

Webs of Words

Webs of Words:
New Studies in Historical Lexicology

Edited by

John Considine

CAMBRIDGE
SCHOLARS

P U B L I S H I N G

Webs of Words: New Studies in Historical Lexicology,
Edited by John Considine

This book first published 2010

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2010 by John Considine and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-1952-2, ISBN (13): 978-1-4438-1952-7

TABLE OF CONTENTS

Introduction	vii
Historical Lexicology	
John Considine	
 Chapter One.....	 1
Diachronic Corpora: Seeing Histories of Words from another Angle	
Karel Kučera and Martin Stluka	
 Chapter Two	 8
Why did Tudor England have no Monolingual English Dictionary?	
Ian Lancashire	
 Chapter Three.....	 24
<i>Dat Boek der Wundenartzstedye</i> : The Low German Translation	
of Hieronymus Brunschwig's <i>Buch der Chirurgia</i> and its Rendering	
of Surgical Lexicon	
Chiara Benati	
 Chapter Four	 56
Cleaning up the Mess: The Case of English <i>Balcony</i>	
Mateusz Urban	
 Chapter Five	 65
Do Savages Get the Blues? William Colenso and the Nineteenth-Century	
Colour Debate	
Jane Samson	
 Chapter Six	 77
<i>Biologisch</i> vs. <i>Biokoffie</i> , <i>Ecologisch</i> vs. <i>Ecotrend</i> : Dutch <i>Bio-</i>	
and <i>Eco-</i> Through the Ages in the Dictionary	
Vivien Waszink	

Chapter Seven.....	94
Masters, Mothers, and Barking Dogs: The Lexical Family of the Words for Grammar in China Tommaso Pellin	
Chapter Eight.....	112
Historical Naming Strategies for Fauna in Trinidad & Tobago English/Creole Lise Winer	
Chapter Nine.....	141
The Third Edition of the <i>OED</i> and Lexical Transmission: Towards a Consistent Research Methodology Mirosława Podhajecka	
Chapter Ten	191
Should Portuguese Dictionaries Register English Words? Isabel Casanova	
Abbreviations.....	200
Bibliography.....	202
Contributors.....	231
Index	234

INTRODUCTION

HISTORICAL LEXICOLOGY

JOHN CONSIDINE

1. Historical lexicology and historical lexicography

What is lexicology, and what is historical lexicology? Christoph Schwarze and Dieter Wunderlich introduced their *Handbuch der Lexikologie* in 1985 by commenting on the increasing interest of linguistic theorists in words in the years before the publication of their book, which had led to “eine eigenständige theoretische Disziplin der Lexikologie” (8), an autonomous theoretical discipline of lexicology. They then proposed that “Die Lexikologie ist die Theorie des Lexikons”—“lexicology is the theory of the lexicon”—and observed that *Lexikon* can mean “vocabulary of a language” or “dictionary”, and that lexicology is the study of the former. Fifteen years later, an English-language textbook suggested that

Lexicology deals not only with simple words in all their aspects, but also with complex and compound words, the meaningful units of language. Since these units must be analysed in respect of both their form and their meaning, lexicology relies on information derived from morphology, the study of the forms of words and their components, and semantics, the study of their meanings. A third field of particular interest in lexicological studies is etymology, the study of the origins of words. (Jackson and Zé Amvela 2000, 1–2)

As Tommaso Pellin remarks in the opening paragraph of his contribution to this volume, it is possible to see two contrasting conceptions of lexicology: either as “the theoretical study of the abstract lexemes of a language and their organization” (as suggested by Schwarze and Wunderlich’s emphasis on *Theorie* and *theoretische Disziplin*), or as the “interdisciplinary study of the concrete usage of the vocables of a language” (as suggested in particular by Jackson and Zé Amvela’s special notice of the highly empirical discipline of etymology). Whether it is

specifically etymological or not, historical lexicology must usually tend towards the more empirical pole.

Just as the word *lexicology* goes back to the eighteenth century (see Considine 2007, viii) but has become increasingly current since the third quarter of the twentieth, so the phrase *historical lexicology* is not a new one, but has quite recently had a new lease of life. It appears to originate in French or German, in the nineteenth century: a review in French of an historical grammar of French (Rabiet 1882, 76) remarks that “si la lexicologie historique a déjà été l’objet de plusieurs traités ... nous n’avions encore aucune grammaire qui nous donnât l’histoire de la syntaxe” and an introduction in German to Romance philology included a short section on “historische Lexikologie” (Gröber 1888). The English phrase occurs at least as early as 1918, in an American article on “lexicological evolution and conceptual progress,” which draws on the then-available fascicules of the *Oxford English Dictionary* to demonstrate that between the early Middle English period and the nineteenth century, speakers of English developed an increasingly rich and subtle vocabulary for describing speech acts, from the twelfth-century and earlier *ask*, *backbite*, *bemoan* and so on to the post-1851 *accredit*, *actionize* and so on, the excerption being described as an exercise in historical lexicology (Clark 1918, 198). In the decades thereafter, however, the phrase is rare: one occurrence (Barbier 1933, 125) is known to me, and comes like its Continental predecessors from the field of Romance philology. Indeed, *historical lexicology* is not attested in the academic journals available through the database JSTOR until after the Second World War.

The first three attestations in JSTOR are all connected with Russian linguistic scholarship, as are all but one of the next ten; the earliest of all these attestations (Matthews 1949, 606) refers to a then-unpublished “Historical lexicology of Russian.” The author of the work whose title was thus translated was V. V. Vinogradov, whose support for Stalin’s rebuttal of the linguistic theories of N. Y. Marr in 1950, together with his appointment in the same year to the directorship of the Linguistics Institute, would make his voice an influential one. One of the points which Vinogradov made in support of Stalin was that the dictator’s “doctrine of the basic word stock laid the ‘Marxist foundations for historical lexicology and the history of word-formation’” (Ellis and Davies 1951, 255). Perhaps Vinogradov brought the phrase *istoričeskaja leksikologija* into wide currency, and perhaps he formed it after French *lexicologie historique*: he was certainly influenced in other respects by French linguistic thought (see e.g. Cowie 1998, 213). The phrase continued to be used by Russian

linguists after the era of Stalin, and a journal called *Russkaja istoričeskaja leksikologija i leksikografija* was founded in 1972.

