in exercises we have used picture materials to generate extended pieces of language for analysis. For the reasons just stated, we have deliberately not attempted to reproduce the kinds of disordered samples that SLT clients would be likely to produce in response

²Phonology is a linguistic system just as much as morphology, syntax and the other topics that we will cover. We note that in the RCSLT Curriculum Guidelines published in 2010, phonology is listed under General Linguistics, and that it is an area assessed in the CELF-4; but, for example, phonology is grouped with ‘articulation’ rather than ‘language’ in the RCSLT Clinical Guidelines (Section 5.3). It is not possible to cover phonology without first providing at least an outline account of phonetics; so inclusion of phonology would increase the size of the text considerably. So although phonology is certainly part of theoretical linguistics, we decided, in the interests of keeping the text to a manageable size, to exclude this topic.

to such materials. Instead, we have chosen to illustrate the concepts that we are trying to impart by using examples from non-disordered language, reasoning that this is appropriate for an introductory linguistics text. Once readers have acquired these concepts, they are in a position to understand texts that focus on disordered samples.

Many of the examples that we have used are modelled on those found in clinical resources. We have avoided directly quoting the resources themselves because of copyright restrictions. It should be borne in mind that the specific examples used in the clinical resources have been validated using normative data; our examples, though they imitate the structure of these resources, have not been normed in this way.

Linguistics is a huge subject area, and as we noted earlier, introductory texts cover a great deal that is not directly relevant to speech and language therapy. In choosing which topics to cover in this text, we have been guided by the concepts that are addressed in speech and language therapy resources. Where we judged it necessary, we have explored additional topics that are not directly addressed in clinical resources, to provide relevant background to those topics that are addressed, or to set them in a wider context.

We recognise that there is a great deal more to communication than we will aim to cover in this text. Particularly when language is disrupted, non-verbal means of communication such as gesture are potentially extremely important, both to the client, and from a professional point of view, to the SLT. But we are concerned here with language, not communication. Communication is a means of conveying meaning (social, emotional, transactional, informational, etc.) between individuals. Language is one form of communication, albeit a highly intricate form that is characterised by a large number of complex rules.

This book will not attempt to impart any clinical knowledge at all. In particular, it will not try to teach students how to administer, score or interpret specific assessments or use therapy materials. For this, students must use the guidance of their clinical educators and the manuals that are provided with clinical resources.

1.3 Some preliminaries

We will continue this chapter by noting a few fundamental assumptions that should be stated explicitly before we begin.

A first observation concerns the notion of ‘standard English’. Our main focus here is core linguistics, and we are using data modelled on the material that will be encountered in the widely available clinical resources that SLTs use. In these resources, target items are almost always framed in terms of a notional ‘standard English’ – the variety of English that is considered not to contain particular dialectal variants. It is obviously of great importance that students of Speech and Language Therapy should acquire theoretical concepts and practical knowledge and skills relevant to language variation, and indeed sociolinguistics, whose province these concepts are,

is a curriculum requirement for Speech and Language Therapy courses. It is not, however, the topic of this book. The core linguistic concepts that we wish to elucidate here can perfectly well be illustrated using standard English, and that is what we will do. But readers should not conclude that this is the whole story.

Clinically, it is important to distinguish between the **processing** (mental operations) involved in **receptive** language (**comprehension**) and **expressive** language (**production**). Although students need to understand the psycholinguistic frameworks that differentiate these activities, the descriptive linguistic frameworks that are relevant are common to them both. Similarly, we are mainly focusing on aspects of language that are common to both the spoken and written forms.

We should also draw the distinction between **competence**, the body of abstract knowledge that a speaker has about the way that language works, and **performance**, the way that a piece of language is produced on a particular occasion, when it is subject to competing processing demands, limitation of resources such as working memory and so on. Our goal here is to provide a background to enable the reader to understand the model of competence that is encapsulated in the resources that SLTs use.

1.3.1 Levels of description in language

Having set out the scope of this book, we now turn to a preliminary discussion of the way that a subject as broad as language can be split into manageable components.

1.3.1.1 *Three aspects of language*

We suggested earlier that language, as one form of communication, was a means of conveying meaning. In fact, it is useful to distinguish at least three aspects of language of which meaning is just one. Let us start by cutting the linguistic cake three ways. We can think of language as involving **meaning**, **form** and **function** (alternatively, **meaning**, **form** and **use**).

Many words in English have more than one **meaning**. This ambiguity is the source of many children's jokes: *Why wouldn't the elephant travel by train? – Because his trunk wouldn't fit in the luggage rack; Why do you always stand on a chair when you sing? – So that I can reach the high notes.*

The **form** dimension involves what words or sentences look or sound like. If the form of a word is disrupted, it may be mispronounced or misspelt. Sentence form may be disrupted by words being used in the wrong order. Think of the way that the Star Wars character Yoda speaks: he consistently produces utterances that violate the standard form of English, saying things like *Take you to him I will*, or *Help you I can*. His meaning is clear, but there is no denying that his form is non-standard.

