



OXFORD

THE OXFORD GUIDE TO
Etymology

PHILIP DURKIN

The Oxford Guide to Etymology

For my parents

The Oxford Guide to Etymology

Philip Durkin

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

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

Oxford New York

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

With offices in

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

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

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

© Philip Durkin 2009

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2009

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,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire

ISBN 978-0-19-923651-0

1 3 5 7 9 10 8 6 4 2

Contents

| | |
|---|-----------|
| <i>Acknowledgements</i> | viii |
| <i>About this book</i> | ix |
| 1 Introduction | 1 |
| 1.1 What is etymology? | 1 |
| 1.2 Some basic concepts: two example etymologies | 3 |
| 1.3 Why study etymology? | 22 |
| 1.4 What an etymologist does | 31 |
| 2 What is a word? Which words need etymologies? | 34 |
| 2.1 What are words? | 34 |
| 2.2 How new words arise | 43 |
| 2.3 Lexicalization | 49 |
| 2.4 Examples of lexicalization | 51 |
| 2.5 Apparent reversals of the process | 56 |
| 2.6 Cranberry morphs | 56 |
| 2.7 Which words need etymologies? | 58 |
| 3 Are words coherent entities? | 61 |
| 3.1 Variety in form and meaning: <i>poke</i> ‘bag, sack’ | 62 |
| 3.2 Do we know precisely when a word’s history begins? Can we assume continuity of use? | 68 |
| 3.3 Homonymy and polysemy | 74 |
| 3.4 How polysemy–homonymy relations can change | 76 |
| 3.5 Merger (or near-merger) in form and meaning | 79 |
| 3.6 Splits in word form | 83 |
| 3.7 A case of merger followed by a split | 86 |
| 3.8 Homonymic clash | 88 |
| 4 Word formation | 94 |
| 4.1 Issues concerning affixation | 95 |
| 4.2 Synonyms, nonce formations, and blocking | 103 |

| | | |
|----------|--|------------|
| 4.3 | Issues concerning compounding | 107 |
| 4.4 | Some other important processes | 111 |
| 4.5 | Arbitrary and non-arbitrary linguistic signs | 123 |
| 5 | Lexical borrowing | 132 |
| 5.1 | Basic concepts and terminology | 132 |
| 5.2 | What constitutes a borrowing from language X into language Y? | 140 |
| 5.3 | Motivation for borrowing: traditional explanations | 142 |
| 5.4 | Examples of borrowing because of 'need' | 143 |
| 5.5 | Borrowing of a new word when a new product of the natural world is encountered | 145 |
| 5.6 | Patterns of borrowing in the history of a language | 149 |
| 6 | The mechanisms of borrowing | 155 |
| 6.1 | Perspectives from contact linguistics | 155 |
| 6.2 | What is basic vocabulary? | 157 |
| 6.3 | Language shift | 161 |
| 6.4 | Borrowing within and between languages | 164 |
| 6.5 | Borrowings from more than one language | 165 |
| 6.6 | Continuing semantic influence and semantic interference | 167 |
| 6.7 | Multiple borrowings from the same source | 169 |
| 6.8 | How can we tell that borrowing has occurred? | 169 |
| 6.9 | Lexical borrowing and code-switching | 173 |
| 6.10 | Some conclusions from chapters 5 and 6 | 177 |
| 7 | Change in word form | 179 |
| 7.1 | Two Germanic sound changes | 180 |
| 7.2 | Examples of English sound changes | 182 |
| 7.3 | Sporadic sound changes | 195 |
| 7.4 | Associative change in word form | 197 |
| 7.5 | Metanalysis | 207 |
| 7.6 | How regular are regular sound changes? | 208 |
| 7.7 | Examples of arguments based on word form | 211 |
| 8 | Semantic change | 222 |
| 8.1 | Meaning change is a common phenomenon | 223 |
| 8.2 | Polysemy and meaning change | 225 |

| | |
|---|------------|
| 8.3 Semantic polygenesis | 228 |
| 8.4 Meaning change in a semantically complex word: <i>quaint</i> | 228 |
| 8.5 Influence from other words | 230 |
| 8.6 Some basic types of change | 235 |
| 8.7 Is semantic change predictable? | 243 |
| 8.8 Some practical examples | 254 |
| 8.9 Arguments based on form and meaning contrasted | 259 |
| 8.10 Etymology and extralinguistic factors | 261 |
| 9 Etymology and names | 266 |
| 9.1 How and why are names different? | 266 |
| 9.2 Two villages called <i>Harvington</i> | 269 |
| 9.3 Change in word form shown by names | 271 |
| 9.4 Which language does a name belong to? | 275 |
| 9.5 Names as evidence for lexis | 277 |
| 9.6 Names as evidence for word meaning | 279 |
| 9.7 Names as etymons | 280 |
| 9.8 Names and non-linguistic history | 281 |
| 10 Conclusion | 284 |
| <i>Glossary</i> | 288 |
| <i>Suggestions for further reading</i> | 297 |
| <i>References</i> | 300 |
| <i>General index</i> | 313 |
| <i>Index of word forms</i> | 326 |

Acknowledgements

As my 'day job' I lead the team of specialist editors researching, writing, and revising etymologies for the new edition of the *Oxford English Dictionary*. As a result my first debt of gratitude must be to all of my colleagues, past and present, for their deep expertise, stimulating questions, and very good company, over (so far) fourteen years of highly enjoyable and rewarding collaborative research.

I am very grateful to the following people and institutions for kind invitations to give lectures and papers which drew extensively on draft chapters of this book: Derek Britton and the Institute for Historical Dialectology at the University of Edinburgh; Hans Sauer and the University of Munich; Ursula Lenker and the University of Eichstätt; Hans Sauer and Ursula Lenker and the Fifteenth International Conference on English Historical Linguistics, Munich, 2008. To the audiences at each of these papers I am grateful for many helpful suggestions and observations, and not least for confirming that there is an audience for a book of this sort.

I am enormously grateful to all of those who have commented on parts of this book in draft: Kathryn Allan, Joan Beal, Alan Durant, Anthea Fraser Gupta, Christian Kay, Peter McClure, Inge Milfull, Nicholas Sims-Williams; and to the following who were generous and brave enough to read drafts of the whole book: Anthony Esposito, Meg Laing, Roger Lass, Anna Morpurgo Davies, John Simpson, and Edmund Weiner. The book as it stands today has benefitted enormously from the keen eyes and still keener intellects of all of these people. Needless to say, all errors and omissions are to be laid, with complete justice, entirely and exclusively at my own door.

I am hugely grateful also to John Davey at OUP for his help, advice, and encouragement at every stage in the process of shaping, writing, and producing this book. Elmandi du Toit, Malcolm Todd, and Lesley Rhodes provided expert input on the production of the book.

I would also like to thank Kathryn Allan a second time, for putting up with me on the many days when even brimming pots of coffee were not enough to focus the mind and drive away irritation. And finally I would like to thank the dedicatees of this book for tolerating a child's at times rather obsessive interest in very old documents and even older words.

About this book

Etymologies appeal to people with a very wide variety of interests and intellectual backgrounds. A very few people, such as myself, spend most of their time researching etymologies. A slightly larger number do so very occasionally. Many, many more people look at etymologies, but have never researched any themselves. Some people will never even have thought of etymologies as things which need to be researched. Particularly when etymologies are encountered in the compressed form found in many dictionaries, they can seem to be a given, rather than the (often very tentative) results of extensive research.

This book is intended for anyone who has taken the important first step of realizing that etymologies are the result of research, and would like to discover something about the nature of that research, and the principles and methodologies which underlie it.