During the second half of the twentieth century, *historical lexicology* and its equivalents came to be used increasingly outside the field of Russian studies. By the end of the century, these phrases were starting to appear in the titles of theses and books: *La formation du vocabulaire gascon de la boucherie: Étude de lexicologie historique et descriptive* (Fossat 1969); *Historische lexikologie zum nordgermanischen Raum* (Jacoby 1990); *Diachronic prototype semantics: a contribution to historical lexicology* (Geeraerts 1997). The association between historical lexicology and lexicography, already evident in the title of *Russkaja istoričeskaja leksikologija i leksikografija*, recurs for instance in the subtitle of a German-language volume devoted to “historische Lexikologie und Lexicographie” (Holtus, Kramer, and Schweickard 1997), and in personal connections such as the eight years on the staff of the *Woordenboek der Nederlandsche taal* which Geeraerts saw as the source of his book (Geeraerts 1997, 5). As noted by Javier Díaz Vera in his introduction to a volume of “Studies in English historical lexicography, lexicology, and semantics” (Díaz Vera 2002a, vi–vii), work on historical English lexicology at the beginning of the twenty-first century was energized by two lexicographical projects, the *Dictionary of Old English* undertaken at the University of Toronto, and what was then the *Historical Thesaurus of English* undertaken at the University of Glasgow (*Historical Thesaurus of the OED* 2009). To these should be added the revision of the *Oxford English Dictionary* (*OED3*: see the sources cited by Mirosława Podhajecka in her contribution to the present volume), which has an etymological element lacking in the two projects mentioned by Díaz Vera.

The First International Conference on Historical Lexicography and Lexicology (ICHLL) took place at the University of Leicester in 2002, organized by Julie Coleman; Díaz Vera’s point about the importance of the Glasgow *Historical Thesaurus* as a stimulus to historical lexicology is illustrated by the fact that Coleman’s doctoral thesis (published as Coleman 1999) was based on data from that project. The first ICHLL was intended as a successor to two round-table conferences of the 1970s, both of which had been primarily concerned with lexicography. Although lexicological papers were presented at the conference, for instance a study of the Old English noun (*ge*)*laðung* and an account of the English element in Croatian maritime technology (Kilpinen 2002; Pritchard 2002), it gave rise to a volume called *Historical dictionaries and historical dictionary research*, with no reference to lexicology in its title and no exclusively lexicological papers in its contents (Coleman and McDermott 2004). Its

successor, which took place two years later at the Palazzo Feltrinelli in Gargnano, under the auspices of the Università degli Studi di Milano, organized by Giovanni Iamartino, gave rise to a volume entitled *Words and dictionaries from the British Isles* (Considine and Iamartino 2007), collecting papers from the conference which bore primarily on the English language. A reviewer pointed out fairly that despite the phrase “the British Isles” in the title, the contents were almost exclusively oriented towards England (Wild 2008, 446). The title’s emphasis on words as well as dictionaries, however, was appropriate: the volume did include lexicological papers (e.g. Pinnavaia 2007). Likewise, the third ICHLL, which took place in Leiden in 2006, organized by Marijke Mooijaart and Marijke van der Wal, included lexicological papers (e.g. Benati 2008) and recognized the separate importance of lexicology in its title, *Yesterday’s words* (Mooijaart and van der Wal 2008).

The ten papers collected in this volume originate in presentations given at or proposed for the fourth ICHLL, which took place in Edmonton, Canada, in June 2008. The principal languages which they address are Chinese, Czech, Dutch, English (including the English / creole of Trinidad and Tobago), German, Māori, Portuguese, and Russian. They are all primarily lexicological: two companion volumes present articles developed from presentations which discussed current projects in historical lexicography and from presentations which discussed the history of dictionaries (Considine 2010a, 2010b). The decision to offer three volumes all of which originate in the Edmonton conference has been taken with two ends in view: firstly, to make each volume as coherent as possible, and secondly, to ensure that contributors had the opportunity to develop their ideas as satisfying articles rather than simply writing up their presentations briefly for publication, as would have been necessary to keep a single-volume collection within bounds. Like other volumes with an ICHLL background, this is meant to be more than a proceedings volume (cf. Considine 2007, viii, and Wild 2008, 450).

2. Overview of this volume

The chapters in the present volume are arranged in a very roughly chronological order of subject matter, although their historical depth tends to be such that there is extensive chronological overlap between them.

The first, Karel Kučera and Martin Stluka’s “Diachronic Corpora: Seeing Histories of Words from Another Angle”, draws on its authors’ work on the diachronic part of the Czech National Corpus, from the fifteenth to the twentieth centuries. It goes to the heart of one of the

lexicological questions to which dictionaries tend not to give a satisfactory answer: when two forms of a lexeme, or two words of closely related meaning, coexist in a language, how may their relative frequencies change over time? As Kučera and Stluka demonstrate, a number of possibilities can be seen playing out in the history of the vocabulary of Czech, from the relatively banal situation in which one variant survives alongside its overwhelmingly more frequent alternative, to various kinds of rivalry, some of which are very difficult to explain (and all the more interesting for that) while others can, at least tentatively, be mapped against developments in the history of the language which have left other traces. The corpus with which they worked was, at the time of their study, a small one, around three million words, and this led them to group their data by centuries in order to smooth out the potentially misleading effects of anomalous individual texts; work with a larger corpus would generate increasingly precise results, enabling the plotting of decade-by decade or conceivably year-by-year shifts in relative frequency.