A speaker may select words with the meaning that they want to convey and pronounce all the words correctly and in the right order, with nothing omitted, and yet the utterance may still be put to an inappropriate **use** or may be interpreted by the

listener as having a **function** other than the one that the speaker intends. Misuse of function often occurs in comedy. In the US sitcom ‘The Big Bang Theory’, one of the characters, Sheldon, is very brainy but struggles with many social conventions including the interpretation of sarcasm. On one occasion he irritates his neighbour, Penny, in several ways including claiming that she snores. ‘You might want to see an otolaryngologist’, says Sheldon, then realising that she does not know what that word means, adds ‘... a throat doctor’. Penny replies, ‘What kind of doctor removes shoes from asses?’, to which Sheldon helpfully responds ‘Depending on the depth, that’s either a proctologist or a general surgeon’. There is no problem here with the meanings of the individual words, or with the form of the words or the sentences. Instead, Sheldon interprets Penny’s utterance as a request for information, rather than recognising its intended function, namely a threat. One could argue that this example is about the ‘meaning’ of Penny’s utterance, but it is not ‘meaning’ in the same sense as was being used earlier with the children’s joke examples, where the alternative meanings of *trunk* or *high* could be found in a dictionary. In the Sheldon/Penny example, there is no dictionary where we can look up the appropriate interpretation of Penny’s utterance – it is all down to the context in which the utterance occurs and Sheldon’s failure to recognise the absurdity of having a doctor whose role is to remove shoes from asses.

These are light-hearted examples, but the three-way distinction between meaning, form and function has a serious point in the context of speech and language therapy, because each of these dimensions can be separately affected in clients with language problems. An example of a clinical problem with meaning might occur when a client is unable to select a word with the intended meaning but selects a meaning-related word instead, such as saying *son* for *daughter*. Clients present with various types of problem with form, from difficulties with articulating speech sounds to omission of words or parts of words (e.g. saying *I go now* instead of *I am going now*) or putting words in the wrong order (e.g. *What you are doing?* instead of *What are you doing?*). A client may correctly interpret the form and word-by-word meaning of *Who do you think you’re looking at?* but not realise that it functions as a threat.

Exercise 1.1

Discuss these examples in terms of meaning, form and function/use:

Stan: I’ve got free tickets for the new James Bond movie – would you like to come?

Archie: Is the Pope a Catholic?

The former President of the United States, George W. Bush, was often satirised for errors that he made when speaking in public. On one occasion he complained about ‘rumours on the Internets’.

Someone intends to say ‘Atlas carried the world on his shoulders’ but says instead ‘Atlas carried the world on his elbows’.

1.3.1.2 A more detailed characterisation of language

The meaning/form/function characterisation of language is useful, but it is not fine-grained enough for all purposes. Linguists identify a larger number of dimensions to language and these dimensions are also recognised in the literature and clinical resources that SLTs use, though the labels that are used sometimes vary. The following are the main elements of the language system as they are conceptualised in this approach:

- **Phonetics:** the physical characteristics of the sounds that are used in language.
- **Phonology:** the sound system of the language.
- **Prosody:** This refers to the sound level of language, and it is related to both phonetics and phonology, but it refers to aspects such as intonation, the ‘melody’ of spoken language, and stress pattern. For example, the word *trusty* meaning ‘able to be trusted’ is stressed on the first syllable, while *trustee* meaning ‘official of a trust’ is stressed on the second syllable but is pronounced in exactly the same way in other respects.
- **Lexicon:** The store of words that the person knows. For each word that a person knows, specifications of the meaning, pronunciation, spelling and grammatical properties are stored in the lexicon.
- **Morphology:** This level of representation is concerned with the internal structure of words. The basic units of morphology are called **morphemes**. Our knowledge of the morphology of English tells us that the word *clinical* consists of two morphemes, *clinic* and *-al*, and that these two elements must occur in this order; it also tells us that the word *clinical* is related to the words *clinic* and *clinician*. It also allows us to work out the meanings of new words that we have not encountered before, that are made up of morphemes that we already know. For example, the term *Bushisms* was coined to refer to the speech errors produced by George W. Bush, making use of three existing morphemes combined in a way that allowed people who had never seen the word before to work out what it meant.
- **Syntax:** This level is concerned with the way that words are combined to produce phrases (e.g. *a whale*), phrases are combined to produce clauses (e.g. *a whale is a mammal*) and clauses are combined to produce sentences (e.g. *A shark is a fish but a whale is a mammal*). It tells us that *The nurses are going on strike* is a permissible sentence of English but that randomly ordering the words to say *Nurses the on going are strike* is not. It also tells us when the use of particular form of a word is dictated by features of another word in the sentence, so that we can recognise that **The nurse are going on strike* contains an error (and how to fix it). Note that in the previous example we have adopted the convention of indicating an unacceptable form by placing an asterisk before it.
- **Semantics:** This system is concerned with the meaning of individual words, phrases and sentences. Our knowledge of semantics tells us, among other things, that *walk*, *skip* and *run* are words related in meaning, that *big* and *small* have opposite meanings, and that *Eric is taller than Iain* cannot be true at the same time as *Iain is taller than Eric*, if *Iain* and *Eric* refer to the same entities in both sentences.