I have attempted to frame this book so that it is addressed most centrally to someone who has an interest in historical linguistics, the study of how languages change and develop over time. Etymology is a part of this wider field, and anyone's understanding of etymology will be greatly enriched by at least some acquaintance with the broader concerns of the discipline as a whole. Readers who are entirely new to this field may find that they get much more out of this book if they read it in conjunction with one of the many excellent general textbook introductions to historical linguistics, such as Schendl (2001) or, in slightly greater depth, Millar (2007, which is a revised edition of Trask 1996) or Campbell (2004); for an excellent introduction to a wide variety of linguistic topics focussing on the vocabulary of English see Katamba (2005).

When deciding what to cover in this book and in how much detail, I have tried to pay particular attention to those areas which are important for etymology but which receive relatively little attention in most introductory books on historical linguistics. Nonetheless, I have also endeavoured to ensure that the book provides a balanced account of all aspects of etymology, especially for readers who are prepared to follow up references to fuller discussions of any topics which may be new or unfamiliar.

Most of my examples will be drawn from English, since this is the one language that any reader of this book will necessarily have some knowledge of. However, my aim has been to assume no particular knowledge about the history of the English language, beyond the explanations and further references given in the text. Drawing examples from the history of English also brings the advantage that I have in many cases been able to make use of very recent research for the new edition of the *Oxford English Dictionary* with which I have been involved personally.

There are no exercises, but at various points in the text I have listed further examples of the phenomena discussed, which readers can pursue if they wish in etymological dictionaries. Access to a good etymological dictionary of English would be of great benefit to anyone reading this book. In particular, access to the full *Oxford English Dictionary*, especially in its online version (www.oed.com), would be of especial benefit, so that many examples given here in summary form can be pursued in greater detail. (The dictionary can be accessed online via most institutional libraries and many public libraries.)

1

Introduction

1.1 What is etymology? 1

1.3 Why study etymology? 22

1.2 Some basic concepts: two
example etymologies 3

1.4 What an etymologist does 31

1.1 What is etymology?

As we will see in this chapter, etymology can tell us that English *friar* was borrowed from Old French *frere* ‘brother’, which in turn developed from Latin *frāter* ‘brother’. It can also tell us, perhaps rather more surprisingly, that Latin *frāter* is ultimately related to English *brother*, and that English *foot* is related to Latin *pēs* ‘foot’ and Armenian *otn* ‘foot’. Just as surprisingly, it can tell us that, in spite of the resemblance in form, English *care* and Latin *cūra* ‘care’ are definitely not related to one another, nor are Latin *deus* ‘god’ and Greek *theós* ‘god’. Etymology can also trace dramatic changes in meaning: for instance, English *treacle* originally had the meaning ‘medicine’, and comes ultimately from a Greek word which originally meant ‘antidote against a venomous bite’; *sad* originally had the meaning ‘satisfied’. How we trace such developments, and what they tell us about linguistic history, will be the topic of this book.

Etymology is the investigation of word histories. It has traditionally been concerned most especially with those word histories in which the facts are not certain, and where a hypothesis has to be constructed to account either for a word’s origin or for a stage in its history. That might be a stage in its meaning history, or in its formal history, or in the history of its spread from

one language to another or from one group of speakers to another. The term is also used more broadly to describe the whole endeavour of attempting to provide a coherent account of a word's history (or pre-history). As we will see in the course of this book, many of the basic methodological assumptions made in etymological research are the same regardless of whether we are looking at well-documented periods of linguistic history or at periods earlier than our earliest documentary records. Indeed, even someone who is primarily concerned only with attempting to solve hitherto unresolved difficulties of word history can only do so by building on the knowledge of many other word histories which have been much more securely established. For this reason, very many of the illustrative examples in this book will come from word histories which are very secure and not in any doubt, since they often provide the surest foundation for further investigation. Nonetheless, we will also look at some rather more difficult cases along the way.

Etymology forms part of the wider field of historical linguistic research, that is to say of attempts to explain how and why languages have changed and developed in the ways that they have. However, it does not concern itself exclusively with a particular linguistic level, as does for instance historical phonology (the study of speech sounds and of their deployment in ways which convey distinct meaning), historical morphology (the study of word forms as used to convey grammatical relationships), historical semantics (the study of the meaning of words), or historical syntax (the study of the meaning relations between words within a sentence). This is not to suggest for a moment that historical phonologists, morphologists, semanticists, or syntacticians never pay any attention to anything other than phonology, morphology, semantics, or syntax respectively. However, etymology is rather different, in that an individual word history will almost never be explicable in terms of only one linguistic level. Typically, some arguments or at least tacit assumptions about word form, probably involving issues of both historical phonology and morphology, will be combined with some arguments or assumptions about word meaning. In fact, etymology can be defined as the application, at the level of an individual word, of methods and insights drawn from many different areas of historical linguistics, in order to produce a coherent account of that word's history. One of the most exciting aspects of etymology is that this sort of detailed work on individual word histories sometimes throws up interesting results which can have a much broader significance in tracing the history of a language (whether that be with regard to phonology, morphology, etc.), especially when we can find

parallels across a group of different word histories. Additionally, it is often crucial that questions of (non-linguistic) cultural and intellectual history are considered in tandem with questions of linguistic history.¹

As well as using the word *etymology* as an abstract noun, we can also talk about *an etymology*, that is to say an account of a word's history. In the next section, we will look at two representative etymologies in some detail, as a practical way of introducing some basic concepts and at the same time some questions and issues which will concern us in much more detail later. The first example involves some very well-documented periods of linguistic history, while the second (which is rather more complex) will offer a first foray into historical reconstruction at a very considerable time depth. Concepts that we will explore include:

- tracing the linear history of a word
- change in word form
- change in word meaning
- borrowing
- genetic relationships between languages
- cognates
- comparative reconstruction
- sound change

1.2 Some basic concepts: two example etymologies

1.2.1 Example one: *friar*

The etymology of the English word *friar* can be sketched very crudely as follows:

Latin *frāter* 'brother'

develops into

Old French *frere* (modern French *frère*) 'brother', also 'member of a religious order of "brothers"'

which is borrowed as

Middle English *frere* 'friar'

which develops into

modern English *friar*

¹ For a short survey of previous definitions of the term 'etymology', accompanied by an adventurous attempt to formulate a fully adequate formal definition, see Alinei (1995).

The symbol ‘>’ is frequently used to stand for both ‘develops into’ and ‘is borrowed as’, and so we can represent the same development in a more ‘shorthand’ way as:

Latin *frāter* brother > Old French *frere* brother, also member of a religious order of ‘brothers’ > Middle English *frere* friar > modern English *friar*

Or we can reverse the arrows, and trace backwards from the modern English word. In fact, this is the style most frequently encountered in dictionaries and in most other scholarship:

modern English *friar* < Middle English *frere* friar < Old French *frere* brother, also member of a religious order of ‘brothers’ < Latin *frāter* brother²

The etymology of the Latin word could also be traced back a lot further than this, and can be linked ultimately with English *brother*, but this requires an acquaintance with some topics which we will investigate in section 1.2.4.