Historical dictionaries can of course make some approach to this sort of frequency mapping: the editor who notices a great wealth of occurrences of a particular form from a given period can comment on this in an entry, as the editors of the *Oxford English Dictionary* did in a number of instances. So, for instance, two figurative senses of *cankered*, “malignant” and “depraved”, are labelled as “Very common in the 16th cent.”, and the form *chiefest* is “very common in the 16th and 17th c., and still frequent in literary use.” However, the density of citation evidence which a historical dictionary presents for a given sense or form can only correspond very roughly with the frequency of that sense or form (see e.g. Brewer 2007a, 112–113). So, the evidence presented in *OED* can sometimes give a misleading picture of the history of a word because it is so loosely linked to questions of absolute or relative frequency. In a discussion of Chaucer’s contributions to English vocabulary, Carter Hailey (2007) argues, inter alia, that even when a word used by Chaucer can be shown to have been used by a previous writer, its frequency may have remained vanishingly low until Chaucer took it up, or reborrowed or reinvented it, and may then have risen significantly. The same—or, in the case of the reinvigoration of obsolete or obsolescent forms, the like—must be true of other widely circulated authors, for instance Walter Scott (*foray* or *scantly*) or perhaps P. G. Wodehouse (*blotto* or *disgruntled*). The movement towards indications of absolute frequency in dictionaries such as the *Trésor de la langue française*, together with the enormous potential explanatory power of elegantly-conceived frequency graphs like Kučera and Stluka’s, suggests a way forward for the lexicological enrichment of

historical dictionaries, provided that they register a language variety for which there is a substantial, and accurately dated, historical corpus.

Kučera and Stluka's evidence begins in the fifteenth century; the next two papers in the collection address sixteenth-century texts. Ian Lancashire's, indeed, might be said to address the absence of a certain kind of sixteenth-century text, as it asks "Why did Tudor England have no monolingual English dictionary?" The question is indeed a striking one: the first free-standing monolingual dictionary of English which was not overtly restricted to the coverage of vocabulary from a single subject-field was Robert Cawdrey's *Table Alphabeticall* of 1604, and in view of the vigorous publication of texts in English in the sixteenth century, the fairly high quality of bilingual and subject-specific lexicography in England in the same period, and the modest scope of Cawdrey's little book, one might well ask why he had no predecessors in the Tudor period, the long century which runs from 1485 to the year before Cawdrey's *Table* was published.

One answer might be that Cawdrey's *Table* is really not a general English dictionary but a *Fremdwörterbuch* of borrowings from Latin, Greek, and French. In that case, the beginnings of general monolingual English lexicography might be redated in either possible direction. They might be seen as later than 1604, in the "Alphabetical Dictionary" of Lloyd and Wilkins (1668; cf. Dolezal 1985) or even in the despecialization of the wordlists of English dictionaries shortly after 1700. On the other hand, it might be argued, as Lancashire has done in the past, that some degree of specialization does not rule out the claim of a given dictionary to be a dictionary of English, and that therefore John Rastell's *Exposiciones terminorum legum anglorum* of 1523 "has a reasonable claim to be the first printed stand-alone monolingual English lexicon, and John Rastell to be our first native lexicographer" (Lancashire 2004, 243). Yet another line of argument, on which Lancashire touches in the present paper, might be that overemphasizing monolingual lexicography confuses the issue: that there were excellent dictionaries of English in the sixteenth century in the form of the English–Latin sections of Latin dictionaries, in a tradition culminating in Rider's of 1588, and that the fact that these explained English words in Latin rather than in English did not become important until the rise of a reading public who were interested in vernacular texts which used a difficult technical vocabulary but were ignorant of Latin. In that case Cawdrey's dictionary, which is meant in particular for readers of Christian texts (see Brown 2001), might be compared to the Hebrew–English and Greek–English dictionaries published a few decades later for the help of Anglophone readers who wanted direct access to the Scriptures but did not know enough Latin to use Hebrew–Latin and Greek–Latin

lexica. Lancashire also remarks in the present paper that the manuscript Old English–modern English dictionaries compiled in the sixteenth century are arguably monolingual; there seem to be no truly monolingual general wordlists of English from the sixteenth century.

Most importantly, however, he here explores an argument which is more about the history of words and attitudes to words than about the history of dictionaries as such. On the one hand, features of the printed vocabulary of sixteenth-century English such as its unstable orthography would have presented a problem to monolingual lexicographers. On the other, the fact that the English vocabulary was growing rapidly might have been expected to stimulate its lexicographical record. Lancashire discusses this growth with reference to its documentation in the *Oxford English Dictionary* and also in the database *Lexicons of Early Modern English*, which he founded and edits, commenting on the ways in which *LEME* evidence supplements, and helps to modify, *OED* evidence, and encouraging his readers to reconsider the conventional arguments for a drastic sixteenth-century increase in the size of the vocabulary. He concludes by reflecting on a final possible reason for the absence of a general monolingual dictionary of English in the sixteenth century: the identification of the vocabulary of English by several contemporaries as defective in comparison to the vocabularies of Latin and even the more prestigious European vernaculars (cf. McConchie 1997, 14–61).

In her paper “*Dat Boek der Wundenartzstedye*: The Low German translation of Hieronymus Brunschwig’s *Buch der Cirurgia* and its rendering of surgical lexicon”, Chiara Benati examines one particular document which illuminates the sixteenth-century lexicology of a European language. This is a version of the *Buch der Cirurgia* which was compiled by Hieronymus Brunschwig and first published in High German in 1497.¹ The vocabulary of Brunschwig’s book, as transmitted through translations into other European vernaculars, influenced those vernaculars, due in part to Brunschwig’s own didactic procedure (for which see also Benati 2008). Benati quotes a typical passage in which he defines his subject as “chirurgia das ist die hantwirckung in der wundertznie” [“chirurgy—that is, hand-working in the doctoring of wounds”], making a compound of *hant* “hand” and *wirckung* “working” to render Latin *chirurgia* “surgery”, which is from Greek *kheirourgia*, a compound of *kheir* “hand” and *ourgia* “force, operation”. This seems to be the original

¹ Brunschwig’s name appears in English-language sources in various forms: *OED*, for instance, calls him variously (Jerome of) Brunswick, Brunswyke, and H. von Brunschweig; Roderick McConchie (1997, 22) refers to him as Hieronymus von Braunschweig.

coinage of this word in German (cf. *OED3* for the use circa 1400 of *hand-working* in English): the first citation for it in *DWB* is from a later medical text, Jakob Rueff's *Trostbüchle* of 1554. The reader of Brunschwig was being taught a vocabulary, and the lesson was learned across Europe. For instance, the English translation of the *Buch der Cirurgia* (1525) gives *OED* its first instances of *fracture* n., *phthisis*, *spatula*, and *uvea*.