- **Discourse:** This term is used in various disciplines that focus on language, but here it is used to explain the way that longer sequences of sentences, such as paragraphs, are structured. For example, in paragraphs it is unusual to continue to refer to individuals by their full name after they have first been mentioned; in subsequent sentences we might use *he* or *she*. In SLT, the term **narrative** is often used when we are concerned with longer sequences of sentences.
- **Pragmatics:** This label is given to a wide range of phenomena that relate to language in use. It includes language function as discussed in the previous section, and this aspect of pragmatics explains the way that we use language to make jokes, be sarcastic, pay a compliment, apologise and so on. It is also concerned with the rules of conversation.

In addition, we could identify **orthography**, the spelling system of the written language. Though many of the aspects of language that we will consider in this book are common to its spoken and written forms, there are some instances where the two differ, and we will identify these as we go along.

Exercise 1.2

Use the terms that we have just introduced to comment on the following examples:

An advertisement of a Mazda car (with a picture of a car parked on the driveway leading to a large mansion): ‘The perfect car for a long drive’.

George W. Bush, who was mentioned earlier, claimed soon after his election that ‘They underestimated me’.

Two more Bushisms: ‘Rarely is the question asked “is our children learning?”’; ‘The literacy level of our children are appalling’.

Judy is sitting in the living room huddled in front of the fire when Howard walks in and sits down, leaving the door wide open. Judy says ‘Were you born in a barn?’ Howard gets up and closes the door.

1.4 How this book is organised

Each chapter will provide an introduction to a specific area of linguistic description that is relevant to the work of a generalist SLT. Chapter 2 will consider the issues relevant to evaluating processing of words and non-words. Chapters 3 and 4 look at semantics, at the word and sentence levels, respectively. Chapter 5 takes a preliminary look at parts of speech. Chapter 6 considers some relevant aspects of morphology (word structure). Chapters 7 to 10 are concerned with syntax (sentence structure). Chapters 11 to 13 are concerned with pragmatics (language in use). Chapters 14 to 16 look at the issues relevant to analyses of discourse or narratives – extended pieces of language.

1.5 Exercises

The text of each chapter is interspersed with practical exercises to help readers to consolidate their learning. At the end of each chapter is a set of exercises that draw on clinical resources (assessments and therapy packages) that readers may be able to access. We also constantly refer to such resources during chapters to show readers that the concepts that we are presenting are relevant to clinical practice. Because many resources have long and complicated names, clinicians tend to refer to them by abbreviations, and where such abbreviations are widely used, we have employed them in the chapter as well. In case students are unfamiliar with them, at the start of each of the following chapters we have listed in full the names of all the resources that we will reference in that chapter, along with abbreviations where appropriate. Full details of the resources that we have referenced are in Appendix B.

Exercises using clinical assessments

- 1.3. The Children's Communication Checklist (CCC-2) is a parent-completed screening instrument that is used to help the clinician to identify areas that need further investigation. Have a look at the following items in the response booklet: 12, 15, 19, 36, 43, 54, 55. Which of the areas listed below do they address (a single item may address more than one)?
- Lexicon
 - Morphology
 - Syntax
 - Semantics
 - Pragmatics
- 1.4. For each of the areas identified in your answer to the previous question, identify an assessment that explores that area in more detail.

Further reading

Introductory text for psycholinguistic topics that will not be covered in detail in this book: Fernandez and Cairns (2011).

To get a feel for the field of Clinical Linguistics, consult the journal *Clinical Linguistics* or the textbook of the same name by Cummings (2008).