Obviously, this is a summary of a series of events in linguistic history. We will now examine each of those events in turn, and to do so we will require a little background at each stage. The Latin language is the direct antecedent of French. That is to say, French, like the other Romance languages (Portuguese, Spanish, Italian, Romanian, etc.), developed from Latin, albeit probably from a form of the language rather different from that reflected by the majority of our literary records. French also shows many borrowings and some structural influences from other languages, especially the Germanic language spoken by the Franks, but its basic line of descent is indisputably from Latin. In the vulgar Latin and proto-Romance varieties which eventually developed into French, the Latin word for ‘brother’, *frāter* (or more accurately its oblique case forms, such as the accusative singular *frātre*) underwent a number of (perfectly regular) changes in word form, resulting in Old French *frere*. Old French is the term used to denote the earliest recorded stage of the French language, up to the early fourteenth century.³ Thus we have our first step:

Latin *frāter* > Old French *frere*

² Some scholars use the symbols ‘<’ and ‘>’ only to link forms related by direct phonetic descent, and use different symbols for processes such as borrowing or derivation, but in this book I will use them to link any two consecutive stages in an etymology.

³ Unusually, in this particular case, an intermediate step in the formal development of the Old French word is recorded in the very early Old French form *fradre* preserved in the *Strasbourg Oaths*, a unique (and very short) document from the year 842 which records (partly in Latin, partly in French, and partly in German) the oaths taken by Louis the German, Charles the Bald, and their followers during a time of conflict.

frere remained the basic word in French for ‘brother’, but it also acquired a secondary meaning denoting the (metaphorical) ‘brothers’ who belonged to various religious orders. This usage in French followed similar use of *frater* in medieval Latin.⁴ The word was then borrowed into English from French. This happened in the Middle English period, the stage of the English language from roughly 1150 to 1500. More accurately, the word was borrowed from the Anglo-French variety of Old French which was used in England in the centuries after the Norman Conquest.⁵ The usual form in Middle English, *frere*, matches the French form exactly, and the pronunciation is likely to have been almost identical in Anglo-French and in Middle English. However, in Middle English the meaning is much narrower, showing only the religious sense and occasionally one or two other metaphorical uses. Thus we have our second step:

Old French *frere* brother, also member of a religious order of ‘brothers’ >
Middle English *frere* friar

It is very common for a borrowed word to show only a very restricted and possibly rather peripheral portion of its meaning when it is borrowed into another language. In this particular instance, it is easy to see why (Anglo-)French *frere* was not borrowed into English with the much more basic meaning ‘brother’: the word *brother* (inherited from the Old English period, and from the Germanic antecedent of English before that) already had that meaning and was in common use, and even in the Middle English period, when very many words were borrowed from French into English, it is relatively uncommon for words with quite such basic meanings as this to be borrowed in place of native words. We will look at this issue in more detail in chapters 5 and 6. In fact English *brother* also had the meaning

⁴ The macrons which indicate vowel length in forms like classical Latin *frāter* are not normally given when citing Latin forms from later than the classical period, although this does not necessarily indicate any change in the vowel length in particular words.

⁵ In this book I use the term ‘Anglo-French’ to denote French as used in England (and elsewhere in Britain) in the centuries following the Norman Conquest. Scholarly practice is divided in this area: ‘Anglo-Norman’ is often used to denote this variety (as in the title of the *Anglo-Norman Dictionary*), but increasingly the broader term ‘Anglo-French’ is used instead, in order to reflect better the varied inputs from different varieties of Continental French which occurred both immediately after the Norman Conquest and in the subsequent centuries: for a useful discussion and further references see Rothwell (2005). For convenience, where a form or meaning belonged to both Insular and Continental French I use the style (Anglo-)French.

‘(fellow) member of a religious order’ in the Old English period on the model of use in Latin, and this meaning continued in the Middle English period (as it does today), reinforced by the similar use in both Latin and French. When *frere* is first found in Middle English it duplicates this meaning, as well as showing the more specialized meaning ‘member of one of the mendicant orders (chiefly the Franciscans, Augustinians, Dominicans, and Carmelites, as opposed to the non-mendicant Benedictines, etc.)’. By the end of the Middle English period a process of semantic specialization took place, with *brother* used in the general sense ‘member of a religious order’ and *friar* in the narrower sense ‘member of one of the mendicant orders’. Thus we might say that the borrowing filled a lexical gap in the vocabulary of English, providing a word specifically for ‘a member of one of the mendicant orders’, although we should perhaps be slightly cautious about such assumptions, since the same gap remained unfilled by any single word in French, even though the two languages were being used in very similar societies. Indeed, Anglo-French and Middle English were being used in precisely the same society. (See section 5.6 for discussion of the different functions of each language.) As we will see later, we can often run into problems of this sort when we attempt to explain word histories in functional terms, although this does not necessarily mean that the attempt is not worthwhile.

In its development from Middle English to modern English the word did not show any further change in meaning, but it did show an unusual change in form. The usually expected modern (British standard) pronunciation of a word which had the Middle English form *frere* would be /fri:ə/ (compare *here*, *deer*) but instead we find /fraɪə/. The same development is found in a small number of other words such as *briar* and *choir*. It probably shows a sporadic phenomenon of vowel raising before a following /r/.

Summary so far We can trace the history of a word’s sound and form. In doing so we are looking for regularity, i.e. developments which are the same as those which happened to the same sounds or combinations of sounds in other words. Where something unexpected or irregular has happened, as with the development of /fraɪə/ rather than /fri:ə/, we will want to find parallels, such as *briar*, etc. Ideally we will want to find an explanation for this as well.

The meaning of the word can also be traced historically. We can see how the meaning broadened in Latin and French, but how the English

borrowing showed only a very narrow component of the donor word's meaning. We can also see how this borrowing fitted into a set of meaning relations with existing words in English (specifically *brother*). The meaning history of this word also shows the importance of factors from non-linguistic history: if we did not know something about the history of the religious orders in medieval Europe we would have considerable difficulty in explaining the historical development in the meaning of this word.

1.2.2 Example two: *sad* from modern English to proto-Germanic

For our next example we will start with the present day and work backwards. Modern English and Middle English *sad* show the reflex or linear historical development of Old English *sæd*. The symbol *æ* which occurs in the written form of this word and of many other Old English words (and some early Middle English ones) represents a front vowel phoneme /a/ (perhaps in fact [æ] rather than [a]) which in Old English was distinct from the back vowel /a/, represented by *a*. (Its italic form *æ* is unfortunately very similar to that of the ligature *œ*, which can sometimes lead to confusion for the unwary.) We could represent this word history as Old English *sæd* > Middle English *sad* > modern English *sad*, but this would be rather artificial, since what we in fact have is a continuous history across all periods in the history of the language.

If we turn to the word's semantic history, a basic dictionary definition of the word *sad* as typically used in modern English is:

Of a person, or his or her feelings, disposition, etc.: feeling sorrow; sorrowful, mournful.

This meaning is first recorded *a*1300 (which stands for 'ante 1300', that is '1300 or a little earlier').⁶ A similar basic dictionary definition for the word's earlier meanings would be:

⁶ Some scholars use 'ante' in the more literal sense 'before', but most, including most dictionaries, use it in the generally more useful sense 'this date or a little earlier'. In this book the dates given for English words, forms, and senses are normally those provided by the *OED*. For words from other languages the data I give is generally drawn from the standard historical or etymological dictionaries of each language. Glosses and definitions of English words are normally based on those in either the *OED* or *The Oxford Dictionary of English* except where otherwise noted, although I have frequently shortened or otherwise adjusted them.

Having had one's fill; satisfied, sated; weary or tired (of something).