The Low German version which Benati studies here, published in 1518 and apparently known only from a single copy in Berlin, does not stand at the beginning of a long tradition of surgical writing in Low German, a language variety whose functions were to become increasingly restricted in the course of the next two centuries. Significant as it is as a monument of early modern Low German, and as an example of the negotiation between the vocabularies of Latin and of a vernacular language variety in early modern technical writing, it is also especially interesting as an example of the relationship between High German and Low German in the sixteenth century. Benati concludes that the maker of the Low German *Boek der Wundenarzstedye* is likelier to have thought of it as a new edition of the High German *Buch der Cirurgia* than as a translation.

Mateusz Urban's "Cleaning up the mess: The case of English *balcony*", in methodological contrast to the three preceding papers, is an etymological study of a single word, or at least a single family of words. The immediate derivation of English *balcony*, borrowed no later than 1618 from Italian *balcone*, is clear enough, and Urban does not linger over it. (It is made even clearer, as Urban notes, when the original pronunciation of the English word, with stress on the second syllable, is considered—the poet Samuel Rogers, who died in 1855, remarked that "'balcōny' makes me sick" [1856, 248: the diacritic indicates an unstressed vowel], on which remark Max Müller commented in 1861 that the new pronunciation was not yet universal, but was more common than the old [1862, 36].)

However, the etymology of the Italian word is much less clear, and correspondingly much more interesting. The "mess" of Urban's title is twofold. On the one hand, a possible cognate of *balcony* and *balcone* is Russian *balagán* "stall, mess", so there is quite literally a mess to be cleared up in the etymological story—and here it is striking that a semantic development from "place in which potentially messy activities are carried out" to "mess" seems to have taken place several times, not only in *balagán* and the Polish *burdel* "brothel, mess" adduced by Urban but also in English *shambles* "butcher's shop" (15th cent.), "slaughter-house" (16th cent.), "disgraceful state of confusion" (20th cent.). On the other hand, the etymology of *balcone* is particularly confused, with four different theories respectively proposing origins in Germanic, two different Persian forms,

or Latin, and a related theory proposing a connection between one of the Persian forms and a widely distributed set of forms in languages ranging from Polish through Russian to Kirghiz and Kazakh, Buriat and Yakut.

In clarifying the possible relationships between these forms, Urban extends the range of this collection to include a vitally important part of historical lexicology, namely etymological investigation—and, in parts of his paper, the romantic and demanding subfield of etymological investigation which addresses *Wanderwörter*, as in A. S. C. Ross's classic *tour de force* on the family represented by English *ginger* (1952) and, nearer the present day, the work of Urban's colleague at Krakow, Marek Stachowski, on etymologies such as those which include English *mammoth* (2000) and *sabre* (2004). But despite the breadth of its subject matter, which necessarily includes not only a great diversity of linguistic forms but also a short architectural history of the balcony, Urban's paper stays faithful to its title and works on clearing up the mess, or confusion, of its subject matter with an elegant concision for which not all etymologists strive (see Malkiel 1993, 67–71 for an account of a counterexample).

Just as Urban's paper examines forms in languages spoken across Eurasia from England to Siberia, so Jane Samson's "Do savages get the blues? William Colenso and the nineteenth-century colour debate" shows a lexicological question posed in the study of ancient Greek being transported in the nineteenth century to New Zealand. Goethe's anti-Newtonian investigations of ideas of colour from the 1790s onwards had led him to ask how colour was understood by ancient Greek writers. He was not the first modern scholar to touch on this question, and did not pretend to be, reprinting the *De coloribus libellus*, an exploration of the ancient words for colour published by Antonius Thylesius (Antonio Telesio) in 1528, as part of the historical section of his *Zur Farbenlehre* (Goethe 1810/1991, 640–653; for Thylesius on colour terms, see Osborne 2002). But Goethe's arguments were vastly more elaborate, and more influential, than those of Thylesius, and the questions Goethe posed were taken up by a number of later writers, not least the statesman and classicist W. E. Gladstone, who proposed that as societies evolve, they develop more complex perceptions of colour. So it was that Homeric Greek became the only dead language discussed—if not as competently as it might have been—in Berlin and Kay's *Basic Color Terms* (1969, 70–71; cf. Sampson 1980, 99 and 249n5).

Samson traces the nineteenth-century development of this debate in a New Zealand context: Māori society, it was suggested, had not evolved far enough at the time of contact with Europeans for its members to have

developed a word for “blue”. The racist undertones visible in this suggestion are perceptible in other lexicological discussions: the paper of 1918 cited above as providing the first known use of the phrase *historical lexicology* in English claims that “savage races” or “savage tribes” do not have the vocabulary to express abstract concepts rather than particular colour words, but its evolutionary perspective and its interest in the language of “savage” humans are of a piece with one of the positions documented by Samson (Clark 1918, 177). Samson is an historian of missionary activity, and she keeps this context of both sides of the debate over Māori colour terms clearly in sight; her paper contributes not only to lexicology but to the history of missionary linguistics (a flourishing area of inquiry: see e.g. Zwartjes and Koerner 2009, 202–205), and comments on the way in which missionary work on languages was to be overlooked or forgotten. The example which she cites as revealing “the occlusion of missionary contributions to the colour debate” is the contribution of W. H. R. Rivers to the reports of the Cambridge Anthropological Expedition to Torres Straits, and this brings us back to Berlin and Kay, for whom Rivers’s work was an important source (Saunders 2000, 84). As Samson remarks in her conclusion, “developmental theories ... formed in the context of modern colonialism” have had a surprisingly long life.

Vivien Wazsink’s “*Biologisch* vs. *biokoffie*, *ecologisch* vs. *ecotrend*: Dutch *bio*- and *eco*- through the ages in the dictionary” looks at two related case studies of the nineteenth- and twentieth-century proliferation of neo-classical compounds, in which elements from Greek and Latin are used to form words in the modern vernaculars which were not present in the ancient languages. After introducing the role of this kind of compound in the Dutch language, where it appears not to have been productive as early as in French and English, Wazsink goes on to give an overview of the productivity of *bio*- and *eco*- in modern Dutch (there is interesting comparative material in *OED3* s.v. *eco*-, an entry revised in 2009²).