2

Words and Non-words

Clinical resources that will be referenced in this chapter:

Boston Naming Test
BPVS – British Picture Vocabulary Scale
CAT – Comprehensive Aphasia Test
CELF-4 – Clinical Evaluation of Language Fundamentals
CNRep – Children’s Test of Non-word Repetition
ERB – Early Repetition Battery
PPVT – Peabody Picture Vocabulary Test
PALPA – Psycholinguistic Assessment of Language Processing in Aphasia
RDLS – Reynell Developmental Language Scales
Renfrew Word Finding Vocabulary Test
VAN – Verbs and Nouns Test
Vocabulary Enrichment Intervention Programme
Western Aphasia Battery

2.0 Introduction

Most native speakers of English, if asked ‘what are the basic units of language?’, would probably say ‘words’. Certainly individual words can be used to great communicative effect. Think of the first utterances of small children, which always consist of single words – ‘More!’, ‘No!’ or ‘Juice!’; despite being so rudimentary, they communicate their message in a very direct way.

In this chapter we will explore the factors that are relevant to the clinical evaluation of a client’s knowledge of words. Although we suggested in the previous chapter that we would be focusing on clinically relevant aspects of linguistics and in general ignoring other related topics that are often covered in introductory linguistics texts,

the factors that are relevant to a discussion of the word knowledge are just as much the province of psycholinguistics, and we will touch on some relevant background to this subject to clarify our discussion. In the interests of brevity, it will not be possible to explore the psycholinguistic background in any detail, but more comprehensive references are signposted at the end of the chapter.

We will also explore some of the kinds of resources that use words and word-like items – non-words. These kinds of materials are sometimes used as measures of language development, because a child’s ability to repeat such items is a useful index of language difficulties. Because some of the same linguistic and psycholinguistic principles apply to both kinds of resource (those that target vocabulary knowledge and those that target repetition as an index of language development) it is appropriate to explore them together in this chapter.

An important aspect of our knowledge of words is our understanding of their meanings. Models of word processing take meaning into account alongside other factors that will be considered in this chapter, but within linguistics as well as psycholinguistics a distinction is drawn between meaning and other factors. Since we will be considering word meanings in the next chapter, this aspect of word knowledge will not be considered in detail in this chapter. Here, we will focus on other factors that are relevant to the processing (production/expression and reception/comprehension) of words, and that sometimes motivate the selection of items in clinical resources.

2.1 Why do SLTs need this knowledge?

Suppose that you have a client with aphasia who has word finding difficulties. Perhaps they can produce certain words but struggle to produce others. Are there principles that can help you to predict what kinds of words are likely to be easier to produce and which will be more difficult? Or consider a 4-year-old girl with language delay; given her age, what words might you expect her to know?

The programme overview of the Vocabulary Enrichment Intervention Programme summarises the importance of this aspect of the language system for children and adolescents. During the school years, children are exposed to tens of thousands of new words. Vocabulary abilities are key to the development of literacy and other language skills. A rich vocabulary improves access to the school curriculum and promotes academic achievement. Children’s vocabulary knowledge is one of the best predictors of the likelihood of escaping the adverse impact of social deprivation early in life. Difficulties with vocabulary are among the most significant problems for children with speech, language and communication needs. Fortunately, this is an area that responds well to direct intervention.

There are many clinical resources that have been designed for assessing vocabulary knowledge and carrying out relevant interventions. If you can identify the principles that typically motivate the design of these resources, it will enhance your

understanding of them. The resources cover several hundred lexical items in total, which may seem like a respectable quantity. But this is probably only a small fraction of the number of words that a typical mature native speaker knows. Clinicians will frequently need to supplement intervention materials with items that they generate themselves, and they need to be aware of the relevant considerations when doing so. Finally, although standardised, norm-referenced resources are generally used for assessment, to some extent materials for intervention may need to be devised or adapted to suit a particular client. For example, a person with aphasia whose hobby is gardening may want to access one set of vocabulary items; another person whose hobby is photography may want to access another. But when designing such customised sets of vocabulary, it is as well to bear in mind the factors that we introduce here.

Also, as noted earlier, poor ability to repeat words and non-words is considered by many researchers to be a useful index of language difficulty. Some assessments focus on this ability, and in order to use and interpret these resources correctly, clinicians need to understand the linguistic and psycholinguistic principles that underpin them.

2.2 Learning objectives

After reading this chapter and doing the exercises listed at the end, you should be able to:

- Explain the difference between lexemes and word-forms;
- Describe some tasks that clinicians commonly use to investigate clients' ability to use words;
- Outline some of the lexical characteristics that are relevant when creating materials for use in such tasks;
- List some resources that can help you to create such materials;
- Explain why non-words are often used in clinical resources;
- Identify items in clinical resources that exemplify the concepts covered in the chapter.

2.3 Words, word-forms and lexemes

The remarks that we made in the introduction assume a common understanding of what we mean when we refer to a *word*. For example, if we asked how many words are in the sentence

The Big Bad Wolf ate two of the pigs but the third pig got away.

most readers would say that there are 15. This reflects the fact that speakers' informal definition of *word* includes the knowledge that, in continuous text, words

are the elements that are separated by spaces (though in spoken language we do not typically pause between words).