If we consider the likely historical development of these meanings, we can hypothesize that the meaning 'weary or tired (of something)' developed from 'satisfied, having had one's fill (of something)', hence showing a metaphorical, narrowed, negative meaning; compare the modern English idioms *to have had enough of something* or *to be fed up with something* for similar developments. Subsequently the sense 'weary or tired (of something)' broadened again (but still with an exclusively negative sense) to 'sorrowful, mournful' in general. Hence we can hypothesize that a meaning development occurred with two main steps:

satisfied, having had one's fill (of something)

[metaphorized and narrowed] > weary or tired (of something)

[broadened] > sorrowful, mournful

We get some further support for the last stage in this hypothesized development when we look at the meanings of the closest relatives of the Old English word, its cognates in the other Germanic languages. The next step back in the history of *sad* can be expressed as follows:

Old English *sæd* is cognate with Old Dutch *sat*, Old Saxon *sad*, Old High German *sat*, Old Icelandic *saðr*, Gothic *saps*, all of which have meanings broadly corresponding to the Old English one, 'having had one's fill; satisfied, sated; weary or tired (of something)'

However, the concept expressed by 'cognate with' needs some unpacking, and we will now look at this in more detail.

1.2.3 Cognates and language families

What does it mean to say that Old English *sæd* (English *sad*) is 'cognate with' the words from Old Dutch, Old Saxon, etc. listed at the end of the previous section? Just as the Romance languages all developed from (some form of) Latin (see section 1.2.2), so English and a number of other languages, which linguists call the Germanic languages, developed from a common antecedent called proto-Germanic. Unlike Latin, we have no historical records for proto-Germanic, but we can reconstruct a good deal of information about it from the evidence of the languages that developed

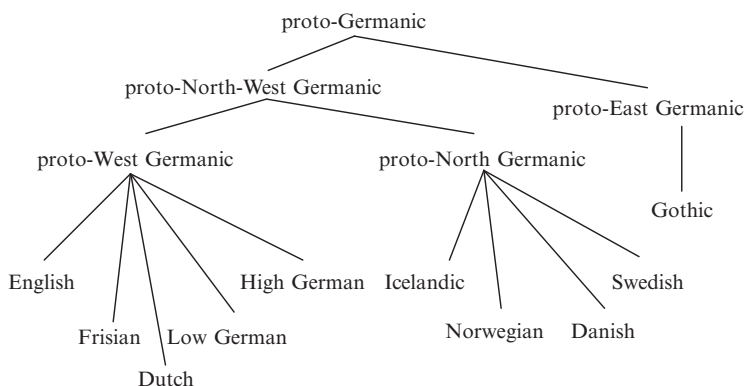


Fig 1.1 The major Germanic languages

from it. The other Germanic languages include Dutch (and hence Afrikaans), German (and hence Yiddish), Danish, Norwegian, Swedish, and Icelandic, as well as others such as Frisian (the closest relative of English, but with very few speakers today) and the extinct language Gothic (which is the Germanic language for which we have the earliest extensive documentary records, in the form of a bible translation dating from the fourth century AD). The cognates of an English word are the words in these other Germanic languages which can be explained as having developed from the same (unrecorded) antecedent word in proto-Germanic.

In fact, we can also identify subdivisions within the larger group of Germanic languages, on the basis of shared innovations that allow us to group the Scandinavian languages together as descendants of a common North Germanic sub-branch and likewise (albeit with rather more rough edges) English, Frisian, Dutch, Saxon/Low German, and High German as descendants of a West Germanic sub-branch. In turn, many scholars would now group together West Germanic and North Germanic as being descended from a shared North-West Germanic sub-branch with shared differences from East Germanic.⁷ Thus the relationships between the major Germanic languages can be represented schematically as in figure 1.1. We can reconstruct a similar tree structure for the major Romance languages, with the difference that in this instance the common ancestor, Latin, is of course attested (figure 1.2).

⁷ See for example Ringe (2006) 213. For a useful introduction to the early Germanic languages, see Robinson (1992).

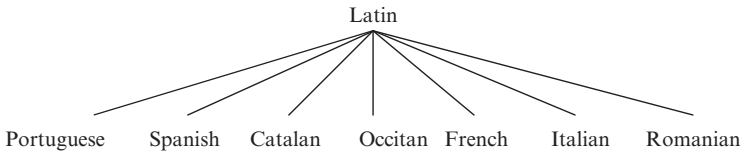


Fig 1.2 The major Romance languages

It is as well to pause for a moment and consider in a little more detail what this concept of a reconstructed antecedent language implies, because it will be crucial to many arguments later in this book. From present-day English to Old English (back as far as the eighth century, or even earlier in runic inscriptions) we have a chain of documents which enable us to trace the history of the English language in reasonable detail. In fact, these documents reflect many different local varieties of the language, showing many divergent developments. Some of these are reflected in different varieties of English today, such as the English of Chicago, or London, or Cape Town. We may analyse these as forming part of larger varieties, such as US English (or perhaps North American English), British English, or South African English. Alternatively we may subdivide them further, by looking for instance at different geographical or administrative areas of London, or at the language of different social classes within the city, or of different age groups, etc. Such variation must have been present throughout the history of English, although in earlier periods the nature and amount of the surviving evidence mean that we can only reconstruct a very limited picture. Modern US English and British English have developed as distinct varieties in different geographical locations from roughly the same antecedent, English as spoken in Britain in the early modern period (usually defined as approximately 1500–1750), but the historical record, as well as the evidence of modern US and British English, shows us that this common antecedent showed considerable internal variation. Similarly English and all of the other Germanic languages developed from a common antecedent (as did French, Spanish, etc. from Latin), but there is no reason to doubt, and every reason to suspect, that Germanic already showed internal variation. (Even though our surviving records for classical Latin are mostly literary and reflect a highly homogeneous literary language, there is indeed some variation in our surviving Latin evidence, and the later evidence of the Romance languages suggests the existence of a good deal of further variation in Latin

which is not reflected in the surviving documentary evidence).⁸ Over the course of time, groups of Germanic-speaking peoples developed distinct communities in different geographical locations (to some of which, like England, they had spread as part of the considerable movements of peoples which occurred in the later stages of the history of the Roman Empire and in the following centuries). As they did so, linguistic differences would have become more pronounced, as different variants from among the existing variation in Germanic came to predominate in different speech communities, and as new variation arose in each speech community.

At the time of our earliest substantial records for English, from several centuries after the Anglo-Saxons arrived in England, there are already important differences between English and its continental relatives, but these clearly took time to develop. We can also trace significant differences between different regional varieties of English in this early period, although the surviving documents leave very many questions unanswered.⁹ The demarcation of the various national languages of modern Europe owes a great deal to geography and, especially, politics. In the sixteenth and seventeenth centuries Scots was well on the way to developing a standard, ‘official’ form, distinct from the English of England, but subsequent political developments led to the adoption in official functions of a highly anglicized variety now usually referred to as Scottish English (although in recent decades as a result of the political process of devolution there have been some interesting developments in the use of Scots once again as an officially recognized variety in some functions). Today Dutch and German are well-defined national languages, sufficiently different from one another that monolingual speakers of either standard language have only an extremely limited degree of mutual intelligibility, but the situation is different among speakers of traditional dialects on or near the geographical boundaries between the two countries: such speakers can with a little effort understand the speech of their neighbours on the other side of the national border, even though one person is speaking something that is classified as a dialect of Dutch and the other something that is classified as a dialect of German. We can say that there is a dialect continuum which crosses the Dutch–German border. Another crosses the French–Italian border, and

⁸ On the degree of regional variation shown by surviving Latin documents from antiquity see Adams (2008).