In the third section of her paper, Wazsink turns to the presentation of these elements which she has undertaken in the *Algemeen Nederlands Woordenboek (ANW)*, a corpus-based, electronic dictionary of the Dutch language since 1970 (the coverage of the great *Woordenboek der Nederlandsche Taal* extends only to 1976). ANW “is not a clone of an existing printed dictionary, but ... truly represents a new generation of electronic dictionaries” (Moerdijk, Tiberius, and Niestadt 2008, 18). Its entry structure naturally has room for all the usual elements of the

² The *OED3* entry for *bio*- is at the time of writing something of a muddle, consisting of elements from *OED2*, the 1993 and 1997 volumes of the *OED Additions Series*, and further additional material of 2007.

dictionary entry—lemma, spelling, pronunciation, definition, and so on—and also provides systematic access to information about the morphology of the word; information about its semantic developments such as specialization and metonymy; and information about its semantic relationships such as antonymy. Moreover, the definition is supplemented by what its editors call a semagram, “the representation of knowledge associated with a word in a frame of ... conceptual structure elements which characterise the properties and relations of the semantic class of a word meaning”, so that the semagram for entries for animals has conceptual structure elements, or “slots”, for sound, colour, size, build, parts, function, and so on (ibid. 19).

Tommaso Pellin’s “Masters, mothers, and barking dogs: The lexical family of the words for grammar in China” discusses the words for grammatical concepts which were developed by writers in Chinese of the nineteenth and twentieth centuries. The concepts themselves were often derived from the Western grammatical tradition: before the nineteenth century, there was a strong and elaborate tradition of lexicography in China, extending to the lexicographical treatment of function words (Yong and Peng 2008, 314), but there was not a tradition of systematic grammatical analysis like those which had developed in India, the Arabic-speaking world, and Greece.

This makes Pellin’s study an investigation of lexical interference as well as of a Sinological topic: Chinese writers on grammar were bound to consider grammatical terminology from the Western tradition, and their own terminology developed under its influence. After surveying the kinds of lexical interference from Western languages which can be detected in the grammatical terminology of a corpus of Chinese texts on grammar published between 1859 and 1924, he comments on these texts individually. The first, the *Lading wenzi* of Angelo Zottoli, is a grammar of Latin for Chinese readers in which the sense “interjection” is expressed by *cuci*, from *cu* “dog that suddenly jumps out from the grass and chases people” plus *ci* “word”, so the barking dogs of Pellin’s title really are there in the grammatical terminology. The early terminological experiments discussed here do not seem to have survived: like the products of lexical interference discussed by Pellin in an earlier paper, they might be seen as “messengers” from a different set of cultures, “and messengers are sometimes welcomed, and sometimes rejected” (Pellin 2008a, 265).

The evidence on which Lise Winer draws in her “Historical naming strategies for fauna in Trinidad & Tobago English/Creole” likewise comes from the nineteenth and twentieth centuries, though it extends much further into the latter than Pellin’s. Much of it was gathered in the three

decades of Winer's work on the *Dictionary of the English/Creole of Trinidad & Tobago* (*DECTT*, Winer 2009). An interviewer who asked Winer what attributes she needed in her work recorded the answer:

“Sitzfleisch”, she says—the patience and persistence to sit on one's behind for hours doing tedious research.

It also helps to have a strong interest in natural history since Trinidad has more flora and fauna than any other Caribbean country. (There are 500 different birds alone, from caciques to cornbirds.) But “you have to be a generalist”, she adds ...her work runs the gamut of “fish, food, forest, frogs and flying things”. (Haldane 2003)

The present paper addresses a topic bearing on natural history: the names given to fauna in Trinidad and Tobago. Winer begins by setting out a general principle for explaining the distribution of imported and indigenous names for fauna. She then goes on to comment on her treatment of zoonyms in the *DECTT*, before offering an authoritative overview of the sources, kinds, and motivations—colour, resemblance, form, and the like—of Trinidadian and Tobagonian zoonyms. This is the latest in a long series of her scholarly contributions to the study of the language of those islands (cf. Winer 2007a), most or all of which must ultimately derive from the *DECTT* project. She remarks of the dictionary that “it is hoped that the inclusion of all common names [for fauna] in the *DECTT* will help to legitimize local zoonyms, and facilitate their use”. The possibility of *DECTT*'s making a difference to language-users in Trinidad and Tobago is likewise suggested in the observation made at its launch by the Creolist Lawrence Carrington, that it “fills a longstanding gap in the cultural self-knowledge and educational resources of this nation” (Carrington 2009). But Winer's work in the present paper and elsewhere has evident relevance beyond a Caribbean context, suggested in her own comments on the possible analogies between her materials from Trinidad and Tobago and the zoonymic terminologies developed in the Englishes of Australia, South Africa, and elsewhere.

The two last papers in the collection address lexicological questions—in both, as in Winer's paper, in a lexicographical context—which bear on continuing developments in the English language and its relationships with other languages. In “The third edition of the *OED* and lexical transmission: Towards a consistent research methodology”, Mirosława Podhajecka discusses the documentation in *OED3* and earlier versions of *OED* of words whose origins are, at least in part, Russian. This is a timely subject. The etymologies in *OED1* were, from the beginning, clear-headed and often impressively learned, benefitting from the rapid progress of

etymological studies in nineteenth-century Europe, the application of this work to the English lexicon in Skeat's excellent *Etymological dictionary* (1882; cf. Malkiel 1993, 31 and Durkin 1999, 2–3), and the consultation of a wide range of experts on individual languages and groups of languages. But their makers were constrained not only by time but by the resources available to them, and *OED3*'s etymological team are making very striking advances on the work of their predecessors (see Durkin 1999). Podhajecka discusses this work and some of the problems it presents, for instance in the classification of more or less naturalized words of foreign origin, with reference to etymologies in the fully revised range in which Russian is cited, concluding that "the quality of the new etymologies is first-class". She also asks what *OED3* is omitting, identifying antedatings, variant forms, and unrecorded words located in the course of searching the Google Books database (for a complementary but very different approach to some of this material, see Leeming 1968–9).