Suppose, however, that we are trying to make a list of the words that a child knows based on a recording of their interaction during play, and the utterance above was one of the ones that we recorded. In this context would we want to count *pig* and *pigs* as two different words? It depends on exactly what theoretical or clinical question we were trying to answer, but if what we really wanted to know about was the number of items in the child's vocabulary, we would probably not count them as different words. We would want to treat the list of words that the child knew as a sort of dictionary or lexicon; indeed, we refer to this list of the words that a person knows as the **mental lexicon**. Just as, in a real dictionary, we would not expect separate entries for *pig* and *pigs*, we would not count them as separate 'words' in this second sense either.

To clarify the distinction between these different senses of *word*, linguists use different terms to refer to them. A word in the 'dictionary entry' sense is termed a **lexeme**; the *lex-* part indicates 'word' (as it does in the word *lexicon*) while the *-eme* part is found in many linguistic labels (e.g. *phoneme*, *morpheme*) and indicates a unit of analysis. For the example we are discussing, linguists would say that there are at least two different **grammatical word-forms** (*pig* and *pigs*) and both associated with the same **lexeme**, **PIG**. Notice that by convention the grammatical word-forms are written in italics and the lexeme in capitals. Though SLTs would not necessarily use these terms, it is useful to be aware of the distinction, partly because it will deepen your own understanding of the concepts around word knowledge, but also for purposes of further reading in this book and elsewhere. We will, however, continue to use the term *word* unless it is useful to make more precise distinctions. In the first part of this chapter, we are mainly concerned with the lexemes that individuals know, rather than the word-forms, although we will need to refer to this distinction again later.

We have already introduced the concept of the mental lexicon, the store of words that an individual knows. It is worth saying a little more about what the mental lexicon contains. Most of the entries in the mental lexicon relate to what we would conventionally call a single word. But the mental lexicon may contain other kinds of unit as well. Occasionally phrases or even whole sentences may be stored, in cases where the group of words has a unitary, idiosyncratic meaning such as occurs in **idioms** like *under the weather* or fixed word combinations like *cold call*. The meaning of forms like this must be stored because if they were analysed by our normal phrase- and sentence-processing mechanisms, an incorrect interpretation would result. The fact that such forms are stored as unitary entities is also clear from the fact that we cannot alter them; for example, a native speaker of English would not say that someone was *beneath the weather* or *below the weather*, even though *beneath* and *below* mean roughly the same as *under*. Therefore idioms or other fixed word combinations are stored in their entirety. But usually we do not store the meanings of phrases and sentences; we work out the meaning of each one afresh when we encounter it, as we discussed in Chapter 1 and will explore in detail in later chapters.

2.4 Testing word processing and related abilities

As we noted in Chapter 1, in this book we are not attempting to explain how to conduct clinical assessments or carry out interventions; our goal is to help you to understand the linguistic structures and concepts that are addressed in these clinical resources. Nonetheless, it is useful to refer to the kinds of tasks that are used clinically, to make the subsequent discussion in this chapter less abstract. We do not attempt here to give a comprehensive picture of the range of tasks that clinicians use, but if you look at the resources that we cite in this and other chapters, you will gain an overview of them.

Some of the most commonly used methods for assessing a client's ability to produce and recognise words involve pictures (or sometimes video, particularly if the words of interest refer to actions rather than objects). To assess expressive abilities, **naming** (or **confrontation naming**) tasks are used; here the client is shown a series of pictures like the ones shown in Figure 2.1, and is asked to name each one. Examples of this kind of expressive test are the Renfrew Word Finding Vocabulary Test and the Expressive Vocabulary subtest of the CELF-4. Alternatively, a person may be asked to read aloud a printed word. This is also sometimes referred to as (word) naming.

To test receptive abilities, **picture selection** is often used; here, a set of pictures is presented simultaneously and the task is to choose one in response to an instruction.

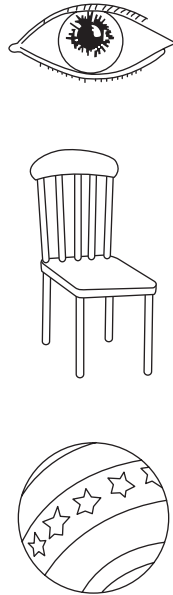


Figure 2.1 Confrontation naming. Adapted from Snodgrass and Vanderwart (1980).
© American Psychological Association.