⁹ For an introduction to the various issues involved see Hogg (2006).

similar cases can be found in many other parts of the world, essentially wherever languages have developed from a common source in adjacent territories.¹⁰

Such dialect continua lead us fairly directly to some limitations in the tree diagrams for the Romance and Germanic languages which I offered above. Diagrams of this type are a good way of representing where the most important shared innovations are found among various dialects in a group, but they have the disadvantage of making linguistic history appear artificially simple and neat. When two speech communities diverge, as represented by the branching on a tree, each takes with it a particular selection of features from the parent language. When further divergences occur subsequently, we may find that a particular feature is retained, quite by chance, in two languages or dialects which the weight of evidence places on completely different sides of the tree. In other cases the same innovation may occur independently in two different places, giving a false indication of inherited similarity. Additionally, where languages or dialects remain in contact, especially when they are spoken in geographically contiguous or overlapping territories, we can find that some features spread by diffusion (i.e. contact) from one variety to another, hence muddling the apparently clean branching shown by a tree. A better metaphor for such diffusion of features through language contact may be the spreading of a wave from a point of origin, rather than the branching of a tree.¹¹

1.2.4 Example two revisited: *sad* from proto-Germanic to proto-Indo-European

If we return to our example of *sad*, we can push this particular word history back further than just to proto-Germanic. The Germanic languages themselves form one branch of a much larger language family which historical linguists call Indo-European, which has numerous other branches, sub-branches, and isolate languages including for example:¹²

¹⁰ For an introductory account of these issues see Chambers and Trudgill (1998) 3–12. On the concept of a traditional dialect see especially Wells (1982) 4–8.

¹¹ For discussions of this issue with reference to the Germanic languages see Trask (1996) 181–7 (also Millar (2007) 225–31) and, at a rather more advanced level, Lass (1997) 139–59. On more general issues to do with language trees see McMahon and McMahon (2005).

¹² For an overview of the Indo-European languages see Fortson (2004).

- the Celtic languages: Welsh, Irish, etc.
- the Italic languages: Latin (and hence the Romance languages), Oscan, Umbrian, etc.
- Greek
- the Balto-Slavonic languages, comprising the Slavonic languages (Russian, Polish, etc.) and the Baltic languages (Lithuanian, Latvian, etc.)
- Albanian
- Armenian
- the Indo-Iranian languages, comprising the Iranian languages (Persian, etc.) and the Indic languages (Sanskrit and hence modern Hindi, etc.)

All of these languages can be shown to have developed from a single parent, proto-Indo-European, although of course all of them show the effects of contact with other languages during their histories. The identification of a shared ancestor for all of these languages rests upon the evidence of regular correspondences of sounds between the various languages, which we will look at in more detail below, and also upon systematic grammatical similarities, which are largely outside the scope of this book.

Many people have attempted to link Indo-European with other language families, but all such attempts remain extremely controversial, and the general view is that no genetic relationship has been reliably established between Indo-European and any other language family.

Precisely when and where proto-Indo-European existed as a spoken language is the subject of a very great deal of debate. This is complicated by the fact that the earliest recorded Indo-European language, Hittite, the oldest documentation for which dates back approximately 4,000 years, belongs to a branch, Anatolian, which probably split from the rest of Indo-European very early. However, what is reasonably certain is that proto-Indo-European began to split into its various daughter languages very much earlier than the date of our earliest documentary records for those languages. It is therefore unsurprising that many of the cognate forms bear little if any superficial resemblance to one another, since we are working at such a great time depth, and centuries of linguistic change lie between proto-Indo-European and even our earliest documentary evidence.

In this section we will trace the history of the word *sad* from proto-Germanic back to proto-Indo-European, and we will examine some of the procedures by which etymologies can be established at this time depth.

In doing so, we will encounter some principles and procedures which are equally applicable to much more recent linguistic history, and which we will investigate mostly from the standpoint of rather more recent linguistic evidence in the remainder of this book. However, reconstruction of linguistic data at a very considerable time depth is one of the big attractions of etymological research for many people, and it is also true that many of the most important aspects of modern etymological research came to fruition in the context of research into proto-Indo-European in the second half of the nineteenth century. We will therefore begin our investigation of the relationship between sound change and etymology by taking a look at how the sound changes known as Grimm's Law and Verner's Law help explain the etymology of *sad*.

By comparing the forms found in the Germanic languages with one another and also with forms in other Indo-European languages, we can reconstruct the proto-Germanic ancestor of *sad* as **saða*.¹³ An asterisk conventionally marks reconstructed forms, i.e. forms which are not actually recorded. **saða-* ends with a hyphen because it is a reconstructed word stem, i.e. the morphological stem to which inflectional endings were then added. In this book I will usually give reconstructions using IPA symbols, but without using square brackets [] implying that they are hypothetical phonetic transcriptions, nor // slashes implying that they necessarily have phonemic status. This is a traditional philological practice, which is useful for three main reasons: (i) we cannot always be certain about the precise phonetic quality of reconstructed sounds; (ii) any past historical sound system almost certainly showed considerable variation in the realization of sounds, which we cannot recover in detail from our historical evidence; (iii) we cannot always be sure whether certain distributions of sounds were phonemic or allophonic in a given historical period.¹⁴ We will look at issues

¹³ The exact phonetic quality and phonemic status of the consonant I have represented here as **ð* is in fact very uncertain. Many scholars choose to use **d* in reconstructions of proto-Germanic forms to represent any sound which may have been either a voiced plosive /d/ or a voiced fricative /ð/. In many modern etymological dictionaries the proto-Germanic form of this particular word is hence represented as **sada-*. However, since the sound in this instance was almost certainly a voiced fricative at an early stage in proto-Germanic, I have used the reconstruction **saða-*, which has the advantage of making the changes from proto-Indo-European to proto-Germanic easier to follow.

¹⁴ For a recent detailed argument for this position see Lass and Laing (2007) §§2.4.2, 8.3.2.

to do with variation and change in any linguistic system in more detail in chapters 3 and 7.

The reconstruction **saða-* depends upon the evidence of the various Germanic languages, and also upon the evidence of forms in other Indo-European languages which can plausibly be referred to the same root form. Most crucially, it depends upon:

- (a) regular sound correspondences between the various languages
- (b) sound changes which can be posited to explain apparent irregularities

To get from proto-Germanic **saða-* to the recorded words Old English *sæd*, Old Dutch *sat*, Old Saxon *sad*, Old High German *sat*, Old Icelandic *saðr*, Gothic *saps* requires just a couple of small steps:

- In West Germanic, proto-Germanic **ð* regularly became the voiced plosive /d/, as in our Old English form *sæd* /sad/ or Old Saxon *sad*. Old Dutch *sat* and Old High German *sat* show subsequent devoicing of this plosive (compare section 2.1.1.3).
- Old English *sæd* additionally shows Old English (and Old Frisian) fronting of West Germanic **a* to /a/.

These are regular, predictable sound changes in a word of this phonological shape in these languages.

This reconstructed proto-Germanic form **saða-* itself shows the reflex of an earlier Indo-European form **səto-*. (The symbol **ə* in this reconstruction represents a sound which was realized as a vowel when it occurred in this position, hence giving rise to vowels in the daughter languages, but which is now generally believed to have resulted from the vocalic realization of one of a series of so-called laryngeal sounds which are hypothesized for proto-Indo-European. They are called laryngeals for historical reasons, although no one in fact knows exactly what their phonetic quality was. This particular laryngeal is sometimes represented as *ə₂* or as *h₂* or as *H₂*, depending on which transcription conventions are being followed. We will return to this topic in sections 1.3.1 and 4.4.1.)