The methodological problems which Google Books presents are an important feature of Podhajecka's paper. On the one hand, it gives access by keyword to an extraordinary wealth of texts, in multiple languages, with publication dates from the seventeenth and even sixteenth century to the present day. On the other, its metadata are often extremely problematic: a controversial article by the linguist Geoffrey Nunberg has called them "a train wreck: a mishmash wrapped in a muddle wrapped in a mess" (2009). Moreover, many texts are inaccessible to some or all users for legal reasons, and as Podhajecka remarks, the accessibility of some texts changes unpredictably. Notwithstanding the shortcomings of Google Books, Podhajecka's findings show what can be done with the resource—at least by a patient and determined researcher who asks it the right questions about relatively uncommon lexical items.

Isabel Casanova's "Should Portuguese dictionaries register English words?" discusses the converse process to that which Podhajecka documents: the borrowing of English words into another language, in this case Portuguese (which is not, by the way, included in Manfred Görlach's *Dictionary of European anglicisms* [2001], making its treatment here all the more valuable). Her argument opens with an account of the academy principle, by which the major officially sponsored dictionary of a language, a dictionary like those produced in the seventeenth and eighteenth centuries by the Académie française, the Real Academia Española, and the Academia Real das Ciencias de Lisboa (now the Academia das Ciências de Lisboa), would document the vocabulary of the prestigious standard variety of a language, omitting or stigmatizing low-status lexical items, and aiming to stabilize the language and to preserve its perceived

purity. Whereas the successive editions of the *Dictionnaire de l'académie française* have maintained this principle, the *Dicionário da língua Portuguesa contemporânea* of the Academia das Ciências de Lisboa (2001) does not: it is generously receptive to anglicisms, “to the great consternation of the Portuguese intellectual elite” (Casanova 2008, 27). Casanova goes on to describe some of the orthographic and phonological features of recent English loans into Portuguese, concluding with a question—“what is the future for the relationship of Portuguese and English in the European context?”—and with the suggestion that “It is down to lexicographers to decide what role a dictionary should play for a language in crisis.”

3. Conclusion: historical lexicology in the present volume

The contributions to the present volume belong to a number of subfields of historical lexicology, and show a number of techniques at work. One might say, to revisit Díaz Vera’s point about the importance of dictionary and thesaurus projects in stimulating work on historical lexicology—or to take up M. A. K. Halliday’s remark (2004, 20) that “the best source of information about lexicology is the dictionary or thesaurus”—that these contributions are variously related to the historical corpus or database, the historical thesaurus, and the historical dictionary. Kučera and Stluka’s study of the historical frequency of morphological variants is based on a carefully selected historical corpus, Lancashire’s study of the sixteenth-century lexicon and the motives for its lexicographical representation makes extensive use of a database of early lexicography, Pellin’s study of lexical interference in a particular semantic field draws on a small corpus of grammatical texts, and Podhajecka’s study of loanwords and their lexicographical treatment draws on Google Books as, in its own way, a huge disorderly corpus. Benati, Samson, and Winer share with Pellin an interest in the interplay of words from given semantic fields, remote as the Low German, Māori, and Trinidadian and Tobagonian Creole of their investigations are from the Chinese materials which he addresses. Waszink, Winer, and Casanova all write as lexicographers, and Urban’s etymological investigation responds in particular to the claims of lexicographers of the last two centuries. Different typologies of work in lexicology in general and historical lexicology in particular would of course bring the contributions together in different patterns.

To return to the contrast drawn by Pellin and quoted at the beginning of this introduction, all ten papers share an interest in empirical evidence rather than in lexicological study at a highly theoretical level, and in the

wide contextualization of the words which constitute this evidence in the social and cultural lives of their users. What Philip Durkin (2009, 237) says of etymology might be said of historical lexicology as a whole and as practiced in this volume: “Few areas of study offer points of contact with so many fields”.

CHAPTER ONE

DIACHRONIC CORPORA: SEEING HISTORIES OF WORDS FROM ANOTHER ANGLE

KAREL KUČERA AND MARTIN STLUKA

Detailed descriptions of changes in the lexicon of a language usually focus on such key points in the history of particular words as their origin (including, if possible, the time of their borrowing or coinage), the times of changes of their forms and meanings, and—where applicable—the time of their going out of use. As a broad generalization it can be said that the existing descriptions of histories of particular words are rather discontinuous, consisting, as a rule, of a limited number of points in time. In contrast, very little is known about periods between these key points, about changes in frequency of words over centuries, and about their coexistence with synonymous words and variants.

The aim of this contribution is to show that even at their present, rather elementary stage of development, diachronic corpora encompassing the entire (written) history of a language seem to demonstrate a significant potential to broaden the traditional focus by making it possible to explore the history of a word as a continuum and to compare it with the histories of other words, variants, or forms. The Diachronic Part of the Czech National Corpus (DCNC: for details see [<http://korpus.cz>](http://korpus.cz)) is used here to show some of the possibilities, even though its exploitation is limited by its current size of about 3 million running words—a size which is too small to provide either adequate coverage of the seven centuries of Czech written texts or reliable information about changes of absolute frequencies of most words and forms. (One should be aware, however, that the DCNC, as well as diachronic corpora of other languages, may never be totally reliable in this respect: their data on history of words, especially the lower frequency ones, will inevitably remain skewed by the unequal representation

of various text types, topics, literary styles etc. in the texts preserved from early language states.)

To minimize the influence of the above-mentioned limitations, (a) the texts of DCNC were organized in 100-year clusters, which proved to be large enough to show the general tendencies rather than differences between individual texts; (b) the main focus has centered round high frequency words and forms, the use of which is largely independent of topics and literary styles; and (c) attention has been confined to comparison of relative frequencies in groups (mostly pairs) of coexisting synonymous words and forms, rather than their absolute frequencies. (To be more specific about (c) and to clarify what the graphs below show: if, for example, the absolute frequencies of the compared words A and B found in the DCNC in the fifteenth century were 60 and 140 respectively, their relative frequencies would be 30% and 70%; these percentages would then be plotted on the graph at the year 1500 and a line would be drawn to link each of the two points with the corresponding points at the years 1400 and 1600.) Fifty probes based on these principles revealed several cases of rather monotonous history characterized by synonyms or variants remaining in a stable state through the whole of their coexistence, with more or less constant frequencies and no marked tendency to change their positions. Figure 1.1 shows a prototypical example of such monotony, extending over the seven hundred years of coexistence of *nikdy* “never” and its infrequent variant *nikda*, which vanished completely from Czech written records around the turn of the nineteenth century, but still existed in spoken Czech in the twentieth century.