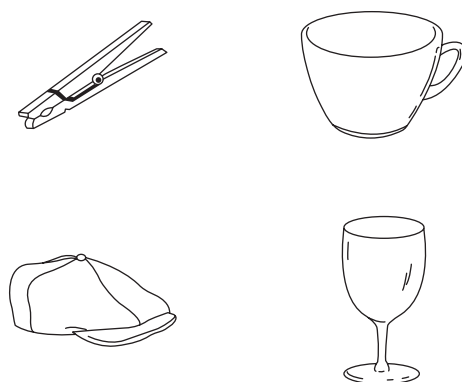


Figure 2.2 Picture selection. Adapted from Snodgrass and Vanderwart (1980). © American Psychological Association.

The clinician says (or writes, or signs) the **target** word, and the client's task is to select it (e.g. point to it) while ignoring the other items (the **distractors**). In the set of pictures in Figure 2.2, if the picture of the cup was the target, the distractors would be the glass, the cap and the peg. The PPVT and the BPVS are examples of this kind of test.

A third kind of task that is used is **repetition**. Such tasks may involve words, non-words, sentences or other sequences, though we are particularly interested in the first two kinds of item in this chapter. As was noted earlier, the ability to perform such repetition accurately is a good index of language development and processing ability. Examples of clinical instruments that use repetition are the ERB and the CAT.

2.5 Principles of selection of items in clinical resources

Only a few of the resources that clinicians use to work on lexical knowledge give an explicit account of the factors that motivated their choice of word items. Here, we will consider the sorts of factors that are relevant.

For picture- or video-based tasks, an obvious requirement is that it should be possible to illustrate the target word in some way. It is much easier to provide a picture that will reliably elicit the response *car*, for example, than, say, *fact* or *way*. Some words, for example *the* or *for*, are probably impossible to illustrate.

Another factor that could be relevant is how early in life a word is typically learned. This is obviously the case for paediatric assessments that require young children to produce a spoken response. It would be unreasonable to expect most 2-year-olds to know the word *system*, for example. A related issue is how early the sounds in the word are acquired, or how easy it is for a young child to produce a particular

sequence of sounds, such as the consonant sequence at the start of the word *straw*. A proper understanding of these factors requires a knowledge of phonology, which is beyond the scope of this book, but see further reading at the end of this chapter.

A further factor that seems to be implicit in item selection in many resources is familiarity; it seems obvious that a client will not be able to produce or understand a word that is unfamiliar to them. Unfortunately for anyone who aims to produce an assessment that can be used by a wide range of people, familiarity is a very individual characteristic. For example, a child who takes part in cooking activities at home might know the names of kitchen implements such as *colander* or *spatula* that are unfamiliar to some other children of the same age. A word that is very familiar to a lawyer, such as *tort*, may be unfamiliar to the average person. Nonetheless, as mature speakers we feel that we have an intuitive knowledge about what words are likely to be familiar to the majority of other speakers (though this knowledge is inevitably coloured by our own lexical experience). So writers of assessments often seem to make a guess at which items are likely to be familiar to their target user group.

Occasionally, assessment manuals provide explicit information about the factors that were involved in the choice of the items used. Examples of assessments that do so are the ERB, a paediatric assessment which includes a test of word and non-word repetition, and the VAN¹ which focuses on word-finding ability. It would be instructive to read the manuals for these assessments, since they discuss several characteristics that have been controlled in their design and which are known to influence lexical processing. Some of the factors that they have taken into account are outlined below. When generating your own intervention materials, it is appropriate to bear these factors in mind.

Imageability: This characteristic is related to the idea that was mentioned earlier that it should be possible to illustrate the objects, actions, etc., that are referred to by words used in tests. Imageability is, however, a broader, meaning-based concept that refers to the extent to which an entity can be perceived by the senses. For example, the word *pineapple* is highly imageable because it can be perceived by nearly all of the senses – it has a distinctive visual appearance, taste and smell, and it can be touched. Other words are far less imageable; it is much more difficult to pinpoint sensory experiences that we might associate with a word like *obedience* or *honesty*, for example. Though some clinical resources, such as the RDLS or the Western Aphasia Battery, provide real objects as part of the package, in practice, because of the tendency to use picture-based tasks in published SLT resources, it is nearly always the visual aspect of items' imageability that is relevant, but presumably there is nothing to stop clinicians from designing their own smell- or taste-based materials! Identifying words that can readily be represented pictorially

¹As noted at the start of the chapter, VAN stands for Verbs and Nouns Test. If you do not yet feel sure what a verb or a noun is, do not worry; this distinction, which is explored in detail in Chapter 5, is not crucial to our discussions here.

is not necessarily a straightforward task, as anyone who has played the game Pictionary can testify.

Imageability has been shown to be an important factor when the task involves the processing of written or spoken words. For example, people with aphasia find it easier to access words with high imageability in various tasks, and the RCSLT clinical guidelines recommend controlling for this factor when selecting materials.