Related words in other Indo-European languages include:

- classical Latin *sat*, *satis* ‘enough’, *satur* ‘satisfied, full’
- Lithuanian *sotus* ‘filling, full, satisfied, substantial’
- ancient Greek *áatos* ‘insatiate’ (showing a negative prefix)

We can see that the meanings of these words help support our hypothesis about the meaning development in the Germanic languages from ‘satisfied, having had one’s fill (of something)’ to ‘weary or tired (of something)’. It is difficult to be certain about the precise relationships between these words. They probably reflect two different variants, **sǣ-* and **sā-*, of a single Indo-European root for which the approximate meanings ‘fill up, (make) replete’ can be reconstructed. In our surviving cognates various different suffixes, **-to-*, **-ti-*, and **-tu-*, have been added to this root. The cognates thus do not represent the reflexes of a single word form, but rather the survivors of an extended word family, derived in various different ways from a common root.¹⁵ The Germanic words probably show what was originally a suffix which formed verbal adjectives, proto-Indo-European **-to-*. The same suffix is probably found in *old* (< proto-Germanic **al-da-*) and *cold* (< proto-Germanic **kal-da-*; compare Latin *gelidus*), and in many Latin words ending in *-tus*. (On roots and their meanings see further sections 4.4.1 and 8.7.3.)

The assumption made in the last paragraph that proto-Germanic **saða-* is likely to have developed from proto-Indo-European **sǣto-* may seem rather startling to anyone who does not have a prior acquaintance with Indo-European linguistics. On the face of it only the initial consonant **s* is common to both forms. However, the development of the vowels is easily dealt with, by the principle of regular sound correspondences. Proto-Indo-European **ǵ* (with the caveats given above) and (short) **o* both regularly develop to **a* in proto-Germanic, thus **sǣto-* > **saða-*. A sound change of this sort is called a merger: the phonetic development of **ǵ*, **o*, and **a* in proto-Germanic led to loss of the distinction between the three separate proto-Indo-European phonemes and merger as the single phoneme **a* in proto-Germanic. Compare Latin *hostis* ‘stranger, enemy’ with its cognate Gothic *gasts* ‘guest’, or Latin *hortus* ‘garden’ with its cognate Gothic *gards* ‘garden’. (Latin *h* and Gothic *g* in these words show the regular development in Latin and in proto-Germanic of proto-Indo-European **g^h*; we will look further at the Germanic side of this in the next paragraph. The modern English cognates of these words are respectively *guest* and *yard*, showing the result of a number of sound changes during the history of English.)

¹⁵ For a specialist readership, the best recent detailed account of the Germanic component of this etymology is provided (in German) by Heidermanns (1993) 458–9; on the Indo-European component see especially Szemerényi (1979).

Probably, on the basis of the evidence of other Indo-European languages, in proto-Germanic the reflexes of proto-Indo-European $*\bar{a}$ and $*a$ merged first as $*a$, with which $*o$ then also merged. Conversely, the proto-Indo-European long vowels $*\bar{o}$ and $*\bar{a}$ merge as $*\bar{o}$ in proto-Germanic.

The explanation for the medial consonant in proto-Germanic $*sa\bar{d}a-$ is a little more complicated, and involves two reconstructed sound changes. Comparison among the Indo-European languages excluding Germanic leads to the reconstruction of three sets of stop consonants: voiceless stops ($*p$, $*t$, $*k$, $*k^w$), voiced stops ($*b$, $*d$, $*g$, $*g^w$), and breathy-voiced stops ($*b^h$, $*d^h$, $*g^h$, $*g^{hw}$). Comparison with the forms in the Germanic languages leads to the conclusion that a series of sound shifts occurred in proto-Germanic:

$*p > *f$

$*t > \theta$ (represented in traditional philological notation as $*\mathfrak{t}$)

$*k > *h$

$*k^w > *hw$

$*b > *p$

$*d > *t$

$*g > *k$

$*g^w > *kw$

$*b^h > *\beta$ (in some environments $> *\mathfrak{b}$)

$*d^h > *\delta$ (in some environments $> *\mathfrak{d}$)

$*g^h > *\gamma$ (in some environments $> *g$)

$*g^{hw} > *\gamma w$ (in some environments $> *gw$)

Thus the voiceless stops became voiceless fricatives, the voiced stops became voiceless stops, and the breathy-voiced stops lost their breathy-voice and probably became fricatives before becoming voiced stops in many environments. Experts in fact differ on many details of this process, especially as regards the proto-Indo-European breathy-voiced stops and also the proto-Indo-European voiced stop $*b$ (which was very rare, and some argue did not exist at all), but this is not of importance for our present purposes.¹⁶ This sound change (or series of changes) is known as Grimm's Law, after the German philologist Jakob Grimm (1785–1863), who compiled with his brother Wilhelm both the celebrated fairy tale collection and the early

¹⁶ The literature on Grimm's Law, and Verner's Law, is vast. For a recent detailed account of the changes see Ringe (2006) 93–116; for particularly useful analyses see also Bynon (1977) 83–6, Collinge (1985) 63–76. See also the discussion in section 7.1 below.

fascicles of the major historical dictionary of the German language. Grimm produced an important early formulation of this sound change, although it had in fact been described earlier by other scholars. An alternative name for this sound change is the Germanic Consonant Shift.

We can illustrate the changes in the proto-Indo-European voiceless stops with the following examples:

$*p > *f$

I-E root **ped-* ‘foot’: ancient Greek *poús* (stem *pod-*), Latin *pēs* (stem *ped-*); Gothic *fōtus*, English *foot*

$*t > *θ$

I-E **tū* ‘you (singular)’: Latin *tū*, Old Irish *tū*; Gothic *þū*, English *thou*

$*k > *h$

I-E root **kerd-* ‘heart’: ancient Greek *kardia*, Latin *cor* (stem *cord-*); Gothic *hairtō*, English *heart*

$*k^w > *hw$

I-E **k^wós* ‘who’: Sanskrit *kás* ‘who’, Lithuanian *kàs* ‘who, what’; Gothic *hwas* ‘who’, English *who*

In the first example here, ‘foot’, Grimm’s Law explains not only the shift of the initial consonant from **p* to **f* but also the shift of the final consonant of the stem from **d* to **t*. However, it will be obvious at a glance that there are other differences between the cognates apart from those explained by Grimm’s Law, even though I have attempted to select forms which have an unusually close mutual resemblance (another of the cognates of English *foot* is in fact Armenian *otn*). In the case of ‘foot’, the Greek, Latin, and Germanic words all have different stem vowels. In this instance the difference is not due to sound changes which have occurred in the daughter languages, but to slightly different etymons in proto-Indo-European: the Greek stem form *pod-* is from proto-Indo-European **pod-*, the Latin stem form *ped-* is from proto-Indo-European **ped-*, and the Germanic forms are from proto-Indo-European **pōd-*. These different etymons are all derived from the root **ped-* by a process known as ablaut which we will look at in section 4.4.1. This also explains the variation between **s̥-* and **sā-* which we encountered above in the etymology of *sad*.

The operation of Grimm’s Law thus explains why proto-Germanic **saða-* < proto-Indo-European **s̥əto-* does not show medial **t*, but it does not explain why it shows **ð* rather than the expected **θ*. This is explained by another sound change known as Verner’s Law, after the Danish philologist

Karl Verner (1846–96), by which the proto-Germanic voiceless fricatives became voiced whenever the accent did not fall on the immediately preceding syllable. (For an analogous situation in modern English, compare *ex'ert* /ɛɡ'zə:t/ with *'exercise* /'ɛksəsaɪz/.) In the ancestor of *sad* the suffix, not the root, was stressed, and hence Verner's Law applied, giving voiced *ð. Later, the accent shifted to the first syllable in all words in proto-Germanic, thus giving the pattern which we find reflected in all of the recorded Germanic languages. Hence, finally, we can explain how proto-Indo-European *sǵ'to would give rise to proto-Germanic *'saða, via the following stages: *sǵ'to > *sa'ta > *sa'tha > *sa'ða > *'saða. We will not do so here, but pre-histories can similarly be reconstructed for classical Latin *sat*, *satis*, *satur*, Lithuanian *sotus*, and also ancient Greek *átos*, and it is this (rather than vague resemblance in form and meaning) which gives substance to the hypothesis that all of these forms are ultimately cognate.

