Review Essays

Recent Trends and Works in Lexicography: A Review

MICHAEL HERRIMAN

The last two decades have been notable for a remarkable increase in the number of English dictionaries published. The increase can in part be explained by the development of a new field known as corpus linguistics, but the phenomenon is probably most attributable to the dramatic surge in the number of people teaching or studying English as a second or other language. This in turn has been a result of the globalization of business and finance, migration trends and the expansion of the EFL business generally[1]. Publishers have catered to a need in producing dozens of "new" dictionaries every few years, many of which are specialized in that they confine themselves to specific areas of interest, or to specific needs or levels of learners.[2] The development of corpora of English in particular has also spurred a new approach to compiling dictionaries. This approach stands in marked contrast to the traditional one, an approach typified by the compiling of the *Oxford English Dictionary* (OED). We might think historically of three approaches to lexicography: the first, that of the scholar defining words based on a personal acquaintance with the language (Samuel Johnson's great dictionary being a notable example); the second, the task seen as representing the meanings of words according to their historical use in written forms of the language (the representative example being the OED); and the last, the compilation of a dictionary based on the actual contemporary use of words in the context of larger strings of language across many different spoken and written situations (the first and perhaps most notable example being the Collins *Cobuild Dictionary*).

The books under review here, three recent publications in lexicography -- one historical and the others contemporary -- serve to focus our attention on issues in lexicography that are important both for linguistics and for English language pedagogy. The fourth, more encyclopedic than lexicological, focuses on the state of English in the contemporary world. The books are:

Simon Winchester. *The Meaning of Everything: The Story of the Oxford English Dictionary*. Oxford University Press, New York, 2003. 259 pages. ISBN 0-19-860702-4

Linda Mugglestone (ed.) *Lexicography and the OED: Pioneers in the Untrodden Forest*. Oxford University Press, Oxford, 2000. ISBN 0-19-823784-7

Geoffrey Leech, Paul Rayson and Andrew Wilson. *Word Frequencies in Spoken and Written English*. Longman, London, 2002. 304 pages. ISBN 0-582-32007-0

Tom McArthur *The Oxford Guide to World English*. Oxford University Press, Oxford, 2002. 501 pages. ISBN 0-190866248-3

The historical work is Simon Winchester's *The Meaning of Everything*, a study of the compilation of the Oxford English Dictionary. The second, *Lexicography and the Art of Dictionary Making*, is an edited work with contributions by prominent lexicographers. Third and perhaps most revealing is Geoffrey Leech, Paul Rayson and Andrew Wilson's *Word Frequencies in Written and Spoken English*. This work is based on the British National Corpus (BNC), a 100 million word collection, 90% of which is based on written texts. Together with these, it is pertinent to consider Tom McArthur's (2002) *The Oxford Guide to World English*, the latest book by this prolific author, editor of *Language Today* and the *Oxford Companion to World English*. These books all deserve to be taken seriously by any whose profession is teaching English as a second or other language. Leech et al. may even cause us to reflect in a radical way on the aims and methods of this activity.

Winchester's volume follows his earlier book *The Surgeon of Crowthorne*, in which he tells of the labors of William Minor, an American Civil War veteran and doctor whose contribution of reference slips to the compilation of the OED exceeded that of any other individual. For 20 years it was not known that Minor had been convicted of murder and confined for life to a prison for the criminally insane in Crowthorne, Berkshire. From his cell, he devoted his life sentence to a search of English literature to supply "slips" that formed the basis of the referents for the headwords of the OED. After reading that book one might have wished for a much fuller historical account of the OED. The book under review is just such a work. Winchester makes it clear that the history of the OED is inseparable from the life of Dr. James Murray, who like Minor, devoted his life to this last and possibly final great work of English literature. As well as the interest it provides in its biographical treatment of Murray, and in detailing the political intrigue that constantly threatened the project, the book serves well our understanding of the principles underlying the OED, in particular, the focus upon word and referent. It is perhaps an unfortunate, but quite understandable fact that most people think of dictionaries as defining words, and believe that the definiens legislates the use of the word defined (the definiendum). Prior to the OED this thinking was perhaps excusable, no better example of it being found in Samuel Johnson's dictionary, which persuaded readers to the great man's own definitions, and prejudices, regarding the referents of words. The OED under Murray had a different aim, one clear from the full title initially proposed – *A New English Dictionary on Historical Principles*. It was to be a dictionary based on documented historical uses of words, in particular concentrating on the origin of a word (its etymology) and subsequent changes in its spelling and usage, all verified by literary references.[3] For this reason we now look at headwords in the OED not to tell us what they mean, but rather to discover their history and development. That focus serves the other (defining) purpose of course, but does so in such a way that we discover the word to be like an organism in a particular environment, evolving according to changes in that environment. Hence we may come to think of words as ecologically responding the the pressures imposed on them by users of the language. In an ever-developing context of usage it makes little sense to think of words as fixed in meaning.