Information about imageability has been collected by researchers in various studies in which the participants have been asked to rate particular words for this characteristic. Some references to published papers that report imageability ratings are given at the end of this chapter. Some information about imageability can be found online in publicly available resources such as the MRC Psycholinguistic Database, which can be found at http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm. This has amalgamated the ratings for words from several studies, and provides ratings ranging from 100 for low imageability to 700 for high imageability. For example, to get the database to propose a list of highly imageable words, go to the first set of tick-boxes (under 'Select the database fields to be displayed in the output') and tick 'Word' and 'Imageability rating'; then scroll down to 'Optionally set upper and/or lower limits for selected word properties' and type a minimum imageability rating, say 500, and a maximum rating, say 700; finally, scroll down to the bottom of the page, and click 'GO'. This returns a few hundred words, such as *crown* and *beach*, with imageability scores in this range. For a particular word, if the database has a score for it, you can discover what that score is by removing the upper and lower limits (i.e. 600 and 700 in the example just given) and under 'simple letter match' type in the word or set of words that you are interested in, one per line; then press 'GO'. A limitation of this resource is that for imageability it only has information for a subset of words, but the database can be used to investigate other factors for which there is more extensive information. If you invest a little time to get to grips with the way that the database works, you will find that you can use it to identify words with various combinations of the characteristics discussed in this chapter. Another resource that contains information about imageability and other characteristics is the Bristol Norms which can be accessed at http://language.psy.bris.ac.uk/bristol_norms.html

Familiarity: As noted above, familiarity is a rather individual characteristic of words: a word that is familiar to you may be unfamiliar to someone with a different cultural background, experience, age, etc. Nonetheless, if we ask large numbers of people to rate the familiarity of words, we can average these ratings and come up with a rough index of how familiar a word is likely to be. Like imageability, familiarity is a factor for which rating scores have been collected by researchers, and these scores can be accessed via some publicly available resources such as the MRC Psycholinguistic Database and the Bristol Norms as previously mentioned. For example, in the Bristol Norms, *clement* and *wigwam* receive low familiarity scores, while *dinner* and *goodbye* receive high scores.

Word frequency: We have pointed out more than once that familiarity is a very individual characteristic, but that an intuitive estimate of this factor does seem to have been a criterion in the selection of items for several tests. A more objectively measurable characteristic that relates to (but is not the same as) familiarity is word frequency. Word frequency can be defined as the number of times that a word appears in a particular sample of language. For example, in the nursery rhyme ‘Hey Diddle Diddle’ (here it is in case you don’t know it!):

Hey diddle diddle,
The cat and the fiddle,
The cow jumped over the moon,
The little dog laughed to see such fun,
And the dish ran away with the spoon.

the word *the* occurs seven times, but the word *spoon* only occurs once, so *spoon* has a lower frequency than *the* in this context. A useful distinction to make is between **types** and **tokens**. To calculate word frequency, we need to count up the number of tokens that occur in a particular piece of text; there are seven tokens of *the* in the rhyme ‘Hey Diddle Diddle’, and one token of *spoon*. The term type refers to distinct words. In this rhyme, 22 different word types, that is 22 different words, occur (*hey, diddle, the, cat, and, fiddle, cow, jumped, over, moon, little, dog, laughed, to, see, such, fun, dish, ran, away, with, spoon*). There are 30 words in total in the rhyme, so some word types are represented by more than one token (specifically, *diddle, the* and *and*). Of course, a word frequency count based on such a tiny sample would not be of any use, but as we will see below, it is possible to access word frequency information that is based on very large samples indeed.

We will say more about resources for getting word frequency in a moment, but first we will continue with our discussion of types and tokens. **Type–token ratio** (or **TTR**) is a measure that is sometimes calculated to work out how diverse a person’s vocabulary is. TTR is calculated by dividing the number of types in a text by the total number of tokens, so for ‘Hey Diddle Diddle’ the TTR is $22/30$, or 0.73, or 73% (though it is desirable to use many more than 30 tokens for an accurate result). A person with a diverse vocabulary will have a higher TTR than a person with a very restricted vocabulary. Suppose that a person produced a short description of a picture consisting of 25 tokens, but used only 5 different word types: here, $TTR = 5/25 = 0.20 = 20\%$. Suppose that another person gave a 40-word description, but they used 16 different word types: $TTR = 16/40 = 0.40 = 40\%$. So even though the descriptions were of different lengths, we can tell that the second person’s vocabulary is twice as diverse as the first person’s. Again, in a real study we would aim to use much larger language samples; although this example illustrates the concept, such small samples would not yield valid conclusions.