We will return to Grimm's Law and Verner's Law in a little more detail at the beginning of chapter 7, but for the time being there are one or two very important general observations which arise from this example. Note that in the preceding paragraph I said that proto-Indo-European *sǵto- 'would give rise to' proto-Germanic *saða-, and not 'could give rise to'. The merger of *ǵ, *o, and *a as *a in proto-Germanic, and the Grimm's Law and Verner's Law changes, are all regular processes, which apply in all cases (where not excluded by specific phonetic environments, which simply involve more precise statement of what the sound change was and in which environments it applied). The standard methodology of comparative linguistics does not permit us to say 'perhaps in this particular instance the merger simply did not happen' or 'perhaps Grimm's Law did not apply to this word' or 'perhaps in this instance an entirely unparalleled change of *ð to *m occurred'. As I have formulated it here, this is an oversimplification, but not a huge one. In chapter 7 we will look at the reasoning behind this in much more detail, and at some important qualifications, but for present purposes it is sufficient to be aware that comparative reconstruction depends upon the regularity of the correspondences and sound changes which are posited: this (as well as general phonetic plausibility, and the existence of parallels in the documented history of languages) is what gives a solid foundation to comparative etymological research.

A useful illustration of this principle is shown by the histories of the words *mother*, *father*, and *brother*. All three words show a voiced fricative /ð/ in modern English. However, in Old English the situation was

rather different: *brōðor* ‘brother’ showed a voiced fricative /ð/, but *mōdor* ‘mother’ and *fæder* ‘father’ both showed a voiced plosive /d/. In proto-Indo-European all three words in fact showed the same termination, **-tēr-* (in the nominative case), which seems typical of terms for family kinship: **mātēr* ‘mother’, **pətēr* ‘father’, and **bhrātēr* ‘brother’;¹⁷ compare Latin *māter* ‘mother’, *pater* ‘father’, *frāter* ‘brother’ (proto-Indo-European **b^h* > *f* in word-initial position in Latin; compare also Sanskrit *bhrātar-*). The explanation for the different outcomes in Old English is the regular operation of Verner’s Law. In the case of *mother* and *father* the stress in proto-Germanic fell on the second syllable, while in the case of *brother* it fell on the first syllable. Thus Verner’s Law applied in the case of *mother* and *father*, but not in the case of *brother*, and so we find that proto-Germanic **brōþēr*, with voiceless fricative **θ*, corresponds to Latin *frāter*, but that proto-Germanic **mōðēr* and **faðēr*, with voiced fricative **ð*, correspond to Latin *māter* and *pater*. In *mother* and *father* the proto-Germanic voiced fricative subsequently became a plosive in West Germanic, just as in the case of *sad*, hence Old English *mōder* (or in fact more commonly *mōdor*, showing variation in the unstressed vowel of the second syllable) and *fæder*. In the case of *brother*, the medial voiceless fricative of proto-Germanic **brōþēr* became voiced in intervocalic position in Old English, hence Old English *brōðer* (again in fact more commonly *brōðor*). Subsequently, in late Middle English, by another sound change, the voiced plosive of *moder* and *fader* developed into a fricative before either /ær/ or syllabic /r/, resulting from reduction or loss of the vowel in the endings *-or*, *-er*. Thus, *mother* and *father* came to have the same voiced fricative as *brother*. So we can see that *mother*, *father*, and *brother* provide a very rare example of how subsequent sound changes can, very occasionally and entirely fortuitously, restore a formal resemblance which had been obscured by a much earlier sound change (figure 1.3). We have also now seen how *brother* and *friar*, discussed in section 1.2.1, are in fact cognate, both being ultimately from proto-Indo-European **bhrātēr*. In the latter case the development was: *friar* < Old French *frere* < Latin *frāter* < proto-Indo-European **bhrātēr*.

¹⁷ In the reconstructions **mātēr* and **bhrātēr* the **ā* in the first syllable shows what is now generally considered to have been the output of earlier **eh₂*, i.e. the vowel **e* followed by a laryngeal which caused colouring and lengthening of the vowel. For a fuller explanation of this see section 4.4.1.

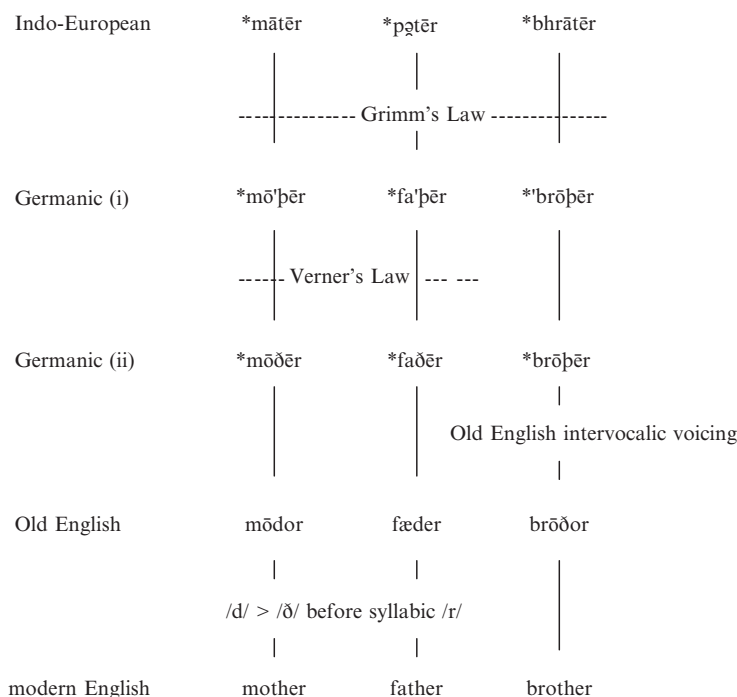


Fig 1.3 *mother, father, and brother* from proto-Indo-European to modern English

1.2.5 Summary

Our initial supposition about the meaning development of *sad* within English was supported by comparison with the meanings of its cognates in other Germanic languages, and ultimately also by the meanings of its cognates elsewhere in Indo-European.

In tracing the word's cognates at a great time depth we have seen the importance of regular sound correspondences and of regular sound changes in accounting for apparent discrepancies. We will return to this topic in more detail in chapter 7.

In the etymologies of both *friar* and *sad*, there is little or no connection between the processes of formal development and the processes of meaning development that we have examined. This is often the case, although there are also cases where form history and meaning history are very closely intertwined, and we will look closely at a number of such cases in chapters 7 and 8.

1.3 Why study etymology?

1.3.1 Etymology, historical and comparative grammars, and dictionaries

Etymology is an essential tool in reconstructing the history of a language, since a corpus of word histories provides a necessary basis for many other aspects of historical linguistic work. Conversely, each individual word history depends for its plausibility on the work that has been done in various subfields of historical linguistics. For instance, someone interested in historical semantics will want to look at the meaning histories of individual words which have been traced through the application of etymology, just as an etymologist will want to draw on the general observations about a whole body of meaning changes and their likely motivations which have been identified by specialists in historical semantics. Each activity informs and enriches the other in a mutually beneficial relationship.