Winchester gives a very detailed account of the background to the dictionary, beginning with a survey of the history of the language itself, showing the extent of change in the language from Old English, through the Norman invasion, and the invention of printing. He surveys the history of dictionaries preceding the OED project, revealing the surprising number of earlier projects, and devotes some space to Samuel Johnson's two-volume work, which preceded the beginnings of the OED by one hundred years. Johnson's dictionary was drawn also from historical works, going back two hundred years, but with meanings based on his own often-idiosyncratic definitions.[4] Johnson's entries of 43,500 head words were superceded only by Webster's first edition (in 1790) which produced 70,000 head words.

Winchester then devotes considerable space to the origins of the idea of the work that was to become the dictionary. The Dean of Westminster, Richard Trench, originally proposed the project to The Philological Society in 1857,[5] as one to correct "…some deficiencies in our English dictionaries". The society, it seems, like many other voluntary institutions of the time, was imbued with the ambitious nineteenth-century ideals of progress, science and the dispelling of ignorance. It was the age of taxonomy, which meant that all objects (including words) must be placed in their appropriate category. One could speculate that the great explorations of the time and the new languages they revealed, were also a spur to thinking about language generally. The project was formally begun in 1860 and through its first three editors the work was known simply as the dictionary, and though we now refer to it as "The Oxford", its association with the press of that university came only after Macmillan (the third proposed publisher) realized that the printing of the dictionary, as it was emerging twenty years later, would be too large an undertaking.

From this background, Winchester then gets into full stride giving the reader an exciting account of the early days of the work, which were beset by problems, many of which were the result of personality clashes between its sponsors and editors. Both parties consisted of extraordinary characters, learned, but capricious and eccentric dilettantes, with egos to match their erudition.[6] One needs to read Winchester to appreciate the extent to which the fine ideal was frequently compromised by caprice. Some stability came to the project when the fourth editor, James Murray was appointed, in 1875. Murray was a member of the Philological Society, an amateur philologist, schoolteacher, and former bank clerk from a humble Scottish background. That he intended to do the job in addition to his school-teaching duties, reveals his immense capability for the job. He was another polymath, but his a Calvinistic background gave him a sense of mission that could be turned towards the immense task of preparing the dictionary. Its previous editor, Frederick Furnivall, had undertaken the task in a highly disorganized way, though following the general plan of the project, which was to back-up the entries with extensive references gleaned from the literature of English back to its earliest days. Furnivall's predecessor, Herbert Coleridge, had envisaged that the project would take four years, a reflection perhaps on his poor capacity for realistic thinking. In the end it took 70 years.

Murray's life became devoted to the dictionary, yet the task outlived him. The bulk of the book is given over to the forty years of his editorship. This story is told with an enthusiasm matching the task that Murray set himself. Under his guidance the job grew to a staggering size. Its main element was the collection of references (the "slips") from published works in English from the earliest available. These were contributed voluntarily by the general public following appeals for readers publicised throughout the English-speaking world. The slips were half-sheets of paper with the headword and a quotation properly sourced and dated. Readers were sent books and assigned to search for the earliest uses of words in any printed form and copy the text surrounding it as an illustrative example of the usage. Changes in usage were to be tracked also. Many thousands of readers were involved and six million slips were ultimately collected and filed in alphabetically organised pigeonholes.[7] These formed the basis for the task of defining or recording usage of the word. The second part of the task, that of providing the meanings for the words, was the task of sub-editors and Murray himself. Most sub-editors worked on one letter only. The senior editors and Murray then reviewed all the entries. Volumes were published as they were finished, in alphabetical order.[8] A massive task still faced the printers at OUP. Winchester tells us that there were 227,779,589 letters and 178 miles of type in the finished set, this in a day when each letter was set in lead type in galleys. The subsequent proofing was another enormous task.[9] The first volume (A to Ant) appeared in 1884. It was thought to contain every English word within that scope.