When word frequency is investigated in psycholinguistic and clinical studies, it is customary to use very large bodies of text containing many different language samples to determine the score for a particular word. Here, ‘text’ is used to refer to a set of samples of language of any type, spoken, written or signed. The group

of samples that is used for a word frequency count, the ‘body of texts’ that the numbers are based on, is called a **corpus**.

Psycholinguists and clinicians have found word frequency to be an important predictor of the speed and accuracy with which a word can be processed. Healthy adults and people with various kinds of acquired language problem can generally process high-frequency words more easily than low-frequency words.

There are many online resources that provide information about word frequency, although not all of them are freely accessible. One useful word frequency resource, which has the advantage of being publicly available (at least in a rudimentary form), is the British National Corpus (BNC), which can be seen at <http://www.natcorp.ox.ac.uk/>. The BNC consists of a body of text (corpus) of around 100 million words; the individual texts in the corpus were drawn from a large range of sources, gathered in the late twentieth century. The written texts include extracts from regional and national newspapers, specialist periodicals and journals for a range of ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, and school and university essays. The spoken texts, which represent about 10% of the total word count, consist of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins. If you go to the BNC link above and type the word *the* into the box labelled ‘Look up’, you will find that this word occurs 6,055,159 times in the 100 million or so words of the corpus, or about 60,000 times per million words. *Spoon* occurred only 797 times in the BNC, or fewer than 8 times per million words.²

The MRC Psycholinguistic Database, which was mentioned in the discussion of imageability, is another publicly available resource that can be used to quantify frequency, although its corpus of texts is much smaller (about one million words) and older (it was collated in the 1960s). However, the MRC Psycholinguistic Database does have the advantage that you can input a combination of factors such as the ones covered in this chapter and get it to suggest sets of words which will yield a word list with particular characteristics.

Another very useful, publicly available tool for getting an idea of word frequency is the Google Ngram Viewer, which you can find at <http://books.google.com/ngrams> (or google it!). ‘Ngram’ refers to letter combinations, or it might be more accurate to say character combinations. The corpus in this case is derived from over 5.2 million books published between 1500 and 2008, which between them cover about 500 billion words in American English, British English, French, German,

²In the publicly available version of the BNC, requesting a frequency count for *pig* will give you the number of instances where the three-letter word *pig* appears in the corpus, plus the number of instances of *pig*'s; it would not include the number of instances of *pigs*. In psycholinguistic studies of frequency effects, for many words it is relevant to take into account all the different grammatical word-forms associated with a lexeme (e.g. add together scores for *pig*, *pigs*, *pig*'s). See Further Reading.

Spanish, Russian, or Chinese. Only words that have appeared in more than 40 of the books are listed in the publicly available database. You can select specific language samples, e.g. British English, and you can restrict your search to particular time windows. You can search for individual words, and see when historically they began to be used by more than a handful of people. You can also input phrases, or sequences that contain non-alphabetic characters like *B&B*. You can compare two or more sequences by typing them with a comma between them. The Google Ngram Viewer does not deliver a numerical result, but provides a graph instead, which can be more useful than a number for some purposes.

Exercise 2.1

Use the Google Ngram Viewer at the previously given link to compare the usage of the terms *cell phone* and *mobile phone* in British English between 1980 and 2008.

Length: Length can be measured in terms of phonemes, syllables or letters. There is some evidence that shorter words are processed more easily than longer words. On the whole, shorter words tend to be of higher frequency than longer words.

Word length is one of the easier lexical characteristics to compute for yourself, though you can also use the MRC Psycholinguistic Database to suggest words of particular length in terms of phonemes, syllables or letters, perhaps in combination with particular frequency ranges.

Exercise 2.2

Use the MRC Psycholinguistic Database to identify a set of five words that are one or two syllables long and have a frequency of less than 10 per million. What are the frequencies of these words in the British National Corpus?

Neighbourhood density: Another characteristic that has been shown to affect lexical processing is called neighbourhood density, which is a measure of how similar a particular word is to other words in the lexicon. For example, thinking for the moment about the spellings of words, we could define neighbourhood density in terms of the number of words that could be produced by changing, adding or deleting a single letter. These other words form the neighbourhood of the target word, and so a word which is similar (according to this definition) to many other words is said to come from a dense neighbourhood, while a word that has few similar neighbours come from a sparse neighbourhood. The word *cat*, for example, comes from a relatively dense neighbourhood; many words can be created by changing, adding or deleting a single letter, including *sat*, *cot*, *can*, *at* and *coat*. Other words, such as *emu*, result in fewer words when we try changing, adding or deleting a single letter; *emu* belongs to a relatively sparse neighbourhood. The same approach is applied to defining neighbourhood in phonemic terms.