Traditionally, etymology has been associated most closely with the construction of historical and comparative grammars. A historical grammar traces the developments in word forms which are found in the history of a language, often also extending into its pre-history. A comparative grammar relates the developments found in one language to those found in cognate languages, to explain the development of two or more languages from a common source using the technique of comparative reconstruction.

We have seen in the case of *friar* an example of how etymology interacts with the functions of a historical grammar:

- Etymological investigation suggests that *friar* shows the continuation of Middle English *frere*.
- A historical grammar identifies parallels such as *briar* and *choir* (themselves the result of other etymological investigations). Ideally, it will also supply an explanation for the unusual form history shown by such groups of words.

Our investigation of *sad* gave an insight into the world of comparative etymology and comparative reconstruction. The identification of regular sound correspondences depends at first upon the investigation of large numbers of potential etymological connections. This may make it possible to identify the regular processes of sound change. If so, our corpus of etymologies can be refined, and some at first apparently attractive connections can be discarded, at least until we can find a new explanation to account for them.

The best illustration of this may be to look at an example of how a sound method may enable us to identify a case of chance resemblance. If we start out, from an entirely uninformed perspective, by looking simply for words which are similar in form and meaning, English *care* and Latin *cūra* ‘care’ might seem attractive candidates for investigation: they overlap completely in their core meaning, and the consonants at least are the same. There is thus more resemblance in both form and meaning than there is between English *sad* and Latin *satis* ‘enough’ or Lithuanian *sotus* ‘filling, full, satisfied, substantial’. However, English *care* is an inherited Germanic word, with a good set of cognates from all branches of Germanic which enable us to reconstruct a proto-Germanic form **karō-*. If we remember Grimm’s Law, we will see that proto-Germanic /k/ is not going to correspond to Latin /k/, and in fact proto-Germanic **karō-* is usually referred to a proto-Indo-European root **gar-* with the meaning ‘to call, cry’. This same root is probably reflected also by Latin *garrīre* ‘to chatter’ (ultimately the base of English *garrulous*). Latin *cūra* shows the regular development of an earlier form **koiśā*, which can be reconstructed on the basis of forms in inscriptions and cognates from other Italic dialects; it has no generally accepted further etymology, but could not conceivably be connected with proto-Germanic **karō-*. In fact some doubts have been raised about the connection of proto-Germanic **karō-* with proto-Indo-European **gar-*.¹⁸ Revised or contested hypotheses are very common in etymological work at this sort of time depth. However, the important point is that a connection with Latin *cūra* remains impossible, even if we have no viable etymology for **karō-*: we do not need to have an alternative explanation in order to reject an impossible etymology.

Latin *deus* ‘god’ and Greek *theós* ‘god’ are another pair of words which are synonymous and have a superficial resemblance in form, but which the methodology of comparative linguistics demonstrates have no etymological connection whatever: the first goes back to proto-Indo-European **deiwós* and the other probably to proto-Indo-European **dhesos*. We can thus make an important generalization: comparative reconstruction provides an essential tool for quickly eliminating very many cases of chance resemblance in form and meaning, just as it identifies many cognates which have little or no superficial resemblance in form or meaning.¹⁹ It also leaves us with

¹⁸ See for instance (in German) Rix (2001) 161.

¹⁹ For an excellent and much more detailed account of these and related issues see Campbell (2003).

very many rather doubtful cases, some examples of which we will examine later.

Sometimes ‘etymology’ has been seen as almost synonymous with ‘comparative reconstruction’, or at least it has been assumed that everything else which an etymologist has to consider is of secondary importance in comparison with the reconstruction of antecedent word forms and the identification of historical sound changes. This will not be entirely the approach adopted in this book, although it should not be forgotten that form history, as reflected in historical and comparative grammars, provides the backbone for nearly all etymological research: we will examine in detail in chapters 7 and 8 how and why it is that arguments based on word form usually provide by far the strongest foundation for etymologies.

Comparative reconstruction has a sister methodology known as internal reconstruction, in which reconstruction is based purely on the data provided by a single language. This is generally much more limited, and also less reliable, than comparative reconstruction, and it will not be a major topic in this book, although it should be noted that methods of internal reconstruction have contributed some important advances in knowledge even in areas such as Indo-European linguistics where the comparative data is relatively rich and plentiful. It tends to be most effective in tracing the origins of morphophonemic relationships, as between English *mouse* and *mice* (see section 7.2.4) or the contrast between voiceless and voiced consonants in German *Rad* and *Rades* (section 2.1.1.3), although even here comparative data is often much more conclusive.²⁰ One very important and justly famous success of internal reconstruction was Ferdinand de Saussure’s identification in the late nineteenth century of a series of hypothetical sounds in proto-Indo-European which he termed (in French) ‘coefficients sonantiques’. These are now generally recognized as a series of so-called laryngeal sounds (although their exact quality is in fact unknown and the subject of much dispute). Hittite documents which began to be deciphered and studied in detail in the early twentieth century, long after Saussure’s initial hypothesis based on internal reconstruction, provided crucial data which confirmed the reconstruction.²¹ We will return to this topic, and to its implications for the sound represented by *ǵ in the proto-Indo-European reconstructed forms given here, in section 4.4.1.

²⁰ For thorough accounts of internal reconstruction see Fox (1995) or Ringe (2003).

²¹ For short accounts of this see for example Fortson (2004) 75–6; also Hock (1991) 545–9, Clackson (2007) 53–61, or Millar (2007) 322–7.

Aside from historical and comparative grammars, etymology is also a crucial scholarly tool in historical lexicography. Historical dictionaries present in linear form the word histories which are treated thematically in grammars: in grammars we can see the connections between the developments shown by individual words, while in historical dictionaries we can see word histories whole and uninterrupted, together with the interplay between form history and meaning history, and at least some information on the influence of extralinguistic cultural and historical factors.

1.3.2 Historical relationships between words

A key function of etymology is that it illuminates the formal and semantic relationships between the words of a language. This is an area where a layman's interests may not be entirely dissimilar to those of a historical linguist, and thus it can be a very good entry point for people who are relatively new to the study of etymology. Indeed, this topic is of particular interest for speakers of a language like English which has seen a good deal of borrowing, and where the semantic relationship between for example *hand* and *manual* 'involving the hand, operated by hand, etc.' is obscured by the absence of any formal relationship between the two words. In this particular instance, the word *manual* is ultimately a derivative formation from a word meaning 'hand', but the word in question is Latin *manus* 'hand' (plus a Latin suffix *-ālis* which forms adjectives with the meaning 'connected with') rather than English *hand*. Latin *manuālis* was borrowed into English (via French) as *manual* in the fifteenth century. For a time it competed with a word with the same meaning which did have a transparent formal relationship with *hand*, namely *handy*. This word today only has the specialized meanings 'convenient to handle or use', 'ready to hand', 'skilful, good with his or her hands', but in early use it also had the meaning 'done by hand, manual'. It is formed from *hand* and the suffix *-y* (which has a function similar to Latin *-ālis*), although this is not the full story: *handy* probably originally arose as a result of reanalysis of the word *handiwork*, which was itself formed much earlier. *handiwork* is not (as we may at first assume) formed from *handy* and *work* but from *hand* and the obsolete noun *geweorc* 'work', which is a derivative of Old English *weorc* 'work' formed with a prefix *ge-* which had a collective meaning (thus 'work collectively') and which was pronounced with a palatal initial consonant /j/, thus /jeweark/. In course of time phonetic reduction occurred in the unstressed medial syllable