The headquarters of the project had moved from London to Oxford as it expanded in size, but its association with Oxford University Press was constantly fraught. Jealous academics and the head of the Press continually sought to intervene and Murray often threatened to resign. That Murray, an outsider with no degree, was able to confront them and have his way is a testament to his capabilities.[10] When Murray died in 1915 the chief-editorship passed to William Craigie, who, with two others, saw the project to its completion. There were 414,825 headwords in the complete set, but as we might expect, in the 70 years it took to complete the project, the language had expanded enormously. A supplement came out in 1933, a further one (of 4 volumes) in 1986, and in 1989 a revised complete OED containing 615,100 words was printed, and at the present time a revision of the whole work is in progress. The project, of course has spawned many other OUP derivatives of itself, *The Shorter OED, The Concise OED, The New OED, The Pocket OED*, and many others (for learners and for regional variants of English) based on the full work and/or supplemented by newly collected corpora.

In this brief review it is impossible to cover the wide ground that the author does. Winchester succeeds in telling a fascinating story, cleverly interweaving the history, the main and minor characters involved, the contributing readers, facts about the dictionary and lexicographical details. It is an excellent work of biography, but as well it is pleasingly serious. Another recent work, however, must be read for a closer academic analysis of the lexicographical basis of the OED. This is *Lexicography and the OED: Pioneers in the Untrodden Forest*, edited by Linda Mugglestone (2000). Its fourteen contributors cover most of the issues that the OED raised initially and tentatively (as the sub-title suggests). The contributions demonstrate the advances made both in scholarship and the conceptualization of language since the OED project began. Yet most contributors acknowledge the immensity of the task begun by the Philological Society and the perspective on language and lexicography that we owe to Murray and his assistants. Some chapters in the book will be of interest to English teachers and applied linguists in particular. These are the chapters on sense and definition (Ch.5; Time and Meaning), morphology (Ch.7; Words and Word-Formation), pronunciation (Ch. 10) and Mugglestone's own contribution (Ch. 11; 'An Historian not a Critic': The Standard of Usage in the OED). The latter chapter covers the controversies surrounding words thought fit to be in a dictionary, and ideas of proper, common and vulgar use. The OED took a prescriptive stance on such matters, descending even to spelling (Murray preferred *rime* and *ax*, to *rhyme* and *axe*, for example). The recognition of English as comprising the spoken language as well as the literary is a product of the last half of the twentieth century. The next work reviewed below represents this. In passing, however, another recent work (Landau, 2001) must be mentioned. Landau in *Dictionaries: The Art and Craft of Lexicography* gives his reader an excellent view of the current state of dictionary compilation. It is particularly relevant for his chapter on the changes to thinking about lexicography brought about by the collection and study of corpora. His conclusion is that "…dictionaries will be better than ever – truer to the language, fuller in descriptions, broader in their range of coverage" as a result of attention to corpora. He also notes that many contemporary publishers take great liberty in their claim to the use made of corpora in their compilation, suggesting that the claim not be taken seriously. This warning should not be taken to refer to the major publishers; especially Collins, Oxford and Cambridge.

The work of Geoffrey Leech et al. might be seen as the counterpart to the OED in a number of ways. Firstly, it looks at contemporary English, then it depends for information on a wide variety of written and spoken texts and lastly, its grammatical tagging is based on distributions – indicating the way in which the words in the lists were used in the sentences or strings in which they were found.[11] The corpus it draws upon is also based on contemporary rather than historical sampling, with the majority of the texts sampled having been published between 1985 and 1994.