Cases like this are quite remarkable from a general linguistic viewpoint, since they do not really correspond with the idea of language as an entity undergoing constant change, but may not be as interesting from the point of view of lexicology and lexicography dealing with the history of words of a particular language, since the long-term linear parallel coexistence provides practically no new information about the history of the words or variants except for the fact of parallelism itself.

However, the probes into DCNC show that the most usual way of coexistence of two or more synonyms or variants is competition, not the type of monotonous parallelism shown in Figure 1.1.

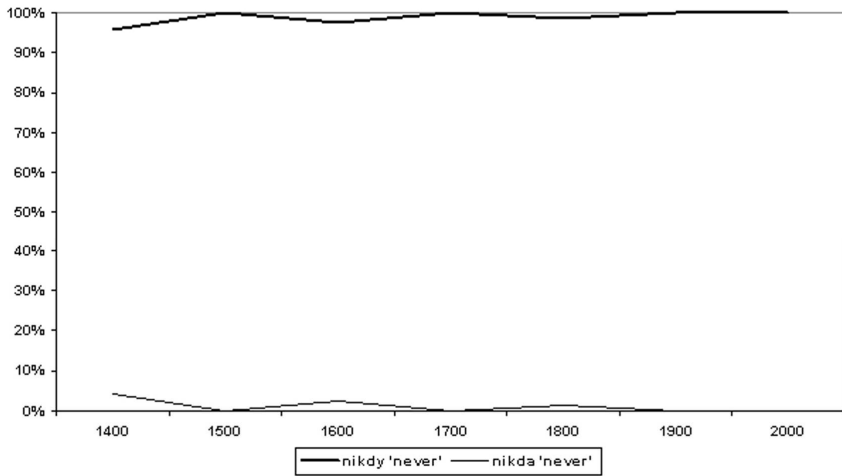


Figure 1.1. Monotony: History of coexistence of the variants *nikdy* and *nikda*

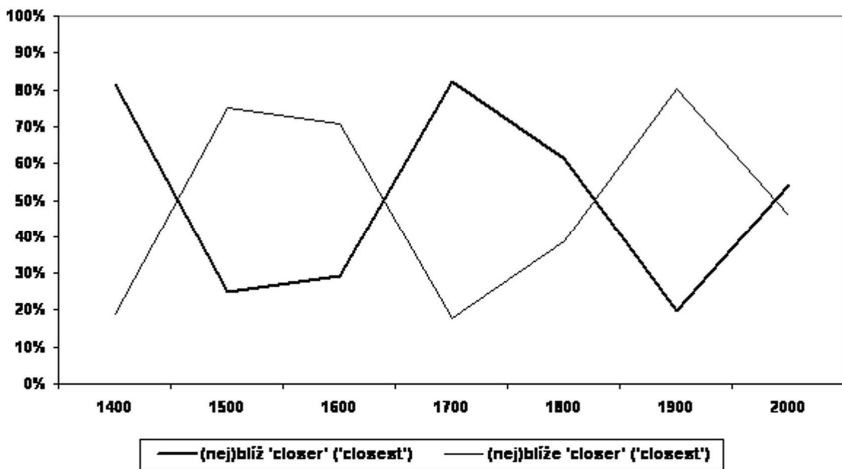


Figure 1.2. Fluctuation: History of coexistence of the variants *blíž/nejblíž* and *blíže/nejblíže* “closer/closest”



Figure 1.3. Fluctuation: History of coexistence of the variants *ač*, *ačkoli* and *ačkoliv* “(al)though”

In some instances, for instance Figures 1.2 and 1.3, the competition may be rather complex, but generally, rivalry between two coexisting words or forms is characterized by much less erratic variation than in Figure 1.2. In such cases the data often reveal as yet unknown long-term tendencies or evident changes, not just undirected fluctuation:

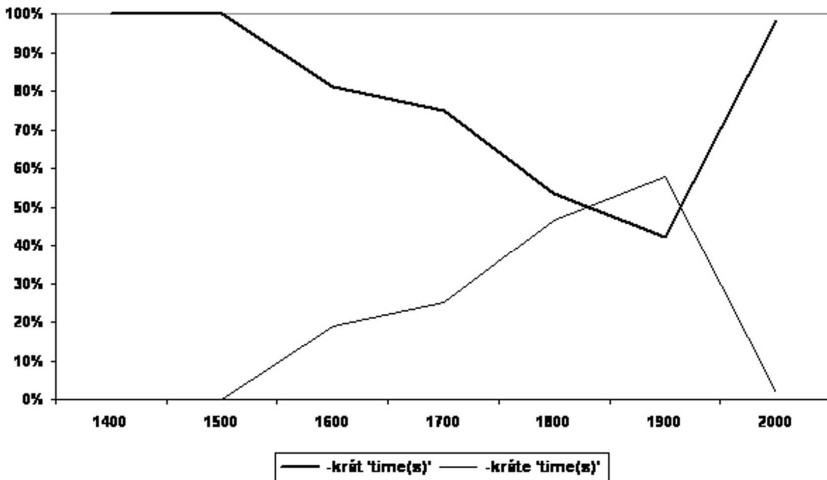


Figure 1.4. Tendencies: History of coexistence of the variants *-krát* and *kráte* “-time(s)”