Hence it gives a view of current or recent English usage. The great interest is in seeing how usage might conflict with ordinary presumptions about the meaning and relationship of words in strings. Further than that, however, it might cause us to reflect on some cherished notions about both language and words. Even if we don't go so far as Humpty Dumpty in telling Alice that words mean what he wants them to mean, we might become more accepting of the idea that the referents of words are not the objects they stand for or definitions in a book, but are rather ideas in our minds that are shared in our social interactions.

Leech's work is based on the British National Corpus (BNC) version 1.0. This corpus incidentally received significant contributions from OUP. Like the other contemporary corpora, it is stored on a computer which means that the tagging of words can be done automatically and collocations of words can be read by simply calling the word onto the screen where it can be represented in all or any of the strings in which it was collected. It is at present, the second largest corpus of English in existence.[12] Unlike the larger one, the BNC is built from a careful sampling of texts. All words are identified in terms of their textual origins. The authors claim:

> It is possible to extrapolate from corpus frequencies to inferences about the language as a whole because the compilers have taken pains to sample different kinds of speech and writing (e.g. conversations, novels, news reporting) broadly in accordance with their representation in everyday language use. (p.1)

The samples based on written text are drawn mainly from non-fictional expository writing (80%), again in contrast to the OED.[13] Books constituted the main source (59%). The spoken component, represented by 10 million words, is based on both conversational (40%) and task-orientated speech. The conversations were gathered by 127 persons representing a spread of ages from 15 years and up, selected from representative groups based on age, sex and social class across the UK, according to population figures. The other part was based on lectures, consultations, sermons, TV/radio and various other public activities in most domains.

The introduction gives a clear idea of the structure of the corpus and should probably be read first even though the temptation is to go straight to the frequencies. In their introduction, the authors detail the statistical procedures that are the key to understanding the significance of the frequency data and the categories used in tagging the words, discuss the lemmatization of the corpus, and illustrate the difficulty of making decisions regarding what is to count as a word. For some words, most nouns perhaps, this is not too difficult a decision to make, but for homographs, abbreviations, spelling variants, ambiguous lemmatization (an example given is of the lemmas of the two verbs *to find* and *to found*) and capitalization, the decision is not straight-forward. Multiword units were particularly difficult (*in spite of* is one cited) and in the frequency list they are treated as single word tokens.[14]

The total number of words in version 1.0 of the BNC is 757,087, but more than half (52.44%) occur only once in the whole (100 million word) corpus.[15] Of words occurring more than ten times though there are 124,002. From this vast collection, the authors have included in their tables only those words with a frequency of 10 instances or more per million.

There are 31 tables in the book. The two main ones consist of alphabetical frequencies for the whole corpus (lemmatized) and a rank frequency list for the whole corpus (not lemmatized).[16] Here is an example from the first table for the verb *drive*.

drive Verb	156	100	94
drive	44	100	95
driven	30	100	95
drives	7	96	94

driving	38	100	93
drove	37	99	90

The frequency for each of the lemmas therefore adds up to the one given for the headword. The verb *drive* will be distinguished from the common noun *drive* (as a headword) and its lemmas (*drive* and *drives*(pl.)), and from a separate headword entry for the noun *driver* (its lemmas being *driver* and *drivers*). The alphabetical frequency list gives data also on the range and dispersion of the word. These data tell us the number of sectors out of a total of 100, which include the word (its range), and the concentration within a text source (its dispersion). In the case of the latter it tells whether a word occurs regularly across its sectors or in bursts or clumps. This is particularly useful information.

Following the main lists of alphabetical and rank frequencies there are 28 lists in various groups. These include lists comparing speech and writing (in alphabetical, rank order and distinctiveness categories), lists of varieties of spoken language and of written language (looking at different contexts of usage), rank frequencies of words within word classes (merged into 10 main classes or parts of speech), and frequency lists of all the tagged grammatical word classes (134 or so classes being identified). For the teacher of English, the rank frequency of words within word classes is a list of some interest, perhaps even ahead of the two main lists. High frequency in writing indicates more formal usage, and in speech, it indicates what James Murray might have called common or even vulgar usage. These data will be useful alongside of the recent studies in text analysis and genre, for example those of Biber.

Do these lists tell us which words to focus on in our teaching? It is interesting to learn that the most frequent common noun is *time*, followed by *year, people, way, man, day, thing, child, Mr, government, work, life,* and *woman*. Their frequencies drop from 1833 per million for *time*, to 1003 for *man* to 631 for *woman*, the gender difference being quite marked as it is in *Mr* (673) vs. *Mrs* (221). The first 100 common nouns very much represent every-day language and are mainly one-syllable words. The authors note that the data confirm Zipf's principle, the speculation that the words most often used are those shortest in terms of syllable length (the explanation being economy of vocal effort).[17] The frequency list for verbs is also interesting. Of the first 70, only 11 are not one-syllable words, and those 11 are all of just 2-syllables. Adjectives are more varied, but adverbs follow the one-syllable pattern and defy the common notion that adverbs are words that end in *–ly*, as only 2 of the 70 most frequent follow this rule.

From this book, and obviously from the BNC as a whole, we might gain a very different perspective on language. The impression is clearly of the language we use everyday.[18] It gives a strong motivation towards use of real material in teaching. As well it might support more of a focus upon words in context, both in their collocational restrictions and their representation in various genre, and support the use of realia in teaching. If one believes in vocabulary learning, the book will be a good guide. Perhaps corpora can also shake up some cherished notions about morphology and grammatical categories. When we think of a language we think of words as its prime representation. This focus may not always serve us, or our learners well. Perhaps it is time to rethink the dominance of words in favor of giving more emphasis to extended groups or strings of words, seeing words as communicatively important in the context of their relation with other words and not as single units. The emphasis on the word is a product of the development of alphabetic literacy, in particular the separation of words in text. Prior to the era of Bede, text showed no boundary between words, leaving us to wonder what early scholars' ideas of morphology were. Overall, we need to restore spoken language to its role as the dominant form of and model for our native tongue. We might think more about this issue in teaching the language too. The corpora approach exemplified by the tagging categories of the BNC, also might lead us to think about grammar more in a distributional sense, getting away from the dominance of traditional prescriptive grammar. Such a grammar will look at words in terms of their relationships with other strings of words. The book does not provide the data for this, but one imagines it will be an area of future interest.

The Oxford Guide to the English Language (2002) follows Tom McArthur's earlier (and edited) *Oxford Companion to the English Language* and draws upon his prodigious publications in the general area over the last 15 years. The book has an unusual structure, being more like a set of index cards for all varieties of the English language throughout the world. Its main source is the *Companion*, but the newer work is more concise and updates much of the earlier work, the product of more than 150 contributors. McArthur begins by noting the further dispersal of English throughout the world making the point that it now has no center. The first chapter is a brief summary of language in general, dialects and varieties of English, language stability and language change, and a curiously sketchy section called the Default Mode, which uses what seem to be unconnected ideas to support the claim that English is the world's default language. Few would challenge this statement. The better part of the book follows. This is the analysis of the history, present state and status of English in the continents. Europe is covered first, yet the discussion is confined to Great Britain and Ireland. The European countries are mentioned only because of the historical influence of their languages on English grammar and vocabulary. The better part of this chapter is the discussion of regional dialects, but its only merit is its contemporaneity, as many others have done the job more thoroughly. In this and the succeeding chapters nonetheless, there are many written text quotations showing variety and change in the language.

The chapters following cover the English language in the Americas, Africa, Asia, and an amalgam comprising Australia, Oceania and Antarctica. The shock of discovering that the frozen continent might have a rich source of English speaking penguins is diminished when we realize the McArthur is really talking about the tips of continents and small islands surrounding Antarctica; the Falklands, for example. Discussion of the other continents also includes details of their native languages and the extent to which these have modified the variety of English spoken there. In these matters the reader is much better served by the recent edition of the book *Learner English*, which though much more limited in range, has a more satisfactory coverage of local influences on English. McArthur's chapters have a very wide, but rather shallow coverage of all the issues raised.