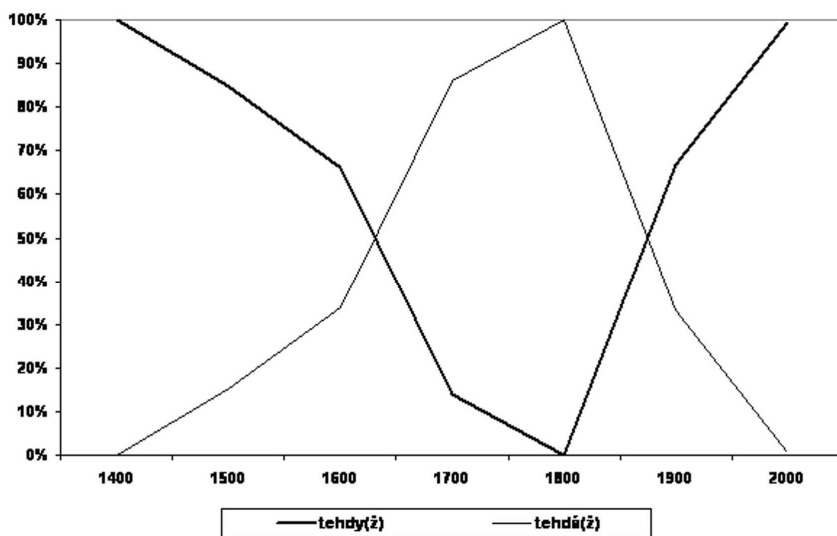


Figure 1.5. Reversal: History of coexistence of the variants *tehdy* and *tehdá* “then”

Maybe this kind of information will be used in practice one day, being included in dictionaries alongside with details about words’ origins or changes in their meanings, but at present—scanty as they are—the data raise historical and lexicological rather than lexicographic questions. Even from the limited number of 50 probes into the DCNC it seems to be evident that competing words sometimes reverse their positions completely, or double-reverse them, with one of them falling out of use (as *tehdy* did, after a gradual 400-year decline, during the eighteenth century, as shown in Figure 1.5), but coming back again and pushing the other member of the pair (*tehdá* in this case) into almost complete oblivion.

New facts like that call for clarification, which we may or may not be able to provide. The case of *tehdy* and *tehdá* seems to be relatively easy to explain: the turn of the nineteenth century, when the two variants started to reverse their positions, was the time of the Czech National Revival movement, when the Czech language was being reformed and enriched to have the necessary vocabulary and other qualities for use not only in everyday communication, but also in prose, poetry, and scientific texts. There are very detailed accounts of technical and literary words being newly formed or borrowed from foreign languages at the time, but what has been unknown is that the histories of some common Czech words were obviously changed then too, with less common or obsolescent words and variants like *tehdy* being chosen as poetic expressions valued for their out-of-the-ordinary stylistic qualities. This explanation must

still be verified through additional observations, but it is already supported by the fact that the fifty rather random probes have revealed four more pairs of words which reversed their positions as rapidly at the same time: *teprve/teprvé* and *teprva/teprvá* “not until”; *nejprv* and *nejprve/nejprvé* “first”; *(ne)jinak* and *(ne)jináč* “(not) in a different way”; and *výše* and *vejše* “higher”. Predictably, individual cases of similar competing pairs of synonyms and variants have also been found to reverse their positions at other times than the National Revival, and to find a plausible explanation for them may be more difficult. However, it does not seem unlikely that random reversals of positions in competing pairs of words or variants happen from time to time simply as part of continuous language variation, and as such many of them may not have any specific explanation at all.

One more fact, in a way similar to these reversals, has been revealed by the probes into the DCNC: a set of eight pairs of competing words and variants has been found, in which each of the pairs shares a similar history with another completely unrelated pair from the same set. An example of two such pairs sharing similar history (*nikdý* and *nikda* “never” and *děl/dýl* and *děle/dýle* “longer”) is shown in Figure 1.6 on the following page. This is another fact demanding explanation, but obviously many more data are needed to find out if sets of similar histories like these arise by mere coincidence or reflect deeper tendencies influencing unrelated words and variants at particular times.



Figure 1.6. Sharing histories: Similar histories of two unrelated pairs of competing variants (*nikdý* and *nikda* “never” and *děl/dýl* and *děle/dýle* “longer”)

The above are all examples of new, as yet unknown facts about histories of words revealed by a rather limited number of probes into one rather insufficient diachronic corpus. They call for interpretation, they raise questions and do not provide answers, but they are facts, and however insignificant their number is at the moment, they seem to open up new vistas. Although further data will have to be accumulated to verify some of the above statements, it seems obvious that diachronic corpora promise to provide a firm basis for identifying both centuries-long trends in the histories of particular words and general tendencies of historical developments, and as such, they should become lexicographers' tools in the future. More emphasis on the historical continuum of language and more emphasis on the study of words as competing and interrelated members of sets of synonyms (as well as other sets) can be named among the most obvious foreseeable contributions of diachronic corpora to historical lexicology and lexicography.

CHAPTER TWO

WHY DID TUDOR ENGLAND HAVE NO MONOLINGUAL ENGLISH DICTIONARY?

IAN LANCASHIRE

The General Editor of the *Oxford English Dictionary* a century ago said that “no one appears before the end of the sixteenth century to have felt that Englishmen could want a dictionary” (J. A. H. Murray 1900, 26). Is James Murray’s comment true, and if so, why is it?

Richard Mulcaster, master of the Merchant Taylors’ school in London, published an English word-list of over eight thousand terms in his *Elementarie* in 1582. In it he distinguished between native or mother-tongue vocabulary and those words that, though borrowed from a foreign language, had been enfranchised in English. Mulcaster proposed that someone use his word-list to make the first monolingual English dictionary.¹ He believed that knowledge of the English language would be central to basic education in the future. Patriotism and a genuine love of English moved him to interrupt a good treatise on elementary education with page on page of sparsely-encoded columns of words.

Mulcaster’s patron—the man who allowed the dedication of the *Elementarie*—was Robert Dudley, earl of Leicester. An oligarchy, the leading members of which were the Queen, her favorite Dudley, and her Treasurer, William Cecil, governed England in the 1580s. Dudley was dedicatee of a hundred books (F. B. Williams 1962, 58) and, after becoming Chancellor of Oxford University in 1564, helped to restore the University Press two decades later. Dudley admired dictionaries. For example, he successfully secured, with Christopher Hatton (Privy Councillor and Vice-Chamberlain of the Household), a monopoly, dated

¹ Printed bilingual dictionaries with English headwords or explanations date from *Promptorium Parvulorum* (1499), for a history of which see Stein 1985, 91–106.