The final chapter is a conclusion regarding World English. Of all, it is the best since much of the material is not available elsewhere. There is not a clear development of theme here however: we jump from a discussion of very large languages, to Gender issues, then the global industry of English teaching and finally, to the most coherent section of all, on what might or could constitute standards of local and world English. His conclusion is that English is so deeply represented on all continents that there is now no centre of the language, and that its increasing global dominance is a threat to many hundreds of small & local languages. These latter claims, if they need to be argued at all, have been put forward much more cogently by others, notably David Crystal. We should remember that this is a reference book, but what one looks for in such a work is a degree of organization and clarity that is lacking here. It bears the mark of a heavily abridged version of something much larger, to whit, the excellent *Companion* referred to above. It is still a worthy addition to the reference section of a library.

CONCLUSION

The books reviewed and mentioned here serve well as an introduction to contemporary lexicology and lexicography. The overall impression we gain is the lexicography is perhaps entering its most productive era yet. The great dictionary projects of Oxford, Cambridge, Collins and New Heritage are expanding still and more strongly so, thanks to the use of corpora. The OED and many others are now available on CD ROM and also through websites, making them much more publicly accessible. Version 2.0 of the BNC is also available, further expanding our knowledge of words in use. The only drawback to this treasury of words is that the language itself is becoming more regionalized, so that as fine a collection as the BNC is, for example, we must realize that it is of British English and hence not representative of World English. This throws us back to the question of what variety of English to teach the EFL learner. In Japan, for example, should we teach British or American English, or the local variety?

For teachers of the language, the most useful work is the Leech (et al.) volume. Most teaching and learning situations will devolve to language of the kind used in contexts heavily represented in the sampling of the BNC, both in text and spoken forms. This work will likely tempt teachers to consult the corpora themselves, which are coming more into the public domain. These will provide interesting information about collocational and sentential contexts for words. As mentioned above, the introductory chapter of Leech contains a valuable and provocative discussion of word formation that can be read in conjunction with Chapter 7 in Mugglestone. Overall the book implicitly challenges the print bias of recent and past lexicography and the pedantic decrees of wordsmiths.

Scholarly, financial and practical considerations suggest the following strategies. Buy Leech (et al.) to place alongside your current dictionary. Order Winchester for your university's library and then borrow it for an illuminating and gripping read. Mugglestone is expensive, but indispensable for your institution's library. You can then borrow it to read the chapters conforming to your interests. McArthur is appropriate for the reference collection, but it may not suffer from overuse. If you desire a similar text for your own library, *The Cambridge Encyclopedia of the English Language* (Crystal, 1995, and now available cheaply in paperback) is much to be preferred for its coverage of the same material (and more) in a vastly superior and engaging form. Finally, as its price declines, you might consider buying the CD of the OED.

REFERENCES

Biber, Douglas. *Variation across speech and writing*. Cambridge: Cambridge University Press. 1988.

Burchfield, Robert W. *The English Language*. Oxford, Oxford University Press 1985 (re-issued in 2002).

Crystal, David. *The Cambridge Encyclopedia of The English Language*. Cambridge, Cambridge University Press, 1995.

Leech, Geoffrey, Paul Rayson and Andrew Wilson. *Word Frequencies in Spoken and Written English*. London, Longman, 2002.

Landau, Sidney I. *Dictionaries: The Art and Craft of Lexicography*, 2nd. edition. Cambridge, Cambridge University Press, 2001.

McArthur, Tom. *The Oxford Guide to World English*. Oxford, Oxford University Press, 2002.

Mugglestone, Linda (ed.). *Lexicography and the OED: Pioneers in the Untrodden Forest*. Oxford, Oxford University Press, 2000.

Swan, Michael and Bernard Smith (Eds.) *Learner English (2nd. Edition)*. Cambridge, Cambridge University Press, 2001.

Winchester, Simon. *The Surgeon of Crowthorne*. London, Viking, 1998.

Winchester, Simon. *The Meaning of Everything: The Story of the Oxford English Dictionary*. New York, Oxford University Press, 2003.

ENDNOTES

1 Some years ago it was estimated that there were about 400 million native speakers of English, but additionally at least that number learning or using the language. There are said to be more than 40 million learning the language in China alone. It is difficult to get precise figures on speaker numbers since it requires a distinction between regular and occasional users and because the numbers increase each year. McArthur in the book under review here gives figures for regular users in 1997, but these cannot take into account the surge of English use in the intervening years, particularly on the Internet.

2 For learners, for example, Oxford University Press publishes at least 12 dictionaries (at last count, 2003), including an *Elementary Learner's Dictionary, an Advanced Learners Dictionary, a Dictionary of Phrasal Verbs, …of Collocations etc*.

3 'Literary works' here refers to any written source, from imaginative literature (novels, plays etc.) to non-fiction (scientific, religious, philosophical writings etc.) and ephemera (newspapers, reports timetables, epitaphs, etc.).

4 Many such definitions could be quoted, but the one most frequently cited is his definition of oats, as; "A grain which in England is generally given to horses, but which in Scotland feeds the people". Johnson's particular dislike of Scotland was often noted by Boswell. One of Johnson's critics said that his dictionary "is merely a glossary of his own barbarisms".

Johnson is often taken to task for definitions which employ terminology more complex than the word being defined. Such a tendency is now avoided in most dictionaries.

5 The term 'philology' in the mid nineteenth century had the meaning we would ascribe to 'linguistics' today.

6 Frederick Furnivall, the second editor, is one such example. Though a lawyer, his inherited wealth allowed him to indulge in the gentlemanly pursuit that the dictionary was at that time. His eccentric worldly pursuits (of much-younger female shop assistants, for one) were undertaken with the enthusiasm he brought to the dictionary, but earned him the moral censure of the clerical backers of the project.

7 Organizing, filing and housing the slips became an additional problem. Some sent abroad to sub-editors were lost for years. Murray employed his numerous children in helping with the filing task.

8 The supplement to the OED begun in the thirties was also completed in alphabetical order. The reader is referred to Burchfield (1985) for an account of this, a part of his larger task in that book.

9 The printers at OUP reckoned the corrections to be the most extensive ever returned to them. Photographs of samples of the text provided in Winchester, show the enormous difficulty of the job of proofing.

10 Murray was subsequently granted an MA from Cambridge and later one from Oxford, the latter it seems, more because it was necessary for him to have library privileges than out of merit.

11 Grammatical tagging refers to the identification of the grammatical category (part of speech) that the word represents in the string in which it appears. In the case of the BNC, 23 main grammatical categories were identified, though some were sub-classified such that 22 subclasses of nouns and 31 of verbs are identified.

12 The largest corpus is the Bank of English at the University of Birmingham with more than 300 million words to draw upon. This, at an earlier stage, formed the basis for the *Collins Co-build Dictionary*, a somewhat revolutionary dictionary based on actual usage, and one well received by the English language-teaching community.

13 In order of predominance the texts cover the domains from world affairs (18%), to social science, leisure, applied science, arts, commerce, natural science, and belief and thought (3.4%).

14 Multiword units are difficult to classify because pf arbitrariness in English orthographic convention. We might wonder why *in spite of* is treated as a three-word unit when it is pronounced as one, while *nevertheless* and *notwithstanding*, for example, would be treated as single-word units? The answers to such puzzles lie deep in the history of derivational

morphology.

15 The list may be considered to be swollen by entries for "words" that would be rejected by many dictionaries and even by Scrabble players. Some examples of such headwords are personal names (*Thatcher*, as in the politician, though the occupation denoted by the same spelling is not included. It is a relief to note that the data must precede the fame of David Beckham), *St*. (2 entries, presumably saint and street) and unclassifiables such as *erm*, and ephemera such as EEC.

16 The alphabetical table gives entries under a headword with its frequency taken as the headword and its lemmas then listed separately (e.g a headword for a verb is entered under the infinitive form with its lemmas being its grammatical inflections).

17 With the exception of one-syllable words which are twice as frequent as all others up to and including 6-syllable words, the frequencies of each class from 2-syllables onwards, is about half that of the class above it.

18 It is also very much a language representing the modern variants of a vocabulary first brought to England between the fifth and tenth centuries by Germanic invaders, and to a lesser extent the Romance languages, especially French, with only minor traces of Celtic and Latin.